Receiver-Initiated Bandwidth Renegotiation Based On Next Generation Internet Protocols: VBR Video Over IP Nets

Wenjun Luo, Magda El Zarki

1. Introduction

The *Internet Engineering Task Force* (IETF) is developing a suite of standards [1,2,3,4,5] called the *Integrated Services Architecture* (ISA), which permits networks to provide Quality of Service (QoS) transport over IP nets. Under such a infrastructure, the *Real-time Transport Protocol* (RTP) can deliver real-time traffic with timing information for reconstruction as well as feedback on reception quality, while the *Resource Reservation Protocol* (RSVP) allows applications reserve network resource dynamically.

For high-end real-time interactive video service, bandwidth (BW) renegotiation [7,8] can greatly improve the video quality by dynamically requiring and assigning BW according to the video bitstream dynamics and user interactions. The on-going work at NEC is based on a native ATM framework. Hence the sender (video server) calculates the required BW and initiates the renegotiation by sending request directly to the ATM switches.

However, RSVP mandates that a resource reservation be initiated by the receiver, because sender-initiated reservation scales poorly for large, dynamic multicast delivery trees and for heterogeneous receivers (e.g., the coexistence of wireless terminal and wired workstation). Receiver-initiated reservation deals with this by having each receiver request a reservation appropriate to itself according to its receiving ability, the channel characteristic and reception quality [6].

RSVP may also make unicast reservation. The sender-initiated approach has the advantage of having the rate control and the renegotiation integrate more efficiently at the sender's side. But it can be shown that the receiver-initiated approach provides the same functionality with comparable responsiveness to the bit rate variation. It also responds much faster to certain user actions that change the resource requirement (e.g., user changes resolution on the fly) especially for multiple hops. Moreover, for the popular scenario of multiple clients retrieving video from the same video server, it also offloads the server from calculating required resource for all the connections, and eliminates the possible congestion in the downstream caused by sender's sending all the renegotiation packets for each connection to the server's frontend router/switch.

Therefore here we propose the study of receiver-initiated reservation for VBR video transport using RSVP and RTP. In section 2, we describe the basic framework. Several important issues that are worthy of research are outlined in section 3.

2. The Basic Framework

In Fig. 1, we outline a typical VBR video session based on the ISA infrastructure and the receiver-initiated renegotiation model.

After the connection setup between the server and the client (including the initial resource reservation), the video bit stream is sent from the video server to the server RTP, which has a traffic shaper to generate Leaky Bucket (LB) conforming traffic, down the protocol stack, through the network and then up to the video client, which is the data path.

During the connection, the receiver renegotiation daemon keeps track of the receiver buffer's occupancy obtained from client RTP. Whenever an abrupt change in the buffer occurs, it recomputes the required BW according to the frame size trace and sends the new reservation request to RSVP, which in turn reserves resource along the path back to the sender. The sender RSVP then passes the new resource information to the sender RTP to update the LB parameters in the traffic shaper. It can be easily shown that the times it takes from detecting the bit rate variation to successfully sending the new resource information to the server is comparable to that of the sender-initiated approach.

The above operation is carried out regardless the control in the server, i.e., the receiver has no knowledge of what happens in the server, they are distributed in a strict sense. If ideally the client

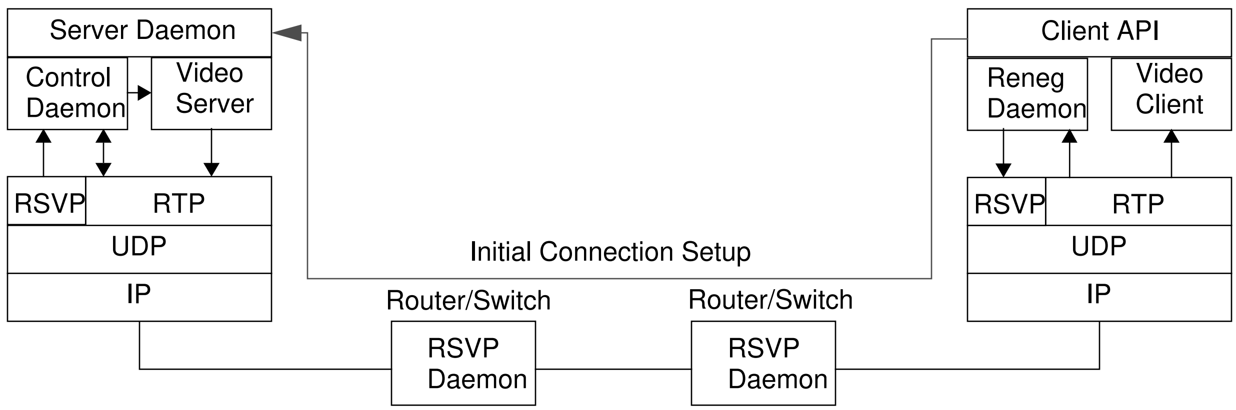

Fig. 1. Receiver-initiated BW renegotiation.

renegotiation module could always accurately request enough resources and it takes no time to do the renegotiation, the input buffer at the server side should never misbehave. However, because of the inevitable end-to-end delay and processing delay for renegotiation requests, it might be necessary for the server to temporarily control the bit rate to avoid input buffer overflow before new resource information comes in.

If the video server allows rate control, a server control engine is implemented at the sender. It keeps track of server RTP's input buffer occupancy and adjusts the quantizer (Q) accordingly to avoid buffer overflow [9]. When the receiver's new reservation comes in, it can switch the Q back to the normal value.

If no rate control is allowed, the absence of rate control in server might result in a high loss rate due to input buffer overflow (loss due to network congestion should be minimum since the output traffic is LB conforming), or a large packet delay ca occur if the sender uses a huge buffer to prevent the overflow. The client can react to the loss and delay with a more aggressive renegotiation approach to respond more quickly to the receiver buffer variation. However, this may result in overreaction and hence over-renegotiation and even instability.

3. Important Issues to study

There are several important issues that need to be studied thoroughly as they have great impact on the system performance.

3.1 Receiver-Initiated Renegotiation Specific Issues

- The relationship between receiver's buffer occupancy, the renegotiation necessity and how much to reserve, especially how to accurately determine the resource requirement. Preliminary buffer analysis has already shown some interesting differences from the sender-initiated approach.
- The form of aggressive renegotiation, and its relationship with packet loss rate and/or delay variation.
- Impact of loss and delay incurred by the network.
- If rate control is allowed in the video server, it could happen that the receiver buffer is OK while the video server is using a higher- or lower-than-specified Q. Certain algorithm needs to be devised in the reneg control engine to keep track of the receiving video quality and the scale the requested resource accordingly.
- RTP optionally uses an input smooth buffer to shape the traffic to certain LB envelope. We studied in detail the two-buffer (an input smoothing buffer plus a token buffer) model, rate control and their impact on video quality in [9]. The integration of that study with this system needs to be looked into.
- In the scenario of multicast, integrate the dynamic MPEG-2 scalability studied in [9] to support heterogeneous receivers.

3.2 Issues Common to Both Sender- and Receiver- Initiated Renegotiation

- Impact of renegotiation delay and input buffer size: This is a common issue for both sender- and receiver-initiated renegotiation. Our initial work shows that these two parameters have critical and profound impact on the video quality. Further work can be carried out to determine a quantitative relationship.
- Statistical multiplexing gain (SMG): In the presence of rate control, there is a direct relationship between renegotiation interval (time between two successive renegotiation request), video quality and SMG (hence cost of connection). For example, the longer the renegotiation interval, the more the rate control has to smooth the video stream, the higher the SMG generally. The knowledge about trade-off between renegotiation interval and SMG might be helpful for the network service provider to determine the tariff model and call admission control. As the renegotiation interval may map indirectly to the cost of the connection, the trade-off between video quality and the interval would be of interest to the client.
- Renegotiation admission control: For the rate control based on MPEG-2 TM5 used in our study [9], a simple source model can be brought up. The concept of equivalent bandwidth [10] can hence be used to determine the admission.

Reference

[1] L. Zhang, etc., *Resource Reservation Protocol - Version 1 Functional Specification*, work in progress, May, 1996.

[2] F. Baker, etc., *Specification of Committed Rate Quality of Service*, work in progress, June 1996.

[3] S. Shenker, etc., *Specification of Guaranteed Quality of Service*, work in progress, June 10,1996.

[4] V. Jacobson, etc., *RTP: A Transport Protocol For Real Time Applications*, RFC 1889

[5] D. Huffman, G. Fernando and S. Kleiman, *RTP Payload Format for MPEG Video*, work in progress, June 1, 1996.

[6] J. Estrin, etc., *Multimedia Over IP: Specs Show the Way*, Data Communications, Aygust 1996.

[7] D. Reinninger and D. Raychaudhuri, *Bandwidth Renegotiation for VBR Video Over ATM Networks*, IEEE JSAC, August 1996.

[8] B. Mark, *Real-Time Estimation and Dynamic Renegotiation of UPC Parameters for VBR Video over ATM*, 7th International Workshop on Packet Video, March 1996, Australia.

[9] W. Luo, M. El Zarki, *MPEG-2 video Transmission over ATM Networks; Quality Control*, To appear in JSAC special issue on real-time video services in multimedia networks, May, 1997.

[10] D. Mitra, etc., *A New Approach for Allocating Buffers and Bandwidth to Heterogeneous, Regulated Traffic in an ATM Node*, IEEE JSAC, August 1995.

Wenjun Luo received the B.S. degree in Electronic Engineering from the Jiao Tong University in 1992, where he was also awarded the honor of “Distinguished Student” at graduation. He received the M.S. degree from the Electrical Engineering Department of the University of Pennsylvania in 1994. He is now a Ph.D. candidate in the same department, where he also serves as a Teaching Assistant and Research Assistant. His current research interests include video compression, error concealment, packet video transmission and QoS control in high speed networks, and multimedia system.

Magda El Zarki (S’81–M’88) received the B.E.E. degree from Cairo University, Cairo, Egypt, in 1979, and the M.S. degree in electrical engineering from Columbia University, New York, NY, in 1981. She received the Ph.D. degree in electrical engineering in December 1987.

She worked from 1981-1983 as a communications-network planner in the Department of International Telecommunications at Citibank, New York, NY. She joined Columbia University in 1983 as a Research Assistant in the Computer Communications research Laboratory where she was involved in the design and development of an integrated LAN testbed called MAGNET. Currently she holds the position of Associate Professor in the Department of Electrical Engineering at the University of Pennsylvania, Philadelphia, where she also holds a secondary appointment in the Department of Computer and Information Sciences and is on the Distributed Systems Laboratory Committee.

Ms. El Zarki is a member of the ACM, IEEE and Sigma Xi. She is actively involved in many of their sponsored conferences and journals. She was the Technical Program Chair of IEEE INFOCOMM’94.