Streaming over Throughput-Limited Paths

Streaming real-time media might increase the congestion of bottleneck links. This may disrupt the transmissions of other users and even delay the delivery of the media stream itself. In the first part of this chapter, we analyze the impact of self-congestion on the performance of a low-latency media streaming system operating over a throughput-limited network path. We present a rate-distortion model which captures both the effect of compression and of late losses due to congestion on the decoded video quality of the system. This model is helpful to determine how close to the physical channel capacity a low-latency video streaming system can operate.

In the second part of the chapter, we present the concept of congestion-distortion optimized (CoDiO) streaming. This is one of the main contributions of this book. Our work builds on the RaDiO scheduling framework, which considers unequal contributions of different portions of a multimedia data stream to the overall distortion. Rather than searching for an optimal schedule for the packets of a stream, which minimizes the expected Lagrangian cost of rate and distortion, we suggest changing metric and replacing rate by congestion, which we define, throughout this book, as expected end-to-end delay. We explain why congestion is better suited to throughput-limited streaming, and describe how, given a simple channel model, CoDiO scheduling is performed. We also present a low-complexity version of our algorithm, CoDiO light, simple enough to run in real-time. Experimental results analyze the performance of CoDiO and CoDiO light compared to a content-oblivious ARQ scheduler. We also analyze how our congestion-distortion schedulers compare to the state-of-the-art RaDiO scheduler.

3.1 Video Encoding for Throughput-Limited Paths

In this section, we present a model to estimate the highest sustainable rate, given a fixed encoding structure, for a video streaming system operating at low latency over a throughput-limited network path.
3.1.1 End-to-End Rate-Distortion Performance Model

In low-latency streaming applications, compressed video is transmitted over a network at a given rate. Typically, it is desirable to achieve end-to-end delays of no more than a few hundred milliseconds. When a packet does not arrive at the receiver by its playout deadline, to avoid interruptions, the decoder conceals the missing information and the playout continues at the cost of higher distortion. Decoded video quality at the receiver is therefore affected by two factors: quantization errors introduced at the encoder while compressing the media stream, and packet loss either caused by transmission errors or due to late arrivals. These two contributions have different characteristics. Typically, the distortion introduced by quantization is evenly distributed across the encoded frames and is determined by the encoding bit-rate. This contrasts with the impact of packet loss which usually introduces decoding errors (i.e., higher distortion) in the frame(s) containing the missing packet(s). Because of the predictive nature of the compressed video stream, this error will propagate to subsequent frames. Usually, these errors tend to decay over time due to intra-macroblock coding and in-loop filtering. Error propagation is eventually stopped when an intra-coded frame is received. Using Mean Square Error (MSE) as the criterion, a video distortion model can be derived based on [31].

The decoded video distortion, denoted by $D_{dec}$, comprises two terms:

$$D_{dec} = D_{enc} + D_{loss},$$  \hspace{1cm} (3.1)

where the distortion for the encoder performance $D_{enc}$ and the contribution from packet loss $D_{loss}$ are described in greater detail in the following.

**Encoder Distortion Model**

The distortion introduced by encoder quantization is reduced when the sequence is encoded at a higher rate. As the coding rate increases, however, the same amount of distortion reduction requires a greater rate increment. In [31], this distortion-rate tradeoff is modeled for H.263 P frame encoding, by a simple formula:

$$D_{enc} = D_0 + \Theta / (R - R_0),$$  \hspace{1cm} (3.2)

where $R$ is the rate of the video stream, and $D_0$, $\Theta$ and $R_0$ are model parameters. Using nonlinear regression techniques, these parameters can be estimated from empirical rate-distortion curves obtained by encoding a sequence at different rates. Interestingly, we have observed that this model can also be extended to sequences encoded by H.264, with more complicated GOPs, incorporating I, B, P pictures as well as SP and SI pictures [135, 198, 199].

---

1 Part of the derivation presented in this section is reproduced with permission from [198] © 2004 IEEE and from [135] © 2005 Elsevier.