QoE-driven Live and On-demand LTE Uplink Video Transmission

Ali El Essaili #1, Liang Zhou #2, Damien Schroeder #3, Eckehard Steinbach #4, Wolfgang Kellerer *5

Institute for Media Technology, Technische Universität München (TUM)
Munich, Germany
1 elessaili, 2 liang.zhou, 4 eckehard.steinbach@tum.de
3 damien.schroeder@mytum.de

* *DOCOMO EURO-LABS, Ubiquitous Networking Research Group*
Munich, Germany
5 kellerer@docomolab-euro.com

***Abstract*—We consider the joint upstreaming of live and on-demand user-generated video content over LTE using a Quality-of-Experience driven approach. We contribute to the state-of-the-art work on multimedia scheduling in three aspects: 1) we jointly optimize the transmission of live and time-shifted video under scarce uplink resources by transmitting a basic quality in real-time and uploading a refined quality for on-demand consumption. 2) We propose a producer-consumer deadline-aware scheduling algorithm that incorporates both the physical state of the mobile producer (e.g., cache fullness) and the scheduled playout time at the end-user. 3) We show that the scheduling decisions in 1) and 2) can be determined locally for each mobile producer. We additionally present an analytical framework for de-centralized scalable video transmission and prove that there exists an optimal solution to our problem. Simulation results for LTE uplink further demonstrate the significance of our proposed optimization on the overall user experience.**

I. Introduction

The production and consumption of multimedia content have changed tremendously over the last few years with the advent of community portals and video sharing websites. Upstreaming of rich multimedia content is more realistic nowadays thanks to the enhanced data rates offered by mobile networks and the high quality video capturing capabilities of mobile devices. Consumers are equally driven towards on-demand consumption which could eventually replace the traditional static TV broadcast [1].

The increasing availability of large screen mobile devices further contributes to this trend. Moreover, video portal services such as YouTube and social web platforms that today serve mostly prerecorded videos will be enhanced with functionality for offering live video streams. Given such distribution channels the popularity of live video capturing and streaming with high quality and resolutions beyond today's phones is expected to challenge in particular the uplink channel of mobile data communication. Mobile networks will have to deal with this vast increase in user-generated content (UGC) given the quite limited resources in the wireless uplink [2].

This paper addresses the above issues by proposing a service-centric approach for uplink resource allocation among multiple mobile video producers. In particular, a video portal plays a central role for our solution by collecting feedback from consumers on their type of video delivery (live or on-demand) and playout time and providing this feedback to the video producers. The uplink resource allocation is optimized to provide video consumers with the best possible Quality of Experience (QoE) within the scheduled playout time taking the limited and time varying uplink resources into account. More specifically, we distinguish between the centralized multi-user resource allocation problem and the distributed optimization of the video content at the mobile terminal.

To the best of our knowledge, this paper is the first one to present a systematic resource allocation and transmission optimization approach for the simultaneous upstreaming of real-time and time-shifted user-generated video content. We target future mobile terminals that are capable of generating scalable video streams. We also assume that these devices can cache parts of the stream for later upload. For a practical mobile environment where the available uplink resources and the channel state of each mobile terminal vary over time, the gap between the total available resources and the required resources for live video transmission is used for the upload of previously cached on-demand data. In summary, the main contributions of our work are:

a) We propose a QoE-based approach for jointly optimizing the uplink transmission of live and on-demand videos.

b) We define deadlines from a consumer perspective (video request time) and a producer perspective (upload time), and include both constraints into our optimization problem.

c) We provide an analytical solution that finds a global optimum for a set of video layers with different deadline constraints.

The rest of the paper is organized as follows. In the next section, we first give an overview of our system and then provide a discussion of related work. In Section III, we present our optimality analysis for the distributed optimization part which is performed by all terminals individually. Section IV describes then our centralized resource allocation framework for LTE uplink. Section V presents the simulation results and Section VI concludes the paper.

978-1-4577-1434-4/11/$26.00 ©2011 IEEE

II. Service-centric Approach

A. System overview

Scalable video coding [3] decomposes a video stream into multiple video layers where the base layer (BL) provides a basic video quality and the remaining enhancement layers (EL) provide a refined video quality. We consider both live and time-shifted upstreaming of scalable video from a video producer to a video portal which acts as an intermediate node for both live consumption and archiving of video streams for on-demand retrieval.

We differentiate two classes of consumers: (1) videos are requested for live consumption acceptable at a low rate or resolution and for later on-demand consumption at a higher quality; (2) videos are requested for on-demand consumption only. The portal provides the video service to both classes of consumers. It makes the information about the live or non-live consumption available to the optimization of the video transmission. It is the goal of the optimization to provide the best possible resource allocation to the video producers while at the same time maximizing the user perceived quality of the consumers across the different classes of consumers described above.

A schematic overview of a system for uplink distribution of live and on-demand user-generated content from a set of video producers to a video portal is illustrated in Figure 1. A standardized feedback on the class of consumers (live or on-demand) is provided by the video portal to the entity performing the part of the optimization which is jointly done for all video streams of one cell (e.g., at the eNodeB). The network, respectively, base station is responsible for allocating the resources for uplink transmission among the video producers in one cell. Complementing the above central optimization entity, a distributed QoE-based optimization is performed by each video producer in the terminal to decide on which video layers to transmit and their respective rates.

For downlink-based services, a pure network-centric approach for optimization is possible as both the network including the base station and the content source are in front of the bottleneck wireless link. However, for uplink-based services the bottleneck is between the content source, i.e., the terminal and the network. Hence, a distributed optimization approach has to be considered involving entities in the network and in the terminal. More specifically, the base station has the information about all users competing for the resources and the information about the consumption type. The terminal can directly influence the source coding and packet transmission.

The benefit for the mobile network operator is a potentially substantial saving in used network resources or the ability to provide higher user satisfaction for a given set of uplink resources. Providing high quality services also to UGC is a challenging task. An optimum resource allocation is difficult to dimension as it purely depends on the user behavior starting to upstream a video. This is different from a server download use-case where some information about the popularity or importance of a video file is potentially available. In this

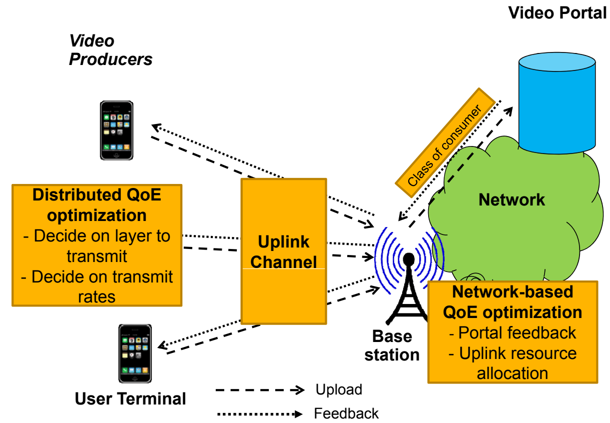

Fig. 1. System image for QoE-driven uplink resource optimization.

paper, we take the knowledge of consumption in terms of live versus video on-demand in order to prioritise the resources. The overall goal is traffic optimization without impacting the user perceived quality of service.

B. Related work

Recently, research on deadline-aware multimedia scheduling is drawing significant attention. A cross-layer packetization and retransmission approach for wireless video transmission under delay-constraints is proposed in [4]. More specifically, the cross-layer optimization problem is solved independently for all packets with a common decoding deadline using a general Lagrangian optimization formulation. While the authors do not consider the potential benefit of transmitting some packets with a later deadline before those with an earlier deadline, it is favorable in our case to schedule layers with a later deadline to guarantee a basic quality in real-time before previously cached enhancement layers with earlier deadline are transmitted for on-demand consumption. As a result, the optimization problem cannot be solved independently as proposed in [4]. In [5], a mathematical framework for channel, deadline and distortion aware scheduling is proposed. The work follows the principle of low-latency media system design where packets are associated with strict deadlines, but it does not further investigate the possibility of sending queued packets for time-shifted consumption. [6] proposes a multi-user gradient-based scheduling framework that considers the streaming of scalable videos with different playback deadlines, and presents a dynamic weighting metric for mitigating the approaching deadline effect. In [7], *Chakareski and Frossard* describe a Rate-Distortion (R-D) optimization framework for scheduling of multiple video streams by considering the importance of video packets. [8] also studies R-D optimized streaming of packetized media and proposes an iterative descent algorithm that determines the optimal transmission policy for each packet. Different from the works of [6], [7] and [8], we optimize the transmission of a video content with both a live and an on-demand delay constraint (i.e., two deadlines). Moreover, [9] describes some generic cases for scalable video coding in mobile environments in which scenarios for on-demand or live transmission of the same content to receivers are considered with different access bit rates and reception capabilities.

III. Distributed QoE Optimization

A. Analytical framework

Suppose that a video generated from a mobile terminal can be decomposed into several video layers (e.g., [3]). More precisely, we assume that a Group of Pictures (GoP) is encoded into a set of video layers with a common deadline (i.e., BL, EL,etc.). Please note that a GoP can be alternatively broken down into individual frames with different deadlines (e.g., [6]). At each scheduling round, we consider a new GoP (with playout deadline d_1) and the layers from previous GoPs which have been cached at a mobile terminal for on-demand upstreaming (with playout deadline $(d_2 - j \cdot t_0)$, $1 \leq j \leq n$, t_0 is the scheduling period). d_1 and d_2 represent the deadlines for live and on-demand video, respectively. If the video is requested for on-demand consumption only, then d_1 equals d_2. In general, we denote by S_L the set of all deadlines and by $S_{L,d}$ one set of video layers with a common deadline d. These sets are illustrated in Figure 2 for a scalable video stream encoded into four layers.

Specifically, we denote by S_k the set of users who have data to send. We define c_k as the instantaneous uplink transmission capacity, and H_k the cache size of user $k \in S_k$. For each video layer l, we use the following notations: L_l is the size, R_l is the instantaneous rate, MOS_l is the Mean Opinion Score (MOS) value and t_l is the required time to transmit the video layer.

The goal of our scheme is to maximize the QoE of all users by determining the set of layers to be scheduled for transmission at each round, and the rate of each scheduled layer. If layer l is scheduled, a_l is equal to 1 and 0 otherwise. $b(l) \in \{0, 1\}$, describes the layer dependencies, i.e., if layer l_i has a higher priority than layer l_j, then $b(l_j)$ is equal to 1 if layer l_i has been transmitted and 0 otherwise. ΔMOS_l denotes the additional improvement of the QoE by transmitting layer l. Our optimization problem is formulated as:

$$\underset{a_l, R_l, \forall S_L, \forall l \in S_L}{\arg\max} \quad \sum_{S_L} \sum_{l \in S_L} a_l \cdot b(l) \cdot \Delta MOS_l(R_l) \tag{1}$$

$$s.t. \quad \sum_{S_L} \sum_{l \in S_L} a_l R_l \leq c_k, \tag{2}$$

$$\sum_{S_L} \sum_{l \in S_L} L_l (1 - a_l) \leq H_k, \tag{3}$$

$$\sum_{l \in S_L} a_l t_l \leq d(S_L), \forall S_L, \tag{4}$$

where, (1) is the QoE-based objective function; (2) corresponds to the rate constraint where the sum of rates of all transmitted layers should not exceed the available capacity; (3) is a necessary condition to avoid cache overflow; the size of cached layers should not exceed the cache size; (4) is a necessary condition to transmit layers by their deadlines; the time to transmit all layers with one common deadline should be less than the deadline.

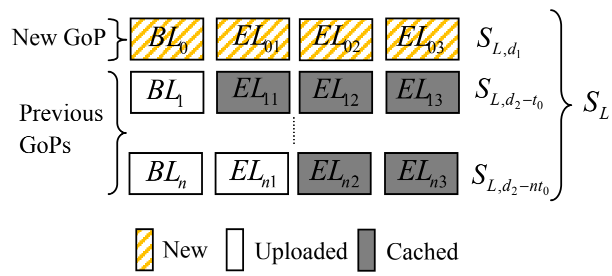

Fig. 2. A set of video layers that correspond to a new GoP and cached layers from previous scheduling rounds at a mobile terminal.

For a given set of resources c_k for user $k \in S_k$, the problem in (1)-(4) can be solved locally for each user. In the next section, we prove that the independent de-centralized optimization can lead to an optimal solution for each user.

B. Optimality analysis

Our optimality analysis is composed of three steps. First we show that the optimization problem in (1)-(4) can be transformed to a Lagrange dual problem. Then we show that the dual problem can be solved by a de-centralized iteration method. Finally we provide a possible way to guarantee that the proposed solution converges to the optimal value.

Step 1: We substitute R_l by $\frac{L_l}{t_l}$ in (1)-(4) and solve for the optimal values of a_l^*, t_l^* and L_l^*. From the definition of the objective function, we can find that (1) is a classical convex optimization problem satisfying Slater's condition [10]. In this case the dual gap is zero, and therefore we can design a decentralized algorithm to achieve the optimum value of (1).

Step 2: A distributed algorithm can be derived through a Lagrange dual problem:

$$\begin{aligned}&L\left(a_l, L_l, t_l, \phi_l, \theta_l, \delta_l\right) = \sum_{S_L} \sum_{l \in S_L} a_l b\left(l\right) \Delta MOS_l\left(\tfrac{L_l}{t_l}\right) \\ &- \sum_{S_L} \sum_{l \in S_L} \phi_l\left(\tau\right)\left(a_l L_l - c_k t_l\right) - \\ &\sum_{S_L} \sum_{l \in S_L} \theta_l\left(\tau\right)\left[L_l\left(1 - a_l\right) - H_k\right] - \sum_{l \in S_L} \delta_l\left(\tau\right)\left[a_l t_l - d\left(S_L\right)\right],\end{aligned} \tag{5}$$

where, ϕ_l, θ_l, and δ_l are the Lagrange multipliers associated with the constraints of (2), (3) and (4), and τ indicates the time index ($\tau \in \mathbb{N}$). Since the constraints of our optimization problem are independent of each other, then (5) satisfies the additivity property [11]. Therefore, we can further decompose the Lagrangian dual problem in (5) into three sub-problems:

$$\begin{aligned}\max: &\sum_{S_L} \sum_{l \in S_L} a_l b\left(l\right) \Delta MOS_l\left(\tfrac{L_l}{t_l}\right) - \sum_{S_L} \sum_{l \in S_L} \phi_l\left(\tau\right) a_l L_l \\ &+ \sum_{S_L} \sum_{l \in S_L} \theta_l\left(\tau\right) a_l L_l - \sum_{l \in S_L} \delta_l\left(\tau\right) a_l t_l.\end{aligned} \tag{6}$$

$$\max: \sum_{S_L} \sum_{l \in S_L} \phi_l\left(\tau\right) c_k t_l. \tag{7}$$

$$\max: -\sum_{S_L}\sum_{l\in S_L}\theta_l(\tau)L_l. \quad (8)$$

In particular, the Lagrange dual function $L_d(\phi,\theta,\delta)$ is defined as the maximum of $L(a_l,L_l,t_l,\phi_l,\theta_l,\delta_l)$ over a_l, L_l and t_l for given ϕ, θ, and δ. So, each video producer can compute an optimizer $a_l^*(\phi,\theta,\delta)$, $L_l^*(\phi,\theta,\delta)$, and $t_l^*(\phi,\theta,\delta)$. Hence, the Lagrange dual problem can be resolved in a de-centralized manner:

$$L_d(\phi_l,\theta_l,\delta_l) = L\left(\begin{array}{c} a_l^*(\phi_l,\theta_l,\delta_l), L_l^*(\phi_l,\theta_l,\delta_l), \\ t_l^*(\phi_l,\theta_l,\delta_l), \phi_l,\theta_l,\delta_l \end{array}\right), \quad (9)$$

where $(\phi_l,\theta_l,\delta_l)$ are the dual variables. Note that $L_d(\phi_l,\theta_l,\delta_l)$ is a convex optimization, and thus we can guarantee that the solution we obtain is the optimum one for (1).

Step 3: Since L_d may be non-differentiable, an iterative sub-gradient method can be used to solve it by introducing the variables ρ_ϕ, ρ_θ, and ρ_δ. Please note that L_d can be solved by a conventional KKT method [11], then we can get:

$$\phi_l(\tau+1) = \left[\phi_l(\tau) + \rho_\phi(\tau)\sum_{S_L}\sum_{l\in S_L} c_k t_l\right]^+, \quad (10)$$

$$\theta_l(\tau+1) = \left[\theta_l(\tau) - \rho_\theta(\tau)\sum_{S_L} L_l\right]^+. \quad (11)$$

In addition, we have $\frac{\partial(6)}{\partial a_l} = 0$ at the optimal solution,

$$\begin{array}{l} \sum\limits_{S_L}\sum\limits_{l\in S_L} b(l)\Delta MOS_l\left(\frac{L_l}{t_l}\right) - \sum\limits_{S_L}\sum\limits_{l\in S_L}\phi_l(\tau)L_l + \\ \sum\limits_{S_L}\sum\limits_{l\in S_L}\theta_l(\tau)L_l - \sum\limits_{l\in S_L}\delta_l(\tau)t_l = 0. \end{array} \quad (12)$$

Hence, we can find that $\delta_l(\tau)$ is a function related to t_l:

$$\delta_l(\tau+1) = \left[\delta_l(\tau) + \rho_\delta(\tau)\sum_{S_L}\sum_{l\in S_L}\left(\begin{array}{c} b(l)\,\Delta MOS_l\left(\frac{L_l}{t_l}\right) \\ -c_k t_l + L_l - t_l \end{array}\right)\right]^+ \quad (13)$$

where $\rho_\phi(\tau)$, $\rho_\theta(\tau)$ and $\rho_\delta(\tau)$ represent the iteration step size. Certain choices of step sizes, such as $\rho_\phi(\tau) = \frac{\rho_1}{\tau}$, $\rho_\theta(\tau) = \frac{\rho_2}{\tau}$, and $\rho_\delta(\tau) = \frac{\rho_3}{\tau}$, where $\rho_1,\rho_2,\rho_3 > 0$ guarantee that this algorithm will converge to the result obtained via joint optimization.

IV. Network-based QoE Optimization

A. LTE uplink model

In this paper, we consider the LTE uplink model originally proposed in [12]. The network-based uplink resource allocation problem is to assign the resources (physical resource blocks (PRBs) in LTE) to the different video producers such that the overall QoE is maximized. We consider long-term optimization periods in the order of seconds. This allows us to integrate our QoE-based optimization with any of the state-of-the art schedulers for LTE uplink without the need to modify the scheduling mechanisms already deployed. Our objective is thus to determine the average resource share (i.e., PRBs) of each user in each optimization round. The underlying uplink scheduler can assign the blocks in a contiguous manner.

We use the link layer model from the 3GPP LTE recommendations [13] to determine the achievable throughput per PRB for a given Signal-to-Noise ratio (γ). The model from [13] approximates the throughput in the uplink, after link adaptation and hybrid automatic repeat request, by an implementation loss $\beta = 0.4$ compared to the Shannon capacity (14). As baseline uplink parameters, it further defines a γ_{min} of -10 dB, a γ_{max} of 15 dB and a maximum throughput Thr_{max} of 2 bps/Hz.

$$Thr = \begin{cases} 0 & \text{for } \gamma < \gamma_{min} \\ \beta\log_2(1+\gamma) & \text{for } \gamma_{min} < \gamma < \gamma_{max} \\ Thr_{max} & \text{for } \gamma > \gamma_{max} \end{cases} \quad (14)$$

The network-based QoE optimization shall adapt to the video transmission of mobile terminals on timescales of seconds. As a result, we do not consider small-scale channel effects but rather a long-term channel quality indicator (CQI) update for each mobile user. At each optimization cycle, we generate a new channel realization that follows the typical CQI distribution of the users in an urban macrocell [14]. It provides an estimate of the average uplink CQI for each user which is sufficient for our link layer model. For a summary of the LTE parameters, please refer to Table I.

B. Uplink resource allocation

To determine the uplink resources (i.e., c_k in (2)) provided to each mobile terminal, we use the long-term abstraction model proposed in [15]. The model defines the actual data rate c_k for user k as a function of its resource share α_k and its maximum achievable rate $C_{max,k}$ if all the PRBs are allocated exclusively to user k, cf. (15).

$$c_k = f_k(\alpha_k) = \alpha_k C_{max,k} \quad 0 \le \alpha_k \le 1, \forall k \quad (15)$$

The objective of the uplink resource allocation is to determine the resource share α_k of each user that maximizes the overall QoE. We denote by U_k, the utility function of user k, as the resulting distributed QoE-based objective function in (1) at the mobile terminal for a given rate constraint c_k. The network-based objective function, that maximizes the sum of utilities of K users, can then be described by:

$$\underset{(\alpha_1,\ldots,\alpha_K)}{\arg\max}\sum_{k=1}^{K} U_k(c_k) \quad \text{subject to} \sum_{k=1}^{K}\alpha_k = 1 \quad (16)$$

Each α_k value corresponds to the fraction of total PRBs assigned to user k. At each optimization round, a new $C_{max,k}$ is estimated using (14) based on the average CQI of user k and passed to the uplink resource allocation function. A greedy resource allocation algorithm, similar to the work in [16], is developed to determine the value of α_k. It is initialized by assigning an equal amount of resources to every user. The algorithm iteratively takes a small amount of resources from the user who is less sensitive to the decrease in resources and

TABLE I
LTE PARAMETERS

Parameter	Value
Carrier frequency	2 GHz
System bandwidth	5 MHz
Number of PRBs	25
Number of subcarriers	300
PRB size	12 subcarriers
Subcarrier spacing	15 KHz
Bandwidth per PRB	180 KHz
Link layer model [13]	see (14)
Channel model	Urban macrocell [14]
CQI averaging cycle	250 ms

assigns it to the user who gets the maximum benefit, until no further improvement in (16) is possible.

Let ΔU_k denote the change of utility for user k due to a change of its resource share $\Delta\alpha$. The greedy algorithm can be expressed as an iterative maximization of the incremental utility values of two users i and j, such that:

$$i = \arg\max_k \quad \{\Delta U_k | \alpha_k \leftarrow \alpha_k + \Delta\alpha\} \tag{17}$$

$$j = \arg\min_k \quad \{\Delta U_k | \alpha_k \leftarrow \alpha_k - \Delta\alpha\} \tag{18}$$

The greedy allocation determines the best overall user experience in a cell by iteratively exchanging the $\Delta\alpha$ and ΔU_k values between the base station and the mobile terminals, respectively. Please note that this provides an upper limit on the system performance if network-based QoE optimization is considered. The uplink resource allocation, however, can be decoupled from the distributed QoE optimization at the mobile terminal. In this case, the actual data rate c_k can be directly determined by the deployed scheduler at the base station, for instance. The optimization problem in (1)-(4) is still solved independently by each mobile terminal.

V. Simulation Results

In our simulation, we consider three users upstreaming the *Football*, *Bus* and *Soccer* test video sequences with CIF resolution, 30 frames/sec and a GoP size of 8 frames. Each sequence is encoded with the H.264 scalable video codec [3] into four video layers. For simplicity, we use a fixed transmission rate for each layer. We assume a linear mapping between Peak Signal to Noise Ratio (PSNR) and MOS [17]. MOS can take on any value between 1.0 (25 dB) and 4.5 (40 dB), which represent the worst and best QoE, respectively. We consider a simulation run of 10 sec and a relative cache size of 0.3. In other words, a user can cache 30% of its video without dropping any layer. The simulation results are based on 25 simulation runs.

We optimize the video transmission for two scenarios: 1) Live + time-shifted: the uploaded videos are optimized for both live and on-demand consumption. 2) Video-on-Demand (VoD): videos are intended for on-demand consumption only. In both scenarios, we assume a VoD delay of 10 sec for all users. The deadline for live video in scenario 1) is 1 GoP (i.e.,

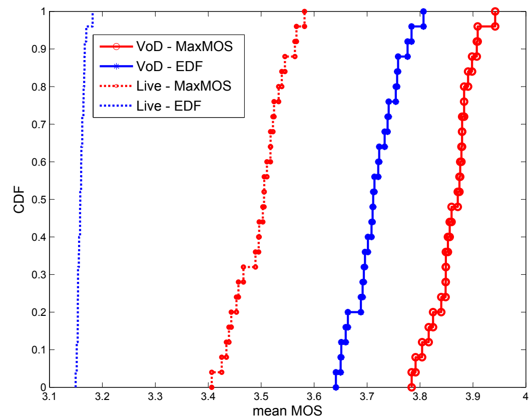

Fig. 3. CDF of the mean MOS for all users in scenario 1 (live + on-demand video).

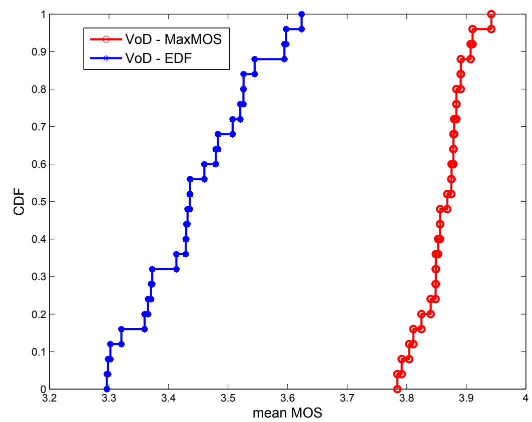

Fig. 4. CDF of the mean MOS for all users in scenario 2 (on-demand video only).

266 ms). To evaluate our proposed approach, we distinguish the uplink resource allocation at the eNodeB (i.e., to find c_k) and the distributed optimization at the mobile terminals (i.e., (1)-(4)). In the following results, we provide an upper limit on the achievable QoE if optimal resource allocation can be made by eNodeB among the three users. More specifically, we consider two schemes:

- Max-MOS: eNodeB is responsible for multi-user resource allocation. For an allocated bit rate, each user performs a de-centralized video layer optimization with the objective of transmitting layers which provide the best MOS improvement. The eNodeB assigns the resources in a greedy way until the sum of QoEs of all users is maximized.
- Priority-aware Earliest Deadline First (EDF): eNodeB serves the user with the earliest deadline. Priority-aware means that each user knows about the video dependencies and orders its layers accordingly. Each user performs a de-centralized video layer optimization with the objective of transmitting the layers with the most imminent deadline first.

Figure 3 shows the cumulative distribution function (CDF) of the mean MOS of all users for the first scenario. The base layer is always sent in real-time and the enhancement layers are transmitted according to the optimization criteria for both schemes. Comparing the dotted curves (live) with the solid curves (VoD) for both schemes, a substantial gain can be achieved by uploading additional cached video layers before playout. In addition, the Max-MOS scheme improves the mean MOS compared to the EDF-based one as it selectively transmits layers which provide highest MOS improvement first. In fact, both schemes should converge to a maximum mean MOS, if the videos are requested after some sufficient time and enough cache is available for layers which are not sent in real-time.

Figure 4 shows the CDF of the mean MOS of all users for the second scenario. More specifically, it shows the average MOS if the videos are requested directly after the uplink transmission stops (VoD delay of 10 sec). Different from the first scenario, the uplink transmission of both base layers and enhancement layers is optimized for on-demand consumption for both schemes. The EDF-based scheme, which transmits video layers according to the earliest deadline criteria, will fail to provide a sustainable video quality for the whole session which results in a reduced average video quality. By comparing with the Max-MOS scheme, a significant gain can be achieved if the video layers are transmitted according to the MOS criteria.

VI. Conclusions

This paper proposes a service-centric approach for uplink resource allocation among multiple video producers based on consumer feedback on the desired playout time available through a video portal. First, we present an analytical framework for scalable video transmission at a mobile terminal which simultaneously maximizes the QoE of different classes of consumers. Next, we validate our proposed approach both theoretically and practically for the uplink of an LTE system. We prove that the de-centralized optimization can lead to an optimal solution for each video producer. Furthermore, we provide by simulation an upper bound on the achievable QoE with optimal resource allocation.

In our future work, we plan to investigate low-complexity approaches for implementing our distributed QoE optimization at the mobile terminal side. We will also study de-centralized approaches for the network-based QoE optimization to reduce exchange of information between the terminals and the base station and thus reduce the overall optimization complexity.

References

[1] F. Gabin, M. Kampmann, T. Lohmar, and C. Priddle, "3GPP mobile multimedia streaming standards," *IEEE Signal Processing Magazine*, vol. 27, no. 6, pp. 134–138, Nov. 2010.

[2] O. Oyman, J. Foerster, T. Yong-joo, and L. Seong-Choon, "Toward enhanced mobile video services over WiMAX and LTE," *IEEE Communications Magazine*, vol. 48, no. 8, pp. 68–76, Aug. 2010.

[3] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1103–1120, Sept. 2007.

[4] M. Van der Schaar and D. Turaga, "Cross-layer packetization and retransmission strategies for delay-sensitive wireless multimedia transmission," *IEEE Transactions on Multimedia*, vol. 9, no. 1, pp. 185–197, Jan. 2007.

[5] A. Dua, C. Chan, N. Bambos, and J. Apostolopoulos, "Channel, deadline, and distortion (CD2) aware scheduling for video streams over wireless," *IEEE Transactions on Wireless Communications*, vol. 9, no. 3, pp. 1001–1011, March 2010.

[6] X. Ji, J. Huang, M. Chiang, G. Lafruit, and F. Catthoor, "Scheduling and resource allocation for SVC streaming over OFDM downlink systems," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, no. 10, pp. 1549–1555, Oct. 2009.

[7] J. Chakareski and P. Frossard, "Rate-distortion optimized distributed packet scheduling of multiple video streams over shared communication resources," *IEEE Transactions on Multimedia*, vol. 8, no. 2, pp. 1011–1020, Apr. 2006.

[8] P. Chou and Z. Miao, "Rate-distortion optimized streaming of packetized media," *IEEE Transactions on Multimedia*, vol. 8, no. 2, pp. 390– 404, April 2006.

[9] T. Schierl, T. Stockhammer, and T. Wiegand, "Mobile video transmission using scalable video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1204–1217, Sept. 2007.

[10] J. He, M. Bresler, M. Chiang, and J. Rexford, "Towards robust multi-layer traffic engineering: Optimization of congestion control and routing," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 5, pp. 868–880, June 2007.

[11] M. Chiang, "Balancing transport and physical layers in wireless multihop networks: jointly optimal congestion control and power control," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 1, pp. 104–116, Jan. 2005.

[12] A. El Essaili, E. Steinbach, D. Munarretto, S. Thakolsri, and W. Kellerer, "QoE-driven resource optimization for user generated video content in next generation mobile networks," *Proc. IEEE International Conference on Image Processing, ICIP 2011, Brussels, Belgium, accepted for publication*, Sept. 2011.

[13] 3GPP TR 36.942 V.8.1.0, "Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Frequency (RF) system scenarios," Dec. 2008.

[14] 3GPP TR 36.814 V.9.0.0, "Evolved Universal Terrestrial Radio Access (E-UTRA); Further advancements for E-UTRA physical layer aspects," March 2010.

[15] A. Saul, S. Khan, G. Auer, W. Kellerer, and E. Steinbach, "Cross-layer optimization with model-based parameter exchange," *IEEE International Conference on Communications, ICC 2007, Glasgow, Scotland*, June 2007.

[16] D. Jurca and P. Frossard, "Media flow rate allocation in multipath networks," *IEEE Transactions on Multimedia*, vol. 9, no. 6, pp. 1227 –1240, Oct. 2007.

[17] S. Khan, S. Duhovnikov, E. Steinbach, and W. Kellerer, "MOS-based multiuser multiapplication cross-layer optimization for mobile multimedia communication," *Advances in Multimedia, article ID 94918*, Jan. 2007.