Quality-Oriented Adaptation Scheme (QOAS) for High Bit-rate Multimedia Streaming

Gabriel-Miro Muntean*, Philip Perry* and Liam Murphy**

* School of Electronic Engineering, Dublin City University, Glasnevin, Dublin 9, Ireland,
  Phone: ++ 353 (0) 1 700 7648, Fax: ++ 353 (0)1 700 5508, E-mail: {munteang, perryp}@eeng.dcu.ie,
  WWW: http://www.eeng.dcu.ie/~{munteang, perryp}

** Computer Science Department, University College Dublin, Belfield, Dublin 4, Ireland,
  Phone: ++ 353 (0) 1 716 2914, Fax: ++ 353 (0)1 269 7262, E-mail: liam.murphy@ucd.ie
  WWW: http://www.cs.ucd.ie/staff/lmurphy/default.htm

Abstract — A Quality Oriented Adaptation Scheme (QOAS) for multimedia streaming in local networks is introduced, and its performance is compared with other existing solutions (TFRCP, LDA+, and non-adaptive). QOAS is designed to balance two opposing goals: providing the highest quality to the end-users, while increasing the network operators’ revenues by increasing the number of simultaneous customers. Simulation results show that for the same average end-user quality, our QOAS system can accommodate a significantly higher number of simultaneous clients while also having higher bandwidth utilization. For the same number of clients, the average end-user quality is always higher for QOAS than for the other solutions studied.

Keywords: Adaptive video streaming, multimedia networking, feedback control, end-user perceived quality.

I. INTRODUCTION

Currently there is a trend in multimedia presentation [1] towards on-demand-based access to rich media and very high quality multimedia to home residences via an all-IP infrastructure [2, 3]. The success or failure of this approach depends on widespread market acceptance, which in turn depends on both the end-user quality of service and the price the users must pay. Network operators and service providers aim for high infrastructure utilization and a large number of customers to increase their revenues. At the same time, customers are interested in receiving high quality streamed multimedia, having access to diverse services, and paying a low cost. This paper analyses how the Quality Oriented Adaptation Scheme (QOAS) [4, 5, 6] balances these opposing goals in comparison to some well-known streaming approaches like TFRCP [7], LDA+ [8], or a non-adaptive approach. This comparison is done in terms of bandwidth utilization, number of concurrent clients, loss rate, and end-user perceived quality.

First some existing adaptive solutions for video streaming are mentioned and then QOAS is described. The simulation model and network topology used for testing are presented, then the test scenarios and their results are analysed. Some conclusions are drawn in the last section, highlighting potential performance benefits of QOAS-based streaming of high quality video to residential users and business premises, and further work directions are indicated.

II. RELATED WORK

Extensive research has proposed different solutions for ensuring a certain Quality of Service (QoS) while streaming video over IP-based networks, including adaptive schemes [9]. The adaptive solution proposed in [10] varies some encoding-related parameters at the server according to feedback from clients that monitor some transmission-related parameters only. The work in [11] describes a layered encoding-based adaptive solution, while in [12] on-the-fly transcoding is used to meet the clients’ requirements. [13] presents a more general solution based on filters deployed in the distribution network, and [14] proposes a receiver driven adaptation scheme based on multicast groups.

Recently, different rate adjustment based solutions for adaptively streaming video have been proposed, such as a protocol that manages its window size in a similar manner to TCP, but does not retransmit lost packets [15]. Limitations include its inflexibility and its problems with time sensitive media. The Loss-Delay based Adjustment algorithm (LDA) [16] uses RTCP reports to estimate round trip delays and loss rates, a packet-pair technique to estimate the bottleneck link bandwidth, and some user-initialized parameters. The enhanced Loss-Delay Adaptation algorithm (LDA+) [8] also makes use of RTCP reports to collect loss and delay statistics, and adjusts the transmission rate in a TCP-like manner subject to equal losses and delays. The Rate Adaptation Protocol (RAP) [11] uses TCP-like packet acknowledgements to estimate loss rates and delays. When there is no loss, the rate is additively increased as a function of round trip delay, otherwise the rate is halved as in TCP. In [7] a TCP-Friendly Rate Control Protocol (TFRCP) is presented, based on a TCP model previously proposed in [17]. When
there are losses, the rate is limited to that computed according to the TCP model, otherwise the rate is doubled. TFRCP’s major problem is that it updates its rate every M time units and changes in traffic that occur on a faster scale are taken into account too late.

Commercial adaptive streaming solutions like Real Networks’ SureStream [18] and Microsoft’s Multimedia Multi-bitrate (MBR) solution [19] are proprietary and detailed technical information has never been revealed. However the available information states that they were specially designed to allow for adaptations at very low bitrates, unlike QOAS which addresses high quality, high bitrate video streams.

Although all these adaptive schemes have shown good results in certain scenarios, their adjustment policies are not directly related to the end-user perceived quality. Also they do not address the balance between the number of simultaneous clients served and their perceived quality.

III. QUALITY ORIENTED ADAPTATION SCHEME FOR HIGH BITRATE MULTIMEDIA STREAMING

Receiver buffering may be helpful in many cases, but it does not always solve multimedia-streaming problems. QOAS was designed to complement receiver buffering in highly loaded delivery conditions. It is based on the fact that random losses have a greater impact on the perceived quality than a controlled reduction in quality [20]. QOAS adjusts the content as well as the transmission rate, increasing or decreasing the quantity of streamed video data by dynamically adjusting its quality. This is done according to feedback information received from the client.

The QOAS-based system architecture (Fig. 1) includes multiple instances of QOAS adaptive client and server applications that bi-directionally exchange video data and control packets through the delivery network [21]. The client monitors some transmission and end-user quality-related parameters, and its Quality of Delivery Grading Scheme (QoDGS) regularly computes scores that reflect the overall quality of the streaming process. These grades are sent as feedback to the server, whose Server Arbitration Scheme (SAS) analyses them and proposes adjustment decisions in order to increase the end-user perceived quality in the reported conditions. Each streaming process involves one server and one client application instance.

For each video streaming process, QOAS involves the definition of a number of different server states (e.g. a five-state model was used for our experimental tests). Each server state is then assigned to a different stream quality. The stream quality versions differ in terms of compression-related parameters (e.g. resolution, frame rate, colour depth) and therefore have different bandwidth requirements. They also differ in end-user perceived quality. During transmission the server dynamically varies its state according to the reported end-user stream quality. 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 viewing conditions, the server gradually increases the quality of the delivered stream.

The Quality of Delivery Grading Scheme (QoDGS) [22] at the client (Fig. 2) monitors and evaluates the effect of the delivery conditions on end-user perceived quality. Its grading is based on monitoring 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 [23, 24]. Monitoring short-term variations helps by learning quickly about transient effects, such as sudden traffic changes. Long-term variations are monitored in order to track slow changes in the delivery environment, such as new users in the system. These short-term and long-term periods are considered, respectively, an order and two orders of magnitude greater than the feedback-reporting interval. The QoDGS also takes into account end-user quality as measured by the moving pictures quality metric Q [25], which maps the joint impact of bitrate and data loss on MPEG2 encoded video streams’ quality onto the ITU-T R P.910 five-point grading scale [26].

The QoDGS regularly computes scores that reflect the quality of delivery as assessed at the client which are then sent to the server. The later bases its quality adaptation mechanism on these scores.

Extensive tests have shown that the design of QoDGS (presented in details in [4]) ensures best results in terms of adaptiveness, link utilization, end-user quality, and stability in local broadband IP networks.

![Fig. 1](image1.png) The architecture of a QOAS-based multimedia delivery system

![Fig. 2](image2.png) QoDGS takes into consideration both traffic-related parameters and end-user perceived quality.
The **Server Arbitration Scheme** (Fig. 3) considers the values of a number of consecutive QoDGS scores received from the client and, by averaging them, 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. Late arrival or non-arrival of feedback messages at the server is considered an indication of network congestion and triggers a quality decrease.

### IV. TESTING RESULTS

The experimental tests consisted of simulations using models for QOAS, TFRCP, LDA+ and non-adaptive (NoAd) streaming, built using Network Simulator 2 [27].

#### A. Network Topology

The “Dumbbell” topology (Fig. 4) used for simulations assumes a single shared bottleneck link with 100 Mbps bandwidth and 100 msec delay. The sources of traffic (server application instances) are located on one side of the bottleneck link, and the receivers (client application instances) are on the other side. 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.

#### B. Simulation Models

Non-adaptive (NoAd) streaming transmits video streams using the highest available rate, regardless of problems that may affect the delivery process (e.g. packet loss, increased delays). During tests a maximum rate of 4 Mbps was used for NoAd streaming.

The equation-based TCP-friendly adaptation scheme TFRCP [7] uses estimates of round-trip delay and loss rates to determine the adaptation policy. When there are losses, the rate is limited to that computed according to the TCP model; in cases of zero loss, the current transmission rate is doubled. The sender can update its rate in intervals of M units (2-5 sec). Our TFRCP implementation had M=5 sec as suggested in [7] for delays greater than 100 msec, as in our setup.

LDA+ [8] is an additive increase/multiplicative decrease algorithm based on estimates of network condition and bandwidth share used. In zero loss periods, the sender increases its rate by a 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 by a value that depends on the current rate and the rate determined by a TCP model. Our implementation of LDA+ used an RTCP feedback interval of 5 sec as suggested in [8].

The QOAS model conforms to the description in section III, with a SAS upgrade period of 6 sec and a downgrade timeout of 1 sec. The QoDGS short-term period was taken as 1 sec, and the long-term period was 10 sec.

#### C. Performance Assessment

In order to assess the performance of QOAS, TFRCP, LDA+ and NoAd while streaming video, they are compared with each other in terms of loss, bottleneck link utilization, perceived quality, and number of clients simultaneously served. End-user quality is computed using the multimedia quality metric (Q) proposed in [25] and expressed using the ITU-T P.910 five-point scale for grading subjective perceptual quality [26].

#### D. Video Clips

Five five-minute long video sequences were selected from movies with different degrees of motion content. The **diehard1** sequence includes a great deal of action, **jurassic3** and **dontsayaword** have average motion content, **familyman** has very little movement, whereas **roadtoeldorado** is a typical cartoon sequence. The clips were MPEG-2 encoded at five different rates between 2 Mbps and 4 Mbps using the same frame rate (25 frames/sec) and the same IBBP frame pattern (9 frames/GOP). Traces were collected,
associated with server states, and stored in a database to be used during simulations. Table I presents some statistics about the 2.5 Mbps versions of these movie clips.

### E. Simulation Scenarios and Results

**1) Single Multimedia Streams**

The multimedia sequence with the most complex motion content - *diehard1* – was streamed over the bottleneck link. Background traffic was generated as in Fig. 5a in order to test the adaptive reactions from QOAS and the effect on the quality of NoAd streaming. This traffic was generated on top of a constant bit-rate traffic of 90 Mbps that ensures high loaded delivery conditions. The durations of both QOAS-based adaptive and NoAd stream deliveries were 50 sec. Next the behaviour of QOAS-based adaptive process is analysed and compared to the NoAd solution.

![Fig. 5a](attachment:image1.png)

Fig. 5a shows both the variations of the background traffic and the reaction of the QOAS adaptive delivery process, indicating very good performance of QOAS in different situations. For example, after 3 sec the background traffic increases by 1 Mbps, giving an overall traffic that exceeds much the available bandwidth. QOAS reacts quickly and reduces the quantity of the transmitted data, avoiding losses. In consequence, as Fig. 5b shows, the end-user perceived quality slightly decreases, but remains above the “good” perceptual level. For the same case, during NoAd transmission, buffering can protect the streaming process for a few seconds, but then high losses severely degrade the perceived quality as shown in Fig. 5c.

QOAS’s asymmetric reaction to events prevents the adaptive scheme from immediately responding to the decrease in background traffic that occurs at 10 sec. Therefore when the background traffic increases again at 12 sec, the adaptive server does not have to adapt. Nevertheless, when the decrease in background traffic is prolonged, the server improves the transmitted stream quality at 15 sec.

The extra increase in background traffic at 33s determines another decrease in the quantity of adaptively transmitted data, further reducing the end-user perceived quality (which is still graded at least “good”). In comparison, the non-adaptively delivered multimedia stream suffered a severe degradation in the end-user perceived quality, which reached the “very annoying” level during certain periods.

**2) Multiple Multimedia Streams**

The simulations involved a number of clients randomly selecting both the movie clip and the starting sequence from within the chosen clip. The resulting video streaming processes began and ended during transitory periods of 50 sec duration, 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.

![Fig. 6](attachment:image2.png)

The results presented in Fig. 6 show that in the NoAd case, an increase of only 4.35% in the number of clients caused a loss rate of just below 1%. When the number of clients was increased by more than 15%, the loss exceeded 10%.

### TABLE I

Statistics for Different MPEG-2 Encoded Multimedia Sequences
(Average Rate of 2.5 Mbits/s)

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Motion Content</th>
<th>Peak/Mean Rate Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>diehard1</em></td>
<td>High</td>
<td>7.43</td>
</tr>
<tr>
<td><em>jurassic3</em></td>
<td>Average-Low</td>
<td>4.38</td>
</tr>
<tr>
<td><em>dontsayaword</em></td>
<td>Average</td>
<td>4.51</td>
</tr>
<tr>
<td><em>familyman</em></td>
<td>Low</td>
<td>3.17</td>
</tr>
<tr>
<td><em>roadtoeldorado</em></td>
<td>Average-High</td>
<td>6.51</td>
</tr>
</tbody>
</table>
severely affect the perceived quality, which drops quickly to the minimum level 1 (“bad”) on the ITU-T R P.910 five-point scale [23].

Under identical conditions, when QOAS was used, an increase of up to 40% in the number of clients (32 viewers) had very little effect on the loss rate, which remained below 0.5%. Fig. 7 shows how for QOAS the resulting end-user quality remained above the “good” level of 4. Increases of up to 70% in the number of clients (39 viewers) resulted in loss rates of around 1%, which did not significantly affect the stream quality which remained above the “fair” level of 3. Further increases in the number of clients caused both an increase in the loss rate and a fall in the perceived quality below the “fair” level, which is considered here as the minimum acceptable quality level.

In comparison, tests using TFRCP streaming achieved only a 13% increase in the number of clients (26 viewers) when maintaining a loss rate below 1% and a corresponding perceived quality around the “good” level. For increases in the number of clients above 17%, the loss rate exceeded 1% and the end-user quality fell below the “fair” level. Given similar increases in the number of clients, LDA+ maintains an average loss rate below 1% and a perceived quality above the “good” level only for 24 clients (4% increase). However it maintained a “fair” end-user quality level for 30 simultaneous clients (30% increase) and loss rates around 1% for all tests performed in highly increased traffic conditions.

According to these tests, the number of simultaneous users served is significantly higher for QOAS in comparison with the other streaming schemes. 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.

In terms of efficient usage of available bandwidth, QOAS was superior at all times to TFRCP and LDA+-based streaming. Using QOAS, the bottleneck link utilization exceeded 95% for 30 simultaneous clients and reached 99% for 40 clients. The values obtained for TFRCP and LDA+ are more modest: around 84% and respectively 87% for 30 simultaneous clients, and 92% and respectively 96% for 40 clients. Under the same conditions, the 100% figures obtained by NoAd came with severe costs in terms of loss and significantly reduced end-users quality.

Table II shows comparative performance figures for the streaming approaches studied when choosing “fair” (3) and “good” (4) quality levels as targets. The increases in the number of clients are computed relative to the NoAd case. Although this paper considers the “fair” level to be the minimum acceptable quality level, further increases in the number of clients could be achieved by using different post-processing techniques to mask the resulting losses that would otherwise severely affect the end-users’ perceived quality.

QOAS facilitates the choice of network load level according to economic, technical, and quality goals. It seems likely that service operators will maximise 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

<table>
<thead>
<tr>
<th>Streaming Approach</th>
<th>QOAS</th>
<th>TFRCP</th>
<th>LDA+</th>
<th>NoAd</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Loss rate (%)</td>
<td>1.39</td>
<td>0.47</td>
<td>1.73</td>
<td>0.53</td>
</tr>
<tr>
<td>Link utilization (%)</td>
<td>95.7</td>
<td>96.4</td>
<td>84.1</td>
<td>87.1</td>
</tr>
<tr>
<td>Number of clients</td>
<td>34</td>
<td>32</td>
<td>27</td>
<td>26</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>
</tr>
</tbody>
</table>

Fig. 6. Loss rate versus increase in the number of clients simultaneously served above a base line of 23.

Fig. 7. End-user average quality versus increase in the number of clients simultaneously served above a base line of 23.

Fig. 8. Bottleneck link utilization using different approaches, while increasing the number of simultaneous viewers.
connection, QOAS could service 320 simultaneous users compared to only 260 using TFRCP, 240 using LDA+, and 230 using NoAd streaming.

F. Simulation Considerations

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.

All these adaptive schemes rely on feedback for achieving high performance and the aim is to inform the server as quickly as possible. A high value of 100 msec delay was deliberately chosen in order to test the schemes in extreme conditions; the results for smaller link delays are similar. For greater delays, feedback-based adaptive schemes may not react in time and end-users’ perceived quality may be affected.

V. CONCLUSIONS AND FUTURE WORK

This paper tests the Quality Oriented Adaptation Scheme (QOAS) against extremely variable multimedia-like background traffic and compares its behavior with a non-adaptive approach. QOAS behavior is then compared to three other solutions for video streaming: TFRCP, LDA+, and non-adaptive (NoAd). This comparison is done in terms of bandwidth utilization, number of concurrent clients, loss rate, and end-user perceived quality.

Simulation results show that QOAS successfully adapts to highly variable and increased traffic delivery conditions maintaining the estimated end-user perceived quality above the “good” perceptual quality level. In identical conditions when using the NoAd approach the end-user perceived quality was severely affected during streaming.

The results presented in this paper also indicate that for the same average end-user quality, QOAS can accommodate a significantly higher number of simultaneous clients while also having higher bandwidth utilization. For the same number of clients, the average end-user quality is always higher for QOAS than for the other solutions studied.

Further work will test the effect of changes in the feedback interval on the performance of these adaptive mechanisms, and their behavior when other types of traffic (e.g. short-lived TCP, long-lived TCP, non-adaptive UCD) share the same link with these adaptive streams. Results of extensive subjective perceptual tests performed with a prototype system that verify the end-user quality results presented here will also be reported.

REFERENCES