Receiver-Initiated Resource Renegotiation for VBR Video Transport

Wenjun Luo
3Com Corp.
wenjun_luo@3com.com

Magda El Zarki
University of Pennsylvania
magda@ee.upenn.edu

Abstract

In this paper we address the important issue of providing QoS for VBR video communications in an efficient manner. We show that efficient transmission of VBR video with a high QoS is feasible when using a receiver-initiated resource renegotiation (RIR) scheme. The scheme for RIR is based on RTP and RSVP. RTP's media specific header is used to send video source information to receivers. Receivers utilize this information to estimate the traffic descriptors. Renegotiations are triggered based on the receiver's buffer status and RSVP is used to renegotiate flow parameters with the network. The performance of the proposed scheme is evaluated via simulations using several 20-minute-long MPEG-2 bit streams. Performance metrics considered are video quality and renegotiation overhead for different receiver buffer sizes and network delays. The results show that the proposed RIR scheme provides high video quality with an average renegotiation interval on the order of seconds, a 5-15 frames receiver buffer and network renegotiation delay below 300 msec. We also investigated call admission control (CAC) schemes for renegotiated VBR services. In particular, we studied the performance of the Central Limit Theorem based and the Chernoff Bound based CAC algorithms, in terms of error in the calculation of the maximum number of admissible connections. Finally, to facilitate the scaling of the renegotiation scheme to large-scale networks, we studied how to reduce renegotiation overhead.

1. Introduction

In existing and currently proposed integrated services frameworks, such as ATM [1][2] and the integrated services architecture (ISA) [3][4] for the IP based networks, call admission control (CAC) and usage parameter control (UPC) are jointly implemented to ensure quality of service (QoS) for admitted connections. CAC is implemented to prevent over-subscription while UPC is used to detect contract-violating traffic streams. Under these mechanisms, the efficient and robust provision of QoS for variable-bit-rate (VBR) video is a challenging issue because VBR video bit streams exhibit significant burstiness at multiple time-scales [5][6].

Currently ATM based networks support VBR services using a static UPC scheme. Thus, the traffic descriptors remain unchanged during the connection's lifetime. Earlier studies have shown that with efficient resource allocation, e.g., the leaky bucket (LB) token rate is equal to the average source rate and the bucket size is equal to the source peak rate, static UPC based on the LB algorithm would tag about 10^{-2}–10^{-3} of the bit stream as nonconforming traffic, which receives no guarantees for delivery [7][8].

Error concealment can be implemented at the decoder to reconstruct the lost picture components via interpolation techniques so as to improve the visual quality of the video sequence under cell losses. However, the resultant cell loss rate of 10^{-2} – 10^{-3} exceeds the limit of 10^{-5} under which typical error concealment techniques can effectively conceal the errors [7].

The above problem is caused by the fact that static LB parameters cannot match the long term correlation or the long bursts over the large time scales that exist in VBR video sequences. One possible solution is to dynamically assign traffic descriptors to match VBR video's high bit rate variation. High video quality can be maintained if the renegotiation scheme can accurately negotiate the necessary traffic descriptors for different scenes in the video sequence. Efficient resource utilization can also be achieved if the assigned token rate is equal to the average source rate during each of the renegotiation periods. The overall average assigned token rate during the connection's lifetime would then be equal to the overall average source rate.

There have been several studies on resource renegotiation for VBR video transport [9][10][11]. These studies all assume resource reservation is requested by the sender. However, the current ATM standard does not include native support for renegotiation. On the other hand, the Internet ISA working group has standardized the Resource

0-7803-5467-2/99/ $ 10.00 © 1999 IEEE

Reservation Protocol (RSVP), which inherently supports renegotiation initiated by the receivers [12]. Recently there have been some studies related to running RSVP on top of ATM [13][14]. We will assume the use of RSVP for the renegotiation purpose and propose receiver-initiated renegotiation (RIR) for VBR video transmission.

In addition to the advantage of easy integration with RSVP, the RIR approach also has the following potential advantages: 1) for unicast services, when there are a large number of connections to a single video server, the RIR approach could simplify and reduce the load of the server by distributing the renegotiation associated computation to the receivers; 2) RIR can adapt to the CPU load of the receiver by changing the bandwidth requirement accordingly; 3) when the receiver is connected to networks without bandwidth guarantees such as Ethernet or a wireless channel, RIR can also adapt to the local traffic load or channel conditions by reacting to packet losses or delay.

This paper is organized as follows: Section 2 outlines the algorithm proposed for RIR and presents performance results. Section 3 investigates the issue of how to provide QoS for renegotiated VBR services. Section 4 studies the practical but important issue of reducing renegotiation overhead. Section 5 presents our conclusions.

2. Receiver-initiated renegotiation

The essential functions of the receiver renegotiation module at the receiver side are first, to detect the necessity to do renegotiation, and second, to estimate future bit rate behavior. Renegotiations are triggered based on the receiver's buffer status. When the module detects the necessity to do renegotiation (the buffer occupancy passes the preset low or high buffer threshold), it *estimates* the required LB token rate λ_s and peak rate λ_p (based on a sliding window of previous frame bit rates with adaptive window size which is adjusted dependent on short- or long- term bit rate variation) and sends the new reservation request in a RSVP packet along the path back to the sender. The soft state property of RSVP is utilized here for renegotiation as the receiver is expected to refresh the resource reservation periodically in RSVP.

After the sender RSVP module receives the new RSVP packet, it notifies the sender rate control module to update the target bit rate and buffer parameters. In this study we assume that source rate control is implemented to ensure that the generated traffic always conforms to the current contract. If no source rate control is implemented, undesirable cell loss due to input buffer overflow could occur. However, as we will show later, source rate control operates only occasionally as long as the renegotiation scheme can reserve resources in an accurate and timely manner and hence high video quality can still be efficiently maintained.

In Figure 1, we show the typical renegotiation dynamics using the above renegotiation algorithm. The MPEG-2 [15] video sequence used is a 20-minute-long segment from the action-packed movie *Terminator II* coded with parameters Q=4, N=12 and M=2. We observe that the renegotiation algorithm can keep up with the high bit rate variations with reasonable responsiveness.

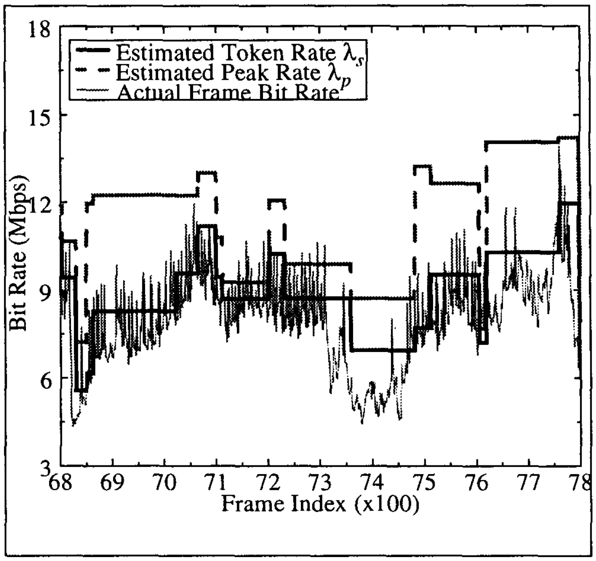

Figure 1. Renegotiation dynamics.

We now present performance results of the receiver-initiated renegotiation scheme described above. In our study we measure the performance of RIR by the following metrics:

- The resultant picture quality of the video sequence reflected by the resultant quantization scales (Q) used to generate the video bit stream: We use the following two criteria to evaluate the video quality: 1) mean quantization score (MQS) computed as

$$MQS = \sum_{i=1}^{5} (6-i) \bullet N_{Q=2i} / N_{Total} \qquad (1)$$

where $N_{Q=k}$ represents the number of frames that are coded with Q equal to k and N_{total} represents the total number frames in the sequence. In our simulation, we set Q_{target} to be 4, hence target MQS value is equal to 4; 2) Average duration of Q equal to Q_{target} and average duration of Q unequal to Q_{target}.

- Accuracy of rate estimation measured by the ratio of the estimated token rate against the actual average video source bit rate.
- Average renegotiation interval (ARI): the longer the ARI the smaller the renegotiation overhead.

Due to the limited space, we only present in Figure 2 the results of MQS and Q duration versus receiver buffer size. The round trip delay is assumed to be 2 frames (66 milliseconds at 30 frames per second) as the U.S.A. coast-to-coast round-trip propagation delay is 42 ms.

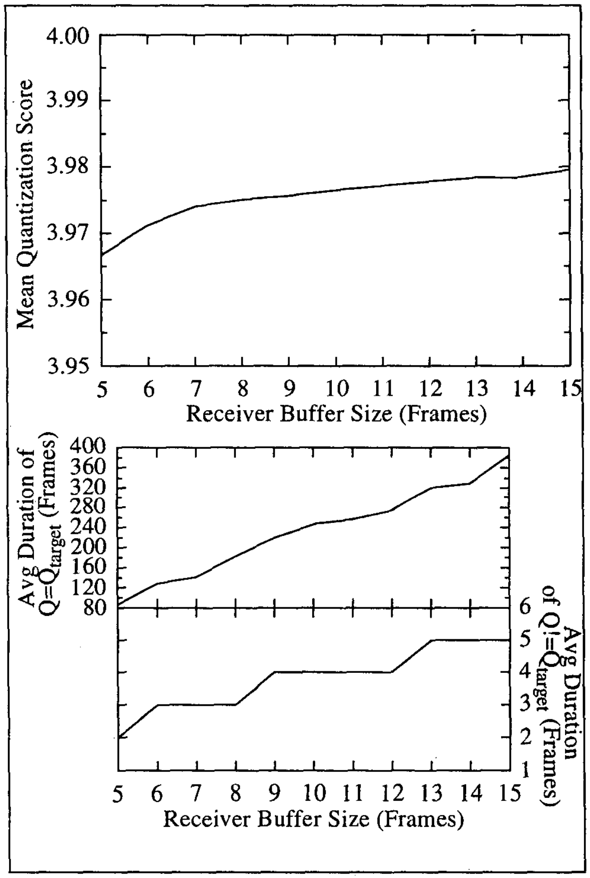

Figure 2. MQS and Q duration versus receiver buffer size.

We observe that both MQS and the average duration of Q equal to Q_{target} increases as buffer becomes larger. Especially the average duration of Q equal to Q_{target} improves drastically from 80 to about 400 frames, which means a degradation occurs every 3-13 seconds and lasts only for 0.05- 0.15 second. Since $Q!=Q_{target}$ usually occur after scene changes and human beings usually do not sense degradation immediately after a scene change (temporal masking effect), the viewing quality remains very high. Simulation results also indicate that the ratio of estimated token rate to the actual average rate is about 1.02, showing high estimation accuracy. The ARI result is also in range with other sender-initiated renegotiation schemes for video proposed in [9][10][11] under the same small buffer setting.

Judging from the above results, we believe that the proposed renegotiation algorithm's performance is reasonably good as it produce very good video quality by matching VBR video's bit rate variations with high accuracy and small renegotiation overhead.

3. Traffic Control Schemes

We now shift the focus to investigating call admission control schemes that can ensure high rate of renegotiation successes and hence provide QoS guarantees for renegotiated VBR video services. For renegotiated VBR video services, one important QoS parameter is the renegotiation failure ratio (RFR) which refers to the fraction of time during which the connection does not obtain the required network resources. In order to maintain a high video quality, the RFR needs to be small and controlled by the CAC procedure. When a new connection request is received at a switch, only when the switch determines that the RFR of the existing connections will not be lower than the specified RFR, can the new connection request be granted. Hence the CAC module needs to evaluate the level of statistical multiplexing during the connections' lifetime and compute the corresponding RFR.

For this purpose, we study the central limit theorem (CLT) and the Chernoff bound based CAC algorithms [16][17]. To measure their performance, we first run a trace-driven simulation (TDS) based on the video traces we have. For three channel capacities, we keep increasing the number of video connections until the resultant RFR exceeded a certain threshold (via simulations, we found out that with an RFR on the order of 10^{-2}, the quality of video can be maintained at a high level). We denote this number as N_{TDS}, which represents the optimum number of admissible connections to maintain a low RFR. Then using the two CAC algorithms, we calculate the number of admissible connections denoted as N_{CLT} and $N_{Chernoff}$ respectively based on the same RFR. The accuracy of the CAC algorithm can then be measured via the error term:

$$error = \frac{N_{CLT}\ or\ N_{Chernoff} - N_{TDS}}{N_{TDS}} \bullet 100 \qquad (2)$$

In Table 1, we present this error term for the video sequences and three different channel capacities. We observe that the CLT based CAC seems to always oversubscribe connections, but the error becomes very small for large channel capacities. This is because the more connections that are multiplexed, the more accurate the CLT approximation. On the other hand, the Chernoff Bound based CAC almost always underestimates the admissible connections, except for a few sequences in the low channel capacity scenario. The absolute value of the error seems to be smaller than that of the CLT based CAC for small channel capacities. It also tends to grow smaller with larger channel capacities. We also observe some differences between different sequences. The CAC is more accurate for sequences with less motion such as *Larry King*. For the sequences with the highest motion (*NCAA Finals*), the CAC is the least accurate.

Table 1. Accuracy of two CAC schemes under differnt channel capacities.

	C=155 Mbps		C=650 Mbps		C=1250 Mbps	
	CLT	Chernoff	CLT	Chernoff	CLT	Chernoff
Terminator II	11.76	5.88	6.25	-3.75	1.21	-2.03
Jurassic Park	10.44	-5.37	6.03	-3.58	1.09	-1.85
The Abyss	9.13	-5.44	5.87	-3.02	1.16	-1.13
Star Wars	12.02	3.59	7.17	-3.84	1.98	-2.17
Indiana Jones	11.24	4.77	6.59	-4.07	1.57	-1.88
Larry King	5.17	-3.92	3.26	-2.11	0.84	-0.67
NCAA Finals	15.74	-7.34	8.07	-4.12	2.46	-3.25

4. Reducing Renegotiation Overhead

Finally, we study a practical yet very important issue for renegotiated services: how to reduce the renegotiation overhead in the presence of multiple connections. For large-scale networks, hundreds or even thousands of VBR video connections can be active at the same time. Processing large amounts of renegotiation requests is not desirable.

The intrinsic nature of scene changes in a VBR video bit stream mostly determines the renegotiation interval. Although it is possible to enforce a longer renegotiation interval in the renegotiation algorithm, this action would result in video quality degradation as more source rate control has to be carried out to compensate for the rate mismatches. Instead of increasing the renegotiation interval for each individual connection, we can try to reduce the total number of renegotiations in the network.

The basic idea is to cluster a certain number of connections and carry out renegotiation on behalf of the cluster instead of dealing with the individual connections one by one. The renegotiation interval is expected to increase due to the fact that after statistical multiplexing, the aggregated bit stream is less bursty than the individual sequences' streams.

Figure 3 presents the renegotiation interval versus the number of connections we cluster together. We can see that the interval ranges from about a second for around 100 connections to 10 seconds for around 1000 connections. The reduction in renegotiation is considerable when compared to our earlier results of doing renegotiation on a per connection basis. For example, for 100 connections, the occurrence of renegotiation is reduced by a factor of 100 since there would be 100 renegotiations per second on the average for the per connection scheme. This simulation result shows that the renegotiation overhead can be substantially reduced and hence makes renegotiation feasible in large-scale networks.

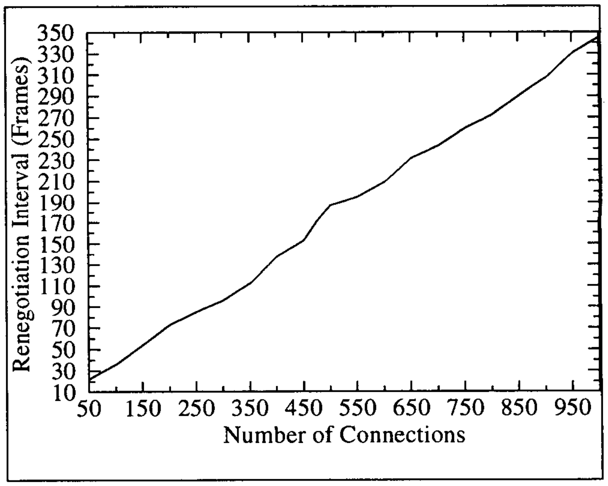

Figure 3. Renegotiation interval versus number of connections.

5. Summary

In this paper we showed that efficient transmission of VBR video with a high QoS is feasible when using receiver-initiated resource renegotiation. The scheme for RIR is based on RTP and RSVP. RTP's media specific header is used to send video source information to receivers. Receivers utilize this information to estimate the traffic descriptors. Renegotiations are triggered based on the receiver's buffer status and RSVP is used to renegotiate flow parameters with the network. The performance of the proposed scheme is evaluated via simulations using several 20-minute-long MPEG-2 bit streams. Performance metrics considered are video quality and renegotiation overhead for different receiver buffer sizes and network delays. The results show that the proposed RIR scheme provides high video quality with an average renegotiation interval on the order of seconds, a 5-15 frames receiver buffer and network renegotiation delay below 300 msec.

In section 3, we investigated traffic control schemes for renegotiated VBR services. We focused on call admission control, which is an important part of the traffic control scheme to provide QoS guarantees. We first studied the impact of RFR on video quality. Via simulations, we showed that with an RFR on the order of 10^{-2}, the quality of video can be maintained at a high level. Then we investigated two particular CAC algorithms: the Central Limit Theorem based and the Chernoff Bound based approximation. In particular, we studied their performance in terms of both error in the calculation of the maximum number of admissible connections and the resultant video quality. We showed that in terms of the error, the Chernoff Bound based CAC has a better performance, however, judging from the resultant video quality, the CLT based CAC performs reasonably well while maintaining a low computing complexity.

To facilitate the scaling of the renegotiation scheme to large-scale networks, two methods are proposed in section 4. A scheme for aggregation of renegotiations without compromising the QoS of individual connections is considered. Simulations show that substantial reduction on renegotiation frequency is possible without significantly affecting bandwidth efficiency.

6. References

[1] ITU Study Group 13, "Traffic Control and Congestion Control in B-ISDN", Perth, Australia, November 1995.

[2] ATM Forum Technical Committee, "Traffic Management Specification Version 4.0", April 1996.

[3] S. Shenker and C. Partridge, "Specification of Guaranteed Quality of Service", Internet Draft, work in progress, July 1996.

[4] J. Wroclawski, "Specification of The Controlled Load Network Element Service", Internet Draft, work in progress, July 1996.

[5] Gunnar Karlsson, "Asynchronous Transfer of Video", IEEE Communications Magazine, Vol. 34, No. 8, pp. 118-126, August 1996.

[6] M. W. Garret, "Contributions Toward Real-Time Services on Packet Networks", Ph.D. thesis, Columbia University, May 1993.

[7] W. Luo and M. El Zarki, "Analysis of Error Concealment Techniques for MPEG-2 Video Transmission over ATM based Networks", in Proceedings of SPIE/IEEE Visual Communications and Image Processing'95, Taipei, Taiwan, May 1995.

[8] E. Rathgeb, "Policing of Realistic VBR Video Traffic in ATM Networks", International Journal of Digital and Analog Communications System, Vol. 6, No. 5, pp.213-226, May 1993.

[9] M. Grossglauser, S. Kashev and D. Tse, "RCBR: A Simple and Efficient Service for Multiple Time-Scale Traffic", Proceedings of SIGCOMM'95, Cambridge, MA, pp. 219-230, September 1995.

[10]D. Reininger and D. Raychaudhuri, "Bandwidth Renegotiation for VBR Video over ATM Networks", IEEE Journal on Selected Areas in Communications, Vol. 14, No. 6, pp.1076-1086, August 1996.

[11]H. Zhang and E. W. Knightly, "RED-VBR: A Renegotiation-Based Approach to Support Delay-Sensitive VBR Video", Technical report, University of California, Berkeley, to appear in ACM/Springer-Verlag Multimedia Systems Journal, 1997.

[12]E. Braden, L. Zhang, D. Estrin, S. Herzog and S. Jamin, "Resource Reservation Protocol (RSVP) - Version 1 Functional Specification", Internet Draft, work in progress, June 1996.

[13]A. Birman, V. Firoiu, R. Guerin and D. Kandlur, "Support for RSVP-based Services over ATM Networks", Proceedings of Global Internet'96, London, England, pp. 140-151, November 1996.

[14]M. Borden, E. Crawley, J. Krawczyk, F. Baker and S. Berson, "Issues for RSVP and Integrated Services over ATM", Internet Draft, work in progress, February 1996.

[15]ISO/IEC 13818-2, "Generic Coding of Moving Picture and Associated Audio", MPEG-2 Draft International Standard, May 1994.

[16]H. Heeke, "Statistical Multiplexing Gain for VBR Video Codecs in ATM Networks", International Journal of Digital and Analog Communications System, Vol. 4, No. 4, pp. 261–268, February 1991.

[17]Z. L. Zhang, D. Towsley, J. D. Salehiy and J. Kurose, "Smoothing, Statistical Multiplexing and Call Admission Control for Stored Video", to appear in IEEE Journal of Selected Areas in Communications, August 1997.