

Ensuring QoE in Contemporary Mobile Networks for Video Content Distribution

Mirghiasaldin Seyedebrahimi, Xiao-Hong Peng

School of Engineering and Applied Science
Aston University, Birmingham B4 7ET, UK
{seyedebm, x-h.peng}@aston.ac.uk

Abstract— A framework that aims to best utilize the mobile network resources for video applications is presented in this paper. The main contribution of the work proposed is the QoE-driven optimization method that can maintain a desired trade-off between fairness and efficiency in allocating resources in terms of data rates to video streaming users in LTE networks. This method is concerned with the control of the user satisfaction level from the service continuity's point of view and applies appropriate QoE metrics (Pause Intensity and variations) to determine the scheduling strategies in combination with the mechanisms used for adaptive video streaming such as 3GP/MPEG-DASH. The superiority of the proposed algorithms are demonstrated, showing how the resources of a mobile network can be optimally utilized by using quantifiable QoE measurements. This approach can also find the best match between demand and supply in the process of network resource distribution.

Keywords—QoE; scheduling; adaptive video streaming; LTE; Pause Intensity

I. INTRODUCTION

With the introduction of the new-generation mobile communication systems, users are experiencing a revolutionary enhancement of the available network capacity. Meanwhile, various types of Apps of the new mobile devices such as Smartphones and Tablets have stimulated the demands for high quality video-related services over mobile communication networks. However, maintaining a satisfactory level of quality perceived by end users in a large scale network remains a considerable challenge.

The currently dominant video formats such as H.264/MPEG-4 [1] and recently developed protocols, e.g. H.265 [2], aim to provide the required levels of compression and scalability for the forthcoming ubiquitous video based services. At the same time, new streaming methods, including the HTTP-based rate adaptive video streaming techniques, have emerged to comply with the nature of network applications and the growing demands of end users [3]. Network operators and Internet-based Over-The-Top content providers are facing the challenges of maintaining operational complexity while trying to meet the user's expectation for better quality [4].

For the practical implementation of a multimedia streaming framework, a range of specifications and techniques must be taken into account, including those for network infrastructures, transmission and resource allocation protocols with various

types of user interfaces. This is especially crucial in a time variant and spectrum insufficient environment such as wireless mobile communications. Consequently, the required infrastructure for the service, the video content specifications as well as the employed protocols are covered in the existing standards for video streaming over mobile communication systems [5].

This paper proposes a method for the control of quality in delivering video streaming services which utilizes the latest developed video streaming techniques. The performance of the last mile's scheduling and link adaptation functions are optimized by exploiting the user's experienced quality through an objective no-reference video quality metric, Pause Intensity [6]. The proposed approach is different from the existing methods in terms of its flexibility to produce a desired balance between fairness and efficiency and take the correlation between the required data rates of the users and their allocated data rates into account. The advantages of the proposed method is examined in the context of 4G-LTE resource allocation function (at the network side) and the client-driven rate adaptive streaming method (at the user-side).

The rest of this paper is organized as follows: a background that justifies the feasibility and benefits of the proposed framework is explained in Section II. The model description and the proposed resource utilization method are discussed in Section III. Section IV is dedicated to the simulation results and analysis and finally the conclusion is provided in Section V.

II. BACKGROUND AND RELATED WORKS

This section discusses the current developments and related works including the overall structure of a content distributor, the nature of impairments in video distribution between the client-side and network elements of a contemporary mobile communication system (i.e. 3GPP standards) in connection to quality assessment.

A. Content distribution solutions

A general infrastructure model of the most common video streaming service including a client-side, a source for the content and its transmission channel is depicted in Fig. 1. A global access to this service involves some elements from the public networks, the operator's infrastructure and the user's last mile access interfaces. Subsequently, the performance improvement of the service requires a global consideration of the elements and protocols of intermediate networks [7].

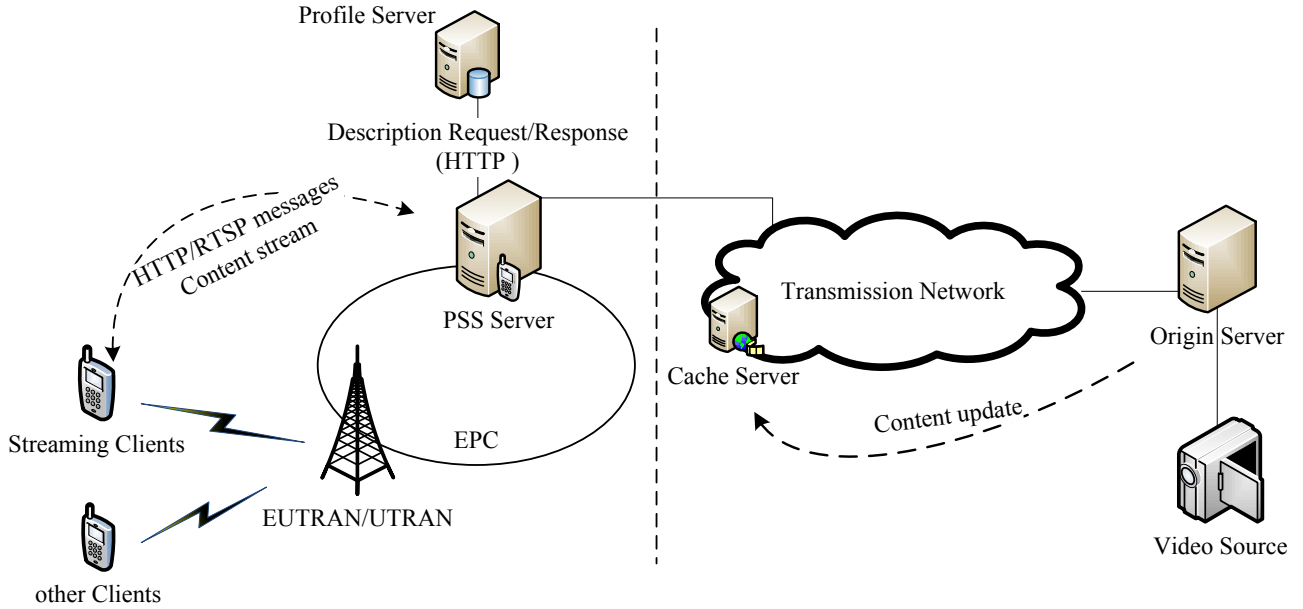


Fig. 1 the network infrastructure of a globally accessible video streaming service over mobile communication and through a distributed network

Distributed system solutions, such as Content Delivery Networks (CDN), are widely used to reduce the complication in quality control for such a global service. A distributed system links up the locally manageable resources with the corresponding globally accessible services. For example, in Fig. 1 the nearest ‘Cache Server’ (edge server) to the user delivers the requested content on behalf of the ‘Origin Server’ to reduce the latency and the load of the traffic to/from the main server.

Some solutions are also introduced specifically for streaming over mobile communication systems. Packet-switched Streaming Service (PSS) is a package of solutions for different aspects of the video streaming over the new generations of the mobile communication systems, e.g. 3G and LTE [5]. PSS defines the functionalities required to support a server-driven streaming solutions (e.g. RTP-based) or a client-driven adaptive streaming (e.g. 3GP-DASH) [8]. The PSS-related protocols aim to provide a packet-based framework which bridges the gap between the trivial downloading protocols and the more complex conversational services. However, with the introduction of the Evolved Packet Core (EPC) in LTE, some aspects of the PSS functionalities for packet and IP-based adaptation are already supported by the system without need for PSS.

The proposed solution in this work is applicable to the clients in Fig. 1 with an HTTP-based and adaptive video streaming content. This content may have been distributed through CDN with or without the help of the PSS functions. A PSS function, such as the user description and profile server in Fig. 1, provides a source of information for matching the content specifications with the user’s device capability and environment. Furthermore, the PSS analysis of the client-side buffer status provides the elements required for the evaluation of the proposed quality metric at the network-side without extra transactions between the client and the network [5].

B. Sources of impairments

Degradation in the fidelity of an image and the discontinuity of service delivery are two major impairments in a video streaming service. The decoding and playback procedures at the client side usually suffer from the impacts of the latency of packet transmission as well as the inefficiency and instability of the network throughput. These impairments are obviously associated with the intermediate network elements and the protocols employed. Error control mechanisms, resource allocation policies, and transport control and routing processes are among those aspects of the network protocol stack that influence the fidelity or continuity of the service.

Two transport protocols, i.e. UDP for RTP (Real-time Transport Protocol) and TCP for the most recent HTTP-based streaming services also have impacts on the transmission impairment [9], [10]. For example, discontinuity during the video playback time (namely pause, buffering or under-run) is a known result of the network performance inefficiency. This service interruption can be caused by the user’s bandwidth limitation, the latency of TCP control mechanisms or its rate throttling. The UDP protocol, however, is less likely to cause discontinuity of the service but its unreliability normally results in degraded image quality.

The quality metric adopted in this work, namely Pause Intensity (PI), is a no-reference metric for continuity assessment and takes both pause duration and pause frequency into account. It is content independent and highly correlated with the subjective QoE metric: MOS (Mean Opinion Score) [6]. Our approach is tailored to tackle the buffering or continuity impairments of HTTP-based and adaptive streaming services using TCP at the transport layer. The PI value can be determined using the network performance, η (throughput), and the required data rate, λ (video code rate) as $PI(\eta, \lambda)$ per user [6]:

$$PI(\eta, \lambda) = 1 - \frac{\eta}{\lambda} \quad (1)$$

In a non-recorded streaming scenario η is always less than or equal to λ , i.e. $\eta \leq \lambda$ and $0 \leq PI \leq 1$. The description of the PI model, buffer paly-pause characterization and associated subjective tests are provided in [6].

C. Quality assessment support in mobile communications

Generally, the assessment metrics for a service are presented in the form of Quality of Service (QoS) or Quality of Experience (QoE). QoS comprises system-oriented and quantitatively defined parameters such as latency, loss or throughput while QoE encompasses user-oriented and qualitatively defined scales such as ‘the opinion and judgment of user about the service’. QoE may also be represented by a quantitative metric which has a proven correlation with the user’s experienced quality. Although several QoS and QoE-based approaches are available for vide-related service assessment, very few of them suit an online and on-the-fly assessment of a streaming service over mobile communication. The employed Pause Intensity (PI) metric in this work provides the required quantitative presentation and the proven correlation with the subjective and QoE-based satisfaction of the user.

The required information for PI evaluation including video code rate, λ , and average network throughput, η , is available at the client-side. This information can be estimated at the network-side through the functions related to the streaming service support in PSS. When the adaptive streaming service is provided independent from PSS (such as the independent 3GP-DASH servers in Fig. 2), this information can still be transferred to the corresponding network element as a part of the provided quality-related transactions [5]. These capabilities are briefly explained below.

The QoE support in a contemporary mobile streaming service is actually an attribute defined as a part of the ‘device capabilities’. ‘Device capability’ and ‘user profile’ are the information which facilitate the server-side negotiation as a matching process at the beginning of the streaming. As it has been shown in Fig. 1, the device capability description is exchanged between the profile server (provided by device manufacturer or mobile operator) and streaming server (i.e. PSS) as a part of the negotiation and matching mechanisms. The attributes related to video streaming service in 3GPP-PSS consist of the ‘QoE support enabler’, the type of the carrier protocol (i.e. RTP/RTSP or HTTP) and the available options for rate adaptation mechanisms.

The mobile device sends the URL of the location where PSS can retrieve the user profile (through the RTSP/HTTP data

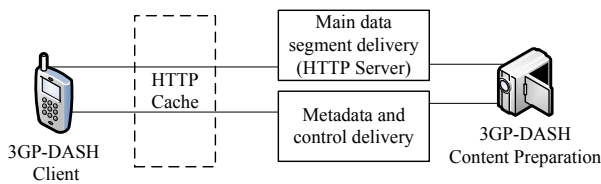


Fig. 2 the infrastructure of an independent 3GP-DASH client-server

units). PSS communicates with the profile server (through HTTP Request/Response) to access the device description information. User can override or add an attribute during this process. The final negotiated attributes will be used by PSS to control the presentation of the streamed media content to a mobile user. The average throughput, initial playout delay, buffer level and metadata Information are the main video-related QoE metrics which are defined by 3GPP and to be included in a quality report. The XML syntax (using the HTTP request with XML in its body) will be used for the transaction of the QoE metrics and reporting protocols.

III. PROPOSED FRAMEWORK

Let an efficiency-oriented resource allocation policy be the default regulator of the available resources in the system:

$$r^* = \arg \max_{\text{all users}} \sum u_k(\bar{R}_k) \quad (2)$$

where $r^{N_{RB} \times N_{UE}}$ is the matrix of the resource allocation and N_{RB} and N_{UE} represent the available resource blocks and the number of users, respectively. $u_k(\bar{R}_k)$ represents the utility function of the allocation process based on the average allocated data rate to user k up to the current allocation round, \bar{R}_k . Based on the specifications of LTE this allocation system can further be described as follows:

- r is a matrix of binary values with each column defining the allocation of the available resources to the corresponding user, $r_k = [r_{k,1}, r_{k,2}, \dots, r_{k,N_{RB}}]^T \in \{0,1\}^{N_{RB}}$.
- Each resource block (180 KHz of bandwidth allocated for one time slot of 0.5ms) will be allocated to just one user, i.e. $r_i \cdot r_j^T = 0 \ \forall i \neq j$.
- All of the resources or a portion of them can be allocated to one or more users in each allocation round, i.e. $\sum_{k=1}^{N_{UE}} \|r_k\|_1 \leq N_{RB}$.
- The allocated data rate to user k in the allocation round i depends on the capacity C_k created by the user in the allocated resources r_k (i.e. $R_k^i = C_k^T \cdot r_k$). The Channel Quality Indicator feedback (CQI) defines this capacity as a function of the channel status, SINR, i.e. $C_k = C_k(CQI(SINR)) \in \mathbb{R}_{>0}^{1 \times N_{RB}}$. CQI actually represents the user’s receiver status (i.e. SINR) and the capability of the user’s device.
- The average allocated data rate to each user in (2) is the weighted average, \bar{R}_k^i , based on the user’s allocation in the current round, R_k^i , and the history of the allocation to that user, \bar{R}_k^{i-1} :

$$\bar{R}_k^i = \left(1 - \frac{1}{t_w}\right) \bar{R}_k^{i-1} + \frac{1}{t_w} R_k^i \quad (3)$$

where t_w is the average window size. The value of t_w must be large enough to filter out the fluctuation of the allocation process. Furthermore, system must be able to capture the average video code rate of the streaming service during the averaging period. A value of $t_w=100ms$ will suffice for LTE with 1ms allocation round and 10ms frame duration. More

detailed explanation of this system and the assumptions are provided in [11].

Assume that the proposed allocation scheme for video streaming users is working as a subroutine of the main allocation function of the system. It means that the desired level of the fairness and efficiency driven by the quality of the service can be applied through a portion of the total available resources, N_{RB}^v . Whenever $N_{RB}^v = N_{RB}$, all of the users are allocated through the same allocation function with entirely shared resources regardless of their corresponding data contents.

Let the quality of the streaming service be represented in the allocation function through its continuity aspect, Pause Intensity (PI), as follows:

$$\begin{cases} r^* = \arg r \max \sum_{k=1}^{N_{UE}^v} u_k(\bar{R}_k, PI) \\ u_k = PI_k^\alpha \cdot \bar{R}_k, \quad u_k \in \mathbb{R}_{\geq 0} \end{cases} \quad (4)$$

where N_{UE}^v is the number of the users with streaming service and u_k is the utility function which reflects the effect of the clients satisfaction (i.e. QoE) through its first term (i.e. PI^α). The first term can also be viewed as the weight for the second term, \bar{R}_k , which represents the efficiency of user for consuming the allocated resources. The value of α defines the desired trade-off between the efficiency and fairness. It further takes into account the correlation between the users' required data rates (i.e. video code rates) and the allocated data rates. A practical algorithm to implement (4) can be achieved by replacing the utility function in (4) with a priority function as follows:

$$\begin{cases} k^* = \arg k \max u_k \\ u_k = PI_k^\alpha \cdot \bar{R}_k = \left(1 - \frac{\eta_k}{\lambda_k}\right)^\alpha \cdot \bar{R}_k \end{cases} \quad (5)$$

TABLE I. LINK ADAPTATION AND MODULATION SCHEME

SINR	CQI	Modulation Order	Code Rate
≤ -6.934	1(*)	2	0.1523
-5.147	2	2	0.2344
-3.180	3	2	0.3770
-1.254	4	2	0.6016
0.7610	5	2	0.8770
2.700	6	2	1.1758
4.697	7	4	1.4766
6.528	8	4	1.9141
8.576	9	4	2.4063
10.37	10	6	2.7305
12.30	11	6	3.3223
14.18	12	6	3.9023
15.89	13	6	4.5234
17.82	14	6	5.1152
≥ 19.83	15	6	5.5547

* Users with SINR lower than a device related threshold will not be scheduled for using the resources

where u_k appears as a priority function and a user with the highest value of u_k will be chosen in each round of the allocation.

Since the video code rate and the throughput of the network are available at the user-side, the PI assessment procedure can be conducted directly as a client-side task. This value can be reported to the relevant network element for scheduling consideration. However, such a solution requires extra control information and transaction between user and network. This usually is not a practical approach and violates the transparency of the service especially from last mile's interface point of view. Alternatively, in a system such as LTE supported by its EPC and/or PSS infrastructures, the incoming video rate from the public network and the outgoing rate toward the users alongside their throughput can be estimated easily. In either way, the required value of PI in (4) and (5) is accessible at the network side.

For the user of a streaming service, the required data rate in the scheduler is proportional to its video code rate. The value of the video code rate in a client-driven streaming technique (e.g. 3GP-DASH) is changing adaptively to match the status of the channel, the available buffer and processing time. This obviously reveals the interaction between the client-side adaptation mechanism and the quality driven allocation scheme at the network-side (i.e. (4) and (5)). For the sake of consistency, the aspect of the rate adaptation mechanism which depends on the network performance, throughput (η), can be recast based on the user's quality metric, PI . Let us take $PI_i > PI_{threshold}$ as a simplified decision making process for the adapted rate of the $(i+1)^{th}$ segment of the streaming data. PI_i represents the assessed value of PI and $PI_{threshold}$ represents the maximum acceptable discontinuity of the service.

Based on the PI definition, this can alternately be shown as $\left(1 - \frac{\eta_i}{\lambda_i}\right) > PI_{threshold}$ which yields the general form of the criterion based on the network performance, η , and video code rate, λ , as $\eta_i < (1 - PI_{threshold})\lambda_i$. We will use this condition as the criterion for increasing or decreasing the video code rate.

TABLE II. SIMULATION SETUP

Parameter	value
No. of Cells	1 (with the first tier interference)
Inter-site distance	2000 meters
Shadowing effect	mean=0, deviation=8 decorrelation distance=25m, inter-site correlation=0.5
Channel model	PedA, speed=3km/h
Bandwidth	5MHz, 20MHz
No. of RBs (per TimeSlot)	25, 100
Subcarrier	15KHz
Range of average SNR	-6 ~ 18 dB (CQI=1~15)
Average video code rate	156kbps ~ 1.5Mbps
No of Users	45
Each scheduling round	One TTI=1ms
Simulation time	10000*TTI (10 s)
Video stream model	Truncated Pareto for packet size and inter-arrival time

IV. SIMULATION RESULTS AND ANALYSIS

A. Simulation setup

A Matlab simulator is developed for examining the proposed algorithm in the context of LTE with the users of an adaptive video streaming service. The LTE system is represented by the required functions from RLC, MAC and PHY layers. These comprise the scheduling and link adaptation functions with the consideration of the standard modulation and code rates, as shown in Table I. This table actually defines the implemented mapping between SINR and CQI (feedback from users to the network) for the lower layer's settings related to the modulation order and channel code rate.

Simulator also takes account of the proper data segmentation function for video streaming data packets. The size of the packet is generated using a truncated Pareto model (for packets' inter-arrival-time and size). No background traffic is considered for the users' data. As it has been shown in Fig. 3, the video code rates initially start with the standard video streaming quality of 780kbps (as in BBC-iPlayer). These rates later increase or decrease (in the range of 156kbps-1500kbps) based on the users' channel status and allocated resources. Users' Head-of-Line packets (HOL) are scheduled in a timely manner with no packet drop due to the delay in the scheduler (i.e. eNodeB).

Table II provides more detail about the simulator and its channel model. The video code rates correspond to the users with different SINRs in the range of the defined CQI for LTE (i.e. 1~15). Users are distributed in one cell considering the interference from the first tier neighboring cells. A pedestrian user's model of fading and the shadowing effect (inter/intra-cell spatial correlation) have been taken into account as well. The subjective quality metric of the service, MOS, has been

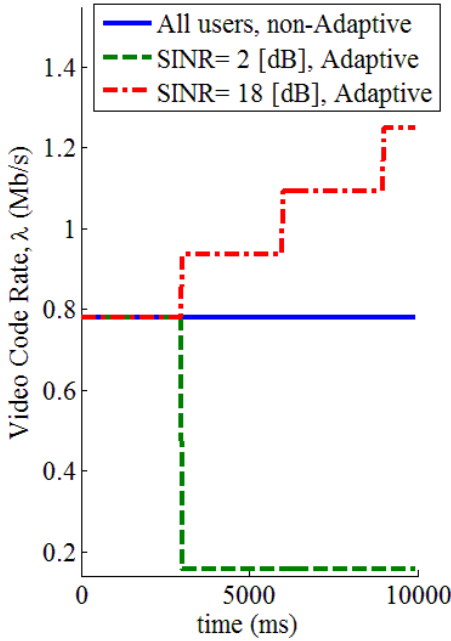


Fig. 3. video code rates in an adaptive video streaming compared to a non-adaptive service. The adapted rates are exemplified by two users with distinctive channel status.

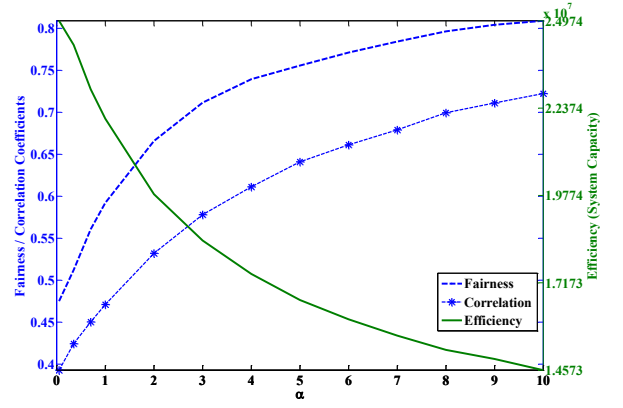


Fig. 4. The achieved trade-off between various aspects of the proposed algorithm vs. different values of α . These include the efficiency of the system, the correlation between required data rates and allocated data rates per user and the achievable fairness among allocated data rates to the users.

estimated through the fitting of the PI values. The relation between MOS and PI achieved from the subjective testing results in [6] has been used for the fitting purpose. MOS values are changing from 1 to 5 and represent the users' perceived quality as Bad, Poor, Fair, Good and Excellent, respectively.

B. Results and Analysis

The overall performance of the proposed algorithm with regards to the trade-off between efficiency and fairness is shown in Fig. 4. The evaluation of the efficiency and fairness is based on the total allocated data rates to the users and the Jain's index of these allocated data rates, respectively. It also depicts the achieved correlation between the required data rates of the users (i.e. their video code rates) and the allocated data rates by the system which is assessed through the Pearson's Linear Correlation Coefficient. These results reveal the main characteristic of the parametric algorithm in (4) and (5) with the capability for achieving a desired trade-off using a certain value of parameter α .

Fairness in the context of scheduling and rate allocation means a fair resource allocation regardless of the exact differences in the users' channel status (i.e. SINR) and device capabilities. However, this policy creates a converse effect on the users which tend to increase their video code rate in an adaptive streaming paradigm. Users with good channel status

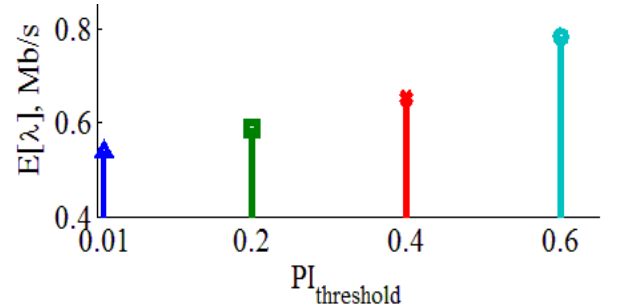


Fig. 5. average video code rate against the quality threshold. The Pause Intensity threshold ($PI_{\text{threshold}}$) represents the acceptable quality from continuity point of view while the average video code rate embodies the image quality/fidelity.

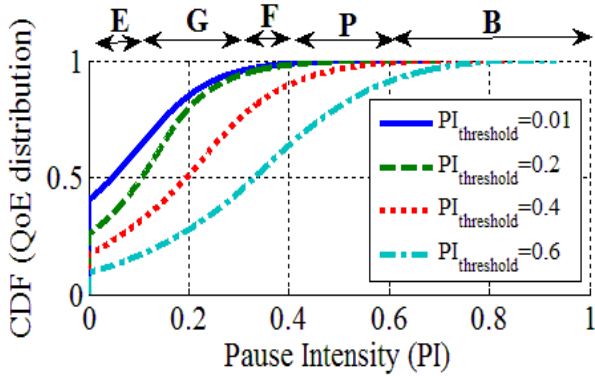


Fig. 6. the effect of different rate adaptation condition ($PI_{threshold}$) over the achieved QoE distribution. PI distribution actually represents the continuity aspect of the QoE. The subjective aspect of the quality is also depicted through the corresponding ranges of MOS values from E(Excellent) to B(Bad).

or with highly capable devices are forced to reduce their video code rate in spite of their capability. The correlation defined in the proposed algorithm is actually the amendment of the allocation in favor of the users with high demands and high capabilities.

As it has been explained in Section III, the video code rate adaptation process can be defined based on the acceptable continuity status (i.e. maximum acceptable Pause Intensity, $PI_{threshold}$). The result of the quality metric involvement in the user-side's rate adaption process is shown in Fig. 5. The average adapted video code rate throughout the simulation period is depicted as a function of the maximum acceptable Pause Intensity, $PI_{threshold}$. The implemented mechanism successfully maintains the adapted video code rates which are low enough to achieve an almost pause-less playback (i.e. $500kbps$ average video code rate for $PI_{threshold}=0.01$). In contrast, with a higher level of tolerable discontinuity for the service the implemented mechanism allows the users with good channel status and high device capability to increase their adapted video code rates (i.e. $800kbps$ average video code rate for $PI_{threshold}=0.6$).

The final distribution of QoE for different values of $PI_{threshold}$, in terms of the continuity, is shown in Fig. 6. Higher expectation for continuity when $PI_{threshold}=0.01$ has resulted in a distribution with more probable 'Excellent' quality. However, as it has been shown in Fig. 5 the average visual quality will be lower in this case. In contrast, service quality is less likely to be 'Excellent' or 'Good' in the case of the higher values of $PI_{threshold}$. The higher value of $PI_{threshold}$ degrades the continuity of the service though it increases the average visual quality (as it has been shown in Fig. 5).

V. CONCLUSION

A rate allocation method with an optimization algorithm for adaptive rate video streaming services in the LTE network has

been presented in this paper. The proposed method takes into account the continuity of the service as a base for rate allocation in the last mile's resource scheduler and for the video code rate adaptation process at the client-side. As demonstrated by the simulation results, this method can properly control the QoE distribution among the users in the network and maintain a desired balance between the efficiency and fairness of the resource allocator. Furthermore, the correlation between the requested rates by the user and the rates provided by the system is also shown to be maximized for the given continuity threshold.

This work provides a framework that integrates the resource optimization process with user's demand in QoE. It has demonstrated that the available network resources can be distributed in such a way where the demand and supply are best matched for users with diverse levels of satisfaction in service provision. And this is done through applying a quantifiable objective QoE metric in both scheduling and rate adaptation processes to collectively address the challenges in resource utilization and quality insurance.

REFERENCES

- [1] ISO/IEC 14496-10:2012 – Information technology — coding of audiovisual objects — Part 10: Advanced Video Coding". ISO. Retrieved 2012-06-11.
- [2] ITU-T, H.265, "High efficiency video coding", International Telecommunication Union, Apr. 2013.
- [3] Cisco Visual Networking Index: Forecast and Methodology, 2013–2018.
- [4] Conviva, 2014 viewer experience report: <http://www.conviva.com/vxr-home/>
- [5] 3GPP TS 26.234: "Transparent end-to-end packet switched streaming service (PSS); Protocols and codecs".
- [6] M. Seydebrahimi, C. Bailey, and X.-H. Peng, "Model and Performance of a No-Reference Quality Assessment Metric for Video Streaming," *IEEE Trans. on Circuits and Systems for Video Technology*, vol.23, no.12, (Dec. 2013), pp.2034-2043.
- [7] H. A. Tran; S. Hoceini, A. Mellouk, J. Perez, and S. Zeadally,, "QoE-Based Server Selection for Content Distribution Networks," *IEEE Trans. on Computers*, vol.63, no.11, pp.2803,2815, Nov. 2014.
- [8] 3GPP TS 26.247: "Transparent end-to-end Packet-switched Streaming Service (PSS); Progressive Download and Dynamic Adaptive Streaming over HTTP (3GP-DASH)".
- [9] S. Chen; B. Shen; S. Wee, and X. Zhang, "SProxy: A Caching Infrastructure to Support Internet Streaming," *IEEE Trans. on Multimedia*, vol.9, no.5, pp.1062-1072, Aug. 2007.
- [10] R.K.P. Mok, E.W.W. Chan, and R.K.C Chang, "Measuring the quality of experience of HTTP video streaming," *2011 IFIP/IEEE International Symposium on Integrated Network Management (IM)*, pp.485,492, 23-27 May 2011.
- [11] M. Seydebrahimi, X.-H. Peng, and R. Harrison, "A Quality Driven Framework for Adaptive Video Streaming in Mobile Wireless Networks," *2014 IEEE Wireless Communications and Networking Conference (WCNC)*, pp.2994-2999, 6-9 April 2014.