Available online at www.sciencedirect.com
SCIENCE DIRECT®

The Journal of Systems and Software 75 (2005) 271–282

The Journal of Systems and Software

www.elsevier.com/locate/jss

An adaptive rate-control streaming mechanism with optimal buffer utilization

Shu-Ching Chen [a,*], Mei-Ling Shyu [b], Irina Gray [a], Hongli Luo [b]

[a] *Distributed Multimedia Information System Laboratory, School of Computer Science, Florida International University, Miami, FL 33199, USA*

[b] *Department of Electrical and Computer Engineering, University of Miami, Coral Gables, FL 33124, USA*

Received 2 February 2003; received in revised form 6 August 2003; accepted 25 October 2003
Available online 24 March 2004

Abstract

In this paper, an end-to-end real-time adaptive protocol for multimedia transmission is presented. The bandwidth is dynamically allocated according to the network status, and the client buffer occupancy and playback requirement. The transmission rate is determined by the quadratic probing algorithm that can obtain the maximal utilization of the client buffer and minimal occupation of the network bandwidth. It is also coupled with a congestion control mechanism that can effectively decrease the packet loss rate during network congestion. We investigate the performance of our quadratic probing algorithm in different congestion levels under both the local area net (LAN) and Internet environments. Performance analysis reveals that our approach is more robust in avoiding overflows and underflows in different network congestion levels, and adapting to the changing network delays. Comparisons are made with the fixed rate approach and the rate by playback requirement approach. The experimental results show that our proposed real-time protocol with the rate adjusting quadratic probing algorithm is efficient in utilizing the network resources and decreasing the packet loss ratios.

Keywords: Multimedia streaming; Protocol; Adaptive transmission rates; Buffer utilization

1. Introduction

Efficient delivery of streaming media over the Internet poses many challenges. Distributed multimedia applications have different Quality of Service (QoS) requirements. For example, the transmission of real-time video requires interactivity, low jitter, low delay and higher bandwidth but can tolerate some transmission errors (Wu et al., 2000). However, the Internet is a best-effort network and dose not provide QoS guarantee for multimedia services. The application must be aware of the conditions of the network and adapts the multimedia transmission to the network conditions. Therefore, it is important to design an adaptive and reliable multimedia streaming protocol that can cope with varying Internet conditions.

Different approaches may be considered to address the QoS requirements. Adaptive rate control is to adjust the bandwidth used by an application according to the existing network conditions. This approach has the advantage of better utilizing available network resources (which change with time) compared to those approaches relying on resource reservation (e.g., RSVP) (Braden et al., 1997). RSVP requires that all intermediate routers have QoS supports. According to Wang and Schulzrinne (1999), adaptive control schemes presented in the literature can be generally classified into three categories: sender-driven, receiver-driven and transcoder-based. Sender-driven adaptation schemes fall into two categories: buffer based and loss based. Buffer based adaptation schemes use the occupancy of a buffer on the transmission path as a measure of congestion (Jacobs and Eleftheriadis, 1998; Kanakia et al., 1995). Loss based adaptation schemes adjust the rate based on the packet loss experienced by the receivers (Busse et al., 1996; Sisalem and Schulzrinne, 1998). In receiver-driven

* Corresponding author. Tel.: +1-305-348-3480; fax: +1-305-348-3549.
E-mail address: chens@cs.fiu.edu (S.-C. Chen).

0164-1212/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.jss.2003.10.031

adaptation, the receivers individually select the transmission of a particular quality according to their needs and capabilities. A number of receiver-driven schemes use a combination of layered encoding, and a layered transmission scheme. The receiver selects a transmission quality appropriate to its requirements and constraints by subscribing to a certain number of multicast groups carrying different layers. The receiver monitors network congestion (based on the parameters such as packet loss and throughput), and adapts to the changes in the network conditions by adding or dropping layers accordingly. Transcoder-based approaches use multimedia gateways at the appropriate locations in the network, which convert through transcoding a high bandwidth transmission into a transmission with the appropriate bandwidth, thus to accommodate groups of poorly connected receivers. The gateway may also use an adaptive rate-control algorithm to adjust its transmission in response to the receiver feedback (Kouvelas et al., 1998).

Efficient delivery of multimedia streams over the Internet also requires that the media react to the network congestion by adapting their transmission rates. Since the routers typically do not actively provide congestion control (Braden et al., 1998), end-to-end congestion control is more recommended for Internet multimedia transmission (Floyd and Fall, 1999). Real-time streams are also expected to share the network bandwidth with the dominant TCP flows to obtain the inter-protocol fairness. Some approaches adjust the transmission rate in an additive increase and multiplicative decrease (AIMD) way which is similar to TCP (Jacobs and Eleftheriadis, 1998; Rejaie et al., 1999). These approaches will result in rather fluctuant transmission rates, which may produce an annoying presentation quality at the receiver. The receiver is required to acknowledge every received packet for the sender to collect information of the network status. The frequent feedback packets will consume the bandwidth and degrade the network congestion. Some approaches use model-based flow control, where the models were developed to calculate the bandwidth as a function of the packet loss ratio and round trip time (Padhye et al., 1998). However, the estimated packet loss ratio may not be suitable for the next time interval, and thus affect the accuracy of the throughput calculation.

Common requirements for multimedia applications have led to the design of the general purposed protocol called the real-time transport protocol (RTP) (Schulzrinne et al., 1996). To satisfy those common requirements, RTP provides the functions such as (1) the ability to communicate the selected coding scheme, (2) the mechanisms to facilitate the application-specific handling of time-stamped data, which enables the receiver to play the data at the appropriate time, (3) the synchronization of multiple media, (4) the indication of packet loss, and (5) the notification to the sender when the loss occurs. Although RTP provides the functionality suited for carrying real-time content and is the primer protocol for real-time applications, it cannot provide any form of reliability or protocol-defined flow/congestion control. However, the existing RTP has a flexible mechanism that leaves many protocol details. This mechanism allows the designers to build up the functionality required by the particular application.

In this paper, we design and implement a novel real-time adaptive multimedia transmission protocol. The central component of the proposed scheme is an adaptive rate-control mechanism. It dynamically changes the server transmission rates, considering the relationships among the server transmission rates, client buffer occupancies, playback rates, network delays and packet loss ratios (Chen et al., 2003; Shyu et al., 2002a,b,c). Since the network bandwidth is scarce in today's Internet, it is meaningful to use minimal bandwidth for each media streaming. Our approach is source rate adaptive, which aims at obtaining a minimal transmission rate according to the client buffer occupancy, network delay, packet loss rates and playback requirement. It can achieve an efficient utilization of network resources such as bandwidth and client buffer at the same time. To ensure the fairly bandwidth sharing with the dominant TCP flows, a congestion control mechanism is incorporated in our optimal bandwidth allocation scheme. We also use the estimated network bandwidth as the upper bound limit for the transmission rates.

The paper is organized as follows. In the next section, the adaptive multimedia transmission protocol is presented and the optimal bandwidth allocation scheme is introduced. Experimental results for both LAN and Internet are given in Section 3. Conclusions are presented in Section 4.

2. The real-time adaptive multimedia transmission protocol

As mentioned earlier, though RTP cannot provide any form of reliability or protocol-defined flow/congestion control, it gives the designers the opportunity to add reliability, flow/congestion control, and functionality for efficient use of the bandwidth. In this paper, a real-time adaptive multimedia transmission protocol is developed with the implementation of the rate adaptive algorithm. This protocol can provide the transmission of the requested video files from the sender to multiple receivers using transport-layer protocol UDP. At the receiver side, the control information such as the packet loss ratio, playback requirement and buffer occupancy are collected and fed back to the sender periodically. The sender uses the feedback information to calculate