Chapter XVI
End-to-End Support for Multimedia QoS in the Internet

Panagiotis Papadimitriou
Democritus University of Thrace, Greece

Vassilis Tsaoussidis
Democritus University of Thrace, Greece

ABSTRACT

An increasing demand for multimedia data delivery coupled with reliance in best-effort networks, such as the Internet, has spurred interest on effective quality of service (QoS) management for multimedia streams. Since today’s multimedia applications are expected to run in physically heterogeneous environments composed of both wired and wireless components, we assess the efficiency of transport-layer solutions for multimedia traffic in heterogeneous networks. In order to quantify the performance on media delivery, we investigate the multimedia application requirements vs. the QoS provided by the underlying network. The chapter also provides means for the perceptual QoS assessment of voice and video streams. In the sequel, we describe some representative end-to-end congestion control schemes, identifying the mechanisms that are most suitable for multimedia traffic. Our analysis is complemented with conclusive performance studies which quantify video delivery, within the context of transport protocol support and efficiency.

INTRODUCTION

In recent years, Internet has been experiencing an increasing demand for multimedia services, typically involving audio and video delivery. Media-streaming applications yield satisfactory performance only under certain quality of service (QoS) provisions, which may vary depending on the application task and the type of media involved. Unlike bulk-data transfers, multimedia flows require a minimum and continuous bandwidth guarantee, while they are also affected by reliability factors, such as packet drops due to congestion or link errors.
Today’s multimedia applications are expected to run in physically heterogeneous environments composed of both wired and wireless components. Wireless links exhibit distinct characteristics, such as limited bandwidth, bit errors, and potential handoff operations. Bit errors typically occur when the signal to interference and noise ratio is not high enough to decode information correctly. Furthermore, wireless channels are hard to model and predict, and designing an error-free communication link generally entails sacrificing significant capacity. QoS requirements in wireless networking essentially remain stringent and complicated, taking additionally into account the influencing mobile device characteristics and limitations. For example, a considerable number of mobile devices offer limited buffer capacities, being unable to smooth the fluctuations in the receiving rate. In this case, the task of smooth media delivery is primarily delegated to the transport protocol.

Several independent mechanisms have been proposed, which normally interact with the transport protocol and provide reliable transmission over wireless links (Balakrishnan, 2002; Hu & Sharma, 2002). Most of them operate on the link layer and generally are considered more efficient than physical-layer techniques, such as spread-spectrum and OFDM modulation, or channel coding. However, link-layer approaches may degrade performance, especially in the presence of highly variable error rates. Local error recovery may alter the characteristics of the network affecting the functionality of higher layer protocols. For example, local retransmission, such as automatic repeat request (ARQ) (Hu et al., 2002), could result in packet reordering or in considerable round trip time (RTT) fluctuations. In addition, concurrent responses from both local and end-to-end error control may result in undesirable interactions, causing inefficiencies and potentially instability. Considering real-time traffic where data packets bear information with a limited useful lifetime, retransmissions are often a wasted effort (Papadimitriou & Tsaoussidis, 2006). In such conditions, unfruitful retransmissions deliver delayed packets which are either discarded, or at the worst they obstruct the proper reconstruction of oncoming packets.

Transmission control protocol (TCP) is basically designed to provide a reliable service for wired Internet. The additive increase multiplicative decrease (AIMD) algorithm (Chiu & Jain, 1989), incorporated in standard TCP versions, achieves stability and converges to fairness when the demand of competing flows exceeds the channel bandwidth. TCP is further enhanced with a series of mechanisms for congestion management, including congestion avoidance (Jacobson, 1988), slow start, fast retransmit, and fast recovery (Stevens, 1997). Despite these features, TCP demonstrates inadequate performance in heterogeneous wired/wireless environments. Tsaoussidis and Matta (2002a) outline three major shortfalls of TCP: (i) ineffective bandwidth utilization, (ii) unnecessary congestion-oriented responses to wireless link errors (e.g., fading channels) and operations (e.g., handoffs), and (iii) wasteful window adjustments over asymmetric, low-bandwidth reverse paths. The difficulty of the task that TCP has to perform is further enhanced, when the protocol provides services for delay-sensitive applications. Standard TCP induces oscillations in the achievable transmission rate, with an adverse effect on the playback quality of multimedia applications. Furthermore, the protocol introduces arbitrary delays, since it enforces reliability and in-order delivery.

In this context, several TCP protocol extensions (Mascolo, Casetti, Gerla, Sanadidi, & Wang, 2001; Tsaoussidis & Badr, 2000; Zhang & Tsaoussidis, 2001) have emerged to overcome the standard TCP limitations providing more efficient bandwidth utilization and sophisticated mechanisms for congestion control. TCP-friendly protocols, proposed in Floyd, Handley, Padhye, and Widmer (2000), Yang, Kim, and Lam (2001), and Yang and Lam, 2000, achieve smooth window