Efficient Delivery of Multimedia Streams Over Broadband Networks Using QOAS

Gabriel-Miro Muntean, Member, IEEE

Abstract—Quality Oriented Adaptation Scheme (QOAS) is compared against other adaptive schemes such as TCP Friendly Rate Control Protocol (TFRCP), Loss-Delay-based Adaptation Algorithm (LDA+), and a nonadaptive (NoAd) solution when streaming multiple multimedia clips with various characteristics over broadband networks. Streaming efficiency is assessed in terms of loss rate, bandwidth utilization, number of concurrent clients and end-user perceived quality. Simulation results show that using QOAS a significantly higher number of simultaneous clients can be served than when using the other schemes given a target average end-user quality. This is while having higher bandwidth utilization. Testing results also indicate that higher performance is achieved when streaming to the same number of clients using QOAS than when other solutions are used.

Index Terms—Adaptive streaming, congestion control, end-user quality, multimedia.

I. INTRODUCTION

WITH the latest significant increase in the penetration of broadband connectivity to home residences and business premises that offers support for streamed high quality multimedia-related IP services (e.g. digital and interactive TV, Video on Demand (VOD), gaming, videoconferencing, etc.), these services are becoming more popular among the users and further contribute to the success of the all-IP infrastructure [1]. However, these IP-based services have important bandwidth requirements and significantly contribute to the total traffic carried by the infrastructure, putting pressure on it.

In order to increase their revenues, both network operators and service providers aim to highly utilize their infrastructure and have large numbers of customers. This should cause traffic congestion and therefore decrease the quality of the multimedia-based services provided. At the same time, customers want to receive diverse services at high quality and low cost. The Quality Oriented Adaptation Scheme (QOAS) [2], [3] was such designed that it successfully balances these opposing goals.

This paper presents testing results that further evaluate QOAS performance in terms of loss rate, link utilization, number of clients simultaneously served and end-user perceived quality. QOAS is compared with other streaming approaches such as TCP Friendly Rate Control Protocol (TFRCP) [4], Loss-Delay-based Adaptation Algorithm (LDA+) [5], and a nonadaptive (NoAd) approach and consistently achieves better results.

Different solutions have been proposed to ensure quality of service while streaming multimedia content over IP networks [6]. Among them, adaptive streaming schemes [7] provide the least complex and the most flexible and easy to deploy mechanisms. Based on where the adaptive decision takes place, there are sender-driven schemes such as TFRCP [4], LDA+ [5], Rate Adaptation Protocol (RAP) [8] or Layer Quality Adaptation (LQA) [9], receiver-driven solutions such as Receiver-driven Layered Multicast (RLM) [10] or Receiver-driven Layered Congestion control scheme (RLC) [11], or transcoding-based solutions (e.g. filters [12]). Unfortunately none of these solutions take into account end-user perceived quality directly in their adjustment process. In contrast, QOAS includes an estimation of end-user perceived quality in its adaptation process, enabling it to achieve higher overall performance.

QOAS principle and algorithm are briefly described next, whereas the results of QOAS simulation-based testing are presented and discussed in the section that follows. The performance analysis of the streaming solutions compared in this paper is presented in Section IV. Conclusions are drawn in the last section of this paper, which also presents the work in progress and indicates some possible future work directions.

II. QUALITY-ORIENTED ADAPTATION SCHEME (QOAS)

A. QOAS—Overview

QOAS is an unicast rate-based adaptive scheme for multimedia streaming that maximizes end-user perceived quality or user Quality of Experience (QoE) in existing delivery conditions [2]. It includes client and server-located components that are involved in the bi-directional exchange of video data and control packets through the delivery network. The client monitors the transmission and user QoE-related parameters using the Quality of Delivery Grading Scheme (QoDGS). QoDGS regularly computes the quality of delivery scores, which are sent as feedback to the server. The Server Arbitration Scheme (SAS) analyzes these scores and proposes adjustment decisions in order to increase user QoE in existing delivery conditions. This is performed as research has shown that viewers of streamed multimedia content prefer controlled reduction in quality to the effect of random losses [13].

QOAS defines a number of different server states that are assigned to a different stream quality during each streaming session. For example a five-state model was used for the experimental tests presented in this paper. The stream quality versions differ in terms of compression-related parameters (e.g. resolution, frame rate, color depth) and therefore have different bandwidth requirements. During transmission the server dynamically varies its state according to the client QoDGS feedback. For example, when the client reports a decrease in end-user quality, the server switches to a lower quality state, which reduces the quantity of data sent. In improved conditions, the

Manuscript received February 16, 2005; revised November 24, 2005.
The author is with the School of Electronic Engineering, Dublin City University, Dublin 9, Ireland (e-mail: munteang@eeng.dcu.ie).
Digital Object Identifier 10.1109/TBC.2006.875612

0018-9316/$20.00 © 2006 IEEE
Fig. 1. Schematic description of QOAS’s adaptation principle.

The client-located **QoDGS** monitors and evaluates the effect of the delivery conditions on end-user perceived quality. Regularly QoDGS grades the multimedia streaming process by computing Quality of Delivery scores (QDscores). The grading process is based on monitoring and assessing both short-term and long-term variations of packet loss rate, delay, and delay jitter, which have been shown to have a significant impact on the received quality. Short-term monitoring is important for learning quickly about transient effects, such as sudden traffic changes, and for quickly reacting to them. The long-term variations are monitored in order to track slow changes in the overall delivery environment, such as new users in the system. QoDGS also takes into account user QoE as measured by the no-reference moving picture quality metric Q [14], which maps the joint impact of bitrate and data loss on video quality onto the ITU-T P.910 five-point grading scale [15]. The calculated QDscores are sent as feedback to the server. More details about QoDGS are presented in [2].

The server-located **SAS** considers the values of a number of consecutive QDscores from the client and, by averaging these values, asymmetrically suggests adjustment decisions. It requires fewer scores to trigger a quality decrease than for a quality increase, ensuring a fast reaction during bad delivery conditions and helping to eliminate its cause. An increase is performed only when the network conditions have improved. This asymmetry helps also to maintain system stability, by reducing the frequency of quality variations. The asymmetry is forced by collecting feedback scores during periods of time of different length for upgrades and respectively downgrades in transmitted quality.

**B. QOAS—Deployment**

Fig. 2 presents the high-level architecture of the system that enables adaptive delivery of multimedia streams. It involves the deployment of the server-side QOAS component at the Multimedia Server level, whereas the client-side QOAS component at the Multimedia Client level. In order to adaptively react fast enough to the highly dynamic variations of the delivery conditions when streaming over wireless networks, there is a need for accurate information from the client at all times. Therefore QOAS employs a very high feedback frequency with small feedback report packets (40 B) that are sent every 100 msec. This value balances the need for the most up-to-date information with the requirement of low overhead. Both the QoDGS short-term and long-term monitoring periods and are respectively set to an order and two orders of magnitude greater than the feedback-reporting interval.

Adaptive decisions must also to be taken quickly and therefore SAS upgrade period was set to 6 sec whereas the downgrade timeout used was 1 sec. These values ensure both protection against any noise that may occur in the grading scheme and the QoDGS’s asymmetry in the grading process.

**III. EXPERIMENTAL RESULTS**

**A. Simulation Models, Setup and Video Sequences**

The experimental tests consisted of simulations using models for QOAS, TFRCP, LDA+ and a nonadaptive solution, built using Network Simulator version 2 (NS-2) [16]. The “Dumbbell” topology (Fig. 3) used for simulations assumes a single shared bottleneck link with 100 Mbps bandwidth and 100 millisecond delay. The other links are over-provisioned so that the only packet drops and significant delays are caused by congestion that occurs on the bottleneck link. The buffering at the bottleneck link uses a drop-tail queue of size proportional to the product of round trip time and bottleneck link bandwidth.

**Non-adaptive (NoAd)** streaming transmits multimedia streams using the highest available rate, regardless of the eventual delivery problems signalized by packet loss, increased delays, etc. During these tests a maximum rate of 4 Mb/s was used.

**TFRCP** [4] uses estimates of round-trip delay and loss rates to determine its adaptation policy. When there are losses, the
rate is limited to the one computed by the TCP model; in cases of zero loss, the current transmission rate is doubled. The sender updates its rate in intervals of M units. The TFRCP model used has \( M = 5 \) sec as suggested in [4] for delays greater than 100 msec, as in this setup.

\textbf{LDA+} [5] adaptation relies on estimates of network condition and each individual stream's bandwidth share. In zero loss periods, the sender increases its rate by the value computed from an estimated bandwidth share rate increase, a bottleneck bandwidth share rate limit and a corresponding TCP rate update. In nonzero loss periods, the server reduces its rate depending on the current rate and a TCP model-computed rate. The LDA+ model used has an RTCP feedback interval of 5 sec as suggested in [5].

\textbf{QOAS model} follows the principle described in Section II and detailed in [2]. It uses a SAS upgrade period of 6 sec and a downgrade timeout of 1 sec. QoDGs short-term period was set to 1 sec, and the long-term period was 10 sec.

Five-minute long video sequences were selected from movies with different degrees of motion content: \textit{DH}—high, \textit{JP}—average, \textit{DW}—average/low, \textit{FM}—low and \textit{RE} (cartoons)—average/high. The clips were MPEG-2 encoded at five different rates between 2 Mbps and 4 Mbps using the same frame rate (25 frames/sec), resolution (320 \times 240) and IBBP frame pattern (9 frames/GOP). MPEG-2 was chosen as major industrial users and providers of digital multimedia including the Digital Video Broadcasting (DVB) consortium [17] make use of it for the delivery of multimedia-based services. Table I presents some statistics about these clips.

\begin{table}[h]
\centering
\caption{Peak/Mean Rate Ratio for All Quality Versions of All Multimedia Clips Used During Simulations}
\begin{tabular}{|c|c|c|c|c|c|}
\hline
Quality Version (average rate) Clip Name & 2.0 Mb/s version & 2.5 Mb/s version & 3.0 Mb/s version & 3.5 Mb/s version & 4.0 Mb/s version \\
\hline
\textit{DH} & 7.48 & 7.43 & 6.31 & 5.65 & 4.06 \\
\textit{RE} & 6.91 & 6.51 & 6.23 & 6.12 & 6.05 \\
\textit{DW} & 5.56 & 4.51 & 4.36 & 4.08 & 3.56 \\
\textit{JP} & 4.83 & 4.38 & 4.04 & 3.71 & 3.41 \\
\textit{FM} & 3.99 & 3.67 & 3.42 & 3.09 & 2.93 \\
\hline
\end{tabular}
\end{table}

\subsection*{B. Simulation Scenarios and Results}

The simulations involved a number of clients randomly selecting both the movie clip and the starting sequence from within the chosen clip. Multiple unicast sessions were established and the resulting video streaming processes began and ended during transitory periods of 50 sec, which were not taken into account when analysing the results. The length of the stable periods taken into account in each case was 150 sec.

The QOAS, TFRCP, LDA+ and NoAd approaches were used in turn as the video streaming method in identical delivery conditions. The number of clients was gradually increased above a base line of 23 in each case and the performance-related data was collected and compared. This number of clients was chosen because it allowed for lossless streaming and maximum end-user perceived quality with each of the four streaming approaches. The number of simultaneous clients was increased so that the delivery network becomes increasingly high loaded. This puts pressure on all streaming solutions and their delivery performance was analyzed. Performance was assessed in terms of average end-user perceived quality, average loss rate and average link utilization in all cases. The end-user perceived quality of streamed multimedia clips was measured by the no-reference moving picture quality metric Q [14] on the ITU-T R P.910 five-point grading scale [15]. Q was computed at regular intervals during the streaming session and the resulting values were averaged in time.

Table II shows comparative performance-related statistics for all tested streaming approaches when choosing “fair” and “good” subjective quality levels as targets for the end-user perceived quality (4 and 3 on the ITU-T five-point scale). The increases in the number of clients are computed relative to the NoAd case. Since no post-processing techniques were applied, the “fair” level was considered here as the minimum quality level of interest. However further increases in the number of clients could be achieved by using for example some error concealment methods to mask the resulting losses that would otherwise severely affect the end-users’ perceived quality.

The results presented in Table II show that for the same average end-user quality, “fair” or “good” in these examples, QOAS can accommodate a significantly higher number of simultaneous clients while achieving higher bandwidth utilization. For example, to maintain a “good” perceptual quality level, by using QOAS 23\% more clients could be served than by using TFRCP, 33\% more clients than by using LDA+, and 39\% more users than by using the NoAd solution. If the goal is to maintain a “fair” average quality level for the clients, the benefit of using QOAS is 26\% greater than TFRCP, 13\% greater than LDA+, and 42\% greater than NoAd. The results are even more impressive if compared to the NoAd scheme as in the table. In terms of efficient usage of available bandwidth, QOAS was superior at all times to TFRCP and LDA+-based streaming, but inferior to NoAd. However when using the latter the end-user perceived quality decreases significantly. Figs. 4, 5 and 6 present in graphical form these comparisons in terms of loss rate, link utilization and number of simultaneous clients.

Comparing the schemes’ performances for the same number of clients, the average end-user quality is always higher for QOAS than for the other solutions tested. Table III presents comparative performance results for these tested schemes obtained during some of the performed tests when streaming multimedia to certain numbers of simultaneous clients. The results are presented in terms of average and standard deviation values.

Successive t-tests assuming unequal variances were performed on the loss rate, link utilization and estimated perceived quality values in order to compare the QOAS’s results with those obtained when TFRCP, LDA+ and NoAd streaming schemes respectively were used. In terms of loss rate, it can be said that there is very significant statistical difference between QOAS and TFRCP (significant level \( \alpha = 0.04 \)), as well as between QOAS and LDA+ (significant level \( \alpha = 0.04 \)) and between QOAS and NoAd (significant level \( \alpha = 0.01 \)) in favor of QOAS. Also in terms of end-user perceived quality QOAS’s results are statistically better than those of TFRCP (significant level \( \alpha = 0.04 \)), LDA+ (significant level \( \alpha = 0.02 \)) and NoAd (significant level \( \alpha = 0.01 \)) respectively. When comparing QOAS with the
other schemes in terms of link utilization, it can be said that there is a statistical difference between QOAS and TFRCP (confidence level of 99.9%) and between QOAS and LDA+ (confidence level of 99%) in favor of the former. It cannot be confirmed any statistical difference between the results of QOAS and NoAd in terms of link utilization.

Both TFRCP and LDA+ seem to perform better for very high loads (when their loss situation behavior is applied) than for an average number of clients when loss and zero-loss periods alternate. In comparison, QOAS has a linear and more predictable response to an increase in the number of clients, which is a significant advantage of the QOAS scheme. In this way QOAS facilitates the choice of network load level according to economic, technical, and quality goals.

However, QOAS was designed for local broadband multiservice IP-networks and therefore it seems likely that it will be used by service providers and network operators in order to maximize their revenues from offering VoD services to an increased number of clients while delivering a target quality level. For example, by scaling these simulation results with the “good” target quality level to a one-gigabit Ethernet connection, QOAS could service 320 simultaneous users compared to only 260 using TFRCP, 240 using LDA+, and 230 using NoAd streaming.

Simulation estimations of end-user perceived quality were confirmed by perceptual TEST results that were reported in [18].

### IV. PERFORMANCE ANALYSIS

Testing has shown that QOAS brings significant performance gains especially in terms of end-user perceived quality. This advantage comes with a cost in terms of extra processing, memory usage and bandwidth used for feedback in comparison with non-adaptive streaming.

QOAS involves extra processing both at the clients and at the server. QOAS clients monitor network-related parameters such as loss, delay and jitter, estimate end-user perceived quality, compute $QoD_{scores}$ and send feedback to the server. QOAS server simply averages the $QoD_{scores}$ received via feedback from the clients and takes adaptive decisions. Like QOAS, both TFRCP and LDA+ involve computation of loss rate and round trip delay and sending feedback to the server. TFRCP and LDA+ servers process the feedback and determine via complex computation the data transmission rate. As QOAS’s computation of the estimated end-user perceived quality and of the $QoD_{scores}$ has similar complexity to that required by TFRCP and LDA+’s algorithms to compute the transmission rate, it could only be noticed that QOAS’s computation is distributed among the receivers, significantly reducing the server load. However the schemes require

### TABLE II

<table>
<thead>
<tr>
<th>Streaming Scheme</th>
<th>QOAS “fair”</th>
<th>“good”</th>
<th>TFRCP “fair”</th>
<th>“good”</th>
<th>LDA+ “fair”</th>
<th>“good”</th>
<th>NoAd “fair”</th>
<th>“good”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loss rate (%)</td>
<td>1.39</td>
<td>0.47</td>
<td>1.73</td>
<td>0.53</td>
<td>1.31</td>
<td>0.50</td>
<td>0.81</td>
<td>0.01</td>
</tr>
<tr>
<td>Link utiliz. (%)</td>
<td>95.7</td>
<td>96.4</td>
<td>84.1</td>
<td>87.1</td>
<td>86.9</td>
<td>93.4</td>
<td>94.7</td>
<td>90.0</td>
</tr>
<tr>
<td>No. clients</td>
<td>34</td>
<td>32</td>
<td>27</td>
<td>26</td>
<td>30</td>
<td>24</td>
<td>24</td>
<td>23</td>
</tr>
<tr>
<td>Increase in no. of clients (%)</td>
<td>41.7</td>
<td>39.1</td>
<td>12.5</td>
<td>13.0</td>
<td>25.0</td>
<td>4.4</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
that these computations are performed at regular intervals that normally are 100 ms intervals for QOAS, and 5 s for TFRCP and LDA+. Yet, the computation of QOAS’s QOAScore is performed incrementally saving significant processing power, unlike the other two adaptive solutions. It is difficult to assess the schemes’ required memory usage, as it is highly dependent on the implementation. Also as the memory becomes cheaper, this issue becomes less significant.

However as all the schemes use feedback and the bandwidth is a highly limited resource, the bandwidth used for sending control data is compared. TFRCP uses acknowledgment for every packet, excessively using available bandwidth. LDA+ uses RTCP packets for feedback carrying information such as loss rate, round trip delay, etc. RTCP feedback usually takes up 5% of bandwidth. Unlike them QOAS sends only a loss rate, round trip delay, etc. RTCP feedback usually takes up 0.1% of the available bandwidth for feedback. This value is suggested by RTP/RTCP in [19] and used by LDA+. Yet, the computation of QOAS feedback accounts for only 0.1% of total bandwidth, which is not significant in the economy of multimedia streaming over broadband networks.

Currently a multicast extension of the QOAS is being devised and testing details will be reported soon. Further work will focus on exploring QOAS’s degree of TCP-friendliness and will compare it to those of the other schemes. This research will also address the relationship between the degree of TCP-friendliness and the adaptiveness of these schemes.

Research that focuses on testing QOAS-based multimedia streaming over WAN and wireless links respectively and assessing the eventual performance benefits is also in progress.

V. CONCLUSIONS AND FURTHER WORK

Performance assessment of the Quality Oriented Adaptation Scheme (QOAS) in comparison with three other solutions for multimedia streaming: TFRCP, LDA+, and a nonadaptive (NoAd) approach is reported in this paper. Simulations that aimed at testing the adaptive behavior while streaming via highly loaded broadband networks show that for the same end-user quality targeted, QOAS can stream to a significantly higher number of simultaneous clients while also having higher bandwidth utilization. When involving the same number of clients, the loss rate is always lower and the average end-user quality is significantly higher for QOAS than for the other streaming solutions taken into account in this paper. These benefits come with a price paid in terms of bandwidth used by feedback messages. However the paper shows that the QOAS feedback accounts for only 0.1% of total bandwidth, which is not significant in the economy of multimedia streaming over broadband networks.

TABLE III

<table>
<thead>
<tr>
<th>Scheme</th>
<th>No. of clients</th>
<th>Loss Rate</th>
<th>Link Utiliz.</th>
<th>Perceived Quality</th>
<th>Loss Rate</th>
<th>Link Utiliz.</th>
<th>Perceived Quality</th>
<th>Loss Rate</th>
<th>Link Utiliz.</th>
<th>Perceived Quality</th>
<th>Loss Rate</th>
<th>Link Utiliz.</th>
<th>Perceived Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>QOAS</td>
<td>23</td>
<td>0.00 0.00</td>
<td>90.04 4.56 0.01</td>
<td>0.00 0.00 89.54 4.56 0.01</td>
<td>0.00 0.00 89.12 4.56 0.01</td>
<td>0.00 0.00 90.04 4.56 0.01</td>
<td>12.34 1.19 99.43 1.00 0.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>26</td>
<td>0.00 0.00</td>
<td>94.34 4.51 0.01</td>
<td>0.53 0.04 87.06 3.86 0.08</td>
<td>2.19 0.18 90.28 1.91 0.17</td>
<td>23.57 3.28 100.00 1.00 0.00</td>
<td>50.00 - 100.00 1.00 0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>27</td>
<td>0.05 0.03</td>
<td>93.68 4.42 0.02</td>
<td>1.73 0.27 84.13 2.58 0.22</td>
<td>4.77 0.61 85.18 1.00 0.01</td>
<td>12.34 1.19 99.43 1.00 0.02</td>
<td>50.00 - 100.00 1.00 0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>32</td>
<td>0.47 0.17</td>
<td>96.38 4.01 0.04</td>
<td>4.82 0.51 85.42 2.62 0.21</td>
<td>1.82 0.53 88.28 1.00 0.02</td>
<td>50.00 - 100.00 1.00 0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>35</td>
<td>1.11 0.31</td>
<td>97.06 3.28 0.09</td>
<td>4.35 0.57 86.18 1.00 0.01</td>
<td>1.59 0.49 91.04 2.87 0.09</td>
<td>50.00 - 100.00 1.00 0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>39</td>
<td>1.38 0.42</td>
<td>99.07 3.06 0.15</td>
<td>2.83 0.35 91.59 1.93 0.18</td>
<td>1.57 0.52 92.88 2.93 0.14</td>
<td>50.00 - 100.00 1.00 0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

REFERENCES


