VBR Video Networks with Deterministic Quality-of-Service Constraints

Jörg Liebeherr

Department of Computer Science

University of Virginia
Outline

- VBR Video
- Deterministic QoS Networks
- Video Traffic Characterization
  - Best Possible and Approximations
- Scheduling and Admission Control
  - Best Possible and Tradeoffs
- Empirical Evaluation with MPEG traces
Video Networks

Video requires Quality-of-Service (QoS) guarantees:

- Delay
- Delay Variation (Jitter)
- Throughput
- Error Rate

QoS guarantees difficult to satisfy if video is variable-bit rate.
MPEG Video Compression

MPEG compression uses three variable-size frame types:

- Intra-coded (I) frames
- Predictive-coded (P) frames
- Bidirectionally predictive-coded (B) frames

⇒ MPEG encoders generate variable-bit rate (VBR) video.
Quality-of-Service Network

- Traffic Characterization
- Admission control and policing mechanisms
Types of Quality-of-Service

- **Deterministic Service** gives worst-case guarantees.
  - The delay of the $k^{th}$ packet on connection $j$
    \[ D_j^k \leq d_j \quad \forall k \]

- **Statistical Service** gives probabilistic guarantees.
  \[ \text{Prob} \left( D_j^k \leq d_j \right) \geq z_j \quad \forall k \]
Statistical Service

Advantages:

- Can achieve high utilization

Disadvantages:

- Stochastic models are complex.
- Traffic policing difficult or impossible.
- Complex admission control.
Deterministic Service

Advantages:

- Policing is well-defined.
- Can consider a "no loss" model.
- Efficient admission control.

Not known:

? Utilization.
In this Talk ... 

Find the best possible utilization with a deterministic service.
Traffic Characterization

For QoS networks with deterministic service we need a **worst-case** traffic characterization.

- $A[t, t + \tau]$ is **actual traffic** in interval $[t, t + \tau]$.

- **Worst-case** characterization of traffic is a function $A^*$ with:
  
  $$A[t, t + \tau] \leq A^*(\tau) \quad \forall t, \forall \tau$$

  $A^*$ is a time-invariant bound of the actual traffic.

- $A^*$ must be sub-additive,
  
  $$A^*(t_1 + t_2) \leq A^*(t_1) + A^*(t_2).$$
What is the best $A^*$?

- Define the “Empirical Envelope” $\mathcal{E}$:

$$\mathcal{E}(\tau) := \max_t A[t, t + \tau]$$

- $\mathcal{E}$ is a time-invariant bound.

- $\mathcal{E}$ is best possible time-invariant bound.

$$\mathcal{E}(\tau) \leq A^*(\tau) \quad \forall A^*$$
Constructing the Empirical Envelope

Trace of VBR Video:

Frame Number

Number of Cells

‘Integral’ of the Trace:

Frame Number

Cumulative Number of Cells

Empirical Envelope $E$
Searching for a "more practical" $A^*$?

1. Use few parameters.

2. Accurately describe actual traffic.

3. Be sub-additive.

4. Enforceable by simple policing mechanisms.

Here: 

‘Leaky bucket’ and ‘Multi-level leaky bucket’ traffic models.
Leaky Bucket

- Describes traffic by a rate $\rho$ and a burst size $\sigma$.

\[ A^*(\tau) = \sigma + \rho \cdot \tau \]
Multi-Level Leaky Bucket

- Describes traffic by multiple rate-burst pairs $(\sigma_i, \rho_i)$.

$$A_j^*(\tau) = \min_i \{\sigma_i + \rho_i \cdot \tau\}$$
Approximations of the Empirical Envelope

- Construct an upper bound for the Empirical Envelope with leaky buckets.

- Result is an approximation of the empirical envelope.
Empirical Evaluation

Determine the maximum utilization on a link …

- … with empirical envelope.

- … with leaky bucket traffic policing.

- Single link with 45 Mbps

- Workload on link is obtained from MPEG traces.
Workload for Empirical Evaluations

- *The Princess Bride*
  - 90-minute MPEG trace.
  - 320x240 pixels per frame.
  - Frame pattern is *IBBPBBPBBPBBPBB*.

- *Advertisements:*
  - 10-minute TV commercials for graphics products
  - 160x120 pixels per frame.
  - Frame pattern is *IBBPBB*.
Scheduling Policy

- Operates at the output port of the switch.
- Decides which packet to transmit next.
Scheduling and Network Utilization

• First-Come-First-Served (FCFS)
  – Simplest, offers only one delay bound.

• Earliest-Deadline-First (EDF)
  – Sophisticated, optimal in terms of schedulability.

• Static Priority (SP)
  – Compromise, offers fixed number of delay bounds.
First-Come-First-Served (FCFS)

Admission Control Test:

- Exact Test:
  \[ d \geq \sum_{j \in \mathcal{N}} A^*_j(t) - t \quad t \geq 0 \]
Earliest-Deadline-First (EDF)

Admission Control Test:

- Exact Test (Liebeherr/Wrege/Ferrari):
  \[ t \geq \sum_{j \in \mathcal{N}} A_j^*(t - d_j) + \max_{k, d_k > t} s_k \quad t \geq 0 \]
  where \( \max_{k, d_k > t} s_k \equiv 0 \) for \( t > \max_{k \in \mathcal{N}} d_k \)
Static Priority (SP)

Admission Control Test:

- Exact Test (Liebeherr/Wrege/Ferrari):

\[
(\exists \tau \leq d_p)
\]

\[
t + \tau \geq \sum_{j \in C_p} A_j^*(t) + \sum_{q=1}^{p-1} \sum_{j \in C_q} A_j^*(t + \tau) + \max_{r>p} s_r
\]

for all \( p, \ t \geq 0 \)
# More Admission Control Tests

<table>
<thead>
<tr>
<th></th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FCFS Exact</strong></td>
<td>( d \geq \sum_{j \in \mathcal{N}} A_j^*(t) - t \forall t \geq 0. )</td>
</tr>
</tbody>
</table>
| **SP Exact**         | \[
(\exists \tau \leq d_p) \ t + \tau \geq \sum_{j \in \mathcal{C}_p} A_j^*(t) + \sum_{q=1}^{p-1} \sum_{j' \in \mathcal{C}_q} A_{j'}^*(t + \tau) + \max_{r > p} s_r \]
|                      | \( \forall p, t \geq 0. \)                                              |
| **EDF Exact**        | \( t \geq \sum_{j \in \mathcal{N}} A_j^*(t - d_j) + \max_{k, d_k > t} s_k \) \( t \geq 0. \) |
| **SP Sufficient 1**  | \( t \geq \sum_{j \in \mathcal{C}_p} A_j^*(t - d_p) + \sum_{q=1}^{p-1} \sum_{j' \in \mathcal{C}_q} A_{j'}^*(t) + \max_{r > p} s_r \) \( \forall p, t \geq d_p. \) |
| **SP Sufficient 2**  | \( d_p \geq \sum_{q=1}^{p} \sum_{j \in \mathcal{C}_q} A_j^*(d_p) + \max_{r > p} s_r \) \( \forall p. \) |
Empirical Evaluations

Questions:

- What is the impact of the scheduling method?
- What is the impact of the accuracy of admission control?
- Same MPEG traces as before.
Conclusions

- Deterministic networks can achieve high utilization for VBR traffic.

- Deterministic VBR can do much better than peak rate allocation.

- Utilization affected by three factors:
  - Traffic Characterization.
  - Scheduling Method.
  - Accuracy of Admission Control.
Maximum Achievable Utilization

# of Connections

Delay Bound (ms)

Lecture

# of Connections

Delay Bound (ms)

Movie

Average Rate

Envelope

Peak Rate

Average Utilization

Average Rate

Envelope

Peak Rate

Average Rate

Envelope

Peak Rate
Achievable Utilization Using Multi-Level Leaky Buckets

![Graphs showing achievable utilization for different delay bounds and numbers of connections for lecture and movie types. Each graph illustrates the relationship between delay bound and the number of connections for 1, 2, 3, 4, and 5 LBs/Envelope.](image)
Achievable Utilization Using Different Schedulers

![Graph showing achievable utilization for different schedulers with movie and lecture connections. The x-axis represents the number of lecture connections, the y-axis represents the number of movie connections, and the lines correspond to different deadlines and peak rates.]

**Graph:**
- **Legend:**
  - **Movie Connections**
  - **Lecture Connections**
  - **Deadline Values:**
    - 200 ms
    - 80 ms
    - 20 ms
    - 10 ms
  - **Peak Rate Values:**
    - 1000 ms
    - 200 ms
    - 80 ms
    - 30 ms

**Legend:**
- **EDF**
### Peak Rate Table

<table>
<thead>
<tr>
<th>Peak Rate</th>
<th>80ms</th>
<th>30ms</th>
<th>10ms</th>
<th>20ms</th>
<th>200ms</th>
<th>1000ms</th>
</tr>
</thead>
</table>

### Deadline Table

<table>
<thead>
<tr>
<th>Deadline</th>
<th>Lecture</th>
<th>Movie</th>
</tr>
</thead>
<tbody>
<tr>
<td>200 ms</td>
<td>1000ms</td>
<td></td>
</tr>
<tr>
<td>80 ms</td>
<td>200ms</td>
<td></td>
</tr>
<tr>
<td>20 ms</td>
<td>80ms</td>
<td></td>
</tr>
<tr>
<td>10 ms</td>
<td>30ms</td>
<td></td>
</tr>
</tbody>
</table>

### FCFS

The chart illustrates the relationship between the number of lecture connections and the number of movie connections, with different peak rates and deadlines. FCFS (First Come, First Served) is indicated at the bottom of the chart.
Achievable Utilization Using Different Schedulers

![Graph showing achievable utilization with different schedulers and deadlines for movie and lecture connections. The x-axis represents the number of lecture connections, and the y-axis represents the number of movie connections. The graph includes lines for different deadlines such as 200 ms for lectures and 1000 ms for movies. The EDF (Earliest Deadline First) scheduler is indicated.]
Static Priorities
Putting it All Together

- Delay bounds for *Lecture* and *Movie*: 30 ms and 50ms
- Benchmark case: Empirical envelope, EDF, exact condition
- “Trade-off” case: three leaky buckets, SP, sufficient condition