CMPSCI 754 Multimedia Systems **Spring 2001**

Lecture 19: 17th April, 2001

Lecturer: Prashant Shenoy *Scribe: Anoop George Ninan*

19.1 Network Support for Next-Generation Applications

In this lecture, we shall look into issues involved in transporting video over high speed networks.

19.1.1 Data Communications - A Historical Perspective

Today, we have different networks that support different types of workloads. For instance we have:

- Voice over telephone networks
- Video over cable TV networks
- Textual data over computer networks

This separation exists because of the fundamental differences in the characteristics of these different workload types. For example, voice and video require bounded transmission delay but may be tolerant to brief loss of information; textual data on the other hand is tolerant to delay, but cannot tolerate transmission error or loss.

Two fundamental mechanisms used to implement the above are *Circuit* and *Packet Switching*. Here we do a brief review.

- **Circuit Switching:** A dedicated physical path is set up from source to destination for the entire call duration. The bandwidth allocated and end-to-end delay are fixed. There is no overhead (i.e., no processing) at intermediate nodes and information delivery is guaranteed to be sequenced.
  Problems with this mechanism are that bursty data may yield poor utilization of resources; short sessions are not cost effective as call set-up requires about 1 second; circuit switching inherently supports only unicast.
- **Packet Switching:** Bandwidth allocation is dynamic. Switching is achieved in a store-and-forward manner and is suitable for traffic with a bursty nature. This mechanism supports both unicast and multicast transmission and permits network resources to be multiplexed among several channels.
  Problems with this mechanism are that there no guarantees are made: end-to-end delay is not fixed and information delivery may not be reliable and sequenced.

19.1.2 Integrated Networks

Given the above overview, we try to reason why we should not have a single *integrated network* that allows for transmission of data with different data characteristics. The levels of integration achieved may span

- *Integrated access and transmission:* in which data with different characteristics is sent over the same wire, but on different frequency bands. Network elements are still distinct and themselves not integrated.

- *Integrated switching:* where the network elements (switches) are themselves integrated, and can yet handle data with different characteristics

As a first step in this direction, we shall investigate protocols to transport video.

19.1.3 Network Protocols for Video

Several applications today require the playback of stored video at clients, over a high speed network. For continuous playback, strict QOS guarantees must be provided in an end-to-end fashion. Variable Bit Rate (VBR) video only complicates this as it exhibits significant, multiple time scale bit rate variability. Means to deal with such variation include reserving resources (bandwidth). Bandwidth may be reserved based on

- *peak rate:* This results in under-utilization of available bandwidth.
- *average rate:* This results in more efficient utilization of available bandwidth, but causes losses when bit rates exceed the average rate.

Reserving resources in a way such that the average rate is within available bandwidth, but the peaks are not, is called ***overbooking***.

Overbooking results in occasional network congestion and hence *packet loss*. Compressed video is however not resilient to packet loss. Ways to deal with this problem include:

- **Packet Loss Recovery**
  - *Re-transmission based schemes (ARQ)*: This is a closed-loop technique (one which uses feedback from the network)
  - *Forward Error Correction (FEC) schemes*: This is an open-loop technique (one which does not use feedback from the network). Here redundant information is sent apriori and this may be used to reconstruct lost information
  - *Multi-level or Multi-resolution encoding with prioritized transmission*: (e.g. send base frames with higher priority than enhancement frames). While this does not deal directly with packet loss, it ensures continuous playback at the client, though at a coarse granularity.
- Avoid packet loss
  - *Traffic Shaping at the source:* Smooth the flow by regulating the rate at the source.
  - *Rate control using network feedback:* Here sources send anticipated requirement into the network and the network intimates sources whether it can handle multiple flows or not.

19.1.4 Smoothing

The goal of smoothing is to regulate rate fluctuations in a video sequence. Smoothing techniques may be classified into 2 categories:

- *Online or dynamic smoothing:* for transmission of live video, for example.
- *Offline or static smoothing:* for transmission of stored video, for example.

The above two techniques differ in the length of video sequences examined before a transmission rate is decided on. i.e., for live video shorter sequences are examined before a rate is decided on, whereas for stored video, longer sequences may be examined.

19.1.5 An Online, Lossless Smoothing Technique

Smoothing may be achieved by employing rate regulation at the sources; but these techniques are *lossy*. The algorithm discussed here smoothes picture-to-picture rate fluctuations in a video sequence. It achieves smoothing in a *lossless* manner as it uses buffering and does not discard information. Buffering is employed to support a lookahead strategy that inspects and detects repeating patterns of I,P and B frames and transmits all such frames at the same rate. The problem with this scheme is that the buffering introduces a delay D.

Hence the key problem here is: *Given a delay bound D, derive a transmission schedule.*

We now see a number of constraints that have to be met in order to achieve the above goal.

Constraints:

- **Delay:** $D \geq (K+1)\tau$
  $t_i = max(d_{i-1}, (i+K-1)\tau)$
  $d_i - (i-1)\tau \leq D$
- **Continuous Service:** $t_i = d_{i-1}$
- **Determination of r_i:**
  - **Deadline Constraint:** $t_i + \frac{\sum_{m=0}^{h} S_{i+m}}{r_i} - (i+h-1)\tau \leq D \Rightarrow r_i \geq \frac{\sum_{m=0}^{h} S_{i+m}}{D+(i+h-1)\tau - t_i}$
  - **Continuous Service:** $d_{i+h} = t_i + \frac{\sum_{m=0}^{h} S_{i+m}}{r_i} \geq (i+h)\tau \Rightarrow r_i \leq \frac{\sum_{m=0}^{h} S_{i+m}}{(i+h)\tau - t_i}$
- **GOAL:** Find the value of r_i that is between the bounds for as large a value of h as possible.

19.1.6 Protocol for Stored Video Transport

Observations: Characteristics of stired video include delay insensitivity and apriori knowledge of bit rate.

Idea: Smooth out fast time-scale and manage slow time-scale. Also combine open-loop and feedback-based control techniques.

Key difference between stored and live video: The delay constraint is less stringent and hence larger smoothing intervals are possible.

Approach: While a simple method would be to convert a VBR stream into a CBR stream by buffering (this method has the disdvantage of large buffering and initiation latency), a better approach would be to use *piecewise CBR transmission*.

Goal: Derive a piecewise CBR schedule that minimizes the rate changes.

19.1.6.1 Reducing Rate Variability

The intuition behind the idea here is as follows:

First, since a CBR transmission is as smooth as possible, in constructing the schedule, consider CBR segments which are as long as possible. Second, when the transmission rate should be increased or decreased to ensure feasibility, change the rate as early as possible. This ensures that the rate change is as small as possible, which is advantageous from the perspective of making the overall schedule as smooth as possible.