# Content and Buffer Aware Scheduling for Video Delivery over LTE


Ahmed Ahmedin, Kartik Pandit, and Dipak Ghosal
University of California, Davis, CA 95616
{ahmedin, kdpandit, dghosal}@ucdavis.edu

Amitabha Ghosh
UtopiaCompression Corporation
Los Angeles, CA 90064
amitabhg@utopiacompression.com



## ABSTRACT

We propose a content and buffer aware Long Term Evolution (LTE) downlink scheduler for allocating radio resources to multiple users requesting Scalable Video Coding (SVC) encoded videos from a server. The scheduler adaptively controls the video quality by assigning physical resource blocks (PRBs) to the users based on the requested video type, device capability, channel quality, and buffer state in the user equipment (UE). The scheduler maximizes the average video quality across all users for a limited number of PRBs, and can be installed in existing LTE systems without additional equipment and very limited signaling overhead.



## Categories and Subject Descriptors

C.2 [**Computer-Communication Networks**]: Wireless communication

## Keywords

LTE downlink; video delivery; content and buffer-awareness


## 1. INTRODUCTION

The evolution of smart mobile devices is driving cellular service providers to introduce more effective techniques and bring high quality services to end users. Multimedia delivery is one of the most rapidly growing services as next-generation cellular technologies, such as LTE and LTE Advanced, continue to offer increasingly higher data rates. It is expected that 70% of the cellular data traffic will be from video by 2016 [1]. As a result, the demand for video traffic is encouraging service providers to rapidly deploy LTE and provide high data rates in both uplink and downlink [2].

LTE is based on orthogonal frequency division multiple access (OFDMA), in which time is divided into LTE downlink frames each of duration 10 milliseconds (ms). Each frame is further subdivided into Transmission Time Intervals (TTIs) with duration 1 ms each. The minimum unit of assignment for a UE is one PRB, consisting of one 0.5 ms time slot and one sub-channel of bandwidth 180 KHz. This enables fine-grained scheduling and radio resource control.

The medium access control layer of an LTE scheduler operates at the TTI level, taking into account the nature of traffic, such as delay tolerance, error resilience, and guaranteed bit rate, via the parameter called quality of service class indicator (QCI). This, however, does not take into account *content characteristics*. For example, streaming a stored video using HTTP at the application layer and TCP at the transport layer may have the same QCI class as some other more elastic applications over TCP. A content aware scheduler in this case will attempt to assign resources to the UEs based on both video characteristics and channel quality.

A number of prior studies have demonstrated the benefits of content aware network optimization [3–7]. In a wireless environment, the link quality can vary drastically, requiring more PRBs for some users to compensate for low bit rates, while fewer for others that have high quality links to guarantee a smooth video playback. We argue that a content aware scheduler is more adaptive to opportunistically utilize the unused PRBs to accommodate new users, or increase the video quality of existing users. A popular standard for video streaming is SVC, which takes into account the capabilities of different types of UEs, ranging from cell phones to more powerful tablets and smart TVs. SVC also provides temporal, spatial, and quality scalability [8] to mitigate the effects of random channel variation, multipath fading, and limited bandwidth faced by different users.

Besides being content aware, we argue that knowing the *state* of the UE buffer enables the scheduler to be more adaptive to network changes. With LTE Advanced, the peak data rates can be very high. As a result, a UE can receive packets at a much higher rate than the actual video playback rate over short periods of time, causing the application layer video buffer to fill up very quickly and, eventually, overflow. If the scheduler knows the UE buffer state, it can reassign the PRBs to other users to increase the network capacity, or enhance the overall video quality. In this paper, we present our ongoing work in designing an adaptive LTE scheduler to optimize radio resources and guarantee high quality video to end users. We propose an extension to our previous work [6] to be *buffer aware*, in addition to being content aware. We discuss the trade-off between requesting high quality videos and buffer overflow/underflow problems.

## 2. SYSTEM MODEL

We consider the LTE downlink of a single eNodeB where multiple users request SVC-encoded videos from a video



*CoNEXT Student Workshop'13,* December 9, 2013, Santa Barbara, CA, USA.



http://dx.doi.org/10.1145/2537148.2537158.

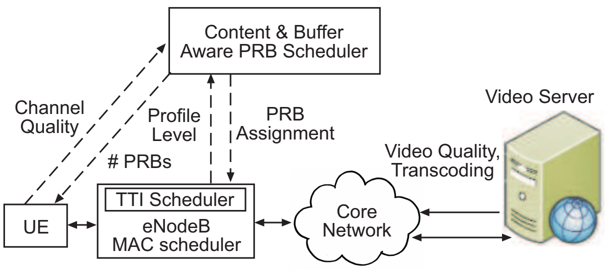


Figure 1: A buffer and content aware scheduler architecture for video delivery over LTE downlink.

server (e.g., NetFlix, YouTube). A schematic diagram of this architecture is shown in Figure 1. Solid lines indicate signaling that already exist between different nodes in the LTE evolved packet system (EPS) core network, whereas dotted lines indicate our new conceptual signaling mechanism [6].

Suppose the eNodeB has $M$ available PRBs to assign to a total of $N$ users, with each PRB having a fixed bandwidth $B$. Each user $i$ has a maximum buffer size $b_{max}$, and can request video packets encoded at a certain SVC enhancement layer. We assume that user $i$ can decode up to a set $L_i = \{l_{ij}\}$ of video profile levels, and that each level $l_{ij}$ requires a certain number, $\alpha_{ij}$, of PRBs depending on the link quality for smooth video playback without buffer underflow. By watching this playback, suppose user $i$ derives a utility $Q_i(l_{ij})$ indicating the video quality. In our evaluation, we quantify the video quality using two full reference metrics: peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) index.

For a given video profile level, a low quality link requires more number of PRBs to achieve the same video quality as compared to a high quality link. However, over-allocating PRBs can lead to the eNodeB running out of resources, or buffer overflow at the UE. This is a substantial overhead, as those unused PRBs can be assigned to other users for increasing their video quality, or admit new users. On the other hand, users always need to request enough number of packets to avoid buffer underflow that may cause video freezing and playback disruptions. This requires continuously monitoring the buffer states and carefully balancing the PRB allocations based on link qualities. The downlink rate required to receive a video at profile level $l_{ij}$ can be computed as [9]:

$$R_i(l_{ij}) = \alpha_{ij} m_i T_i B \log_2 \left(1 + \frac{P g_i}{n_0}\right), \quad (1)$$

where $P$ is the transmit power of the eNodeB; $g_i$ is the channel gain from the eNodeB to user $i$; and $n_0$ is the noise power. We assume that each user $i$ uses a forward error correction (FEC) code for protection, with coding rate $T_i$ and modulation scheme $m_i$, and that the channel gain $g_i$ can be estimated from channel quality indicator (CQI) measurements. We assume that there is a monotonic, one-to-one relationship between the utility $Q_i(l_{ij})$ and the corresponding rate $R_i(l_{ij})$.

In our proposed mechanism, at the beginning of each session, user $i$ reports its maximum buffer size, channel quality, and decodable profile levels with the required rates. At any given point in time, suppose the amount of data stored in the buffer for user $i$ is $b_i$, which is initialized to zero at the beginning of each session.

## 3. PRB ASSIGNMENT PROBLEM

The PRB scheduler decides the profile level and the number of PRBs to be assigned to each user based on the UE's decoding capability, its buffer state, available number of PRBs, and link quality. The goal is to maximize the average video quality across all users, which we formulate as an integer linear program (ILP). We propose that the eNodeB assigns the PRBs only to those users (say, $N_{<b_{th}}$ of them) who have buffered data less than a certain threshold, $b_{th} = \mu b_{max}$ ($\mu$ typically set to 0.75), and also admits $N'$ new users so long as the existing users are satisfied. In other words, additional PRBs are not assigned to a user if it has enough buffered data to playout. The eNodeB updates the UE buffer state every frame based on hybrid automatic repeat request (HARQ) signaling. The ILP can be written as:

$$\begin{aligned}
\text{maximize} \quad & \sum_{i \in (N_{<b_{th}} \cup N')} \sum_{l_{ij} \in L_i} x_{ij} Q_i(l_{ij}) \\
\text{subject to} \quad & \sum_{i \in N_{<b_{th}} \cup N'} \sum_{l_{ij} \in L_i} x_{ij} \alpha_{ij} \leq M \\
& \sum_{l_{ij} \in L_i} x_{ij} = 1, \quad \forall i \in N_{<b_{th}} \cup N' \\
\text{variables} \quad & x_{ij} \in \{0,1\}, \quad \forall i \in N_{<b_{th}} \cup N', \ \forall l_{ij} \in L_i \\
& N' \in \mathbb{N}_0,
\end{aligned}$$

where $x_{ij}$ is a decision variable that is 1 if profile level $l_{ij}$ is assigned to user $i$, and 0 otherwise; $\mathbb{N}_0$ is the set of natural numbers including zero. The first constraint ensures that the number of PRBs assigned does not exceed the total number of available PRBs, and the second constraint chooses exactly one profile level for each user. This problem is NP-hard because of the integer variables $x_{ij}$ and $N'$.

## 4. SOLUTION APPROACH

We propose to solve the ILP using a multiple choice Knapsack approach similar to [6], with new techniques to determine the set $N_{<b_{th}}$, and the number $N'$ of new users that may be admitted in the system without sacrificing the quality for existing users. We propose that in each optimization cycle, the scheduler reduces the number of assigned PRBs one at a time for users in the set $N \setminus N_{<b_{th}}$, so that TCP/IP does not suffer from severe congestion. The scheduler also continuously updates the buffer states for all users, and, if the buffered data is less than $b_{th}$, will add those users in the set $N_{<b_{th}}$.

## 5. CONCLUSION AND ON-GOING WORK

We propose a new LTE scheduler to opportunistically assign PRBs for video streaming applications. Our goal is to maximize the average video quality based on content characteristics, buffer states, and link quality. We are currently studying the best way to update the buffer states at the PRB scheduler, and also how to change the PRB assignment for users with enough buffered data. In addition, we are developing a way to consider the eNodeB buffer in the scheduling process, as well as implementing a complete system simulation in ns-3.

## 6. REFERENCES


[1] Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2012-2017. Technical report, 2012.

[2] Driving LTE Adoption: Mass-Market Pricing and a Large Device Ecosystem Emerge as Key Factors. Technical report, 2013.

[3] E. Haghani, N. Ansari, S. Parekh, and D. Colin. Traffic-Aware Video Streaming in Broadband Wireless Networks. In *IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1–6, 2010.

[4] K. Pandit, A. Ghosh, D. Ghosal, and M. Chiang. Content Aware Optimization for Video Delivery over WCDMA. *EURASIP Journal on Wireless Communications and Networking*, 2012(1):1–14, 2012.

[5] X. Zhu and B. Girod. Distributed Media-Aware Rate Allocation for Wireless Video Streaming. *IEEE Transactions on Circuits and Systems for Video Technology*, 20(11):1462–1474, 2010.

[6] A. Ahmedin, K. Pandit, D. Ghosal, and A. Ghosh. Exploiting Scalable Video Coding for Content Aware Downlink Video Delivery over LTE. In *International Conference on Distributed Computing and Networking (ICDCN), 2014 (accepted)*.

[7] S. Sharangi, R. Krishnamurti, and M. Hefeeda. Energy-Efficient Multicasting of Scalable Video Streams over WiMAX Networks. *IEEE Transactions on Multimedia*, 13(1):102–115, 2011.

[8] 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*, 17(9):1103–1120, 2007.

[9] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, 1991.