Affordable delay based quality selection for HTTP adaptive video streaming

Stefania Colonnese*, Francesca Cuomo*, Konstantin Miller†, Vincenzo Sapio*, Adam Wolisz†
*DIET at University of Roma "Sapienza", Rome, Italy {name.surname}@uniroma1.it
†Technische Universität Berlin, Department of Electrical Engineering and Computer Science, Berlin, Germany

***Abstract*—We present a quality-selection policy for Quality of Experience (QoE) demanding video streaming in wireless networks. The proposed policy predicts the TCP throughput and adapts video segment requests in order to assure high QoE by taking into account the client buffer level. We introduce the concept of Affordable delivery Time (AT) and we design a buffer-based algorithm, hereafter referred to as Buffer-based lolyPOP (BPOP). The AT accounts for the ideal chunk download time which equals the chunk playout time, as well as for the client buffer status. In a nutshell, the buffer-based AT favors the download of higher quality video chunks when the client has more buffered data than a pre-established target buffer occupancy. Conversely, AT inhibits higher quality video chunks download when the buffer is below the targeted occupancy.**

Keywords*—*HTTP Adaptive Streaming, Quality selection, Throughput prediction, Buffer based.

I. Introduction

HTTP Adaptive Streaming (HAS) is a key technology enabling downloading-while-playing videos from a server to client with an adaptation to the network conditions [1]. Video servers store different versions of the video and the client, by means of HTTP GET requests, downloads video segments, also named *chunks*, at different video qualities q. HAS allows switching from a quality level to another on the basis of the network (e.g., throughput) and system conditions (e.g. client buffer level) so as to provide uninterrupted services in challenging environments [2] where the throughput randomly and suddenly changes.

The Quality of Experience (QoE) perceived by end users of HAS services is related to a few metrics observed during the whole video playback, such as the average quality level, the number of quality switching and the number and the duration of video playback interruptions (stalls).

In recent literature, several quality selection approaches can be found (see [3] and the therein references for a thorough review), each one leveraging a few system parameters and differently trading off the improvement of the average quality level with the reduction of the number of quality switching and of the stalls. The segment-aware quality selection algorithm SARA [4], considers the segment size variation in addition to the estimated throughput and the current buffer occupancy to accurately predict the time required to download the next segment. However, SARA makes a deterministic use of the known/estimated system parameters, such as buffer status, next fragment throughput, and it exhibits somewhat limited performance in presence of throughput random variations. The work in [3] designs a novel prediction-based algorithm called Low-Latency Prediction-Based Adaptation (LOLYPOP) able to support quality-based adaptation of quality switching, measured in terms of required probability with a transport latency on the order of a few seconds.

Inspired by the approach in [3], in this paper we present a quality selection policy that selects the requested video quality based on the notion of Affordable delivery Time (AT). The AT accounts for the ideal chunk download time, which equals the chunk playout time, as well as for the client buffer status. We named our approach Buffer-based lolyPOP (BPOP).

II. System Modeling and BPOP Algorithm

We consider an HAS architecture where, at the server side, multiple versions of a source video content are stored, each encoded for reception at its target video quality level q, parsed in chunks corresponding to τ seconds of video. The chunk size $s_i^{(q)}$ varies with the time i and the quality q.

Initially, and after any rebuffering, the buffer is typically filled at a given initial value b_0; then, in order to correctly play the video, the client should receive, every τ seconds, a chunk available in the receiving buffer.

We extend the LOLYPOP algorithm in [3] by computing the affordable time delivery for a generic chunk in accordance to the buffer status at time i, namely b_i. When b_i exceeds the target occupancy b_{tg}, the delivery may introduce a short delay with respect to the ideal delivery time τ by consuming a small amount of buffer, whereas when $b_i < b_{tg}$, the download should occur in a reduced time. We then introduce the concept of Affordable delivery Time: $AT_i = \tau + \Delta T(b_i)$, being $\Delta T(b_i)$ the deviation of the sustainable delay with respect to the ideal constant delivery time τ.
The value of $\Delta T(b_i)$ is computed as:

$$\frac{\Delta T(b_i)}{\tau} = (b_i - b_{tg}) \cdot w(b_i - b_{tg}) \tag{1}$$

where $w(b_i - b_{tg})$ is the parametric weighting function:

$$w(b_i - b_{tg}) = \begin{cases} \left(\dfrac{b_i - b_{tg}}{b_{min} - b_{tg}}\right)^p, & b_i \in [b_{min}, b_{tg}] \\ \left(\dfrac{b_i - b_{tg}}{b_{max} - b_{tg}}\right)^p, & b_i \in [b_{tg}, b_{max}] \end{cases} \tag{2}$$

which ranges in $[0, 1]$ and measures the deviation of the buffer status b_i from the target occupancy b_{tg} with respect to the minimum and maximum required buffer load b_{min} and b_{max}.

The time needed to download the i-th video chunk at quality q in case of a throughput ρ_i is $s_i^{(q)}/\rho_i$. The chunk

978-1-5386-0728-2/17/$31.00

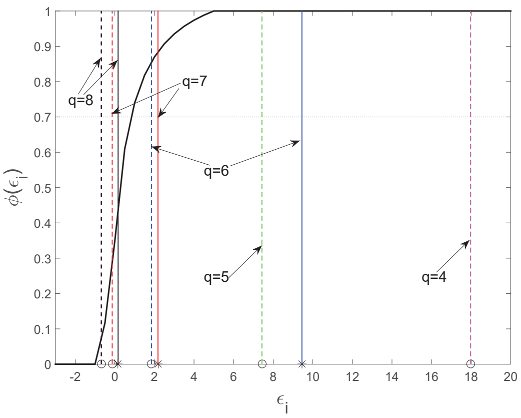

Fig. 1. Cumulative distribution function of the throughput prediction normalized estimation error ($b_i > b_{tg}$); values $\epsilon_i^{(q)}, q = 8, 7, 6$ (star markers, BPOP), values $\hat{\rho}_i \cdot \tau / s_i^{(q)} - 1, q = 8, 7, 6, 5, 4$ (circle markers, LOLYPOP).

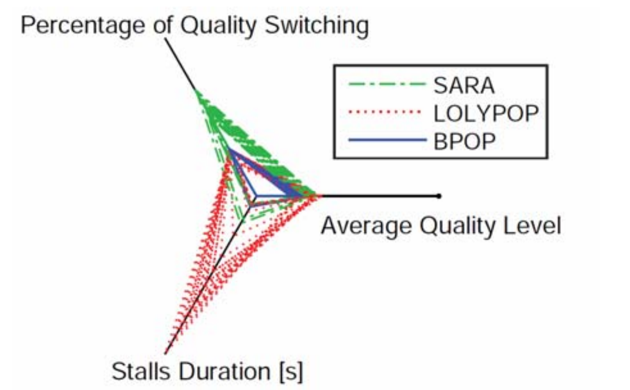

Fig. 2. QoE comparison for i) BPOP ($\Pi^{(SW)} = 0.3$, $P_{Fail} = 0.3$, $b_{min} = 3$, $b_{max} = 15$, $b_{tg} = 6$ and $p = 1$ in Eq. (1)); ii) LOLYPOP [3] ($\Pi^{(SW)} = 0.3$, $P_{Fail} = 0.3$); iii) SARA [4] ($I = 6$, $B_\alpha = 10$, $B_\beta = 15$ and $B_{max} = 17$;throughput predicted as the harmonic average on a 2-chunk window). For all three algorithms $\tau = 2\ s$, $b_0 = 3$.

delivery is to be considered timely as long as the download time is within the AT, i.e. $s_i^{(q)}/\rho_i \leq AT_i$. Thereby, in BPOP we identify the probability of being compliant with the AT as:

$$P_i^{(q)} = \mathbf{Pr}\Big\{\frac{s_i^{(q)}}{\rho_i} \leq AT_i\Big\} = \mathbf{Pr}\Big\{\epsilon_i \leq \frac{\hat{\rho}_i\ AT_i}{s_i^{(q)}} - 1\Big\} = \phi_i\left(\frac{\hat{\rho}_i\ AT_i}{s_i^{(q)}} - 1\right). \quad (3)$$

where we approximate the true unknown throughput ρ_i with a predicted value $\hat{\rho}_i$, and characterize the normalized prediction error $\epsilon_i = (\hat{\rho}_i - \rho_i)/\rho_i$ by inline updating the prediction error cumulative distribution function $\phi_i(\epsilon_i)$. See Fig. 1 for a comparison of the intercept values of the CDF in case of LOLYPOP and BPOB, $\phi(\hat{\rho}_i\ AT_i/s_i^{(q)} - 1)$ are systematically larger than the intercept values $\phi(\hat{\rho}_i\ \tau/s_i^{(q)} - 1)$ since $b_i > b_{tg}$ and AT_i accounts for a tolerated buffer depletion. The converse would happen for $b_i \leq b_{tg}$.
Once computed $P_i^{(q)}$ for all available Q qualities at the server, BPOP chooses the highest quality for which $P_i^{(q)}$ exceeds a threshold $1 - P_{Fail}$ (P_{Fail} identifies the maximum timely download failure probability); the overall switching behavior is finally constrained by the threshold on the percentage of quality transitions, namely $\Pi^{(SW)}$, to avoid too many switches.

To sum up, the BPOP firstly, for the i-th video segment, computes: $\epsilon_i^{(q)} = \hat{\rho}_i\ AT_i/s_i^{(q)} - 1 = \hat{\rho}_i\ [\tau + \Delta T(b_i)]/s_i^{(q)} - 1$, for each and every quality $q = 1, \cdots Q$. Secondly, it selects for the ith chunk the quality:

$$q_i^{(BPOP)} = \begin{cases} q^* & \text{for } \pi_i^{(SW)} \leq \Pi^{(SW)} \\ \min\left\{q^*, q_{i-1}^{(BPOP)}\right\} & \text{for } \pi_i^{(SW)} > \Pi^{(SW)} \end{cases} \quad (4)$$

where

$$q^* = \arg\max_q \left\{q\ s.t.\ \phi_i(\epsilon_i^{(q)}) \geq 1 - P_{Fail}\right\} \quad (5)$$

denotes the maximum quality compatible with the affordable time AT.

III. Performance Evaluation

We compare BPOP, LOLYPOP and SARA, for different configurations of the parameters of the quality selection algorithms. For each parameter settings we consider 5 different runs using 5 different real throughput Wi-Fi traces and randomly selecting a sequence of 900 chunks (1800 s) out of 2 real video traces [5].

On each run, we collect the visual quality, the percentage of quality switching and the number of stalls observed, and plot them as triangle vertices in Fig.2; the best QoE performance are obtained on the smallest triangles in the plot. We notice that BPOP almost uniformly outperforms the other methods in terms of the three quality metrics, since it exploits the buffered chunks while maintaining a good margin over the buffer depletion state.

IV. Conclusions and Future Work

In this work, inspired by a throughput prediction based algorithm (LOLYPOP) [3] to select the video quality in HAS services, we propose a Buffer-based lolyPOP (BPOP). The BPOP quality-selection algorithm leverages predictions of TCP throughput distributions and combines it with the knowledge of the client buffer status by introducing the concept of affordable delivery time (AT). We show by numerical simulations that BPOP i) improves the average selected quality and ii) stabilizes the buffer occupancy, reducing the number of stalls.

References

[1] M. Seufert, S. Egger, M. Slanina, T. Zinner, T. Hobfeld, and P. Tran-Gia, "A survey on quality of experience of HTTP adaptive streaming," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 1, pp. 469–492.

[2] A. H. Zahran, J. J. Quinlan, K. K. Ramakrishnan, and C. J. Sreenan, "Impact of the LTE scheduler on achieving good QoE for DASH video streaming," in *IEEE LANMAN*, pp. 1–7, June 2016.

[3] K. Miller, A.-K. Al-Tamimi, and A. Wolisz, "QoE-Based Low-Delay Live Streaming Using Throughput Predictions," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 13, pp. 4:1–4:24, Oct. 2016.

[4] P. Juluri, V. Tamarapalli, and D. Medhi, "SARA: Segment aware rate adaptation algorithm for dynamic adaptive streaming over HTTP," in *IEEE ICC workshop*, pp. 1765–1770, June 2015.

[5] P. Seeling and M. Reisslein, "Video Transport Evaluation With H.264 Video Traces," *IEEE Communications Surveys Tutorials*, vol. 14, pp. 1142–1165, Fourth 2012.