A Comparison-Based Study of Quality-Oriented Video on Demand

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

***Abstract*—The Quality Oriented Adaptation Scheme (QOAS) is used for high bit-rate multimedia streaming in local broadband multi-service all-IP networks. It balances two opposing goals: providing high quality multimedia-based services to end-users, and increasing the infrastructure utilization and number of customers simultaneously served. Extensive objective testing results presented in this paper show that QOAS achieves high performance in terms of end-user perceived quality, loss rate, throughput, link utilization, and number of customers simultaneously served. These results were obtained even in highly loaded and variable delivery conditions caused by traffic of different types, sizes, and variation patterns. QOAS performance was assessed stand-alone and in comparison with other existing solutions, adaptive and non-adaptive.**

***Index Terms*—Adaptive video streaming, end-user perceived quality, feedback control, grading scheme.**

I. Introduction

FOR THE NEAR future, a sustained growth in the number of broadband connections to residential users and business premises is expected (e.g. exceeding 67 million in Europe [1] and 300 million worldwide [2] in 2007), as part of a trend towards multi-service all-IP networks [3], [4], [5]. Broadband all-IP networks provide a low cost, high bandwidth infrastructure that enables the distribution of rich content services based on very popular IP-based applications, such as digital and interactive video and audio—including video on demand (VoD) and voice over IP (VoIP)—high-speed data transfer, gaming, shopping, and banking.

The success of this trend depends on widespread market acceptance OF IP-based broadband services, which in turn relies heavily on the users' quality of experience and on the price that the end-user must pay. On the one hand network operators and service providers aim for high infrastructure utilization and for simultaneously serving a large number of customers, both to increase their revenues and eventually to reduce the price of services for their customers. On the other hand, customers are interested in receiving high quality streamed multimedia and in having access to a large variety of services. And since the large majority of the services that customers desire includes multimedia data, which has high bandwidth requirements and timing constraints, the technical solution used for their distribution has to accommodate these constraints.

Manuscript received January 25, 2006; revised September 30, 2006.

G.-M. Muntean is with the School of Electronic Engineering, Dublin City University, Glasnevin, Dublin 9, Ireland (e-mail: munteang@eeng.dcu.ie).

P. Perry and L. Murphy are with the School of Computer Science and Informatics, University College Dublin, Belfield, Dublin 4, Ireland.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBC.2006.886451

In this context the Quality Oriented Adaptation Scheme (QOAS) [6] was proposed as an application-level end-to-end adaptive solution for streaming high bit-rate multimedia content. QOAS targets the distribution of high quality multimedia-based services to customers via local broadband all-IP networks.

This paper evaluates the performance of QOAS in balancing customers' need for high quality of service and service providers' and network operators' goals of increased infrastructure utilization and more customers served simultaneously. The paper presents results of extensive tests that involve QOAS and demonstrates its performance even when the multi-service all-IP distribution network is subject to very high traffic load of different types, sizes, and variation patterns. QOAS performance is assessed from different points of view: end-user perceived quality, loss rate, throughput, link utilization, and the number of customers simultaneously served. QOAS testing was comprehensive and included objective testing, using a simulation model. QOAS was compared to some well-known adaptive streaming approaches such as TFRCP [7] and LDA+ [8], a non-adaptive approach, and a (hypothetical) ideal adaptive scheme.

In the next section, major solutions for providing some level of Quality of Service (QoS) including some other adaptive multimedia streaming solutions are mentioned and then QOAS is briefly presented. The results of QOAS simulation-based testing are presented and discussed in the section that follows, before the performance analysis of the streaming solutions compared in this paper is made. Conclusions are drawn in the last section of this paper, which also presents the work in progress and indicates some possible directions for further work.

II. Related Work

Finding solutions for providing a certain level of quality for multimedia-based services delivered over best-effort IP networks has attracted extensive research.

Different approaches were proposed and among the best known are those based on bandwidth over-provisioning, traffic engineering, QoS architectures and adaptive solutions.

Bandwidth over-provisioning [9] looks at allocating statically more bandwidth than the expected network peak requirements. This increases the probability of avoiding congestion-related problems, but provides no guarantees, especially during peak-hours and with very bursty traffic, while wasting resources most of the time.

Traffic engineering involves planning, design, monitoring and management of networks and their traffic in order to allow for the most efficient transport possible. Although traffic engineering-based solutions (e.g. Integrated Services (IntServ) [10], Differentiated Services (DiffServ) [11], Multiprotocol Label Switching (MPLS) [12], etc.) have unchallenged merits, there are significant concerns regarding their complexity, deployment costs and some other issues such as security.

QoS architectures provide a unifying framework for different aspects concerned on QoS provisioning, ranging from user requirements through operating system and hardware characteristics to network capability and performance [13]. Proposals include the Lancaster QoS-Architecture (QoS-A) [13], OSI QoS Framework Model [14], [15], Tenet Approach [16], etc.) which can provide good results, but are complex and relative expensive to deploy.

Adaptive schemes [17], [18] for multimedia streaming take the distribution networks as they are and provide the least complex and the most flexible mechanisms for providing certain level of QoS in existing network conditions. Among the adaptive schemes, bandwidth-smoothing techniques [19] help in lightly loaded networks as they vary the transmission rate of bursty traffic without affecting the transmitted content, averaging it out. However in highly loaded conditions, tougher measures need to be taken and they include variation in the amount of data to be transmitted. These adaptive schemes adjust the bandwidth used by the applications according to the existing network conditions, increasing or decreasing both the transmission and content encoding rates.

The adaptive streaming schemes were classified [17] according to where the adaptation takes place in: sender-driven solutions such as TFRCP [7], LDA+ [8], RAP [20] or LQA [21], receiver-driven solutions such as RLM [22] or RLC [23], and transcoding-based solutions [24], [25].

The authors of [22] proposed and the ones of [23] improved a receiver-driven adaptation scheme based on multicast groups that allowed the clients to directly select the desired multimedia quality in the absence of feedback. On-the-fly transcoding is used in [24] to meet the clients' requirements, whereas [25] presents a more general transcoding-based solution that relies on filters deployed in the distribution network to match the quality level required by clients. Among the sender-driven schemes, the adaptive solution proposed in [26] varies some encoding-related parameters at the server to adjust the bit-rate of transmitted multimedia data according to feedback from clients that monitor some parameters related to multimedia transmission only, while the work in [21] describes a layered encoding-based adaptive solution.

Recently, different rate adjustment sender-driven 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 [27]. Limitations include its inflexibility and its problems with time sensitive media. The Loss-Delay based Adjustment algorithm (LDA) [28] 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) [21] 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 [29]. 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 may be taken into account too late.

Commercial adaptive streaming solutions like Real Networks' SureStream [30] and Microsoft's Multimedia Multi-bitrate (MBR) solution [31] 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 bit-rates, unlike QOAS, which addresses high quality high bit-rate video streaming.

Although all these adaptive schemes are supported by 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 link utilization and the number of customers served and their perceived quality.

III. The Quality-Oriented Adaptation Scheme (QOAS)

As with any adaptive scheme for multimedia streaming, the Quality Oriented Adaptation Scheme (QOAS) [6], [32], [43] relies on the fact that random losses have a greater impact on the end-user perceived quality than a controlled reduction in quality [33]. Therefore the end-to-end sender-driven adaptation mechanism employed by QOAS controls the adjustment of both the quality of the streamed multimedia content and the transmission rate so that it maximizes the end-user perceived quality in existing delivery conditions. This **intra-stream adaptation** controls the quality, and consequently the quantity of streamed multimedia-related data and is based on information received from the client.

The **QOAS-based system architecture** includes multiple instances of the end-to-end QOAS adaptive client and server applications. These exchange video data and control packets through the IP-based delivery network. The QOAS client continuously monitors some transmission parameters and estimates the end-user perceived quality, and its *Quality of Delivery Grading Scheme* (QoDGS) regularly computes Quality of Delivery scores $(\mathrm{QoD}_{\mathrm{scores}})$ that reflect the multimedia streaming quality in current delivery conditions. These grades are sent as feedback to the QOAS server, whose *Server Arbitration Scheme* (SAS) analyzes them and proposes adjustment decisions to be taken in order to increase the end-user perceived quality in existing conditions.

The **QOAS adaptation principle** is schematically described in Fig. 1. For each QOAS-based multimedia streaming process, a number of different quality states are defined at the server (e.g. the experimental tests have involved a five-state model). Each such 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, color depth) and therefore have different bandwidth requirements. They also differ in end-user perceived quality. The difference between the average bit-rates of these different quality streams is denoted as

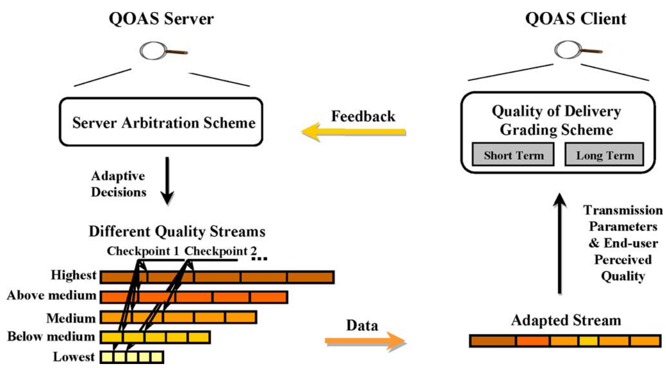


Fig. 1. Schematic description of QOAS's adaptation principle.

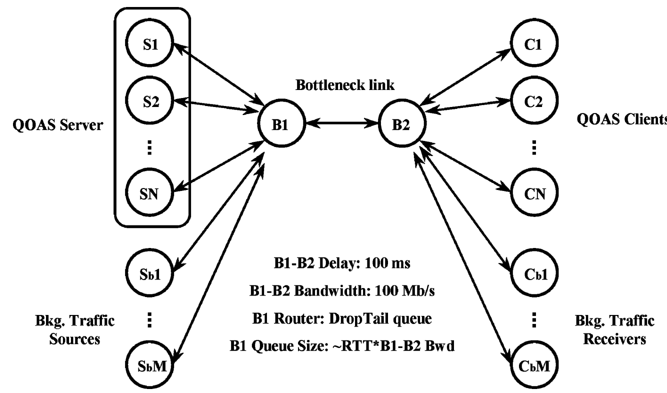


Fig. 2. "Dumbbell" simulation topology includes a bottleneck link, N communicating QOAS server and client application instances and a number of background traffic source-sink pairs.

"adaptation step". During data transmission the client-located QoDGS computes $(\mathrm{QoD}_{\mathrm{scores}})$ that are sent via feedback to the QOAS server, which dynamically varies its quality state based on suggestions made by SAS. When the delivery conditions cause excessive delays and/or loss the client reports a decrease in end-user quality and the server switches to a lower quality state, reducing the bit-rate of the streamed multimedia. Consequently this may reduce the delays and the loss, increasing the end-user perceived quality. If the QOAS client reports improved streaming conditions, the server increases the quality of the delivered stream. These switches to higher and lower quality states respectively are performed gradually with the granularity of the QOAS adaptation step. The smaller the adaptation step, the less noticeable to the viewer is the effect of the bit-rate modification. However the higher the adaptation step, the faster is the convergence of the algorithm to the bit-rate best suited in existing network conditions.

The client-located **Quality of Delivery Grading Scheme** (QoDGS), described in detail in [6], evaluates the effect of the delivery conditions on end-user perceived quality. It monitors both short-term and long-term variations of packet loss rate, delay and delay jitter, which have the most significant impact on the received quality [34], [35] and estimates the end-user perceived quality. The end-user quality is estimated using the no-reference moving pictures quality metric Q [36], which maps the joint impact of bit-rate and data loss on video quality onto the ITU-T R P.910 five-point grading scale [37].

The **Server Arbitration Scheme** assesses the values of a number of consecutive $(\mathrm{QoD}_{\mathrm{scores}})$ received as feedback in order to reduce the effect of noise in the adaptive decision taking process. Based on these scores SAS suggests adjustment decisions. This process is ***asymmetric***, requiring fewer $(\mathrm{QoD}_{\mathrm{scores}})$ to trigger a decrease in the server's quality state than for an increase. This ensures a fast reaction during bad delivery conditions and helps to eliminate its cause. The increase is performed only when there is enough evidence that the network conditions have improved. This asymmetry helps also to maintain system stability, by reducing the frequency of quality variations.

When QOAS is used to stream multimedia to multiple viewers, an **inter-stream adaptation** scheme [6] complements the intra-stream adaptation and aims for a finer adjustment in the overall adaptation process. The inter-stream adaptation is responsible for preventing QOAS-based adaptive processes from reacting simultaneously to variations in the delivery network. It selectively allows some of the QOAS-based sources of multimedia data to react to the received feedback, in a step-by-step process, achieving near optimal link utilization and long-term fairness between the clients.

IV. Objective Testing—Setup

QOAS was proposed as an adaptive solution to deliver high quality multimedia-based services to home residences and business premises via local broadband multi-service all-IP networks. In order to verify and validate its performance, **objective testing** was employed, involving Network Simulator 2 (NS-2) [38] and NS-2-built simulation models.

These simulations aim at showing that the QOAS-based solution achieves significant performance in different delivery conditions and subject to cross traffic of various types, sizes and variation patterns. Next the network topology, multimedia clips, simulation models and performance assessment are presented.

A. Network Topology

The "Dumbbell" topology used for simulations is presented in Fig. 2. It assumes a single shared bottleneck link (B1–B2) with 100 Mb/s bandwidth and 100 ms latency. The sources of traffic (QOAS server application instances) are located on one side of the bottleneck link, and the receivers (QOAS client application instances) are on the other side. The other links are over-provisioned so that the only packet loss and significant delays are caused by congestion that occurs on the bottleneck link. The buffering at the B1–B2 link uses a drop-tail queue of size proportional to the product of round trip time (RTT) and bottleneck link bandwidth (labelled "B1–B2 Bwd" in Fig. 2).

B. Performance Assessment

The QOAS performance is assessed stand-alone and in comparison with other schemes in terms of loss rate, throughput, bottleneck link utilization, end-user perceived quality, and number of customers simultaneously served. End-user quality is computed using the no-reference moving pictures quality metric Q [36] and expressed using the ITU-T R P.910 five-point scale for grading subjective perceptual quality [37]. The comparison is done with a non-adaptive solution, with TFRCP and LDA+ adaptive schemes and with an ideal adaptive solution that would achieve maximum bandwidth utilization with zero loss at any moment. The models used during simulations are described next.

C. Simulation Models

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

The equation-based ***TFRCP*** [7] 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 $\mathrm{M} = 5$ s1 as suggested in [7] for delays greater than 100 ms, as in this setup.

LDA+ [8] 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 s as suggested in [8].

An ***ideal adaptive scheme*** would successfully adapt to changing network conditions and determine an output rate that matches the available bandwidth at any moment yielding the best end-user quality possible in existing conditions. In consequence the model built for this ideal adaptive scheme for streaming multimedia achieves 0% loss and 100% link utilization data at all times.

The ***QOAS model*** conforms to the description given in Section III, with a SAS upgrade period of 6 s and a downgrade timeout of 1 s. The QoDGS short-term period was set to 1 s, and the long-term period was 10 s.

NS-2 built-in models generate all the background traffic used during testing.

D. Multimedia Clips

Five five-minute long video sequences were selected from movies with different degrees of motion content representative for their type. 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 Mb/s and 4 Mb/s (adaptation step is 0.5 Mb/s) using the same frame rate (25 frames/s), the same IBBP frame pattern (9 frames/GOP) and resolution 320 × 240. Traces were collected, associated with the different server quality states, and stored in a database to be used during simulations. Table I presents some statistics about the different quality encoded versions of these clips.

TABLE I
PEAK/MEAN BIT-RATE RATIO OF ALL MPEG-2 ENCODED QUALITY VERSIONS OF THE CLIPS USED DURING SIMULATIONS

Clip	2.0 Mb/s	2.5 Mb/s	3.0 Mb/s	3.5 Mb/s	4.0 Mb/s
diehard1	7.48	7.43	6.31	5.65	4.06
roadtoeldorado	6.91	6.51	6.23	6.12	6.05
dontsayaword	5.56	4.51	4.36	4.08	3.56
jurassic3	4.83	4.38	4.04	3.71	3.41
familyman	3.99	3.67	3.42	3.09	2.93

V. OBJECTIVE TESTING—SCENARIOS AND RESULTS

Considering the setup described in the previous section, next simulation scenarios and results of the simulations are presented.

A. Single Clip Streaming With Various Cross Traffic

The first set of tests aims at assessing the performance of a single multimedia stream delivery using QOAS in highly loaded traffic conditions and in the presence of various cross traffic, commonly expected in multi-service IP networks. The NS-2 simulations that lasted between 200 s and 400 s used the "Dumbbell" topology presented in Section IV-A and involved *diehard1*—the highest motion content multimedia clip introduced in Section IV-D. Background traffic of different types, shapes and variation patterns was generated using NS-2 models. UDP: constant bit-rate (CBR) and variable bit-rate (VBR) and TCP: long-lived FTP-like and short-lived WWW-like traffic are two main classes of traffic taken into account. Traffic with different sizes and variation patterns was considered within each of these classes. First column from Tables II, III, IV and V indicates the number of traffic sources and their average bit-rate expressed in Mb/s (e.g. 4 × 0.4 describes four sources of 0.4 Mb/s each). The second column presents the characteristics of the generated background traffic (e.g. 0.001 s on—0.1 s off indicates an exponential traffic generated with such an on-off pattern). This background traffic is taken into account on top of a CBR traffic of at least 95.5 Mb/s. This CBR traffic simulates a well-multiplexed result of multiple traffic sources as it happens in real world situations after statistical multiplexing. The 95.5 Mb/s bit-rate was such chosen to allow for minimum remaining bandwidth for transmission of both multimedia traffic and different types of other background traffic. This background traffic puts the highest pressure possible on multimedia streaming and ensures highly loaded traffic conditions for the tests performed. In these conditions, QOAS and the ideal adaptive scheme models were tested and their performance compared according to the principles stated in Section IV-B.

1) UDP-CBR as Background Traffic: Some multimedia streaming solutions use smoothing techniques in order to reduce the burstiness of the traffic generally associated with multimedia deliveries, whereas some others use CBR encoding to produce a flat rate output stream that would be easily transmitted over IP networks. Also if the volume of traffic is very large, even if consisting of different types and shapes of individual flows, it is subjected to statistical multiplexing that produces an overall CBR-like output. The effect of CBR traffic is studied in this section, taking into consideration different

TABLE II
STATISTICAL RESULTS FOR 400 s-LONG MULTIMEDIA STREAMING AGAINST *UDP-CBR PERIODIC* BACKGROUND TRAFFIC WITH DIFFERENT VARIATION PATTERNS ON TOP OF 96.0 Mb/s CBR TRAFFIC

		Avg. Bitrate (Mb/s)				Avg. Percv.Qual. (Q)				Total Bw. Utiliz. (%)				Loss Rate (%)			
Size of Periodic Traffic (Mb/s)	Characteristics	QOAS	Ideal	TFRCP	LDA+	QOAS	Ideal	TFRCP	LDA+	QOAS	Ideal	TFRCP	LDA+	QOAS	Ideal	TFRCP	LDA+
1 x 0.5	20s on - 40s off	3.670	3.833	3.696	3.807	4.521	4.548	4.405	4.268	99.804	100.0	99.822	99.901	0.0	0.0	0.10	0.42
1 x 0.5	30s on - 60s off	3.737	3.848	3.690	3.737	4.532	4.550	4.356	4.156	99.840	100.0	99.800	99.803	0.0	0.0	0.16	0.65
1 x 0.5	40s on - 80s off	3.764	3.838	3.646	3.744	4.537	4.548	4.302	4.171	99.873	100.0	99.760	99.830	0.0	0.0	0.19	0.50
1 x 0.7	20s on - 40s off	3.496	3.580	3.585	3.614	4.490	4.505	4.233	4.143	99.858	100.0	99.726	99.752	0.0	0.0	0.31	0.42
1 x 0.7	30s on - 60s off	3.520	3.581	3.490	3.540	4.495	4.505	4.183	4.081	99.876	100.0	99.627	99.636	0.0	0.0	0.34	0.90
1 x 0.7	40s on - 80s off	3.331	3.555	3.476	3.511	4.458	4.501	4.168	3.997	99.721	100.0	99.861	99.601	0.0	0.0	0.36	1.04

TABLE III
STATISTICAL RESULTS FOR 200 s-LONG MULTIMEDIA STREAMING AGAINST *UDP-CBR STAIRCASE* BACKGROUND TRAFFIC WITH DIFFERENT VARIATION PATTERNS ON TOP OF 95.5 Mb/s CBR TRAFFIC

		Avg. Bitrate (Mb/s)				Avg. Percv.Qual. (Q)				Total Bw. Utiliz. (%)				Loss Rate (%)			
Size of Staircase Traffic (Mb/s)	Characteristics	QOAS	Ideal	TFRCP	LDA+	QOAS	Ideal	TFRCP	LDA+	QOAS	Ideal	TFRCP	LDA+	QOAS	Ideal	TFRCP	LDA+
4 x 0.4	Up 40s long steps	3.592	3.617	3.260	3.214	4.508	4.512	3.978	3.947	99.905	100.0	99.979	99.888	0.00	0.0	0.63	0.77
4 x 0.6	Up 40s long steps	3.085	3.031	2.724	2.565	4.300	4.391	3.368	3.522	99.945	100.0	99.999	99.954	0.09	0.0	2.62	1.34
4 x 0.4	Down 40s steps	3.096	3.696	3.206	3.195	4.269	4.525	3.830	4.199	99.689	100.0	99.763	99.794	0.02	0.0	0.99	0.17
4 x 0.6	Down 40s steps	2.808	3.296	2.814	2.701	4.097	4.451	3.807	3.987	99.886	100.0	99.946	99.913	0.04	0.0	0.67	0.35

TABLE IV
STATISTICAL RESULTS FOR 200 s-LONG MULTIMEDIA STREAMING AGAINST *UDP-VBR EXPONENTIAL* BACKGROUND TRAFFIC WITH DIFFERENT VARIATION PATTERNS ON TOP OF 95.5 Mb/s CBR TRAFFIC

		Avg. Bitrate (Mb/s)				Avg. Percv.Qual. (Q)				Total Bw. Utiliz. (%)				Loss Rate (%)			
Size of Expon ential Traffic (Mb/s)	Characteristics	QOAS	Ideal	TFRCP	LDA+	QOAS	Ideal	TFRCP	LDA+	QOAS	Ideal	TFRCP	LDA+	QOAS	Ideal	TFRCP	LDA+
1.0	0.001s on - 0.1s off	3.651	3.658	3.450	3.564	4.518	4.519	4.153	4.048	99.942	100.0	99.807	99.881	0.00	0.0	0.30	0.66
1.0	0.01s on - 0.1s off	3.649	3.659	3.464	3.489	4.517	4.519	4.186	4.036	99.935	100.0	99.874	98.526	0.00	0.0	0.34	0.57
1.0	0.1s on - 0.1s off	3.601	3.639	3.493	3.527	4.509	4.516	4.069	4.090	99.926	100.0	99.793	98.500	0.00	0.0	0.63	0.49
0.8	0.001s on - 0.1s off	3.736	3.824	3.616	3.727	4.532	4.546	4.192	4.140	99.849	100.0	99.827	99.867	0.00	0.0	0.18	0.44
1.0	0.001s on - 0.1s off	3.651	3.658	3.450	3.564	4.518	4.519	4.153	4.048	99.942	100.0	99.807	99.881	0.00	0.0	0.30	0.66
1.2	0.001s on - 0.1s off	3.433	3.445	3.290	3.444	4.478	4.481	4.126	4.024	99.950	100.0	99.835	99.904	0.00	0.0	0.31	0.65

TABLE V
STATISTICAL RESULTS FOR MULTIMEDIA STREAMING AGAINST *TCP LONG-LIVED (FTP) & SHORT-LIVED (UDP)* BACKGROUND TRAFFIC ON TOP OF 75 Mb/s AND 95.5 Mb/s CBR TRAFFIC RESPECTIVELY

		Avg. Bitrate (Mb/s)				Avg. Percv.Qual. (Q)				Total Bw. Utiliz. (%)				Loss Rate (%)			
Type and size of TCP Traffic (Mb/s)	Duration (s)	QOAS	Ideal	TFRCP	LDA+	QOAS	Ideal	TFRCP	LDA+	QOAS	Ideal	TFRCP	LDA+	QOAS	Ideal	TFRCP	LDA+
50 x FTP (~22.0)	200	3.042	3.140	3.173	3.210	4.394	4.417	4.286	4.306	98.423	100.0	99.999	99.999	0.00	0.0	0.09	0.11
54 x FTP (~22.5)	200	2.701	2.729	2.024	2.000	4.291	4.309	3.532	3.702	98.425	100.0	99.999	99.945	0.04	0.0	0.77	0.49
40 x WWW (~0.5)	250	3.802	4.016	3.366	2.839	4.543	4.575	4.070	3.819	99.690	100.0	99.670	99.512	0.00	0.0	0.90	0.87
50 x WWW (~1.0)	250	3.505	3.587	3.230	3.166	4.492	4.507	3.948	4.126	99.803	100.0	99.644	99.828	0.01	0.0	0.82	0.38

variation shapes such as periodic and staircase, with different frequencies and step sizes.

a) CBR periodic: The step size of the CBR periodic background traffic variation is set to 0.5 Mb/s in the first set of tests and to 0.7 Mb/s in the second set while the periodicity of these variations is varied as indicated in the "Characteristics" column from Table II. These traffic variations are performed on top of a 96 Mb/s CBR background traffic, which aims at creating highly loaded delivery conditions. Table II presents a comparison of performance results of the QOAS, TFRCP, LDA+ and ideal adaptive multimedia streaming in these conditions. The performance is presented in terms of throughput, average end-user perceived quality, loss rate and link utilization.

In all the cases when either the CBR background traffic varied periodically with steps of 0.5 Mb/s (comparable to the QOAS adaptation step) or with steps of 0.7 Mb/s (much higher than the adaptation step) and regardless of the variation periodicity, QOAS successfully adapted to the change in the background traffic. This adaptation was performed with almost the same frequency as the variation of the cross traffic, completely avoiding packet loss. Consequently the end-user perceived quality achieves very high values, which are between 4 and 5, "good" and "excellent" respectively on the subjective quality scale and little lower for the cases with step sizes of 0.7 Mb/s that require two QOAS quality adaptation steps to be performed. The fact that QOAS's end-user perceived quality is less than 1% below that of the ideal adaptive scheme indicates very good QOAS performance. Even in the case when the highest network load was simulated, QOAS still achieves an end-user perceived quality above the "good" perceptual level, 0.3 higher than that of TFRCP and 0.46 higher than that of LDA+. At the same time it is significant that QOAS's link utilization is very close to the ideal 100% for the majority of time and even when it temporarily varies, it does not decrease below 99%.

It is important to note that the link utilization here is for the total link and includes the efforts of the CBR periodic traffic.

b) CBR staircase: The background traffic is increased in four steps of 0.4 Mb/s and 0.6 Mb/s, lower and higher respectively than the QOAS adaptation step of 0.5 Mb/s and is added to a 95.5 Mb/s CBR traffic that creates highly loaded network delivery conditions. The QOAS's reaction is then tested when decreasing the cross traffic with similar steps. Details about the generated cross traffic and the consequent average bit-rate, end-user perceived quality, loss rate and link utilization for QOAS, TFRCP, LDA+ and the ideal adaptive scheme are given in Table III.

Regardless of the background traffic step size, its variation in staircase up manner triggered an immediate step-by-step QOAS downward adaptation and its step-wise decrease—a delayed reaction from QOAS that conservatively reacts to improved delivery conditions. This is shown clearly by the average bit-rate and consequently end-user perceived quality that are lower in these tests than those corresponding to the tests with staircase-up background traffic variation. The QOAS adaptations triggered by the staircase down variations in background traffic achieve almost zero loss similar to when the traffic step-wise increases with steps greater than the QOAS adaptation step. In consequence the end-user perceived quality is between the "good" and "excellent" perceptual levels for the whole duration of these streaming processes. When some loss occurs, it is for short periods of time (on average 1.75 s) and although the end-user perceived quality decreases during these short lossy periods, its average is still maintained well above the "good" subjective level.

The performance of the QOAS is highlighted when the resulting end-user perceived quality is compared with that of TFRCP, LDA+ and the ideal adaptive scheme. On average the ideal adaptive scheme achieves higher end-user perceived quality, but that obtained by QOAS is only 1% adrift. However QOAS outperforms both TFRCP and LDA+ in all tests, including the one that simulates very difficult delivery conditions (staircase up background traffic variation with 4 steps of 0.6 Mb/s). In this test QOAS maintains viewer's perceived quality above the "good" quality level, whereas the other adaptive schemes fail to do so.

In these simulated conditions, the link utilization when using QOAS for streaming is very close to 100% and the average loss rates less than 0.1% at all times (unlike those achieved by TFRCP and LDA+ that are on average 1.2% and 0.6%, respectively), which indicate a highly significant adaptation result.

2) UDP-VBR as Background Traffic: The majority of multimedia streaming solutions produce very bursty output traffic especially MPEG-encoded streams, due to the different compression ratios achieved for their I, P and B frames. In this context, the effect of UDP-VBR background traffic is studied next, taking into consideration different situations in terms of average bit-rate and degree of burstiness.

a) Constant average bit-rate and variable burstiness: To examine the effect of the VBR burstiness on the QOAS-based streaming, the burstiness of the exponentially generated background traffic is varied while the bit-rate is maintained constant at 1 Mb/s. The traffic burstiness is varied by modifying the on-off characteristics of the exponential traffic as indicated in the second column of Table IV. The values indicated in the table refer to average durations of the exponentially distributed 'on' and 'off' times, rather than fixed durations as in the case of CBR periodic background traffic. This traffic is sent across the bottleneck link along with a 95.5 Mb/s CBR traffic that simulates well multiplexed other traffic. The QOAS is assessed in terms of adaptation to cross traffic variation, throughput, estimated end-user perceived quality, loss rate and bottleneck link utilization and compared to TFRCP, LDA+ and ideal adaptive solutions.

QOAS adapts successfully in all these delivery conditions including the one with highly bursty UDP-VBR background traffic (0.001 s on—0.1 s off), achieving no loss. In these conditions QOAS quality variations are slow, not following the bursty VBR traffic variations on which it has limited dependency. In consequence QOAS maintains an approximately stable end-user perceived quality, which is above the "good" subjective level at all times. The detailed statistical results presented in Table IV show that QOAS performance results are within 1% from the ones obtained by the hypothetic ideal adaptive scheme. It is significant to notice that QOAS achieves an average end-user perceived quality (measured on the 1–5 scale) higher with 0.3 than that of TFRCP and with 0.4 than that of LDA+. QOAS also outperforms the other adaptive schemes in terms of average loss rates (TFRCP and LDA+ loss rates are on average 0.3% and 0.6%, respectively). QOAS's link utilization is within 0.1% from the maximum of 100%, being very close to it at all times. Again, this includes the beneficial efforts of the bursty traffic, smoothed by the buffer at node B1.

b) Constant burstiness and variable average bit-rate: Maintaining a constant high level of burstiness, the bit-rate of the UDP-VBR background traffic is varied in order to study its effect on the QOAS-based adaptation. This traffic is sent across the bottleneck link on top of a 95.5 Mb/s CBR traffic that creates high loaded delivery conditions. Details about the characteristics of the exponentially generated VBR traffic are given in Table IV.

Again QOAS successfully adapted and achieved no loss, regardless of the pressure put on the bottleneck link by increasing the average bit-rate of the VBR background traffic. As expected this increase triggered a decrease in the average bit-rate of the QOAS-transmitted multimedia stream, from 3.74 Mb/s to 3.65 Mb/s and 3.43 Mb/s respectively. As a consequence the end-user perceived quality slightly decreased but remained significantly above the "good" perceptual quality level, varying from 4.53, to 4.52 and 4.48 respectively. The results presented in Table IV show how the QOAS's throughput, end user perceived quality, link utilization and loss rates are all within 1% from the values achieved by the ideal adaptive scheme and they are much better than those achieved by TFRCP and LDA+.

3) TCP as Background Traffic: The large majority of Internet traffic today consists of file transfers that use TCP as the transport protocol and among the most popular applications that have based their functionality on TCP are FTP applications employed for file transfers and WWW applications used for content viewing. These applications were chosen because they are representative for two types of TCP-based traffic: **long-lived** and **short-lived**. The former is characterized by long duration processes that produce in general slow-changing traffic, whereas the latter is responsible for highly variable traffic, of short durations and therefore very bursty. The effect of these traffic types on QOAS-based multimedia streaming is studied next and is assessed in terms of throughput, end-user perceived quality, loss rate and link utilization.

a) Long-lived TCP (FTP-like traffic): Two sets of tests involving 50 and 54 simultaneous FTP flows are generated using the NS-2 built-in models. They are transmitted on top of a 75 Mb/s CBR background traffic in order to create both variable and highly loaded network delivery conditions. QOAS, TFRCP, LDA+ and an ideal scheme were used in turn to stream multimedia and Table V offers both details about the background traffic and testing results.

When streaming multimedia it is important to achieve the highest possible end-user quality by sending as much multimedia data in as timely a manner as possible. On the other hand the other services, including TCP should not suffer due to the lack of bandwidth. When multimedia traffic was competing for bandwidth with long-lived TCP, QOAS gracefully balanced its aggressiveness and TCP friendliness. The effect was that both the long-lived TCP and the QOAS adapted and shared the available bandwidth. Due to QOAS adaptation, packet loss was completely avoided when 50 FTP flows were generated and in consequence QOAS achieved a very high end-user perceived quality, reaching an average of 4.39, which is very close to the 4.42 computed for the ideal adaptive scheme and higher than those for TFRCP and LDA+ in the same conditions. However, when 54 FTP flows are transmitted, increasing the load on the bottleneck link, QOAS experiences some loss and the end-user perceived quality decreases to an average of 4.29. The higher loss rates recorded for TFRCP and LDA+ (0.77% and 0.49% respectively) determine consequent lower end-user perceived quality levels. In this context, it is significant to note that QOAS outperforms with 21% and 15% TFRCP and LDA+ respectively in terms of end-user perceived quality. Unlike for the other adaptive schemes, QOAS's quality value is still well above the "good" perceived quality level and within 1% from that achieved by the ideal adaptive scheme. During these tests, the loss rate experienced during QOAS-based streaming is less than 0.1% and the link utilization within 1.5% from the ideal, showing significant performance.

It is important to note that the link utilization figures include the beneficial contribution of the TCP flows. Also the ratios between the average bit-rates achieved by the QOAS-based system and the ideal adaptive system are on average around 97% which suggests that the flow control of QOAS is very close to that of the ideal adaptive scheme.

b) Short-lived TCP (WWW-like traffic): Two sets of tests involve 250 s long multimedia streaming over a bottleneck link that carries also background traffic. This traffic consists of 40 and 50 WWW sessions respectively and 95.5 Mb/s CBR traffic that ensures highly loaded delivery conditions (see Table V). The WWW traffic is generated using the NS-2 built-in models and has the following characteristics, considered typical for a WWW session by researchers in the WWW area [39], [40]: inter-session time—exponentially distributed with an average of 2 s, number of web pages retrieved during a session—constant and equal to 5, time between consecutive pages—exponentially distributed with an average of 2 s, number of objects within a page—constant and equal to 10, time between consecutive requests for objects of the same page—exponentially distributed with an average of 0.01 s and size of the objects—Pareto distribution with average 10 KB and shape 1.2.

In these tests QOAS achieved very good adaptation and outperformed both TFRCP and LDA+, as the results presented in Table V show. During QOAS-based multimedia streaming very little loss (less than 0.01%) was experienced even in such bursty delivery conditions caused by the WWW background traffic. Consequently the end-user perceived quality was very high, between the "good" and the "excellent" subjective levels, achieving an average of 4.54 and 4.49 respectively. These values are within 1% of those experienced by the ideal adaptive scheme in the same conditions. Unlike for QOAS when TFRCP and LDA+ were used for streaming, loss rates of an average of 0.8% determined lower end-user perceived quality that decreased to the "fair" quality level. QOAS's link utilization, although highly variable due to the burstiness of this background traffic, was on average within 0.3% of the maximum 100%, whereas that achieved for the other adaptive schemes exceeded 99%.

During these tests, regardless of the background traffic type, shape and size, the QOAS scored highly in terms of throughput, end-user perceived quality, loss rate and bottleneck link utilization in comparison to the other adaptive schemes and even to the ideal scheme. The adaptation was so successful that the QOAS streaming maintained loss rates of less than 0.1% in all cases, although the delivery network was fully loaded. It is significant to mention that all the perceived quality scores are above the "good" perceptual level (4 on the ITU-T R. P.910 1–5 scale) being in almost all cases within 1% from the ideal (in only one case the score is 3% adrift). Fig. 3 compares QOAS and the other adaptive schemes when streaming multimedia in the most difficult simulated delivery conditions determined by the major background traffic types in terms of end-user perceived quality. It could be clearly seen how QOAS outperforms the other adaptive schemes in all situations and is very close to the performance of the ideal scheme.

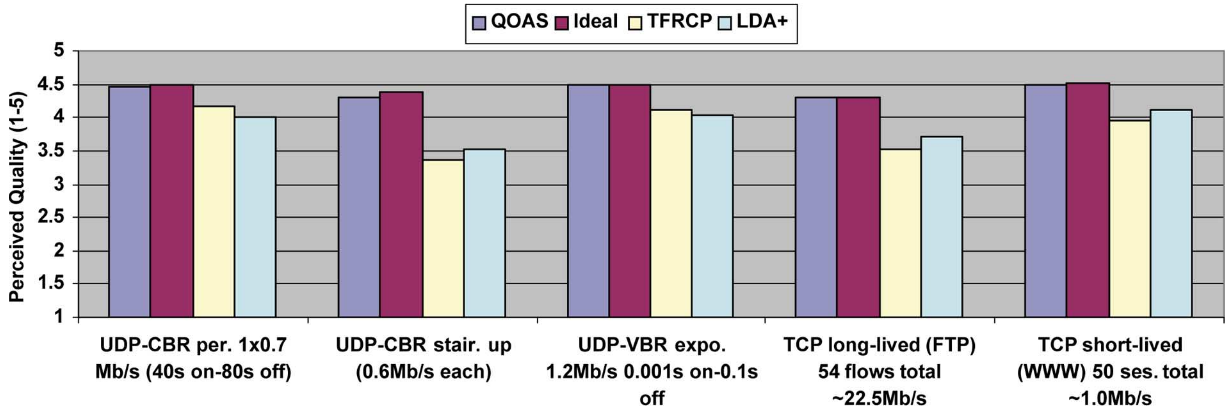


Fig. 3. Comparison between end-user perceived quality when streaming multimedia with QOAS, ideal, TFRCP and LDA+ respectively in the most difficult conditions from the simulated cases.

The bottleneck link utilization also reaches very high levels, with QOAS making use of more than 99.5% of the bandwidth resources in the large majority of tests and even in the two remaining cases the available resources are less than 1.5% from being fully used. All these results indicate high performance for QOAS and show good adaptations regardless of the cross traffic.

B. Multiple Multimedia Streaming With No Cross Traffic

This set of tests aims at assessing the performance of a high number of simultaneous multimedia stream deliveries using QOAS (with its inter-stream adaptation capability enabled) and other schemes. QOAS, TFRCP, LDA+ and non-adaptive (NoAd) approaches were used in turn as the streaming method, and the number of clients was gradually increased above a base line of 23 in each case. This number of clients was chosen because it allows for lossless streaming and maximum end-user perceived quality in all the four cases. The NS-2 simulations performed involved the 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 s 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 s.

The results presented in Fig. 4 and Fig. 5 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%, severely affecting the end-user perceived quality which drops quickly to the minimum level 1 (“bad”) on the ITU-T R P.910 five-point scale.

When QOAS was used for streaming, a 40% increase in the number of clients (32 viewers) had very little effect on the loss rate, which remained below 0.5% and Fig. 5 shows how the resulting end-user perceived quality remained above the “good” perceptual level. Increases of up to 70% in the number of clients (39 viewers) resulted in loss rates of around 1%, which triggered a consequent end-user perceived quality level of “fair”. Further increases in the number of clients caused a higher loss rate and a fall in the end-user quality below the “fair” level, considered here the minimum limit of interest.

Tests using TFRCP streaming achieved only a 13% increase in the number of clients (26 viewers) when maintaining a loss

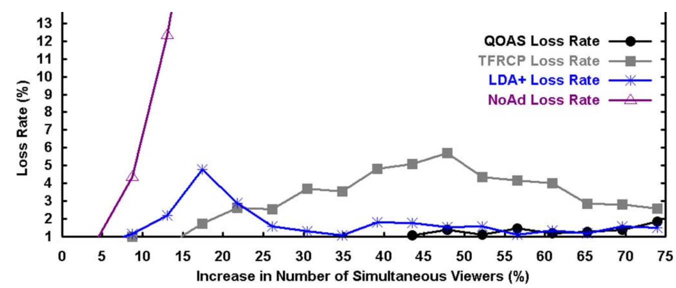


Fig. 4. Loss rate versus increase in the number of clients simultaneously served above a base line of 23.

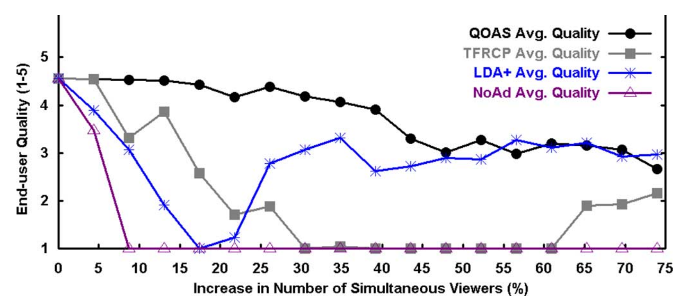


Fig. 5. End-user average quality versus increase in the number of clients simultaneously served above a base line of 23.

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 delivery conditions.

In these conditions, the ideal adaptive streaming scheme would achieve 100% bandwidth utilization, 0% loss rate and an average bit-rate per customer computed by dividing the available bandwidth (100 Mb/s) to the number of simultaneous customers. For example 40 customers could stream on average multimedia at a rate of 2.5 Mb/s each and would have their end-user perceived quality at a level of 4.24.

TABLE VI
COMPARISON BETWEEN QOAS, TFRCP, LDA+ AND NOAD

	QOAS		TFRCP		LDA+		NoAd	
Quality	3	4	3	4	3	4	3	4
Loss rate (%)	1.39	0.47	1.73	0.53	1.31	0.50	0.81	0.01
Link utilization (%)	95.7	96.4	84.1	87.1	86.9	93.4	94.7	90.0
Number of clients	34	32	27	26	30	24	24	23
Increase in no. of clients (%)	41.7	39.1	12.5	13.0	25.0	4.4	-	-

In conclusion, the number of simultaneous users served is significantly higher for QOAS in comparison with the other streaming schemes. For instance in order to maintain a "good" quality level, by using QOAS 23% more clients could be served than by using TFRCP, while the benefit is 33% when using LDA+, and 39% when using the NoAd solution. If a "fair" average quality level is to be maintained for the clients, the benefit of using QOAS is 26% greater than TFRCP, 13% greater than LDA+, and 42% greater than NoAd. Only the ideal scheme outperforms QOAS, streaming multimedia at "good" quality level to up to 50 viewers, but it is very unlikely that such a scheme will be ever built.

In terms of usage of available bandwidth, QOAS was superior at all times to both TFRCP and LDA+ and more details are presented in [32].

A comparison between streaming approaches when choosing "fair" (3) and "good" (4) end-user quality levels as targets is shown in Table VI that includes a presentation of increases in the number of clients 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.

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 particular it is noted that their estimated end-user perceived quality first decreases sharply and then increases with the increase in the number of simultaneous clients. The main reason for this behavior can be the fact that in average loaded network conditions some of the flows adapt whereas others are affected by loss that affects the end-user perceived quality. In highly loaded conditions all the adaptive streams adjust their bitrate to the available bandwidth, therefore decreasing the loss rate and consequently the end user perceived quality increases. 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 (Fig. 6).

VI. 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 and other adaptive streaming schemes.

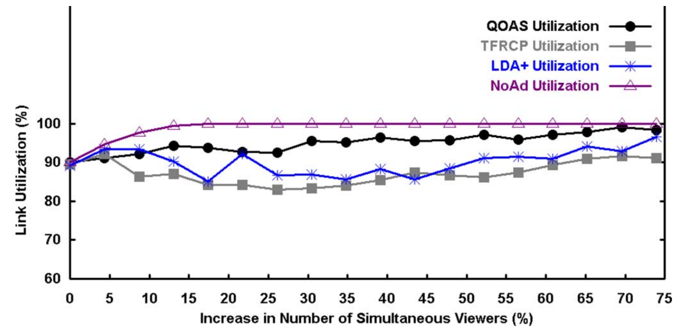


Fig. 6. Bottleneck link utilisation using different approaches, while increasing the number of simultaneous viewers.

QOAS involves extra processing both at the clients and at the server. Like QOAS, both TFRCP and LDA+ involve computation of loss rate and round trip delay and sending feedback to the server. As QOAS's computation of the estimated end-user perceived quality and of the $(\mathrm{QoD}_{\mathrm{scores}})$ has similar complexity to 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 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 $(\mathrm{QoD}_{\mathrm{scores}})$ 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 memory becomes cheaper, this issue will be 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 acknowledgements for every packet, excessively loading the bandwidth. LDA+ uses RTCP packets for feedback carrying information such as loss rate, round trip delay, etc. RTCP feedback usually takes up to 5% of bandwidth. Unlike them QOAS sends only a $(\mathrm{QoD}_{\mathrm{scores}})$ at a time as feedback, saving significant bandwidth. If RTCP packets are used, with standard values for the headers' sizes and a 4-Byte payload, the feedback packet size becomes 40 Bytes long. For a very low inter-feedback transmission time of 0.1 s the bandwidth used by QOAS feedback for a single client becomes 400 Bytes/s. For over 300 customers that could be served simultaneously via a gigabit Ethernet infrastructure feedback will consume only 0.1% of the available bandwidth. This value is significantly lower than the 5% used by LDA+. More details about the performance analysis are presented in [43].

VII. CONCLUSIONS AND FURTHER WORK

This paper shows how the Quality Oriented Adaptation Scheme (QOAS) can be used to stream high bit-rate multimedia in multi-service all-IP delivery networks. This is done so that the customers' need for high quality is balanced with the network operators' and service providers' goal of achieving high infrastructure utilization and simultaneously serving more customers. Extensive objective tests using simulation models have tested QOAS stand-alone and in comparison with other solutions such as TFRCP, LDA+, an ideal adaptive scheme and a non-adaptive mechanism in different delivery conditions.

The objective tests have assessed the effect on QOAS performance of background traffic of various types and sizes and with different variation patterns commonly found in multi-service IP networks. QOAS showed very good performance in terms of end-user perceived quality, loss rate, throughput and link utilization, and was very close to the performance of a hypothetical ideal adaptive scheme. In terms of the number of customers served from an existing infrastructure, by scaling the 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 non-adaptive streaming. These results are also confirmed by the very good subjective testing results presented in [44].

QOAS assumes that no error control mechanisms are employed during multimedia streaming. Further work could study the effect of some error control techniques such as error concealment in conjunction with QOAS in achieving even better performance. Future research could also propose a multicast extension to QOAS so that the infrastructure utilization can be further increased when common content is delivered to a large number of customers (e.g. live streaming).

REFERENCES

[1] J. H. Bakkers, "European Broadband Access Services Market Analysis," 2002-2007 IDC, April 2003 [Online]. Available: http://www.idc.com/getdoc.jhtml?containerId=BT03K

[2] ARC Group, Broadband Access, "Worldwide Market Analysis and Strategic Outlook 2002–2007," January 2002.

[3] L. Harper, White Paper, Cisco Systems, "Building the Next Generation Digital Video Network," July 2003 [Online]. Available: http://www.webtorials.com/main/resource/papers/cisco/paper28.htm

[4] J. Horrobin, Cisco Systems, White Paper, "Delivering MPEG-2 Video Services Over a Multiservice IP Network," [Online]. Available: http://www.webtorials.com/main/resource/papers/cisco/paper10.htm

[5] S. Dravida, D. Gupta, S. Nanda, K. Rege, J. Strombosky, and M. Tandon, "Broadband access over cable for next-generation services: A . switch architecture," *IEEE Communications Magazine*, vol. 40, no. 8, pp. 116–124, August 2002.

[6] G.-M. Muntean, P. Perry, and L. Murphy, "A new adaptive multimedia streaming system for all-IP multi-service networks," *IEEE Transactions on Broadcasting*, vol. 50, no. 1, pp. 1–10, March 2004.

[7] J. Padhye, J. Kurose, D. Towsley, and R. Koodli, "A model based TCP friendly rate control protocol," in *Proc. ACM Int. Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV)*, Basking Ridge, NJ, USA, 1999, pp. 137–151.

[8] D. Sisalem and A. Wolisz, "LDA+ TCP-friendly adaptation: A measurement and comparison study," in *Proc. ACM NOSSDAV*, Chapel Hill, NC, USA, June 2000.

[9] W. Effelsberg and R. Steinmetz*, Video Compression Techniques*. Heidelberg, Germany: dpunkt-Verlag, 1998.

[10] B. Braden, D. Clark, and S. Shenker, "Integrated services in the internet architecture: An overview," RFC 1633, June 1994 [Online]. Available: http://www.ietf.org/rfc/rfc1633.txt

[11] S. Blake *et al.*, "An architecture for differentiated services," RFC 2475, December 1998 [Online]. Available: http://www.ietf.org/rfc/rfc2475.txt

[12] E. C. Rosen, A. Viswanathan, and R. Callon, "MPLS architecture," RFC 3031, January 2001 [Online]. Available: http://www.ietf.org/rfc/rfc3031.txt

[13] D. Hutchison, A. Mauthe, and N. Yeadon, "Quality-of-service architecture: Monitoring and control of multimedia communications," *IEEE Electronics and Comms. Eng. Journal*, vol. 9, no. 3, pp. 100–106, 1997.

[14] *Information Technology—Quality of Service: Framework*, TU-T Recommendation X.641, Dec. 1997.

[15] *Information Technology—Quality of service: Framework*, ISO/IEC 13236, International Standards Organisation, December 1997.

[16] A. Banerjea, D. Ferrrari, B. A. Mah, M. Moran, D. C. Verma, and H. Zhang, "The tenet real-time protocol suite: Design, implementation, and experiences," *IEEE/ACM Trans. on Networking*, vol. 4, no. 1, pp. 1–10, Feb. 1996.

[17] D. Wu, Y. T. Hou, W. Zhu, Y.-Q. Zhang, and J. M. Peha, "Streaming video over the internet: Approaches and directions," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 11, no. 3, pp. 282–300, 2001.

[18] X. Wang and H. Schulzrinne, "Comparison of adaptive internet multimedia applications," *IEICE Trans. on Communications*, vol. E82-B, no. 6, pp. 806–818, June 1999.

[19] W. Feng and J. Rexford, "Performance evaluation of smoothing algorithms for transmitting prerecorded variable-bit-rate video," *IEEE Trans. Multimedia*, vol. 1, no. 3, pp. 302–312, Sep. 1999.

[20] R. Rejaie, M. Handley, and D. Estrin, "RAP: An end-to-end rate-based congestion control mechanism for real-time streams in the internet," in *Proc. INFOCOM*, New York, USA, March 1999, vol. 3, pp. 1337–1345.

[21] ——, "Layered quality adaptation for internet video streaming," *IEEE Journal on Selected Areas of Communications (JSAC)*, vol. 18, no. 12, pp. 2530–2543, Dec. 2000.

[22] S. McCanne, V. Jacobson, and M. Vetterli, "Receiver-driven layered multicast," in *Proc. SIGCOMM*, Stanford, California, USA, August 1996.

[23] L. Vicisano, J. Crowcroft, and L. Rizzo, "TCP-like congestion control for layered multicast data transfer," in *Proc. IEEE INFOCOM*, San Francisco, CA, USA, March 1998, vol. 3, pp. 996–1003.

[24] L. Wang, A. Luthra, and B. Eifrig, "Rate control for MPEG transcoders," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 11, no. 2, pp. 222–234, February 2001.

[25] N. Yeadon, F. Garcia, D. Hutchison, and D. Shepherd, "Filters: QoS support mechanisms for multipeer communications," *IEEE Journal on Selected Areas in Computing (JSAC)*, vol. 14, no. 7, pp. 1245–1262, Sept. 1996.

[26] J. C. Bolot and T. Turletti, "Adaptive error control for packet video in the internet," in *Proc. ICIP*, Lausanne, Switzerland, Sep. 1996, pp. 25–28.

[27] S. Jacobs and A. Eleftheriadis, "Providing video services over networks without quality of service guarantees," in *WWW Consortium Workshop on Real-Time Multimedia and Web*, Sophia Antipolis, France, Oct. 1996.

[28] D. Sisalem and H. Schulzrinne, "The loss-delay based adjustment algorithm: A TCP-friendly adaptation scheme," in *Proc. ACM NOSSDAV*, Cambridge, UK, July 1998.

[29] J. Padhye, V. Firoiu, D. Towsley, and J. Kurose, "Modelling TCP throughput: A simple model and its empirical validation," in *Proc. ACM SIGCOMM*, Vancouver, Canada, October 1998, pp. 303–314.

[30] RealNetworks, "SureStream," [Online]. Available: http://www.realnetworks.com

[31] Microsoft, "Windows Media," [Online]. Available: http://www.microsoft.com

[32] G.-M. Muntean, P. Perry, and L. Murphy, "Performance comparison of local area video streaming systems," *IEEE Communication Letters*, vol. 8, no. 5, pp. 326–328, May 2004.

[33] G. Ghinea and J. P. Thomas, "QoS impact on user perception and understanding of multimedia video clips," in *Proc. of ACM Multimedia*, Bristol, United Kingdom, 1998, pp. 49–54.

[34] L. Zhang, L. Zheng, and K. S. Ngee, "Effect of delay and delay jitter on voice/video over IP," *Computer Communications*, vol. 25, no. 9, pp. 863–873, June 2002.

[35] J.-C. Bolot and T. Turletti, "Experience with control mechanisms for packet video in the internet," *ACM Computer Communication Review*, vol. 28, no. 1, pp. 4–15, January 1998.

[36] O. Verscheure, P. Frossard, and M. Hamdi, "User-oriented QoS analysis in MPEG-2 video delivery," *Journal of Real-Time Imaging*, vol. 5, no. 5, pp. 305–314, October 1999.

[37] "Subjective Video Quality Assessment Methods for Multimedia Applications," ITU-T Recommendation P.910, September 1999.

[38] "Network Simulator-2," [Online]. Available: http://www.isi.edu/nsnam/ns/

[39] B. Krishnamurthy, C. Wills, and Y. Zhang, "On the use and performance of content distribution networks," in *Proc. ACM SIGCOMM, Internet Measurement Workshop*, San Francisco, USA, Nov. 2001, pp. 169–182.

[40] ——, Preliminary Measurements on the Effect of Server Adaptation for Web Content Delivery AT&T Labs and Research, Technical Report TD-59VNB8, April 2002.

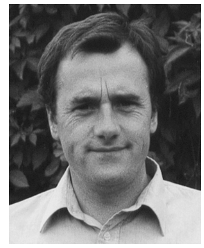
[41] NIST Net, [Online]. Available: http://snad.ncsl.nist.gov/itg/nistnet

[42] Internet2 QoS Working Group, "Network QoS Needs of Advanced Internet Applications—A Survey," 2002.

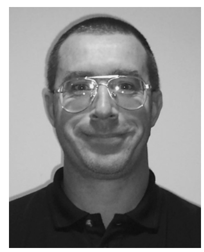
[43] G.-M. Muntean, "Efficient delivery of multimedia streams over broadband networks using QOAS," *IEEE Trans. on Broadcasting*, vol. 52, no. 2, pp. 230–235, June 2006.

[44] G.-M. Muntean, P. Perry, and L. Murphy, "Subjective Assessment of the Quality-Oriented Adaptive Scheme," *IEEE Trans. on Broadcasting*, vol. 51, no. 3, pp. 276–286, September 2005.

Gabriel-Miro Muntean is a Lecturer in the School of Electronic Engineering, Dublin City University, Ireland, where he obtained his Ph.D. degree in 2003 for research in quality-oriented adaptive multimedia streaming over wired networks. He was awarded the B.Eng. and M.Sc. degrees in Software Engineering from the Computer Science Department, "Politehnica" University of Timisoara, Romania in 1996 and 1997 respectively. Dr. Muntean's research interests include QoS and performance issues of adaptive multimedia streaming, and personalized eLearning over wired and wireless networks and with various devices. Dr. Muntean has published many papers in prestigious international conferences and journals, as well as a book and a book chapter. Dr. Muntean is reviewer for important international journals, conferences and funding agencies. He is member of IEEE.



Philip Perry (M'92) is a Senior Research Fellow in the Performance Engineering Laboratory, with responsibilities in both the School of Computer Science and Informatics at University College Dublin, Ireland and the School of Electronic Engineering at Dublin City University, Dublin, Ireland. He obtained his Ph.D. in Microwave Engineering from the Department of Electronics and Electrical Engineering at University College Dublin, Ireland in 1998. Dr. Perry obtained his Master's degree at the University of Bradford, Yorkshire, England (1989) while his primary degree is from the University of Strathclyde, Glasgow, Scotland (1987). His current research interests are focused on time constrained applications and enabling technologies for mobile systems.



Liam Murphy obtained his B.E. in Electrical Engineering from University College Dublin in 1985, and Masters and Ph.D. degrees in Electrical Engineering and Computer Sciences from the University of California, Berkeley, in 1988 and 1992 respectively. He is currently a Senior Lecturer in Computer Science at University College Dublin, Ireland, where he is Director of UCD's Performance Engineering Laboratory. Dr. Murphy has published approximately 100 refereed journal and conference papers on various topics, including multimedia transmissions, dynamic and adaptive resource allocation algorithms, and software development. His current research projects involve mobile and wireless systems, computer network convergence issues, and web services performance. He is a Member of the IEEE.