Anticipatory Buffer Control and Resource Allocation for Wireless Video Streaming

Sanam Sadr∗† and Stefan Valentin†

∗Department of Electrical and Computer Engineering, University of Toronto, Canada
†Bell Labs, Alcatel-Lucent, Germany

ssadr@comm.utoronto.ca, stefan.valentin@alcatel-lucent.com

Abstract—This paper describes a new approach for allocating resources to video streaming traffic. Assuming that the future channel state can be predicted for a certain time, we minimize the fraction of the bandwidth consumed for smooth streaming by jointly allocating wireless channel resources and play-out buffer size. To formalize this idea, we introduce a new model to capture the dynamic of a video streaming buffer and the allocated spectrum in an optimization problem. The result is a Linear Program that allows to trade off buffer size and allocated bandwidth. Based on this tractable model, our simulation results show that anticipating poor channel states and pre-loading the buffer accordingly allows to serve more users at perfect video quality.

I. INTRODUCTION

Recent studies show that video streaming is the dominating traffic in the mobile Internet [1]. This enormous traffic results from streaming news, entertainment clips and live events to mobile users. Video applications demand higher data rate and are delay sensitive based on the type of the video. Non-real time streams have the properties of known duration and data rate and can be buffered. The main objective is to provide a smooth video stream by ensuring that the video play-out bit rate to prevent outage in the near future. Due to the obvious trade-off between the instantaneous video play-out bit rate to prevent outage in the near future. Due to the obvious trade-off between the instantaneous and the future channel condition. The rational choice of the proposed scheme is as follows: as the user moves away from the serving base station (BS), the channel state decreases. The channel state hence the video quality will improve if the user performs a hand-over to the adjacent cell with enough resources. However, outage will be inevitable if the next cell is overloaded. There are two ways to address this issue, if the serving base station has knowledge of the desired future channel condition based on the video user’s trajectory: (a) delaying the hand-over or (b) pre-loading the buffer. Both essentially suggest to allocate more resources from the current BS to the video user. However, delaying the hand-over means allocating resources to the user while it is farther away from the BS resulting in higher transmit power for the same level of modulation and lower spectral efficiency. Therefore, we exploit prediction instead by pre-loading the buffer under good channel condition when a low channel gain is anticipated. Consuming the buffer then assures smooth video streaming until the improved channel state provides a stable filling rate again.

The challenge in doing so arises from three main factors: 1) scarce resources from the serving BS; 2) the user’s tendency to switch to another video without watching the pre-loaded one; 3) the user experiencing undesired channel condition or traversing an overloaded cell. The first two dictate limiting the allocated rate to the minimum video rate requirement while the third factor obligates the BS to allocate more than instantaneous video play-out bit rate to prevent outage in the near future. Due to the obvious trade-off between the mentioned factors, the technical question we attempt to solve is: how much pre-loading is enough to guarantee a smooth video?

The contributions of this paper are as follows: first, we propose a model for the buffer of the video user to keep track of the allocated, played and remaining bits in the buffer while meeting the video bit rate constraint. Second, we use this model to formulate an optimization problem to adjust the video user’s requested rate per time slot to have enough buffered video on the one hand while consuming the minimum spectrum on the other. After summarizing the related work in Section II, we describe the proposed buffer control and resource allocation scheme in Section III. The simulation results of the proposed scheme are presented in Section IV. The paper is concluded in Section V.

II. RELATED WORK

There is a large body of work on techniques for adaptive video streaming. At the application layer, channel-aware prefetching schemes invoke a traffic burst at high channel gain [3], [4]. At the link layer, cross-layer schedulers to jointly adapt video quality and wireless resource allocation have been proposed [5], [6].
Compared to this work, our approach differs two-fold. First, it does not adapt the video quality but adjusts the size of the video play-out buffer along with the wireless data rate. Unlike [5] and [6], this enables to trade off buffer size with the amount of allocated resources. Second, our adaptation scheme is anticipative. Unlike any of the above approaches, our joint buffer-rate allocation is based on a prediction of the user’s average channel gain. Using this prediction enables our scheme to plan its resource allocation ahead. This enables to compensate for upcoming channel outages (e.g., when a user drives through a tunnel) by pre-loading the video buffer in advance. Although this idea of anticipation has been applied for software interfaces [7] and cognitive radios [8], it has not been proposed for media streaming so far.

Further benefits of our work are its generality and low computational complexity. Being based on bit rates, our buffer model captures arbitrary video and audio codecs, while being independent of subjective quality metrics. This level of tractability is not provided by the Utility-based formulation in [6]. Finally, our buffer model enables to incorporate the dynamics of streaming media traffic into a Linear Program. Such low complexity is not provided by the combinatorial approach in [5].

III. ANTICIPATORY BUFFER CONTROL AND RESOURCE ALLOCATION

The proposed approach is based on the fundamental assumption that the average channel gain is predicted and known for the video user over the next $T$ time slots which is referred to as the look-ahead window. The optimization problem is formulated in the context of orthogonal frequency division multiple access (OFDMA) scheme as in Long Term Evolution (LTE) standard. There are $N$ subchannels - physical resource blocks (PRBs) in LTE - available in the system each with a bandwidth of $B$. We assume a cellular communication system with the BS at the centre of each cell. A frequency reuse of 1 is assumed among the cells resulting in an average interference power within each cell. The video user requires an average data rate of $R$ bits/s. For convenience, the rate achieved on a subchannel is assumed to be given by the Shannon capacity. The gap function, however, is used to modify the effective power within each cell. The video user requires an average of $d$ bits per time slot and $z_t$ represents the bits already in the buffer carried over from the previous time slot to the current one. The video user requires an average of $N$ bits per time slot in the play-out buffer to smoothly play the video. The total number of bits in the buffer at any time slot is therefore $r_t + z_t$. Outage occurs if this number is less than $V$; otherwise, each time slot can be of arbitrary duration with the required video bit per time slot proportioned accordingly.

![Buffer State](image)

Fig. 1. Buffer model: at each time slot, the allocated and the carry-over bits are adjusted to have a sum no less than the required video bit rate to avoid outage. The carry-over is limited by the maximum buffer size while the allocated rate is constrained by the link budget.

$V$ bits are played and the excess, $z_{t+1} = \lceil (r_t + z_t) - V \rceil^+$, will be carried over to the next time slot. $Z$ is the maximum number of carried over bits which indirectly limits the length of the buffered video. The buffer has only $r_1$ bits in the first time slot since $z_1 = 0$. In this model, the outage is avoided if:

$$\begin{cases} r_1 - z_2 = V, \\ r_t + z_t - z_{t+1} = V, \quad t = 2, 3, ..., T - 1 \\ r_T + z_T = V \end{cases}$$

which in the matrix form can be written as:

$$Ax - V = 0,$$

where $x = [r_1, r_2, \ldots, r_T, z_2, \ldots, z_T]^T$, $A$ is the $T \times (2T - 1)$ matrix given by:

$$A = \begin{bmatrix} I & B \end{bmatrix},$$

$I$ is the identity matrix of the order $T$ and $B$ is given by:

$$B = \begin{bmatrix} -1 & 0 & 0 & \cdots & 0 \\ 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}.$$
minimum load on the system is formulated as:

$$\min_{x_k} \sum_{k=1}^{K} \sum_{t=1}^{T} w_{k,t}$$

subject to:

- **C1**: $Ax_k - V_k = 0$ for all $k$.
- **C2**: $0 \leq \sum_{k} w_{k,t} \leq N$ for $t = 1, 2, 3, \ldots, T$.
- **C3**: $0 \leq z_{k,t} \leq Z_k$ for $t = 2, 3, \ldots, T$ and all $k$.
- **C4**: $r_{k,t} = w_{k,t} I_{d} B \log_2 \left( 1 + \frac{P h_{k,t}^2}{N}\right)$, where $C1$ is the no-outage constraint for each video user with $V_k$ video bit rate requirement. $w_{k,t}$ is the number (can be a fraction) of subchannels required for user $k$ at time $t$. $C2$ is the constraint on the total available number of subchannels at each time slot. $C3$ ensures that the number of carry-over bits does not exceed $Z_k$ which limits the buffer size for each user. Finally, $C4$ incorporates the channel prediction in the optimization problem. $h_{k,t}$ denotes the estimated average channel gain for video user $k$ at time slot $t$. $P$ is the total transmit power and $\sigma^2$ and $B$ are the noise power and the bandwidth of each subchannel, respectively. $I_d$ is the average intercell interference power.

The problem formulated in (3) minimizes the amount of spectrum required to meet each video user’s bit rate requirement considering predicted average channel condition in the upcoming time slots. This amount of spectrum, $\{w_{k,t}\}_{t=1}^{T}$, is directly related to the rate to be requested, $\{r_{k,t}\}_{t=1}^{T}$, at each time slot to avoid outage.

Figure 2 illustrates how our approach would be integrated into a wireless communication system. The set of estimated required rates provided by our scheme would be passed to a scheduler as new rate constraints for each user. While trying to fulfill these rates for the video users, the scheduler optimizes the frequency-time allocation to all its users considering instantaneous channel state.

**IV. PERFORMANCE STUDY**

In this section, we evaluate the performance of the proposed anticipative buffer and resource allocation. The numerical results for a single video user are presented as the proof of concept. We then show the performance of the scheme when multiple video users request streaming video.

**A. Scenario and Parameters**

A two-cell scenario is illustrated in Fig. 3. The video user moves away from BS1 and is admitted by BS2 experiencing first decreasing and then increasing channel gain. The look-ahead window (assumed to be smaller than the length of the video) is 16 seconds over which the user leaves cell 1 and is admitted by cell 2 at the speed of 30 meters per second roughly the moving speed of the highway. This simulation setup has been chosen to capture the essence of the proposed scheme based on channel prediction. It can also emulate a scenario where the user is traversing an overloaded cell with the possibility of good reception in the future based on its trajectory. The corresponding average channel gain is shown in Fig. 4.

The path loss between the BS and the user accounts for outdoor propagation given by [9, Table A.2.1.2-3]: $PL = 128.1 + 37.6 \log_{10}(d) + L_s$ where $d$ is the distance between the serving BS and the user in kilometers. Without loss of generality, we assume that the user connects to the BS with the highest average received power. $L_s$ accounts for shadowing and is modelled by log-normal random variable with standard deviation of 10dB. Table I lists the parameters used in the simulations, unless otherwise stated.

**B. Proof of Concept for a Single Video User**

The positive effect of buffering can be clearly seen in Fig. 5. The figure shows the total amount of spectrum required by the video user for the whole duration of $T$ time slots in the same setup but versus different buffer size constraints $Z$. The required spectrum is normalized by the total number of PRBs available in the system. In the worst case with zero buffer size, the user can not pre-load. Therefore, it has to receive the required video bit rate at each time slot even in poor coverage.
| Parameters                      | Value       |
|--------------------------------|-------------|
| Channel bandwidth             | 10 MHz      |
| Resource Block $B$            | 180 kHz     |
| Total Number of PRBs          | 50          |
| Transmit Power                | 46 dBm      |
| Antenna gain                  | 0 dB        |
| Antenna                       | 1 x 1       |
| Noise PSD                     | -174 dBm/Hz |
| NF                            | 9 dB        |
| I                             | -149 dBm/Hz |
| Cell diameter                 | 500 m       |
| Minimum distance from BS      | 35 m        |
| Video bit rate                | 1.5 Mbits/s |

Table I

![Graph](image1)

**Fig. 5.** Total required spectrum versus buffer size.

The buffer status and the excess number of bits carried over at each time slot are shown in Fig. 7. Four phases can be identified in both figures for the user with anticipatory buffer control: 1) In the first time slot, the user pre-loads the buffer.

channel condition hence consuming the highest total amount of spectrum. As the buffer size increases, the user can pre-load more when closer to BS1; hence a smaller number of PRBs is needed. This however, will saturate at some point where the user pre-loads most of the video and plays later. Although desirable from the spectral efficiency point of view, it seems undesirable due to user’s unpredictable behaviour as discussed before. In other words, it will be a waste of resources, if the user switches to another video without watching the pre-loaded one. Hence, it is preferable to limit the maximum buffer size.

Figure 6 shows the allocated bits and the corresponding required number of PRBs over time as the user moves. $T_d = 0.167 s$ and $V = 250$ Kbits per time slot corresponding to $R = 1.5$ Mbits/s video rate. Two cases are plotted: 1) non-zero buffer size and 2) zero buffer size. The latter essentially captures the performance of a user with no possibility of pre-buffering either due to video time constraints or no knowledge of future channel condition. As expected, the video is pre-loaded when the user is close to BS1 to prevent the outage at the cell edge with the worst channel gain. This behaviour minimizes the required spectrum given a maximum buffer size.

![Graph](image2)

**Fig. 6.** Allocated bits $r_t$ and spectrum $w_t$ to the video user versus time. $T_d = 0.167 s$ and $V = 250$ Kbits per time slot. $Z = 5V$ for the pre-buffering user.
2) In the subsequent time slots, it keeps the buffer full while playing the video. The cost of the full buffer increases as the user gets further away from the BS. 3) The increase in the required spectrum and the knowledge of improved future channel gain result in allocating less bits at the cell edge and playing the pre-loaded buffer instead. While the buffer remains full in the previous two phases, it is used up during this phase. 4) Finally, the user is admitted to cell 2, the channel gain increases and with this knowledge, the user only receives $V$ bits per time slot to play the video without any buffering. At this stage, the user does not need to pre-load since a good average channel gain is predicted. As clear in both figures, the main gain is achieved by playing the video during the time the undesired channel condition is experienced.

C. Performance for Multiple Video Users

We use the same two-cell scenario with a total number of $K_v$ video users requesting service from BS1. The inter arrival time between the users is modelled by an exponential random variable with mean of 0.58 seconds. After the BS1 receives a service request, it estimates the required spectrum as in (3). If the system is not overloaded, the user is admitted and the estimated available spectrum over the look-ahead window is updated. In the mean time, the resources are scheduled among the admitted video and already existing users at each time slot. The service request at the required video quality is rejected if the estimated required spectrum is not available.

Figure [8] shows the number of video users that can be supported versus the total number of service requests. Available number of PRBs $N = 15$ to count for the existing load of each cell. Anticipatory buffer control and resource allocation allows for servicing higher number of video users by adjusting the requested rates considering the future load. Such consideration is not possible in a system with no future channel prediction. Hence, although admitted by the system, either the number of video users experiencing no outage decreases or the video user is served with a lower video quality.

V. CONCLUSION

In this paper, we proposed a buffer control and resource allocation scheme for the streaming of non-real time video in cellular networks.

The proposed formulation has three important benefits. Firstly, it takes advantage of the future user trajectory for a period of time referred to as “look-ahead” window, reported by the user or predicted and available at the serving BS. This information is translated into the average channel gain and is used in formulating the optimization problem. Such information gives an insight about the possibility of undesired channel condition which in turn mandates the serving BS to increase the allocated data rate for pre-buffering. However, the increase in the data rate is in control; the manner and timing to do so is adjusted consuming the minimum bandwidth for smooth video streaming.

Secondly, our scheme is based on an objective quality measure, i.e., average video bit rate. This makes our formulation general and relevant, while ignoring subjective quality metrics which may be affected by the video content or time of the day. We showed that with the increase in the buffer size for the user with the undesired future channel condition, we can save more resources but at the risk of the user switching the video before watching the pre-loaded one. However, given a maximum buffer size, the scheme optimizes the rate allocation for a smooth video streaming.

Finally, the proposed scheme is based on linear programming, which makes the implementation easy and fast. In the next step, we will study the trade-off between the error associated with the channel prediction and the achieved gain.

REFERENCES

[1] Cisco, “Cisco visual networking index: Global mobile data traffic forecast update, 2011-2016,” White paper, Feb. 2012.
[2] R. Pantos and W. May, “HTTP live streaming,” IETF, Informational Internet-Draft draft-pantos-http-live-streaming-07, Sep. 2011.
[3] S.-F. Chang and A. Vetro, “Video adaptation: Concepts, technologies and open issues,” in Proc. of the IEEE, vol. 93, no. 1, pp. 148–158, Jan. 2005.
[4] M. Reisslein and K. Ross, “A join-the-shortest-queue prefetching protocol for VBR video on demand,” in Proc. of International Conference on Network Protocols, Oct. 1997, pp. 63–72.
[5] M. Draexler and H. Karl, “Cross-layer scheduling for multi-quality video streaming in cellular wireless networks,” in Proc. of Int. Wireless Communications & Mobile Computing Conference (IWCMC), Jul. 2013, to appear.
[6] J. Huang, Z. Li, M. Chiang, and A. K. Katsaggelos, “Joint source adaptation and resource allocation for multi-user wireless video streaming,” IEEE Trans. Circuits and Syst. Video Technol., vol. 18, no. 5, pp. 582–595, May 2008.
[7] M. Nadin, “Anticipatory computing,” Ubiquity, vol. 2000, Dec. 2000.
[8] I. Tadrous, A. Eryilmaz, and H. E. Gamal, “Proactive resource allocation: Harnessing the diversity and multicast gains,” Arxiv Computing Research Repository, vol. abs/1110.4703, 2011. [Online]. Available: http://arxiv.org/abs/1110.4703.
[9] 3GPP, “Further advancements for E-UTRA physical layer aspects,” 3GPP, Tech. Rep. TR 36.814 V9.0.0, Mar. 2010.