Timely Delivery Versus Bandwidth Allocation for DASH-Based Video Streaming Over LTE

Stefania Colonnese, *Member, IEEE*, Sergio Russo, Francesca Cuomo, *Senior Member, IEEE*, Tommaso Melodia, *Member, IEEE*, and Izhak Rubin, *Fellow, IEEE*

***Abstract*—We study HTTP video streaming over a wireless access cell, such as LTE. We derive a closed form expression for the probability of timely delivery of the streamed video packets as a function of the bandwidth allocated to the user. The probability expression uses application, transport, and physical layer parameters relating to the employed video stream rate and to the quality of the communications channel. We validate the analytical results via numerical simulations based on the execution of real video traces and with LTE channel quality index trace modeling. The attained results allow the system designer, scheduler, and manager to calculate the bandwidth level that should be allocated to maintain acceptable system performance, expressed by two quality-of-experience (QoE) metrics, namely the video rate and the probability of timely delivery of video packets. The model provides a rationale and tools for performing admission procedures, bandwidth pricing policies, and cell dimensioning designs.**

***Index Terms*—4G mobile communication, Mobile video streaming, Quality of Experience.**

I. Introduction

Mobile video streaming applications generate a dominating fraction of traffic flows downloaded over 3G platforms, and are going to rapidly expand over Long Term Evolution (LTE) systems [1]. Streaming applications rely on a service architecture molded on the Dynamic Adaptive Streaming over Http (DASH) ISO/3GPP Standard [2]. Under DASH, several copies of the same video, encoded at different rates, are stored on a content server, and are reliably downloaded in sequences of “chunks,” or fragments, at the user’s request. After an initial buffering stage, the client starts reading chunks from the buffer, decoding them and playing them out. At the same time, it requests new chunks at a quality (rate) that is compatible with the estimated throughput.

HAS architecture presents peculiar challenges when applied over a mobile wireless system framework. Reliable protocols over error-prone wireless channel induce chunk dependent delays, due to either channel quality fluctuations or intrinsic chunk size fluctuations around the nominal average value. Occurrence of persistent consecutive delays in downloading single chunks cause buffer depletions at the client, which may lead to video playout interruptions (stalls). Intrinsic HAS protocol adaptation techniques react and allow switching away from a stored video quality, but induce annoying quality fluctuations [3] as well as additional delays due to the slow reaction time of higher layer protocols.

In [4], the authors show that the base station can dynamically adapt the wireless (e.g., LTE) bandwidth resources that are allocated to a stream by taking into account the actual chunk size and the communications channel quality observed at the mobile user. Notwithstanding this, in real systems, wireless bandwidth resources tend to be at short supply. Consequently, it is essential to study the impact of the supported video stream QoE level on wireless bandwidth requirements.

The study of a key QoE metric, the probability of timely delivery, is motivated by: i) the specific challenges imposed by HAS over LTE systems [5]; ii) the impact that video playout stalls, caused by the accumulation of chunk delays, may have on the perceived quality [6]; for example, see [6] relating to the impact caused by stalls on the user’s experience while watching YouTube video streams on a mobile device. Motivated by these factors, we derive in this letter a closed form mathematical expression for calculating the probability of timely delivery of DASH based streamed video chunks as a function of the downlink LTE bandwidth that is allocated to the mobile user.

II. System Model

We consider a video server and a video client running a DASH video streaming process. The video client is run at the User Equipment (UE). At the server side Q traces ($q = 1, \dots, Q$) of a given video program are stored, every trace has the same length of K chunks but a different QoE level q and a consequent average video rate R_q. A requested flow, at a prescribed q, is streamed as a sequence of K chunks of size of $\lambda(k, q)$ bits, $k = 1, \cdots K$. Let us denote by τ the video chunk playout time at the client side. During the streaming service, the instantaneous level r_k of the application layer rate required for the timely transmission of the k-th chunk, $k = 1, 2, \dots, K$ is given by $r_k = \lambda(k, q)/\tau\ bps$.

In general, the throughput rate provided at the application layer is a fraction $\alpha < 1$ of the actual wireless Data Link (DL) rate $r_k^{(DL)}\ bps$. This latter depends on the overhead that is added above the DL, by the associated network and transport layer protocols. Hence, a mobile user that requests a streaming service at QoE = q, requires a DL rate $r_k^{(DL)} = r_k/\alpha$ to transmit the k-th chunk.

As for the LTE DL rate, the Base Station (BS) monitors the levels of the Channel Quality Indicators (CQIs) observed and delivered by each UE. The CQI is an integer valued parameter in $\{1 \cdots 15\}$. Based on the CQI levels provided by each active UE, the BS proceeds to select the transmission power

Manuscript received October 31, 2015; revised December 14, 2015; accepted January 3, 2016. Date of publication January 8, 2016; date of current version March 8, 2016. The associate editor coordinating the review of this paper and approving it for publication was W. Xiang.

S. Colonnese, S. Russo, and F. Cuomo are with the DIET, University of Rome Sapienza, Rome 00184, Italy (e-mail: francesca.cuomo@uniroma1.it).

T. Melodia is with the Department of Electrical and Computer Engineering, Northeastern University, Boston, MA 02115, USA.

I. Rubin is with the Electrical Engineering Department, University of California, Los Angeles (UCLA), Los Angeles, CA 90095-1594, USA.

Digital Object Identifier 10.1109/LCOMM.2016.2516001

level, and the employed modulation and coding scheme. This setting determines the realized spectral efficiency level, which in turn provides for the calculation of the bandwidth resources that must be allocated for streaming the flow to the mobile user [7]. The overall bandwidth B_{LTE} ranges in [1.08–18] MHz.

Channel quality variations and BS activated rate adaptations occur at a time scale of the order of $T_c \approx 10-100\, ms$, which is much smaller than the chunk downloading time (that can be of the order of $\approx 1-2\, s$). Let us denote by L the number of intervals within the download process during which the CQI parameter is assumed to be approximately constant. If the bandwidth assigned to the user is kept constant within the k-th chunk, the rate is proportional to the average spectral efficiency $\overline{\eta(c)} = \frac{1}{L}\sum_{l=0}^{L-1}\eta(c_l)$ observed during the chunk's transmission period. Therefore, given the observed CQI level and the consequent attained spectral efficiency value, the radio bandwidth required for the timely downloading of the k-th chunk is calculated as $B_k = r_k^{(DL)}/\overline{\eta(c)}Hz$.

In reality, the system typically imposes limitations on the maximum frequency bandwidth that can be used. We denote by B_s the maximum bandwidth assigned to the streaming service. The bandwidth to be assigned to the user is then chosen as: $\widetilde{B_k} = \min(B_k, B_s)Hz$. This translates to an application rate of $\widetilde{r}_k = \widetilde{B_k}\cdot\overline{\eta(c)}\cdot\alpha$ and a time duration taken to download the chunk $\tau_k = \lambda_k/\widetilde{r}_k\, s$.

When the download time τ_k is longer than the temporal interval τ at which chunks must be read at the receiver buffer, a delay level $\delta_k = \max(\tau_k - \tau, 0)\, s$ is induced.

Positive δ_k delays lead to the depletion of the buffer at the receiving mobile node. Since total buffer depletion would lead to service interruptions (stalls), the frequency of occurrence of non-zero delay events should be kept low, and the percentage of video chunks that are timely delivered is relevant to the quality level of video stream reception experienced by the user. The fraction of timely delivered video chunks jointly depends on both the video stream characteristics and the bandwidth assigned by the BS manager, as well as on the quality level of the communications channel experienced by the user.

III. Probability of Timely Delivery of Video Streamed Chunks

We compute the Probability of Timely Delivery (PTD), denoted henceforth as $\mathcal{P}_{\text{TD}}$, as a function of the bandwidth assigned to the streaming service B_s. It is related to the statistical characterization of the application layer video traffic. From the above analysis, we note that the probability of a delay event is equal to the probability that the download time τ_k exceeds τ and the chunk download time τ_k is proportional to the chunk size λ_k. As far as the chunk size λ_k is concerned, in the literature it is observed that video packets corresponding to a playout time of one to few seconds are often well modeled as being distributed according to heavy-tailed probability density functions (pdf) such as Gamma pdfs, in the following Γ, whose mean and variance levels depend on the configured quality q value [4]. We thus use the following characterization:

$$p_\Lambda(\lambda;\mu_q,\sigma_q^2) = C_q \cdot \lambda^{(\mu_q^2/\sigma_q^2-1)} \cdot e^{-(\mu_q/\sigma_q^2)\lambda} \cdot u_{-1}(\lambda) \quad (1)$$

where μ_q is the mean value of the chunk size, σ_q^2 is the corresponding variance value and $C_q = \left(\mu_q/\sigma_q^2\right)^{\mu_q^2/\sigma_q^2}/\Gamma\left(\mu_q^2/\sigma_q^2\right)$.

We next consider a user downloading chunks from a stream at a quality q, experiencing an average spectral efficiency $\overline{\eta(c)}$ level, and been assigned a maximum bandwidth that is set equal to B_s. Since probability that a chunk delay is observed is equal to the probability that the chunk size exceeds $\lambda_{max} = \overline{\eta(c)}B_s\alpha\tau$, the $\mathcal{P}_{\text{TD}}$ can calculated as:

$$\mathcal{P}_{\text{TD}}(B_s) = 1 - \int_{\lambda_{max}}^{\infty} p_\Lambda\left(\lambda;\mu_q,\sigma_q^2\right)d\lambda = D_\Lambda\left(\lambda_{max};\mu_q,\sigma_q^2\right)$$
$$= \frac{\left(\mu_q/\sigma_q^2\right)^{\mu_q^2/\sigma_q^2}}{\Gamma\left(\mu_q^2/\sigma_q^2\right)} \cdot \int_0^{\overline{\eta(c)}B_s\alpha\tau} \lambda^{(\mu_q^2/\sigma_q^2-1)} \cdot e^{-(\mu_q/\sigma_q^2)\lambda}d\lambda \quad (2)$$

The cumulative density function $D_\Lambda(\lambda_{max};\mu_q,\sigma_q^2)$ is expressed in terms of the lower incomplete gamma functions $\gamma(\xi,x) = \int_0^x \lambda^{\xi-1}\cdot e^{-\lambda}d\lambda$; thereby, we rewrite $\mathcal{P}_{\text{TD}}$ as:

$$\mathcal{P}_{\text{TD}}(B_s) = \frac{\left(\mu_q/\sigma_q^2\right)^{\left(\mu_q^2/\sigma_q^2-1\right)}}{\Gamma\left(\mu_q^2/\sigma_q^2\right)} \cdot \int_0^{\overline{\eta(c)}B_s\alpha\tau} \lambda^{\left(\mu_q^2/\sigma_q^2-1\right)}\cdot$$
$$e^{-\left(\mu_q/\sigma_q^2\right)\lambda}d\left(\mu_q/\sigma_q^2\right)\lambda$$
$$= \Gamma\left(\mu_q^2/\sigma_q^2\right)^{-1}\gamma\left(\mu_q^2/\sigma_q^2, \mu_q/\sigma_q^2\cdot\tau\,\alpha\,\overline{\eta(c)}\,B_s\right) \quad (3)$$

We note that Eq. (3) relates the maximum bandwidth B_s level allocated to a user to the resulting probability $\mathcal{P}_{\text{TD}}$ of timely delivery of the downloaded chunk. It thus compactly summarizes the action of the different layers involved in the communications protocol architecture. Specifically, it is parameterized by application layer parameters $(\mu_q, \sigma_q^2, \tau)$, the network and transport protocol overhead related parameter α, and the physical layer parameter $\overline{\eta(c)}$.

Eq. (3) allows us to evaluate the bandwidth B_s level that must be allocated to the transport of the stream in order to guarantee a quality q value, and assure a specified $\mathcal{P}_{\text{TD}}$ level to users that experience various CQI (e.g. cell edge user, cell center user, etc.)[1]. Furthermore, it can be used to estimate the number of users $\mathcal{N}(c,q)$ that can be accommodated within an overall B_{LTE} bandwidth for a given CQI c level and an intended average video quality q value, when the prescribed $\mathcal{P}_{\text{TD}}$ level is set to a constant (e.g, 95%) value for all users. This can in turn be exploited to device a pricing policy for servicing a mix of streams that impose different video quality requirements.

It is noted that the above analysis applies in case of no stalls. The occurrence of playout stall events induces the client to activate a re-buffering phase, whose duration and required rate are implementation-dependent and imply additional bandwidth requirements. For reasonably high values of $\mathcal{P}_{\text{TD}}$ the additional bandwidth requirements are negligible. Finally, it is assumed that the delay is generated at the radio access network where we can assume that the bottleneck is (i.e., the wireless link).

[1]Notice that given R_q, τ, μ_q and σ_q^2 can be readily computed as $\mu_q = R_q\tau$ and using the rule of thumb $\sigma_q \approx 0.7\mu_q$

TABLE I
HP QUALITY LEVELS AND CORRESPONDING VIDEO STATISTICS

q	1	2	3	4	5
R_q(HP) $kbps$	1192.11	527.7	273.65	150	81
μ_q(HP) kB	298	132	68.4	37.7	20.3
σ_q(HP) kB	191	91.1	48.3	26.5	13.8

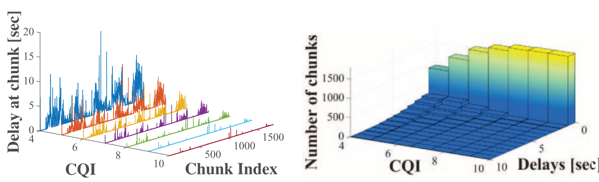

Fig. 1. Sequences and histograms of delays for HP in case of $q = 1$ and $B_s = 2.7\ MHz$.

IV. NUMERICAL RESULTS

In this section we evaluate the precision of the mathematical expression of $\mathcal{P}_{\text{TD}}$, as given by Eq. (3), with that observed by carrying out numerical computations as obtained by simulations. We first consider one real video trace [8], namely *Harry Potter* (HP), 86352 frames long; the sequence is H.265 encoded at 24 fps at $Q = 5$ different average rates R_q, as reported in Table I; each chunk corresponds to a playout time $\tau = 2$s; the HP sequence is built by 1799 chunks. The overhead due to lower layer protocols is set by observing that TCP and MAC layer retransmissions across error prone channels may often occupy up to 20% of the physical layer capacity. The net throughput at the TCP layer is therefore computed by setting $\alpha = 0.8 \times 0.8 = 0.64$. We note that under lower communications quality operations, the net TCP throughput efficiency may drop below 50% [9], so that lower α levels may then be employed. We consider the transmission of the chunks over a LTE channel characterized by randomly varying CQI levels. In the first reference scenario, the CQI observed during the video session is a realization of a stationary process with constant spectral efficiency expected value $\text{E}_{CQI}\{\eta(c)\} \in \{\eta(4), \ldots \eta(15)\}$. The delay sequence δ_k is then computed for different qualities $q = 1, \cdots 5$ and by setting $\overline{\eta(c)}$ equal to different values of $\text{E}_{CQI}\{\eta(c)\}$.

An example of the sequences of the observed delays δ_k (measured in seconds) versus the chunk index k is exhibited in Fig. 1 (left) for the $q = 1$ HP video, at $B_s = 2.7\ MHz$, and for different values of $\text{E}_{CQI}\{\eta(c)\}$. Fig. 1 (right) also shows the histograms of the delays collected at each CQI throughout the transmission of the whole sequence. Specifically, each histogram presents one bin representing the number of chunks exhibiting zero delays, i.e., $\lambda < \lambda_{max}(q, c)$, and a set of bins representing the number of occurrence of nonzero delays, i.e., chunks having $\lambda \geq \lambda_{max}(q, c)$; apart a normalization factor, the bins amplitudes follow the tails of the corresponding Gamma distributions above the threshold value $\lambda_{max}(q, c)$.

In Fig. 2, we plot the theoretical values of the $\mathcal{P}_{\text{TD}}(B_s)$ computed by using Eq. (3) (solid lines) for the HP video at different CQI levels, and the frequency of occurrence $f_{\text{TD}}(B_s)$ of timely delivered chunks measured by simulating the transmission over different bandwidth levels B_s ranging in $(0, 4.5]\ MHz$.

In Fig. 3, we plot $\mathcal{P}_{\text{TD}}(B_s)$ computed by using Eq. (3) at two CQI levels $CQI = 8$ (solid lines) and $CQI = 12$ (dash-dot lines) for different H.265 encoded high resolution videos at $q = 1$, namely HP, *Finding Neverland* (FN), (84192 frames long), *Lake House* (LH) (86352 frames long), *Speed* (SP) (86352 frames long), *Blue Planet* (BP) (61008 frames long). Over each point, we highlight the error $f_{\text{TD}}(B_s) - \mathcal{P}_{\text{TD}}(B_s)$. We recognize that the $\mathcal{P}_{\text{TD}}$ is always evaluated with a good accuracy (i.e., with a small relative error), and the error further reduces in the bandwidth range of interest for video applications, i.e., for $\mathcal{P}_{\text{TD}} \geq 0.5$.

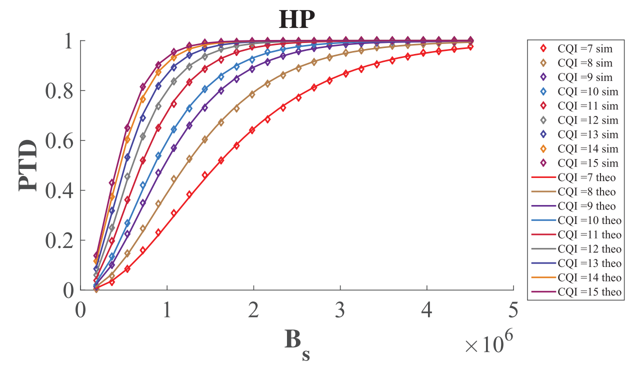

Fig. 2. $\mathcal{P}_{\text{TD}}$ versus the assigned B_s: theoretical results and simulations for different CQIs in case of $q = 1$, HP sequence.

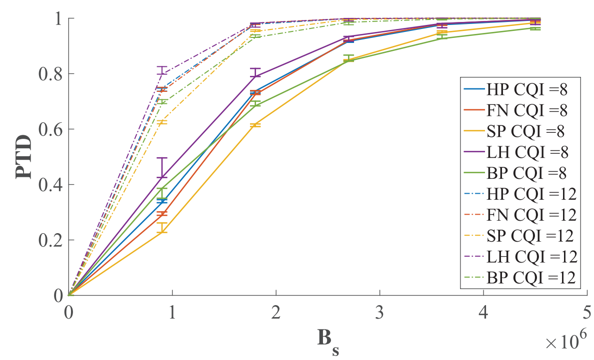

Fig. 3. $\mathcal{P}_{\text{TD}}$: theoretical results and observed errors for five different video sequences at $q = 1$ for CQI = 8 and 12.

Furthermore, we compute the percentage relative error $E(B_s) = (f_{\text{TD}}(B_s) - \mathcal{P}_{\text{TD}}(B_s))/f_{\text{TD}}(B_s) \times 100$. We present in Table II the mean and peak of $E(B_s)$ observed for B_s ranging in $(0, 4.5]\ MHz$ and for different CQIs; peaks are observed for low values of B_s, i.e., for small values of $\mathcal{P}_{\text{TD}}$, whereas for larger values of B_s the error decreases leading to a significantly smaller mean error. Both the mean and peak of $E(B_s)$ decrease for increasing CQI. Thereby, in the bandwidth/CQI range of practical interest, the metric $\mathcal{P}_{\text{TD}}$ tightly captures the relation between application layer parameters and lower layer parameters, including the user's reported CQI.

In the second reference scenario, we consider the real CQI traces acquired at a pace of $T_c = 200\ ms$ analyzed in [10]. Therein, the authors report their sample cumulative distribution function. Herein, based on these results, we infer two probabilistic models $\mathcal{M}_1$, $\mathcal{M}_2$ of the CQI samples, by specifying a CQI probability mass functions (pmf) $\mathcal{M}_1$, characterizing a trace with lower CQIs (e.g., users in a bad channel condition) and a CQI pmf $\mathcal{M}_2$ characterizing trace with higher CQIs (e.g., users in a good channel condition); details are given in Fig. 4.

TABLE II

RELATIVE PERCENTAGE ERROR (MEAN÷PEAK) OVER A BANDWIDTH RANGE (0, 4.5] MHz IN CASE OF 5 DIFFERENT VIDEOS AT $q = 1$

Video	μ_1 $[kB]$	σ_1 $[kB]$	R_1 $[kbps]$	CQI 7		8		9		10		11		12		13		14		15	
HP	298	191	1192.11	0.87	5.00	0.42	3.20	0.40	2.34	-0.01	0.65	-0.15	-0.68	-0.29	-1.45	-0.09	-0.51	-0.09	-0.22	-0.14	-0.45
FN	308	198	1230.8	-0.68	-9.56	0.89	4.09	1.56	8.14	1.02	6.73	0.55	4.86	0.15	1.94	0.07	1.75	-0.03	0.63	-0.13	-0.39
SP	363	221	1452	2.45	13.89	2.30	13.13	0.98	5.64	0.57	3.48	-0.24	-0.80	-0.28	-1.84	-0.22	-1.40	-0.38	-1.68	-0.20	-0.76
LH	264	223	1052	4.12	14.84	2.51	14.26	1.37	12.00	0.90	9.78	0.33	5.85	0.13	3.31	-0.40	-1.80	-0.64	-1.65	-0.64	-1.67
BP	328	282	1310	0.88	4.12	-0.93	-10.22	-0.79	-6.03	-0.16	2.79	0.49	2.40	0.34	2.04	0.53	3.76	0.16	2.02	0.28	2.58

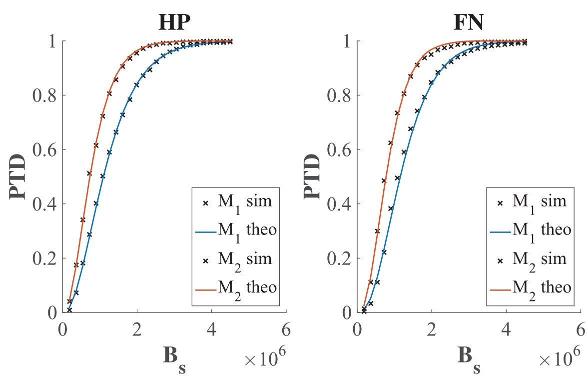

Fig. 4. $\mathcal{P}_{TD}$ versus the assigned B_s: theoretical results and simulations for the pmfs $\mathcal{M}_1$, $\mathcal{M}_2$ (HP left and FN right); $\mathcal{M}_1$: $p_{CQI=7} = 0.08$, $p_{CQI=8} = 0.5$, $p_{CQI=9} = 0.3$, $p_{CQI=10} = 0.12$, 0 otherwise; $\mathcal{M}_2$: $p_{CQI=6} = 0.07$, $p_{CQI=7} = 0.07$, $p_{CQI=8} = 0.07$, $p_{CQI=9} = 0.15$, $p_{CQI=10} = 0.15$, $p_{CQI=11} = 0.2$, $p_{CQI=12} = 0.06$, $p_{CQI=13} = 0.06$, $p_{CQI=14} = 0.06$, $p_{CQI=15} = 0.11$, 0 otherwise.

TABLE III

NUMBER OF ADMISSIBLE USERS STREAMING HP AT $\mathcal{P}_{TD} = 95\%$ FOR TWO DIFFERENT VALUES OF B_{LTE}

(a) $B_{LTE} = 4.5\ MHz$

CQI	q 1	2	3	4	5
4	0	0	1	3	6
5	0	1	2	5	8
6	0	1	3	6	12
7	1	2	4	8	12
8	1	3	5	8	12
9	1	3	6	12	25
10	2	4	8	12	25
11	2	5	8	12	25
12	2	6	8	12	25
13	3	6	12	25	25
14	3	8	12	25	25
15	4	8	12	25	25

(b) $B_{LTE} = 18\ MHz$

CQI	q 1	2	3	4	5
4	1	3	7	12	25
5	2	5	10	20	33
6	3	7	14	25	50
7	4	9	16	33	50
8	5	12	20	33	50
9	7	14	25	50	100
10	8	16	33	50	100
11	10	20	33	50	100
12	11	25	33	50	100
13	12	25	50	100	100
14	14	33	50	100	100
15	16	33	50	100	100

We then analyze the delays introduced in the transmission of the video streams over randomly varying channels by generating $L = \tau/T_c = 10$ samples per chunk, and computing the average spectral efficiency $\overline{\eta(c)}$ per chunk.

In Fig. 4, we exhibit the average values attained in 10 Monte-Carlo runs of the above described experiments, and the theoretical counterparts obtained for the same values of spectral efficiency, namely, $\overline{\eta(c)} = 1.49$, and $\overline{\eta(c)} = 2.22$ respectively. We observe the simulation results to fit very well the results obtained by using our mathematical model.

We finally discuss how Eq. (3) can be exploited to study cell dimensioning and admission procedures. Table III shows the number of users $\mathcal{N}(c, q)$ that can be accommodated within an overall B_{LTE} bandwidth for a given CQI c and a video quality q, when the $\mathcal{P}_{TD}$ is set to 95% for all the users sharing the same bandwidth. As expected, under a high system bandwidth level, typically referring to wide cell coverage, several users can be accommodated at medium to high video rates. In turn, under a low system bandwidth level, the system is able to accommodate a small number of low quality videos streams (e.g. surveillance videos) or a smaller number of higher quality videos (e.g. few multicasted videos).

V. CONCLUSIONS

The percentage of video chunks that are timely delivered during a video streaming downlink transport service is relevant to the quality experienced by the user. In this letter, we provide the closed-form expression of the probability of timely delivery of video chunks for DASH video streaming over a wireless cell (such as LTE) as a function of the downlink wireless bandwidth level allocated to the user. The result, validated by numerical simulations, can be used to infer the number of users that can be accommodated in such a wireless access (LTE) system, given a Quality of Experience (QoE) requirement that includes as metrics the probability of timely delivery and the received video stream quality level; it can also be used to perform admission procedures, bandwidth pricing policies, and cell dimensioning designs.

REFERENCES

[1] M. Rinne and O. Tirkkonen, "LTE, the radio technology path towards 4G," *Comput. Commun.*, vol. 33, no. 16, pp. 1894–1906, 2010.

[2] I. Sodagar, "The MPEG-DASH Standard for multimedia streaming over the Internet," *IEEE Multimedia*, vol. 18, no. 4, pp. 62–67, Apr. 2011.

[3] G. Tian and Y. Liu, "Towards agile and smooth video adaptation in dynamic HTTP streaming," in *Proc. 8th Int. Conf. Emerging Netw. Exp. Technol. (CoNEXT'12)*, 2012, pp. 109–120.

[4] I. Rubin, S. Colonnese, F. Cuomo, F. Calanca, and T. Melodia, "Mobile HTTP-based streaming using flexible LTE base station control," in *Proc. IEEE 16th Int. Symp. World Wireless Mobile Multimedia Netw. (WoWMoM)*, Jun. 2015, pp. 1–9.

[5] T. Stockhammer and M. Luby, "Dash in mobile networks and services," in *Proc. IEEE Vis. Commun. Image Process. (VCIP)*, Nov. 2012, pp. 1–6.

[6] M. Seufert, F. Wamser, P. Casas, R. Irmer, P. Tran-Gia, and R. Schatz, "YouTube QoE on mobile devices: Subjective analysis of classical vs. adaptive video streaming," in *Proc. IEEE Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Aug. 2015, pp. 43–48.

[7] G. Ku and J. Walsh, "Resource allocation and link adaptation in LTE and LTE advanced: A tutorial," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 3, pp. 1605–1633, Aug. 2015.

[8] P. Seeling and M. Reisslein, "Video transport evaluation with H. 264 video traces," *IEEE Commun. Surveys Tuts.*, vol. 14, no. 4, pp. 1142–1165, Oct. 2012.

[9] J. Huang *et al.*, "An in-depth study of LTE: Effect of network protocol and application behavior on performance," in *Proc. ACM SIGCOMM*, 2013, pp. 363–374.

[10] J. Fajardo, I. Taboada, and F. Liberal, "Analysis of CQI traces from LTE MIMO deployments and impact on classical schedulers," in *Wired/Wireless Internet Communications*, M. C. Aguayo-Torres, G. Gomez and J. Poncela, Eds., 2015, Switzerland: Springer International Publishing, vol. 9071, pp. 60–73.