Bystander: QoE Perception for Dynamic Video Streaming from Encrypted Traffic

Jialin Zhang, Hongyun Zheng, Yongxiang Zhao, Yuchun Guo
School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing, China

***Abstract*—In this paper we introduce a new application layer objective QoE metric and propose a machine learning approach to infer the introduced metric for QoE estimation based on features abstracted from encrypted traffic. Experimental results show that our method can infer video users QoE with an accuracy higher than 80%.**

I. Introduction

Video streaming has become the main application on the Internet, which accounts for nearly 58% of the total Internet traffic in 2018 [1]. It is critical for network operators to understand Quality of Experience (QoE) perceived on user sides. However, assessing QoE is challenging since not only it subjective, but also application-specific, and the operators do not have access to applications at terminal device to get information on objective metrics impacting QoE. Instead, operators have to passively monitor network traffic to infer objective QoE metrics, which tends to be more and more complicated by many video being transmitted on HTTP Adaptive Streaming (HAS) and encrypted [2].

One way to estimate HAS video QoE is using machine learning methods to learn the relationship between network traffic and application layer objective QoE metrics [3], [4]. Stalls during video playback is important factor impacting QoE [5]. Since stalls typically happen due to empty playback buffers, the buffer occupancy or length is usually used as an objective QoE metric [6]. However, we find that not only the absolute buffer length but also the variation tendency of buffer length is related to QoE. Thus we use the combination of them as a new application layer objective metric to indicate user QoE. We then propose an approach to infer the QoE metric for estimating QoE from encrypted network traffic.

Our contributions in this paper are as follows:

- We introduce a new application layer objective QoE metric, and propose a machine learning method to predict the metric for assessing QoE from encrypted traffic. Experimental results show that our method can infer video users QoE with an accuracy higher than 80%.
- Our approach is based on the statistical characteristics of network traffic. It not only can work without client/server's involvement, but also be applied to encrypted traffic.

II. New QoE metric

Fig.1 depicts the process of filling video buffer during a video playback. According to the buffer occupancy and its variation direction, we use a new metric, called buffer-dynamic, to describe the buffer states. The buffer goes through four states. In Climbing (CL) state the buffer occupancy keeps increasing from zero to the max threshold. In Swing (SW) state the buffer length oscillates between the max and min threshold. In Close-Empty (CE) state the buffer length keeps falling beyond the min threshold but still higher than zero. When the buffer is empty, it goes into Empty (EM) state. Correspondingly the player states are shown in the bottom of Fig.1. Events change the state of a buffer and that of a player. For example, when the buffer becomes empty it switches to EM state, and at the same time the player goes into stalling state. The reason why we use buffer-dynamic to describe buffer states is that we find it is the combination of buffer length and its variation direction rather than buffer length alone that exactly indicates the player state and users viewing experience. From Fig.1 it can be seen that the buffer occupancy would be the same in different buffer states and player states, which correspond to different users QoE. For example, with the occupancy of l_1 the buffer would stay in CL or CE state. When the buffer is in CL state, the video player is either starting or re-starting. In such a condition the player keeps initializing thus the video is frozen. In contrast when the buffer stays in CE state the player is playing video. During this status even though the buffer length is below the min threshold, the player would keep playing unless event happens (*e.g.*, buffer empty) to forcing the player to change state. It is similar to the case where the buffer size is l_2 , the buffer states can be CL or SW, and the player stays in starting

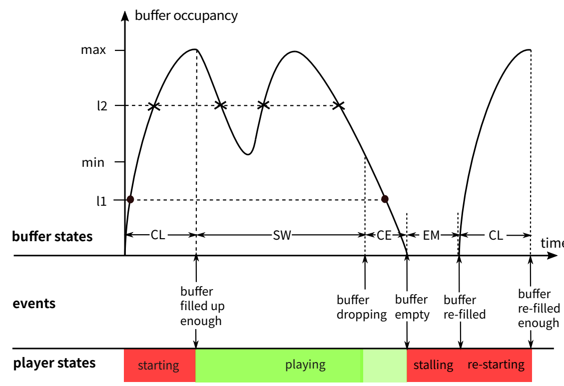

Fig. 1: The process of filling video buffer and associated events which change buffer and player state

or playing state correspondingly. This observation motivates we to use the metric of buffer-dynamic as QoE metric.

III. QoE Assessment

We propose a machine learning method to infer the QoE metric of buffer-dynamic from encrypted video traffic.

In the training phase operators play videos and collect sufficient videos traffic in the network as well as ground truth of video QoE metrics directly from the video player. For each video traffic, operators extract features and use them to train the classifier model for QoE metric. In the testing phase, operators extract features for a test video from its packet trace and use the machine learning models to infer QoE metric.

Table 1 shows the features we extracted from traffic. These features are separately computed from upstream and downstream traffic. Based on the client and server IP addresses, and timestamp at which request packet is sent, request interval can be obtained. In order to identify each video chunk, we rely on TCP headers. Each chunk is transferred in multiple TCP packets from the server to client. The client uses the same ACK numbers for all received TCP packets from the same chunks. Thus we use the ACK numbers as chunk ID to separate video chunks. We further sum up the number of bytes of packets from the same chunk to get byte count of each video chunk. We also compute the size difference between two adjacent chunks. The difference can be zero, positive or negative value, which reflects the variation direction of the next chunk size.

TABLE I: Features extracted.

Features	Description
Request interval	The time difference between two adjacent instants of sending requests. (Upstream)
Chunk bytes	Bytes of video chunk from the server to client. (Downstream)
Dif of chunk bytes	The size difference between two adjacent chunks. (Downstream)

IV. Results

We set up a controlled testbed in lab environment to collect data. A PC, equipped with Ubuntu Kylin 16.04 LTS, is used as video client and simultaneously runs Wireshark software to monitor video traffic. The client selects and plays video files via dash.js, an open source DASH video player. We use TC (Traffic Control) module of the ubuntu kernel to limit the network bandwidth accessed by PC.

The video file is Big Buck Bunny with 10 different representations and a frame rate of 30fbps, which lasts about 10 minutes. The collected data contains 16 video traffic traces, totally 38,338 samples, including 1694 of CL, 32472 of SW, 2380 of CE and 1792 of EM. We randomly select 70% of samples for model training, and the remaining are used for testing. We utilize random forest classifier contained in the machine learning toolkit Sklearn for classification.

Table 2 exhibits the classification results. It shows that except CE other three buffer states can be recognized with a precision rate above 80% while recall rate of all buffer states is at least 84%. The reason why identifying CE state with lower accuracy is perhaps that CE is so near to SW that some of SW samples are easily misclassified as CE. As is shown in the confusion matrix in Table 3, although only 3.8% of SW samples are misclassified as CE, the absolute number is close to the number of CE samples classified correctly, which are 381 and 595 respectively.

TABLE II: Results of model evaluation.

Buffer states	Classes	Precision	Recall	F1-score
CL	0	0.87	0.93	0.90
SW	1	0.99	0.95	0.97
CE	2	0.60	0.84	0.70
EM	3	0.80	0.94	0.87

TABLE III: The confusion matrix from evaluation.

Original label	Predicted label			
	CL	SW	CE	EM
CL	92.8%(464)	4.6%(23)	1.6%(8)	1.0%(5)
SW	0.56%(54)	94.98%(9287)	**3.89%(381)**	0.57%(56)
CE	0.85%(6)	6.7%(48)	**83.8%(595)**	8.6%(61)
EM	1.94%(10)	1.56%(8)	2.14%(11)	94.36%(485)

V. conclusion

It is crucial and challenging for network operators to estimate video users QoE. We combine the video buffer length with its variation direction and introduce a new application layer metric to indicate video users QoE objectively. We extract features from the statistical characteristics of encrypted traffic and propose a machine learning methods to learn the relationship between network traffic and the new application layer objective QoE metric. The experimental results show that this approach can infer users QoE with an accuracy higher than 80%. In future work we will improve the results by exploiting features and balancing sample data.

Acknowledgment

This work was supported in part by the National Science Foundation of China under Grant No.61572071, 61872031.

References

[1] 2018. Technical Report - SANDVINE: The 2018 Global Internet Phenomena report. https://www.sandvine.com/phenomena.(2018).

[2] Dimopoulos G, Leontiadis I, Barletros P, Papagiannaki K, "Measuring Video QoE from Encrypted Traffic", Internet measurement conference, 2016, pp.513-526.

[3] Hongyun Zheng, Yongxiang Zhao, Xi Lu, Rongzhen Cao, "A Mobile Fog Computing-Assisted DASH QoE Prediction Scheme", Wireless Communications and Mobile Computing, vol. 2018, https://doi.org/10.1155/2018/6283957.

[4] M. H. Mazhar, Z. Shafiq, "Real-time video quality of experience monitoring for https and quic", Proc. of IEEE INFOCOM, pp. 1-9, 2018.

[5] Dobrian F, Sekar V, Awan A, Stoica I, "Understanding the impact of video quality on user engagement", Computer Communication Review, 2011, 41(4), pp. 362-373.

[6] Krishnamoorthi V, Carlsson N, Halepovic E, Petajan E, "BUFFEST: Predicting Buffer Conditions and Real-time Requirements of HTTP(S) Adaptive Streaming Clients", MMSys17, 76-87. https://doi.org/10.1145/3083187.3083193.