Streaming Media via TCP: Approaches to Bandwidth Sharing and Rate Adjustment

Jack Brassil and Puneet Sharma
Hewlett-Packard Laboratories
Palo Alto, CA 94304
{jtb,puneet}@hpl.hp.com

Supporting multimedia applications is one requirement driving the development of a differentiated services Internet. One challenge multimedia application developers face is that streaming audio and video places multiple competing requirements on the underlying transport mechanism. For example, sharing network resources with data applications encourages media applications to use either TCP-friendly rate control mechanisms [1] or TCP itself. Yet a multimedia application's transport requirements are generally poorly matched to the transport service provided by a single TCP connection.

Two areas where this mismatch is particularly evident are:

- **Bandwidth Sharing**

  Researchers have shown that additive-increase multiplicative-decrease (AIMD) congestion control algorithms approximately realize proportional fair rate sharing of bottleneck links [2], [3]. A consequence of this result is that competing continuous media TCP connections sharing a bottleneck link can receive bandwidth shares that are not well matched to their individual needs, resulting in what receivers may perceive as vastly different service quality. As a simple example, if two independent streams sharing a bottleneck have the same RTT, then both will receive an equal proportion of the available bandwidth. However if one of the two media streams was encoded for transmission at a higher bit rate, a receiver might perceive that this stream suffered disproportionately worse service quality.

- **Coarse-Grain Rate Adjustment**

  Streaming video is frequently encoded at multiple bit-rates to support rate adaptation for transmission over either congested or low bandwidth links. For example, Real Network's Surestream technology permits multiple bit-rate encoding and midstream rate adaptation. Though such a stream is rate-adaptable (at either source or transcoding proxy), the number of permissible rates is typically few, and the granularity of rate adjustments is often large. TCP's aggressive probing for available bandwidth is poorly suited to a multirate source's ability to make only large, discrete rate adjustments. Further, a media application has no effective means of communicating its bandwidth needs to TCP, while the transport layer provides little help to the application seeking a large rate increase. In fact, an attempt to increase rate in the face of insufficient available bandwidth can result in lower perceived service quality than no attempt at all.

In this paper we propose techniques to address these mismatches for the transmission of stored (on-demand) media using TCP. We begin by examining the advantages of transmitting a single stream over multiple (actual or virtual) TCP connections. Multiple connections allow bandwidth sharing in proportion to the target media encoding rate – one possible measure of fairness – and also mitigate large congestion induced window size reductions, which receivers perceive as disruptive. Use of multiple TCP connections has been considered in many contexts. [1]

[1] Both multiple TCP connections between common endpoints as well as noncompliant TCP implementations have been used extensively as an aggressive 'throughput acceleration' technique for data applications, particularly on networks with large bandwidth-delay products.

To achieve weighted fair sharing in a differentiated services environment, Crowcroft [4] has examined a modified TCP congestion control algorithm designed to behave as a collection of N virtual TCP connections. Achieving weighted fair sharing of a bottleneck link is particularly appealing in the context of multimedia flows encoded at unequal rates.

We next show how multiple TCP connections can be most effectively used to support rate adaptation for multi-rate video sources. In particular, we argue that temporarily accelerating the delivery of data encoded at a lower rate is an effective means of determining whether adequate bandwidth is available for a large rate increase. We show that one benefit of this approach is that when insufficient bandwidth is available for a rate increase, allowing the source to continue transmitting at the lower rate can reduce perceived service disruption.

References

[1] H. Adiseshu, G. Parulkar, and G. Varghese, “The TCP-Friendly Web Site,” http://www.psc.edu/networking/tcp_friendly.html.

[2] Kelly, F., A. Maulloo, D. Tan, “Rate Control in Communication Networks: Shadow Prices, Proportional Fairness and Stability,” *Journal of the Operational Research Society*, 49 (1998), 237-252.

[3] Hurley, P., J. Le Boudec, P. Thiran, “A Note on the Fairness of Additive Increase and Multiplicative Decrease,” *Proceedings of ITC-16*, Edinburgh, June 1999.

[4] Crowcroft, J., P. Oechslin, “Differentiated End-to-end Internet Services using a Weighted Proportional Fair Sharing TCP,” *ACM Computer Communications Review*, 28 (1998), 53-67.

[5] Tan, W., A. Zakhor, “Real-Time Internet Video Using Error Resilient Scalable Compression and TCP-Friendly Rate Control,” *IEEE Transactions on Multimedia*, vol. 1 no. 2, (1999), 172-186.