Streaming Video Detection and QoE Estimation in Encrypted Traffic

Bogdan Iacob
Advisor: Kajó Márton
Seminar Future Internet WS2017/2018
Chair of Network Architectures and Services
Departments of Informatics, Technical University of Munich
Email: bogdan.iacob@in.tum.de

ABSTRACT

With security and privacy as key matters of evolving importance in the field of online content distribution, Over The Top (OTT) providers of video streaming services shifted their focus towards end-to-end encryption. At the same time, the demand for online video content has been dramatically increasing, as a result of the ever-growing number of users and the expanding popularity of video streaming services. While end users benefit from the improvement in privacy, the change to encrypted video traffic and the high market demands raise new challenges for Internet Service Providers in the quest of monitoring service performance and maintaining a competitive level of quality.
This paper describes current research that addresses the challenges of detecting streamed video and estimating the Quality of Experience (QoE) in encrypted traffic. In approaching these challenges, various methodologies, relying on network traffic analysis and machine learning-based QoE classification models, are presented.

Keywords

Encrypted traffic; Network measurements; Video streaming; Quality of Experience; Passive monitoring; Machine learning

1. INTRODUCTION

The need for improved online security increased substantially side by side with the growth of the Internet. As a result, the expansion of HTTPS across the World Wide Web has gained a strong momentum. Along this trend, big players in the online video streaming industry switched to encrypted traffic in delivering their video services. One of the biggest OTT providers of video streaming services, YouTube, began the process in 2011 and managed to roll out encryption using HTTPS to 97 percent of its total traffic by 2016 [18]. In a similar manner, Netflix's Internet television network switched to encrypted traffic in 2016 [16].
Furthermore, Cisco forecasts the global IP video traffic to increase at a rapid pace in the next 4 years, to an expected 82 percent of all consumer Internet traffic by 2021 [2].
Consequently, the importance of good QoE estimations has risen considerably, while the QoE management has become an even bigger challenge for ISPs and mobile operators, which rely their measurements mostly on deep packet inspection [16].
QoE is needed in order to allow ISPs and mobile operators to assess and improve the quality of their services in order to remain competitive.

The end-user QoE is determined based on the evaluation of certain Key Performance Indicators (KPIs), such as initial start-up delay, stalling, average video quality in terms of bitrate and resolution, as well as quality variations. While these KPIs could be easily extracted from video traffic based on HTTP, by performing passive network measurements, the use of HTTPS prevents the inspection of packet headers and payload data. In attempt to address this challenge, research presented in this paper proposes approaches based on Machine Learning algorithms to statistically relate network- to application-level metrics, as a solution to detect QoE degradation.
The rest of the paper folds into 3 main chapters. Chapter 2 presents current mechanisms for detecting streamed videos and provides a brief analysis with regard to their performance and accuracy, as well as their impact on the Net Neutrality principles. Chapter 3 approaches the process of QoE estimation, describing the KPIs' classification and providing a comparison of several machine learning algorithms used in the referenced research. Finally, Chapter 4 outlines the state of the art and future trends of QoE estimation and management development.

2. STREAMING VIDEO DETECTION

The first step in assessing the QoE is the detection of streamed videos. This section describes approaches to detect YouTube and Netflix traffic. These approaches rely on the DNS Lookup method, as well as on fingerprinting the content based on the video streaming protocol.

2.1 DNS Lookup

As the bitrate and content specific information from encrypted traffic cannot be obtained through classic deep packet inspection techniques, the DNS Lookup method can be used to at least identify video packets, in case specific server IP addresses are known in advance. Research from [14] focuses exclusively on YouTube traffic and identifies video traffic by checking the server IP address from either the DNS response packets or TLS handshake messages. In [14], a video server IP list is built and constantly updated by checking for a specific string pattern in the DNS response or SSL/TLS handshake packets. These records are associated with a video streaming and are removed from the list if they are not hit within a week.

2.2 Fingerprinting

doi: 10.2313/NET-2018-03-1_01

Fingerprinting YouTube Traffic
The current implementation of the YouTube video service is based on two adaptive streaming modes, Apple HTTP Live Streaming (HLS) and MPEG Dynamic Adaptive Streaming over HTTP (DASH), both of which adopted SSL/TLS encrypted transmission. As a first step, the streaming mode has to be distinguished. In approaching this challenge, [14] propose a solution, which relies on the fact that each mode starts a video streaming session with the transfer of certain video index files. On one hand, the index file, DASH uses the "initsegment" file together with the related video files, which are stored on the same server. On the other hand, for HLS, the index file "manifest" and the associated video files are stored on different servers [14]. A second difference between the two streaming modes resides in the fact that, unlike DASH, HLS ClientHello messages of the SSL/TLS handshakes contain the specific string "manifest.googlevideo.co". Thirdly, the first segment after the TLS handshake differs in size, depending on the mode. Hence, in HLS, a large video block is sent right after the handshake, while in DASH, a small 1.5KB segment is sent from the video server.

Fingerprinting Netflix Traffic
The technology used by Netflix for browser-based streaming, relies on first encoding the videos as variable bitrate (VBR), followed by streaming them using DASH via Microsoft Silverlight [16]. According to [16], the combination of VBR and DASH can produce unique fingerprints for each video. Furthermore, this pattern is particularly observable for Netflix videos, as they have a higher amplitude of the bitrate variation, compared to other streaming services. The implementation proposed by [16], analyzes Netflix traffic, by leveraging two tools: adudump and OpenWPM.
The former is a program that can infer the sizes of the application data units (ADUs) transferred over each TCP connection, using the TCP sequence and acknowledgement numbers. Being able to distinguish successive video segments of an HTTPS-protected Netflix stream, adudump provides the input to the identification algorithm, as can be seen in Table 1.
OpenWPM is a framework used to conduct web measurements, being basically an automated browser, based on Firefox and Selenium, which takes a list of URLs as input and visits them sequentially.
In the first step, the video database is created using the two tools to gather the unique information for each Netflix video, by extracting fingerprints from the first 20 seconds of playback. In the second step, live traffic is logged from the network by extracting ADUs larger than 200,000 bytes, sent from a server on port 443. Finally, the live information is compared to the existing entries in the database, by performing a kd-Tree Search. Results of the success rate and detection time, obtained by [16], can be seen in Figure 1, respectively Figure 2.

2.3 Net Neutrality
While on the one hand, the goal of QoE in mobile traffic is to optimize the end user's perceived quality and increase the level of satisfaction, on the other side, traffic management measures impact other dimensions like neutrality and privacy. Not only the QoE monitoring and measurement are challenging tasks, but also the QoE control and optimization constitute a demanding duty. In order to fulfill clients'

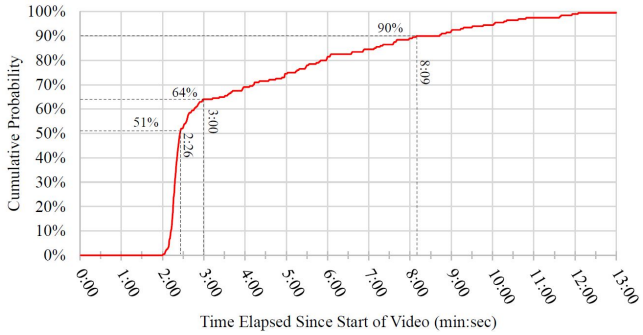


Figure 1: Cumulative probability of identifying a video before a specified amount of time has elapsed.

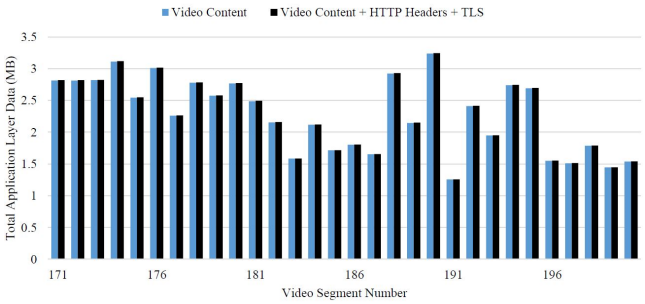


Figure 2: Netflix video overhead due to HTTP headers and TLS (Home, 3830 kbps encoding).

expectations and maintain their market shares on a competitive market, ISPs have to manage resources in a way that the QoE is not negatively impacted, while achieving cost efficiency. In a heterogeneous environment with diverse usage contexts, limited bandwidth and changing conditions, net neutrality is often affected [8].
From the technical point of view, in order to address performance issues and cope with congestions, ISPs use traffic prioritization schemes. As a result, communication services classified to be more valuable to the end user, like voice calls, are routed with higher priority than other services [8]. However, congestion issues can be solved through infrastructure updates to expand capacity.
Moreover, [10] describe also other prioritization schemes, stronger related to economic and financial arguments, rather than technical ones. Although wireless ISPs like AT&T and Verizon justify these restrictions as being related only to capacity issues, they offer monthly data plans that differentiate between the applications that are allowed to access the Internet. As a result, the tethering functionality of smartphones can be used only with a monthly data plan that allows it.
According to [6], ISPs justify using borderline practices for the traffic management for the following reasons: to ensure security, to relieve congestions, to ensure adequate QoS for selected traffic. The first reason is used when justifying blocked traffic, the second one to justify shaping on file-sharing traffic, while the third one to justify exclusive QoS for the ISPs own Voice over IP (VoIP) traffic.
In conclusion, the net neutrality issue remains an open debate point [7], as in absence of detailed regulatory policies,

ISPs continue to use the same traffic management practices, while research like [8] admittedly falls short in providing feasible solutions.

3. QOE ESTIMATION

Estimating the QoE from encrypted traffic is based on several KPIs, which constitute the ground input for the QoE estimation models. The most important KPIs [17], as well as the most common machine learning techniques are described below.

3.1 Key Performance Indicators

Initial Delay

The initial delay or start-up delay refers to the time span between a user's video request and the actual playback begin. This delay comprises two components, the network delay and the initial buffering delay. The first delay involves the required time for sending the request to the server, as well as receiving the first segments and is influenced by factors like server response times, DNS lookups and CDN redirections [3]. The second delay refers to the necessary time to fill the initial buffer required to permit a seamless playback [4].
As stated in [11] and [19], the initial delay has a small impact on the Mean Opinion Score (MOS), which is only marginally influenced by the length of the video stream, as can be seen in Figure 3.

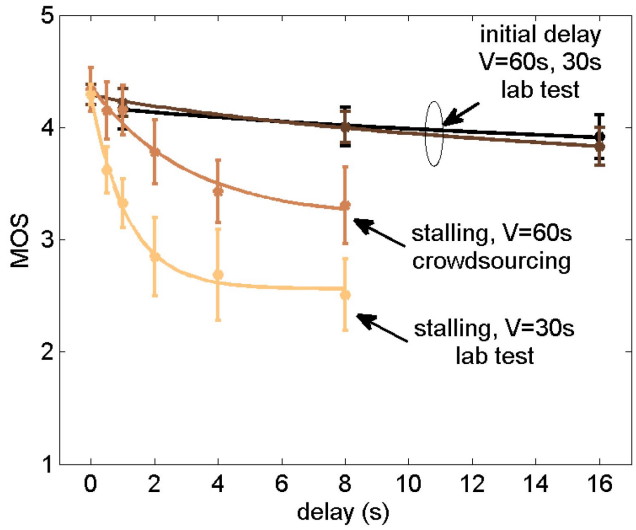


Figure 3: One stalling vs. initial delay for YouTube QoE for videos of duration V = 30 s and V = 60 s, respectively.

Stalls

A stall occurs when the content consumption rate exceeds the average download rate due to a limited network throughput. This forces the playback to freeze until the buffer is refilled with sufficient content segments. The severity of stalls is influenced by their duration and frequency during a video playback.
As reported by [12], the combination of the two factors strongly influences the MOS, being the most relevant KPI, with the highest impact on the QoE. Furthermore, as can be seen in Figure 4, sector AB of the chart is mainly a dark color region, indicating a low MOS, while the light color regions, indicating a high MOS, shown in sectors CA and BC are the result of zero packet loss rate, respectively zero packet delay.

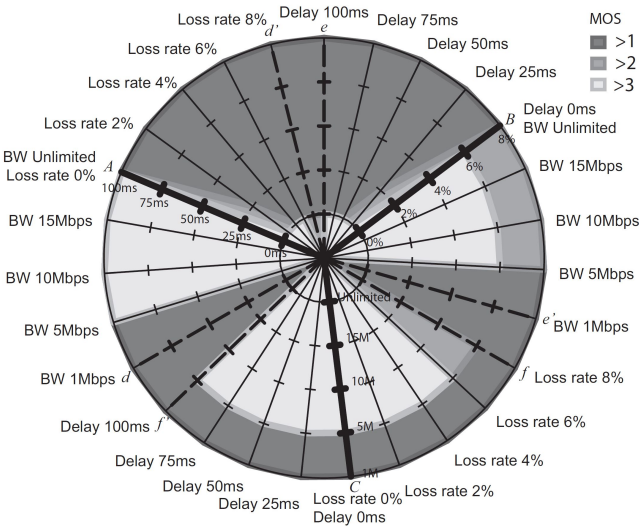


Figure 4: The radar chart mapping network QoS with QoE.

Average Representation Quality

This metric represents the average quality of all streamed video segments during a video session and is the result of the overall media throughput, measured in bits per second. However, this KPI is relevant only for video sessions based on HTTP Adaptive Streaming (HAS).
According to [9], a high average representation quality is correlated to superior user experience, as can be observed in Figure 5.

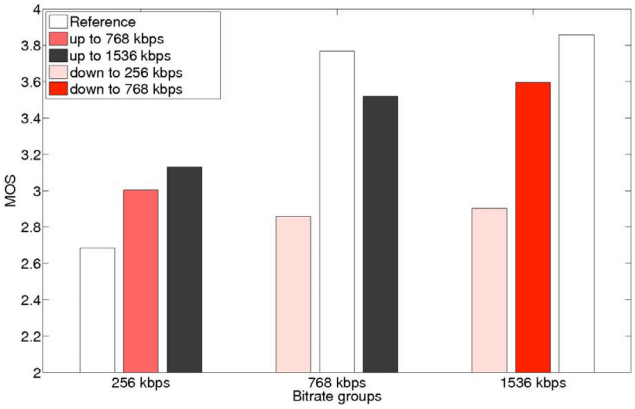


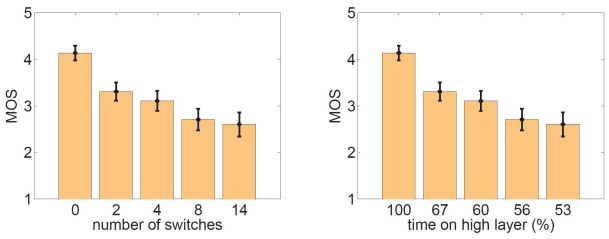
Figure 5: Bit rate switching for H.264 in WiFi.

Representation Quality Variation

Finally, the changes in the representation quality throughout a video session can significantly affect the overall QoE. This metric is based on two components, the frequency of quality changes and their amplitude. The former counts the number of quality switches during a video session and is influenced by the changing network conditions. The latter defines the

doi: 10.2313/NET-2018-03-1_01

difference in quality between two consecutive video segments in the attempt to avoid buffer depletions or to increase the video quality when the network conditions allow it.
As presented in [5], studies performed in mobile networks have shown that the representation quality variations have a high impact on the overall QoE, as can be seen in Figure 6.



(a) Number of switches (b) Time on highest layer

Figure 6: Frequency and time on highest layer changed simultaneously.

3.2 Machine Learning

Several studies propose machine learning, as a technique to quantify the correlation between QoS and QoE. Machine learning methods are used to predict the QoE based on inference rules extracted from a set of measurements reflecting the network state and the user's perception [1].
In [14], the authors propose a Machine Learning-based bitrate estimation system, which parses the bitrate information from IP packet level measurements. In order to assess the QoE based on the bitrate of video segments, a decision tree classifier is used. In the first step, YouTube traffic is identified based on the DNS Lookup method. After the HTTPS adaptive streaming protocol has been identified, the bitrate identifiers are extracted and the KPIs are computed in order to assess the vide QoE, as can be seen in Figure 7.
The solution proposed by [13] focuses on mobile devices and is based on a model built using encrypted network traffic and information collected on a client device, as illustrated in Figure 8. The process consists of three steps, data collection, data processing, and model building. Data are first collected by leveraging the YouTube Iframe API [13], which is run on the client to monitor application-level data and extract relevant KPIs. On the one hand, network traffic is captured in order to calculate traffic features like the average throughput or inter-arrival time. On the other hand, the collected application-level data, like the initial delay and stalling duration, are stored into log files, which are uploaded to the YouQ server. Based on the collected information, 33 traffic features for each video are derived to form a dataset for training the model.

Finally, a high, medium or low quality level is assigned to each video. In order to assess the quality levels, different algorithms were run, using the Waikato Environment for Knowledge Analysis (Weka), which is a suite of machine learning software written in Java.

3.2.1 *Common Machine Learning Algorithms*

This section briefly describes the principles of the most common classification algorithms used in [13] and provides an

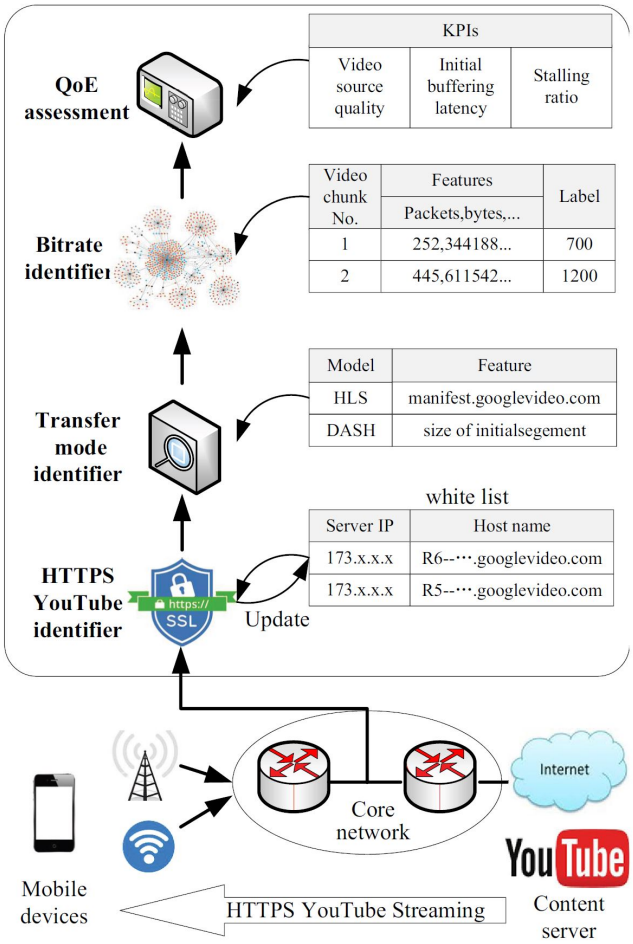


Figure 7: Architecture of QoE assessment system.

overview of the classification results.

OneR
OneR is a simple learning algorithm that creates a rule for each predictor in the data and then selects the rule with the fewest prediction errors as its single rule.

Naive Bayes
Naive Bayes is a supervised learning algorithm that uses every pair of features, as being equally relevant and statistically independent.

J48
J48 is an algorithm used to generate a decision tree, based on a top-down strategy to split the dataset on each attribute depending on its value.

SMO
The sequential minimal optimization (SMO) algorithm is based on finding a hyperplane that maximizes the minimum distance to the training set.

Random Forest
The Random Forest is a classifier that operates by constructing a multitude of decision trees and distribute instances in

doi: 10.2313/NET-2018-03-1_01

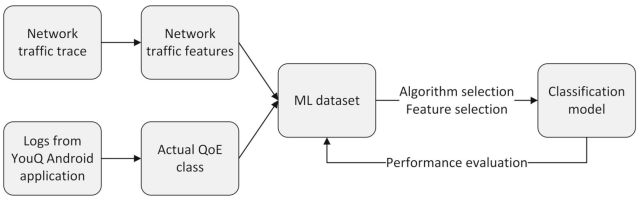


Figure 8: Approach for QoE classification based on network features.

Table 1: adudump trace of Home (3830 kbps encoding). These are segments 171-180 from Figure 2.

Timestamp	Local PC	Dir.	Netflix Server	Size (B)
1471357732.77583	134.240.17.111.31177	>	198.45.63.167.443	756
1471357736.70148	134.240.17.111.31177	<	198.45.63.167.443	2817667
1471357736.77902	134.240.17.111.31177	>	198.45.63.167.443	756
1471357740.89304	134.240.17.111.31177	<	198.45.63.167.443	2816159
1471357740.97057	134.240.17.111.31177	>	198.45.63.167.443	756
1471357744.45695	134.240.17.111.31177	<	198.45.63.167.443	2822089
1471357744.53453	134.240.17.111.31177	>	198.45.63.167.443	756
1471357748.76052	134.240.17.111.31177	<	198.45.63.167.443	3117490
1471357748.83926	134.240.17.111.31177	>	198.45.63.167.443	756
1471357752.72718	134.240.17.111.31177	<	198.45.63.167.443	2548098
1471357752.80466	134.240.17.111.31177	>	198.45.63.167.443	756
1471357756.87447	134.240.17.111.31177	<	198.45.63.167.443	3014236
1471357756.95195	134.240.17.111.31177	>	198.45.63.167.443	756
1471357760.48768	134.240.17.111.31177	<	198.45.63.167.443	2263764
1471357760.56593	134.240.17.111.31177	>	198.45.63.167.443	756
1471357764.73616	134.240.17.111.31177	<	198.45.63.167.443	2782180
1471357764.81363	134.240.17.111.31177	>	198.45.63.167.443	755
1471357768.73659	134.240.17.111.31177	<	198.45.63.167.443	2577683
1471357768.81421	134.240.17.111.31177	>	198.45.63.167.443	756
1471357772.97218	134.240.17.111.31177	<	198.45.63.167.443	2770492

the class the most decision trees agree on.

3.2.2 Classification Results

As can be seen in Figure 9, the accuracy of the selected algorithms ranges between 74.62% and 80.18% in classifying the videos according to the 3 QoE classes.

Algorithm	Selected features	Accuracy
OneR	*throughputMedian*	74.62%
Naïve Bayes	*avgPacketSize, averageInterarrivalTime, minimalInterarrivalTime, sizeThroughTimeMedian, push, interarrivalTimeThroughTimeMedian, initialThroughput2*	77.35%
SMO	*maximalSizeThroughTime, minimalInterarrivalTime, sizeThroughTimeMedian, maxThroughputThroughTime, dupack, effectiveThroguhput*	77.35%
J48	*minimalInterarrivalTime, avgInterarrivalTimeThroughTime, sizeThroughTimeStdDev, interarrivalTimeThroughTimeMedian, throughputMedian*	78.20%
Random Forest	*avgPacketSize, minimalSizeThroughTime, push, initialThroughput10, minThroughputThroughTime, interarrivalTimeThroughTimeMedian, dupack, effectiveThroughput*	80.18%

Figure 9: Classification results.

4. FUTURE TRENDS

While many studies rely on machine learning techniques to estimate the QoE, most of them use only a limited number of influence factors, while few approaches map simultaneously the impact of multiple factors to obtain a multidimensional QoE model [13].

Future work also focuses on developing the existing platforms in order to increase the degree of support for different devices, operating systems, network types and other environment variables [15] .

5. CONCLUSION

While the field of online content distribution has been continuously gaining momentum, the importance of good QoE has significantly increased. In order to remain competitive, ISPs need to adapt their QoE detection techniques for encrypted traffic, as the big OTT providers of video streaming services already use end-to-end encryption to a big extent. Existing studies approach this challenge relying on the current video streaming protocols and ML techniques, but are limited to specific platforms, devices, operating systems and types of networks. While the presented ML methods show promising results, there is still room for improvements in the areas of performance and accuracy.
On the other side, current work shows that the legislation lacks consistency in regulating the QoE and traffic management to preserve principles of the net neutrality.

6. REFERENCES

[1] S. Aroussi and A. Mellouk. Survey on machine learning-based qoe-qos correlation models. In *2014 International Conference on Computing, Management and Telecommunications (ComManTel)*, pages 200–204, 2014.

[2] Cisco Visual Networking Index: Forecast and Methodology, 2016-2021. June 6, 2017. `https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/complete-white-paper-c11-481360.pdf`.

[3] G. Dimopoulos, I. Leontiadis, P. Barlet-Ros, and K. Papagiannaki. Measuring video qoe from encrypted traffic. In *Proceedings of the 2016 Internet Measurement Conference*, pages 513–526, 2016.

[4] T. Hossfeld, S. Egger, R. Schatz, M. Fiedler, K. Masuch, and C. Lorentzen. Initial delay vs. interruptions: Between the devil and the deep blue sea. In *2012 Fourth International Workshop on Quality of Multimedia Experience*, pages 1–6, 2012.

[5] T. Hossfeld, M. Seufert, C. Sieber, and T. Zinner. Assessing effect sizes of influence factors towards a qoe model for http adaptive streaming. In *2014 Sixth International Workshop on Quality of Multimedia Experience (QoMEX)*, pages 111–116, 2014.

[6] S. Jordan. Some traffic management practices are unreasonable. In *2009 Proceedings of 18th International Conference on Computer Communications and Networks*, pages 1–6, Aug.

[7] S. Jordan. Four questions that determine whether traffic management is reasonable. In *2009 IFIP/IEEE International Symposium on Integrated Network Management*, pages 137–140, June 2009.

[8] S. Jordan. Traffic management and net neutrality in wireless networks. *IEEE Transactions on Network and Service Management*, 8(4):297–309, December 2011.

[9] B. Lewcio, B. Belmudez, A. Mehmood, M. WÃd'ltermann, and S. MÃűller. Video quality in next generation mobile networks x2014; perception of time-varying transmission. In *2011 IEEE International*

Workshop Technical Committee on Communications Quality and Reliability (CQR), pages 1–6, 2011.

[10] L. Martinez, O. S. J. Alvarez, and J. Markendahl. Net neutrality principles and its impact on quality of experience based service differentiation in mobile networks. In *2015 Regional Conference of the International Telecommunications Society (ITS)*, 2015.

[11] R. K. Mok, E. W. Chan, X. Luo, and R. K. Chang. Inferring the qoe of http video streaming from user-viewing activities. In *Proceedings of the First ACM SIGCOMM Workshop on Measurements Up the Stack*, pages 31–36, 2011.

[12] R. K. P. Mok, E. W. W. Chan, and R. K. C. Chang. Measuring the quality of experience of http video streaming. In *12th IFIP/IEEE International Symposium on Integrated Network Management (IM 2011) and Workshops*, pages 485–492, 2011.

[13] I. Orsolic, D. Pevec, M. Suznjevic, and L. Skorin-Kapov. A machine learning approach to classifying youtube qoe based on encrypted network traffic. *Multimedia Tools and Applications*, pages 1–35, 2017.

[14] W. Pan, G. Cheng, H. Wu, and Y. Tang. Towards qoe assessment of encrypted youtube adaptive video streaming in mobile networks. In *2016 IEEE/ACM 24th International Symposium on Quality of Service (IWQoS)*, pages 1–6, 2016.

[15] N. Radics, P. SzilÃągyi, and C. VulkÃąn. Insight based dynamic qoe management in lte. In *2015 IEEE 26th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pages 1746–1751, 2015.

[16] A. Reed and M. Kranch. Identifying https-protected netflix videos in real-time. In *Proceedings of the Seventh ACM on Conference on Data and Application Security and Privacy*, pages 361–368, 2017.

[17] Subjective audiovisual quality assessment methods for multimedia applications. December 3, 1998. `http://handle.itu.int/11.1002/1000/4538`.

[18] YouTube's road to HTTPS. August 1, 2016. `https://youtube-eng.googleblog.com/2016/08/youtubes-road-to-https.html`.

[19] G. Zhai, J. Cai, W. Lin, X. Yang, W. Zhang, and M. Etoh. Cross-dimensional perceptual quality assessment for low bit-rate videos. *IEEE Transactions on Multimedia*, 10(7):1316–1324, Nov 2008.