



HAL
open science

Context-aware mobility resource allocation for QoE-driven streaming services

Majed Haddad, Imen Triki, Rachid El-Azouzi, Afef Feki, Marouen Gachaoui

► **To cite this version:**

Majed Haddad, Imen Triki, Rachid El-Azouzi, Afef Feki, Marouen Gachaoui. Context-aware mobility resource allocation for QoE-driven streaming services. 2016 IEEE Wireless Communications and Networking Conference (WCNC), Apr 2016, Doha, Qatar. pp.1-6, 10.1109/WCNC.2016.7565146 . hal-03485142

HAL Id: hal-03485142

<https://hal.science/hal-03485142>

Submitted on 17 Dec 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Context-Aware Mobility Resource Allocation for QoE-Driven Streaming Services

Imen Triki*, Majed Haddad*, Rachid El-Azouzi*, Afef Feki[†] and Marouen Gachaoui*

*CERI/LIA, University of Avignon, Avignon, France

[†]Mathematical and Algorithmic Sciences Lab France Research Center, Huawei Technologies Co. Ltd. France

Abstract—Streaming services come with their own challenges and technical issues that still need to be addressed for satisfying the target quality of experience (QoE) of the end-users in mobile environments. In this paper, we explore the idea of combining users' context information with the packed prefetching process features to enhance users' QoE in heterogeneous networks. More specifically, we propose a scheduling mechanism for video streaming traffic, in which the access to the network resources is restricted to users with a signal-to-noise plus interference ratio (SINR) above a given threshold. This scheme benefits from the fact that, as users are in permanent motion, they may experience different SINR values during the same video streaming session offering the opportunity to only schedule users with good channel conditions. The proposed scheduling approach (subsequently referred to as context-aware mode switching (CAMS)) not only allows to achieve overall network spectral efficiency improvement, but also guarantees fairness and QoE among users. Our simulation results show that CAMS achieves almost 1 bit per second per Hertz gain compared to the conventional scheduler (without CAMS), and up to 87% improvement in the probability of no starvations when users move at 40 kmph.

Index Terms—LTE, heterogeneous network, video streaming, mobility, spectral efficiency, fairness, quality of experience, startup delay, starvation, rebuffering delay.

I. INTRODUCTION

With the development of smartphones, tablets and laptops from one side, and the emergence of new opportunities in media productions such as digital video and social networking from the other side, video streaming has become one of the most prominent internet services. According to recent researches, today video traffic represents a significant fraction of internet traffic and is expected to exceed 80% in 2019 [1]. Though nowadays most of the videos in the social media are prerecorded, in the future, the demand for live video streaming will become more important. This leads to several research challenges which call for new radio resource management (RRM) paradigms. In response to such challenges, a variety of video compression and streaming techniques have been proposed to provide a seamless QoE to the end-user [2]–[4]. In this context, new measures for evaluating the QoE are defined and used for network optimization. For example, the QoE for video streaming is translated into starvation, which correspond to low user satisfaction levels. In addition, to cope with this large traffic demand, the deployment of low power base stations (BSs) in dense urban areas is an efficient solution

to complement the existing macro cellular architecture with higher capacity levels. Such a heterogeneous network (HetNet) consists of macro cell base station (MBS), coexisting with low power small cell base stations (SBS) [5]. HetNets improve spatial reuse and coverage, thus allowing cellular systems to achieve higher data rates, while retaining the seamless connectivity and mobility for end-users. An accurate measure of the QoE is required to evaluate and improve the way users perceive the quality of video streaming playback. This is even more challenging in the context of HetNet where the available channel usually fluctuates in a wide range. Thus, content aware transmission of scalable multimedia is essential in order to guarantee the target performance for the end-user.

The analysis of the YouTube workload [6]–[9] emphasizes the correlation between access patterns and human behaviors (e.g., time-of-day, day-of-week). The main objective of this paper is then to propose a context-aware resource allocation for QoE-driven streaming services, pertinent for the mobile users. Most of the existing works on scheduling and user selection schemes are only valid for slow varying fading scenarios. Nevertheless, in mobile environments, such an assumption generally fails especially with high speed users, as the channel state information (CSI) varies over a faster time scale. To overcome this hurdle, we propose an efficient resource allocation mechanism which profits from the users' mobility context along with the packet prefetching process of the streaming video player by considering the mobility as an opportunity rather than an obstacle [10], [11]. In fact, in a large system with users fading independently and channel varying faster (having different dynamic range due to their perpetual mobility), there is likely a user with a very good channel at any time during the same streaming session. Note that it has been proven in [12] that long term total throughput is maximized by always serving the user with the strongest channel.

In this paper, a context-aware mode switching approach is proposed to allow users switching from active to inactive states and vice-versa during their streaming session: Active users will stream video packets till having their playback buffer filled, whereas inactive users will play their already prefetched packets (see Figure 1). To do so, we fix a threshold SINR value and only serve users with SINR higher than this threshold.

Unlike videos delivered using a deterministic channel with constant delay and rate, and very limited data drops (such as digital TV broadcasting), videos delivered over IP networks are much more problematic since they are supported by a non-

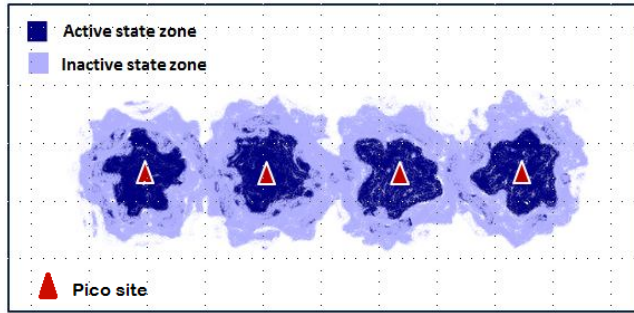


Fig. 1. The context-aware mode switching (CAMS) approach.

guaranteed bite rate (GBR) resource type due to intermittent usage of network resources [9], [13]. Notice that a non-GBR bearer is used for best-effort traffic such as file downloads, as it does not have dedicated bandwidth. The radio channel conditions can change consistently during a video streaming (e.g., due to mobility), which may degrade considerably the user viewing experience and thus the user engagement. As it is hard to guarantee a high quality of service (QoS) for video streaming services, research interests are today being steered toward end-user quality of experience which differs from a user to another depending on a set of factors. In the literature, many quality metrics have been defined [14]. In this work, we will focus on the three most commonly known metrics: the start-up delay, the probability of starvation and the rebuffering delay. According to a recent research in [14], 20% of video streaming viewers experience a rebuffering ratio of 10%, and 14% of them have to wait 10 seconds to start playing the video.

The paper is organized as follows. Section II presents the network model adopted including the HetNet topology, the users mobility and the video streaming traffic. Section III describes the proposed CAMS approach. In Section IV, we describe the simulations set up as well as the obtained results compared to baseline schemes. Section V concludes the paper.

II. THE NETWORK MODEL

In this work we adopt a road traffic topology with MBSs and SBSs distributed over a road (see Figure 2). We assume that we have K users moving along 3 lines with two speed classes: low speed users driving at v_{min} and high speed users driving at v_{max} (see Figure 3). During a typical streaming session, users are subject to many handover processes to switch from one cell to another. The handover triggering condition we adopt in our study is known as the Event A3, defined as a reporting triggering event. The event A3 corresponds to the situation where the measured Reference Signal Received Power (RSRP)¹ on a neighboring cell is higher than the measured RSRP on the serving cell by a predefined offset [15].

¹RSRP is a measure of the received signal strength of a cell at a user equipment (UE) and it is measured based on the strength of predefined reference signals that cells broadcast.

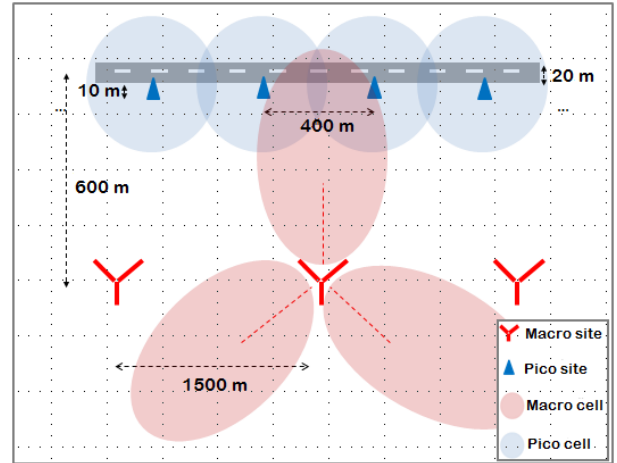


Fig. 2. The network topology.

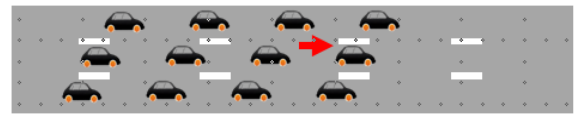


Fig. 3. Users' mobility model.

The Video Streaming Traffic

In this work, we consider the classical (chunk-based) video streaming where the video file is transmitted under only one resolution (e.g., 720p [16]). We do not consider adaptive video streaming (e.g., Dynamic Adaptive Streaming over HTTP (DASH)) since it would be very CPU and time consuming due to frequent switching events between different video resolutions caused by to the high variability in channel conditions in the high mobility context. Notice that, although most of video sharing web sites are nowadays using adaptive streaming techniques such as dynamic adaptive streaming over HTTP (DASH), they do not have abandoned the classical video streaming. That said, the analysis provided in this paper can be easily extended to the case of adaptive streaming.

In order to ensure standard-compliant implementation, three main parts are taken into account: the streaming server, MBS/SBS and the video player. Each video is stored in a streaming server as a bench of frames where each frame is divided into 8 slices (or packets) of equal sizes. When asked for a streaming delivery, the server sends periodically the video slices to the end-user through the evolved Node B (eNodeB). The eNodeB identifies the user to serve and starts sending the video bit by bit depending on his channel quality indicator (CQI). The video player gathers the received bits in its buffer to form the video frames. When a given number of frames –corresponding to the playback buffer length (hereinafter denoted by x_0)– is reached, the player starts reading the video without stopping the downloading process. When it sees no frames in its buffer, the player stops reading the video and carry on the downloading process until the same

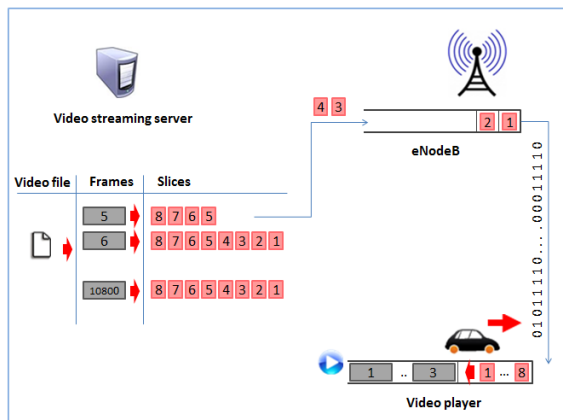


Fig. 4. The video streaming traffic.

number of frames is reached again (see Figure 4).

A starvation occurs when the player buffer is empty before the video playback is completed. Accordingly, the probability of starvation P_s is measured in terms of number of starvation distribution over a video streaming session. When a starvation occurs, the player stops playing the video to prefetch x_0 frames again. The time needed for the prefetching is defined as the rebuffering delay.

III. THE CONTEXT-AWARE MODE SWITCHING (CAMS) APPROACH

Mode switching approach has been already widely used by a variety of services and mobile applications, including elastic services such as e-mail and instant messaging. The key idea is that instant messaging is only occasionally available when the user is in a covered zone. One can send a friend an offline message that will be transmitted effectively the next time he is under a covered zone.

Figure 1 describes the CAMS approach. We assume that each cell is classified into two coverage zones: a zone with good radio conditions at the cell center (active zone) and a zone with bad radio conditions at the cell edge (inactive zone). We note that CAMS only “filters” video streams through the initial buffering and the mobility of users. Each UE measures his SINR, and depending on whether his SINR is below or above a given threshold, he remains active or inactive during the next scheduling time slot. More specifically, when the user’s SINR exceeds a given threshold, he will be allocated resource blocks (RBs) by the scheduler based on other resource requests in the cell, otherwise the user is not scheduled. These two states are respectively defined as active and inactive states. The merit of this approach lies in the fact that good SINR users will be given more chance to store packets in their player buffers than bad SINR users with a lower packet loss rate, yielding a better user’s experience. Obviously, although some users may stay silent, there is high probability they become active during the next scheduling period within the video streaming session, due to their mobility. In fact, channel

conditions of mobile devices are time-varying and location-dependent due to fading and shadowing.

The impacts of this mechanism are three-fold:

- Unused RBs by bad SINR users will be more efficiently used by good SINR users. In fact, these released radio resources will be employed by users experiencing good channel conditions before they fall in bad channel conditions,
- This allows good SINR users to rapidly store the video packets, giving the chance for other users in the queue to be scheduled afterwards,
- This will not condemn bad SINR users during their whole streaming session: Users are in perpetual mobility and thus the chance they experience a good channel condition is high. Moreover, by virtue of a buffering mechanism, users are enabled to keep track of the packets they have already stored in previous active states.

To conclude, CAMS approach allows for better radio resource management, provides improved overall mobile-broadband performance, and allows operators to maintain a more seamless user experience.

In addition, CAMS can be used with minimal modifications on the existing scheduler. A simple way to do this consists in adding a “filter” behind the conventional scheduler to restrict the set of UEs to be served at each transmission time interval (TTI), which corresponds to 1 ms in LTE. By doing so, bad SINR users are excluded from being served. But this highly depends on the scheduling metrics and the utility functions adopted by the scheduler. In our context, we make use of the alpha fair scheduler [17] which allocates network resources proportionally to users’ throughput. Additional references about scheduling algorithms related to QoE can be found in [18], [19].

IV. SIMULATION RESULTS

We use the LTE Wien simulator, implemented with Matlab and published by the TU Wien’s Institute of Telecommunications under an academic non-commercial use license [20]. This tool implements a set of 3GPP LTE standard specifications. It uses simplified models that capture the essential characteristics of the standard with high accuracy and low computational complexity. In this study, we focus only on downlink transmissions. The main event is the transmission of a subframe from the eNodeB to the end-user, which lasts 1 ms and corresponds to a subframe duration in the LTE standard.

We consider a HetNet with macro and small sites as described in Table I. The video streaming is configured as mentioned in Section II. The videos’ duration is equal to 120 seconds with high definition (HD) 720p quality and a playing rate of 60 frames/s. For benchmarking purposes, we propose to compare the proposed scheduling scheme with the conventional case where no mode switching is configured.

In what follows, we will evaluate the performance of CAMS by illustrating network performance along with QoE metrics. Notice that the SINR thresholds set in our experiments are estimated through an empirical study. For the sake of clarity

TABLE I
SCENARIO CONFIGURATION.

Parameters	Macro sites	Small sites
Bandwidth (in MHz)	20	20
Frequency (in GHz)	2.14	2.14
RB bandwidth (in KHz)	180	180
Tx-Power (in Watts)	40	1
α -fair scheduler	$\alpha=0.6$	$\alpha=0.6$
Inter-sites-distance (in meters)	1000	400
Number of sites	7	12

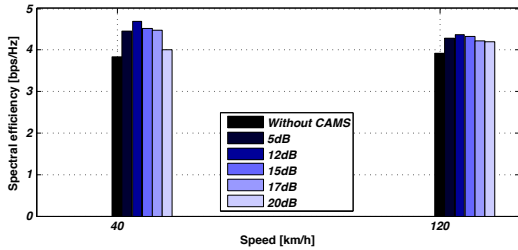


Fig. 5. Spectral efficiency.

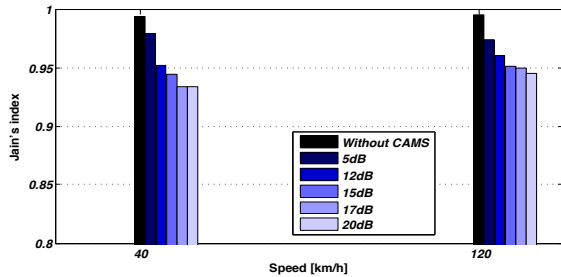


Fig. 6. Jain's fairness index.

and consistency, we will only show results corresponding to SINR thresholds with a significant performance gain.

A. Spectral Efficiency

The spectral efficiency is defined as the total number of bits allocated by the network to a given user during the simulation length per available bandwidth [21].

Figure 5 depicts the variation of the average spectral efficiency with the SINR threshold for two speed classes: low vehicular velocity with $v_{min} = 40$ kmph and high vehicular velocity with $v_{max} = 120$ kmph.

We first notice that the SINR threshold does not impact the spectral efficiency in the same way for the two speed classes. Results show that the spectral efficiency varies more significantly from one threshold to another at low speed class. This can be explained by the fact that, when a user moves slowly, he spends more time at active states, which highlights the importance of the SINR threshold. Interestingly, we notice a maximum spectral efficiency for a SINR threshold of 12 dB, with almost 1 bps/Hz more than the reference scheme (without CAMS) at 40 kmph corresponding to a gain of 23%. This can be justified by the fact that bad SINR users

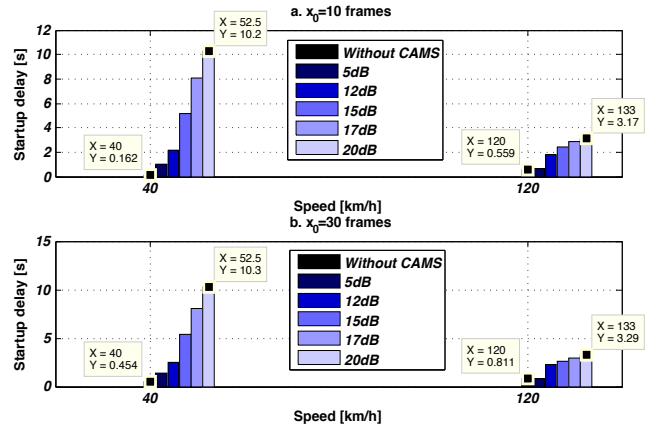


Fig. 7. The startup delay.

have given up their RBs to good SINR users to be more efficiently exploited. In fact, when a user has bad channel conditions, he can not exploit all the RBs he has been allocated by the network. Leaving these resources to users with better channel conditions is more efficient. Moreover, when the SINR threshold increases, the scheduler becomes more selective and the allocated RBs become more efficiently used. However, when the SINR threshold becomes very high, users are not scheduled and the number of allocated RBs decreases, which explains the maximum spectral efficiency obtained with a SINR threshold of 12 dB.

B. Fairness

In addition to the spectral efficiency evaluation, we examine the fairness degree among the users. A commonly used metric is Jain's fairness index [22] which measures the fairness among all users in the system. It is given by

$$J(r_1, r_2, \dots, r_K) = \frac{(\sum_{k=1}^K r_k)^2}{K \cdot \sum_{k=1}^K r_k^2} \quad (1)$$

where K is the number of UEs and r_k is the throughput of user k divided by the number of the RBs assigned to this user over the streaming session.

Figure 6 illustrates the Jain's fairness index computed for the case of the proposed CAMS scheme and compared to the conventional scheduling procedure (without CAMS) for low and high speed and different SINR thresholds. With a standard alpha-fair scheduler ($\alpha = 0.6$), the Jain's index is very close to 1. Although we have integrated CAMS, it is clearly seen that the Jain's index remains very close to 1, which proves that CAMS ensures fairness among all users.

So far, we have studied commonly used key network performance indicators. With the same goal evaluating the performance of the proposed approach, let us now give some leads towards QoE indicators.

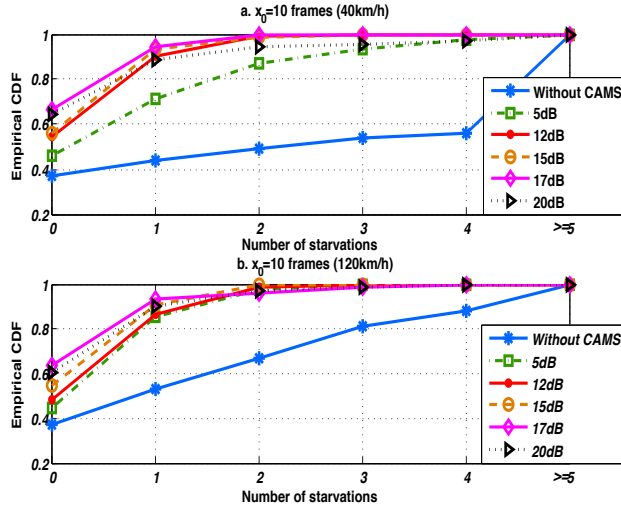


Fig. 8. Empirical CDF of starvation with $x_0 = 10$ frames.

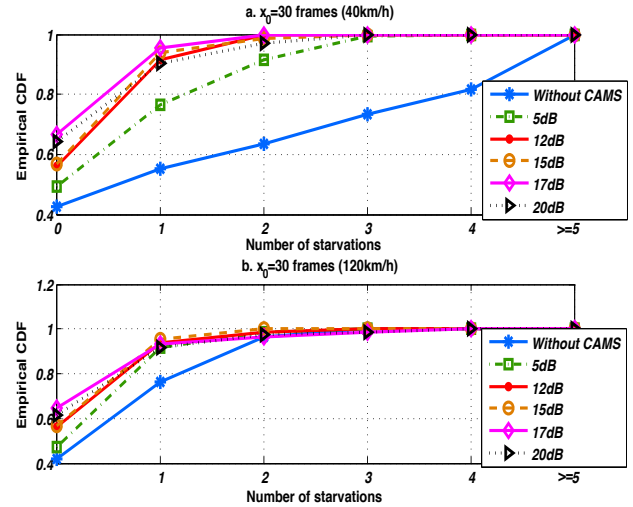


Fig. 9. Empirical CDF of starvation with $x_0 = 30$ frames.

C. Startup Delay

The startup delay is defined as the waiting time needed by the player to initially prefetch x_0 frames to start playing the video. In Figure 7, we show the variation of the startup delay with speed and SINR thresholds for two different values of x_0 . In both figures, we remark that the startup delay decreases with the speed. This is even more significant in the case of highly selective filters. This is due to the fact that, when we use CAMS, the waiting time for starting the video becomes longer as the SINR threshold increases. This impact is greater for low speeds than for high speeds. In fact, when users move faster, they are more likely to be in active states and spend less time in inactive states giving a startup delay larger for low speeds. Now, when the filter becomes more selective, the users become less likely to be in an active states yielding an increase in the waiting time. To sum it up, we can conclude that CAMS performs better at high speeds.

Comparing the two figures with the two different values of x_0 , we notice that the waiting time is more important for a higher value of x_0 , especially with low selective filters. Hence, CAMS becomes better suited when used with a big playback buffer in order to not degrade the QoE in terms of the startup delay.

D. Probability of Starvation

In Figures 8 and 9, we show the empirical cumulative distribution function (CDF) of the number of starvation experienced for different SINR thresholds and different startup buffer lengths. Each of the four graphs below corresponds to a given x_0 and speed. We remark that the performance are approximately the same for low and high speed. Figure 8.a and Figure 8.b show that when users move fast, they

are less likely to experience high number of starvations. Another noteworthy observation that emerges from numerical evaluations is that, when we use CAMS, the probability of having no starvation increases for high selective filters. In particular, we notice a gain of 87% at 40 kmph and 70% at 120 kmph for a SINR threshold of 17 dB compared to the case without CAMS.

The same interpretations hold for $x_0 = 30$ in Figure 9.a and Figure 9.b. From here, we also notice that the probability of no starvation slightly increases as x_0 increases, especially for the case without CAMS.

E. Rebuffering Delay

To proceed further with the QoE analysis, we plot the duration of the n^{th} starvation for $n = \{1, 2, 3, 4\}$ and using the same parameters. Figure 10.a and Figure 10.b are for $x_0 = 10$ frames, whereas Figure 11.a and Figure 11.b are for $x_0 = 30$ frames. Figure 10 (for $x_0 = 10$ frames) shows that when users move faster, the waiting time for prefetching packets is shorter. The same analysis as for the startup delay variation can be done. When we compare Figure 10 and Figure 11, we see that the waiting time becomes more important as the value of x_0 increases especially for low selective filters. This suggests that it is better to use CAMS with a high buffer length in order to not greatly degrade the QoE in terms of the rebuffering delay. Notice that this is even better for high speeds.

V. CONCLUSION

Video streaming services have grown to become increasingly popular over the recent years. In this paper, we have designed a context-aware mobility resource allocation scheme that allows to intelligently exploit the playout buffer at the application layer for streaming videos, and the multi-user

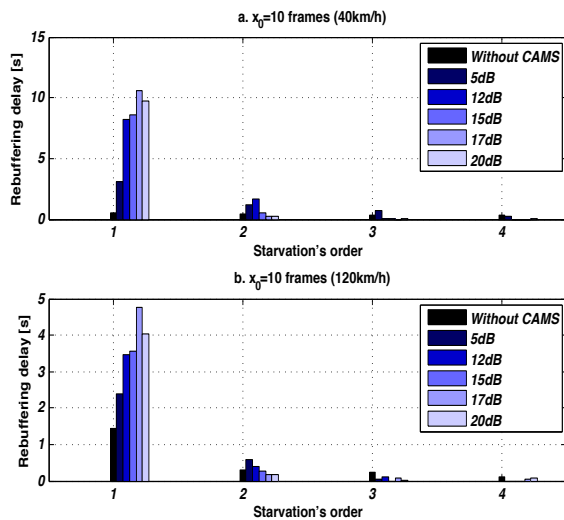


Fig. 10. Rebuffering delay with $x_0 = 10$ frames.

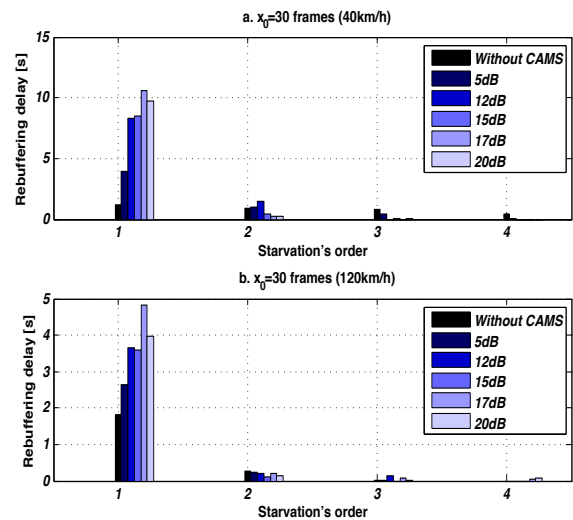


Fig. 11. Rebuffering delay with $x_0 = 30$ frames.

diversity at the physical layer. The proposed CAMS scheme not only enables the wireless network to achieve spectral efficiency improvement but also guarantees a high fairness and QoE among users. The QoE has been evaluated with three parameters: the startup delay, the probability of starvations and the rebuffering delay. Simulation results have shown that, in order to globally improve the users' QoE, the proposed CAMS mechanism is better suited for high speed users provided that a wise choice of the SINR threshold is considered. These results offer hope that such a simple and accurate scheduling approach can be easily deployed, as CAMS does not only improve the users' QoE, but also leads to better network performance. As a future work, we plan to analytically study the optimal values of the SINR threshold in terms of end-users QoE and network performance.

REFERENCES

- [1] Cisco, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2014-2019," *White paper*, Feb. 2015.
- [2] 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 Systems for Video Technology*, vol. 18, pp. 582-595, May 2008.
- [3] Y. Xu, S.-E. Elayoubi, E. Altman, and R. E. Azouzi, "Impact of flow-level dynamics on qoe of video streaming in wireless networks," in *INFOCOM*. IEEE, 2013.
- [4] Y. Shen, Y. Liu, Q. Liu, and D. Yang, "A method of qoe evaluation for adaptive streaming based on bitrate distribution," in *IEEE International Conference on Communications, ICC Sydney, Australia, Workshops Proceedings*, 2014.
- [5] A. Damnjanovic, J. Montojo, Y. Wei, T. Ji, T. Luo, M. Vajapeyam, T. Yoo, O. Song, and D. Malladi, "A survey on 3GPP heterogeneous networks," *IEEE Wireless Communications*, vol. 18, no. 3, pp. 10-21, 2011.
- [6] P. Gill, M. Arlitt, Z. Li, and A. Mahanti, "YouTube traffic characterization: A view from the edge," in *In: Proc. of IMC*, 2007.
- [7] M. Cha, H. Kwak, P. Rodriguez, Y. yeol Ahn, and S. Moon, "I Tube, You Tube, Everybody Tubes: Analyzing the world's largest user generated content video system," in *In Proceedings of the 5th ACM/USENIX Internet Measurement Conference (IMC07)*, 2007.
- [8] M. Zink, K. Suh, Y. Gu, and J. Kurose, "Characteristics of YouTube network traffic at a campus network - measurements, models, and implications," *Computer Networks*, vol. 53, no. 4, pp. 501 - 514, 2009.
- [9] M. Haddad, E. Altman, R. El-Azouzi, T. Jiménez, S. E. Elayoubi, and S. B. Jemaa, "A Survey on YouTube Streaming Service," in *Valuetools*, Paris, France, 2011.
- [10] P. Gupta and P. Kumar, "The capacity of wireless networks," *Information Theory, IEEE Transactions on*, vol. 46, no. 2, pp. 388-404, Mar 2000.
- [11] M. Grossglauser and D. Tse, "Mobility increases the capacity of ad hoc wireless networks," *Networking, IEEE/ACM Transactions on*, vol. 10, no. 4, pp. 477-486, Aug 2002.
- [12] R. Knopp and P. A. Humblet, "Information capacity and power control in single-cell multiuser communications," in *Communications, ICC IEEE International Conference on*, 1995.
- [13] A. Rao, Y.-S. Lim, C. Barakat, A. Legout, D. Towsley, and W. Dabbous, "Network Characteristics of Video Streaming Traffic," *ACM CoNEXT*, Dec. 2011.
- [14] X. Liu, F. Dobrian, H. Milner, J. Jiang, V. Sekar, I. Stoica, and H. Zhang, "A case for a coordinated internet video control plane," in *SIGCOMM*, Helsinki, Finland, 2012, pp. 359-370.
- [15] S. Sesia, I. Toufik, and M. Baker, *LTE - The UMTS Long Term Evolution: From Theory to Practice*, 2nd ed. Wiley, August 2011.
- [16] Live encoder settings, bitrates and resolutions. <https://support.google.com/youtube/answer/2853702?hl=en>.
- [17] J. Mo and J. Walrand, "Fair end-to-end window-based congestion control," in *International Symposium on Voice, Video and Data Communications*, 1998.
- [18] K. Piamrat, K. Singh, A. Ksentini, C. Viho, and J. Bonnin, "Qoe-aware scheduling for video-streaming in high speed downlink packet access," in *Wireless Communications and Networking Conference (WCNC), IEEE*, April 2010.
- [19] A. Alfayly, I. Mkwawa, L. Sun, and E. Ifeakor, "Qoe-based performance evaluation of scheduling algorithms over lte," in *Globecom Workshops, IEEE*, Dec 2012.
- [20] J. Ikuno, M. Wrulich, and M. Rupp, "System level simulation of lte networks," in *Vehicular Technology Conference (VTC 2010-Spring), IEEE 71st*, May 2010.
- [21] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2004.
- [22] D. C. R. Jain and W. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer system," *DEC Research Report TR-301*, Sep. 1984.