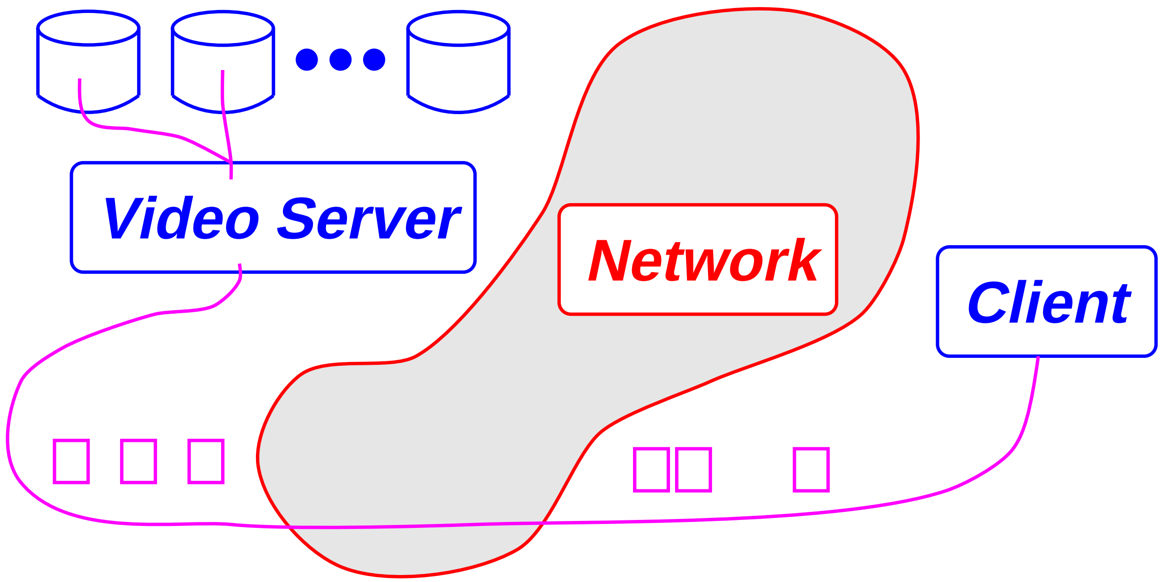
A New Class of Packet Schedulers for Quality-of-Service Networks

Jörg Liebeherr

Department of Computer Science
University of Virginia
jorg@cs.virginia.edu

Motivation

- Transmission of video and audio over packet-switched networks.



- Requires *new networks and protocols*.

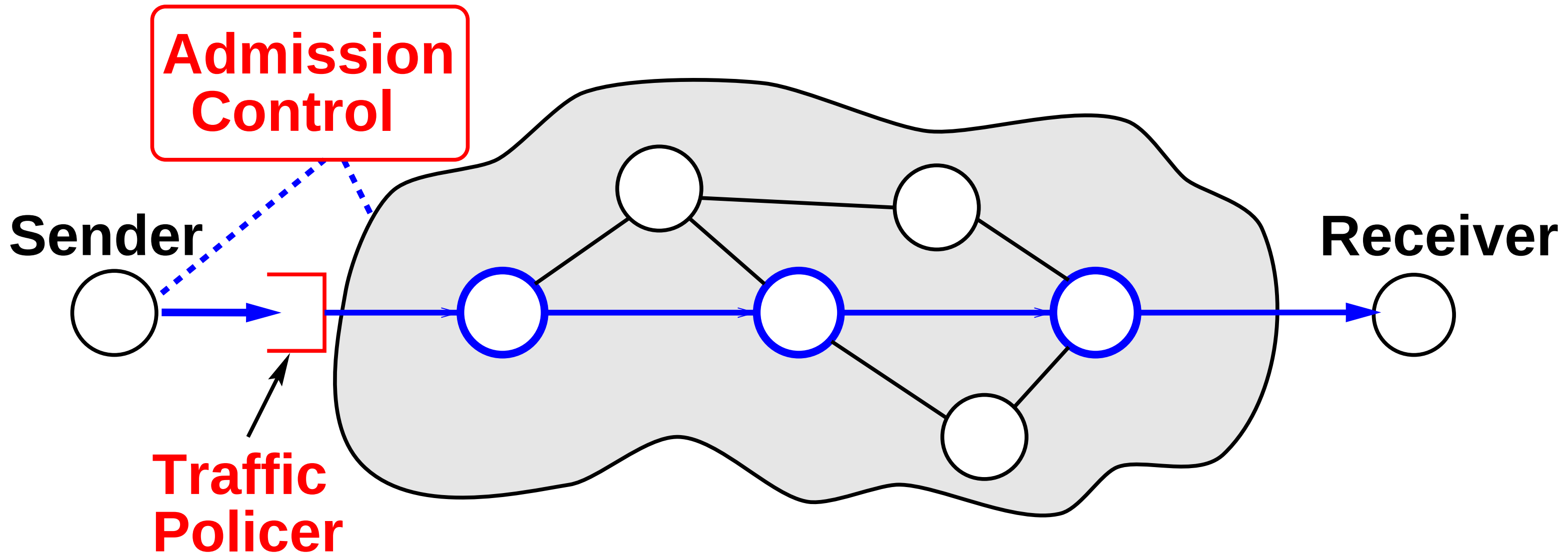
Overview

- Background
- QOS Networks
- *Rotating Priority Queues (RPQ) Scheduling*
- *Rotating Priority Queues Plus (RPQ$^+$) Scheduling*
- Conclusions

Quality-of-Service

- Video and audio need *Quality-of-Service (QoS)* guarantees:
  - *delay*
  - jitter
  - throughput
  - loss rate
- A *deterministic service* gives worst-case guarantees.

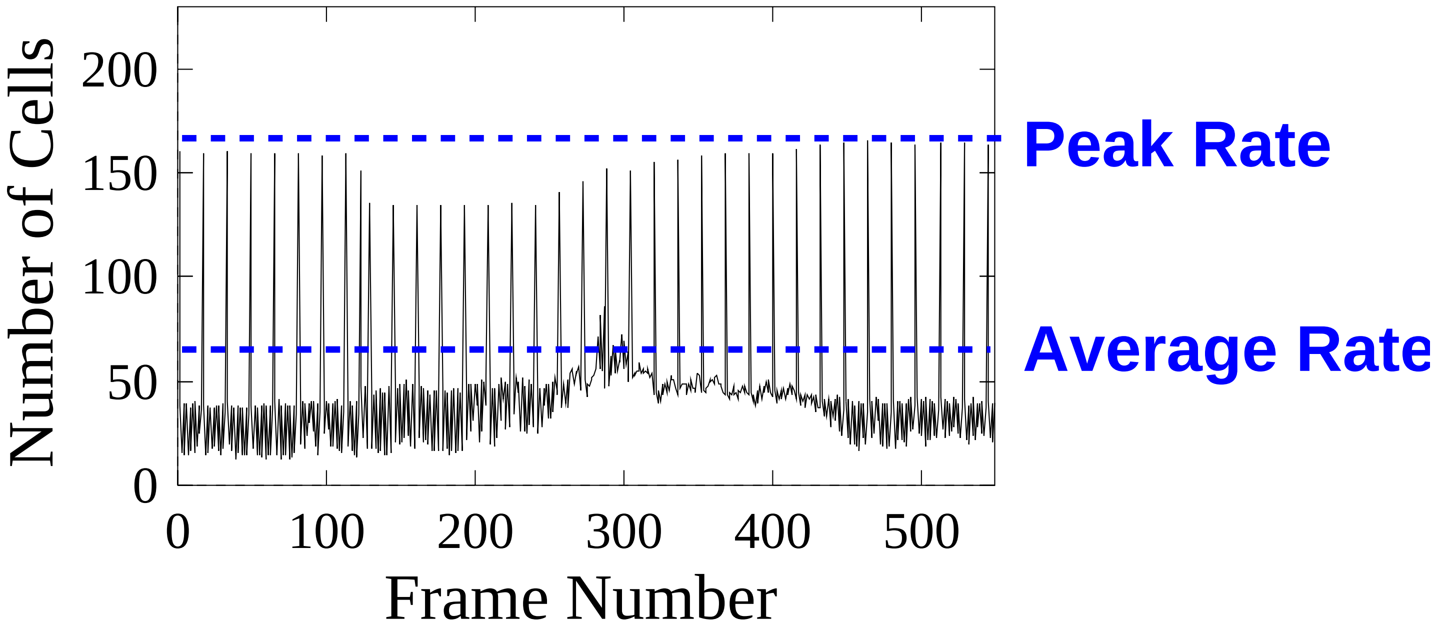
Multimedia Networks



- Multimedia connections have *QoS* and *traffic* parameters.
- Multimedia networks need *resource reservation.*

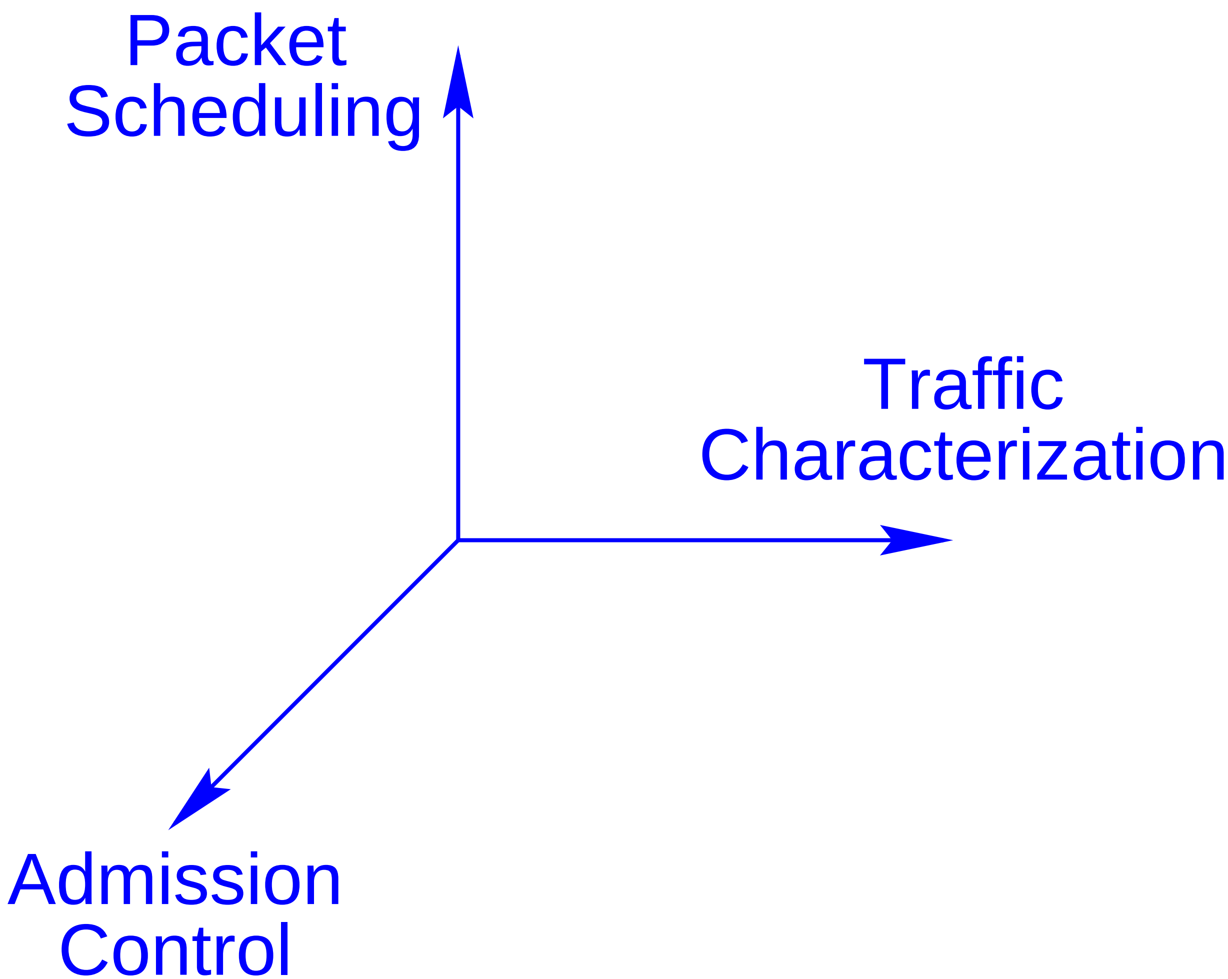
Why is Resource Reservation Difficult?

- Compressed digital video has a *variable bit rate*.

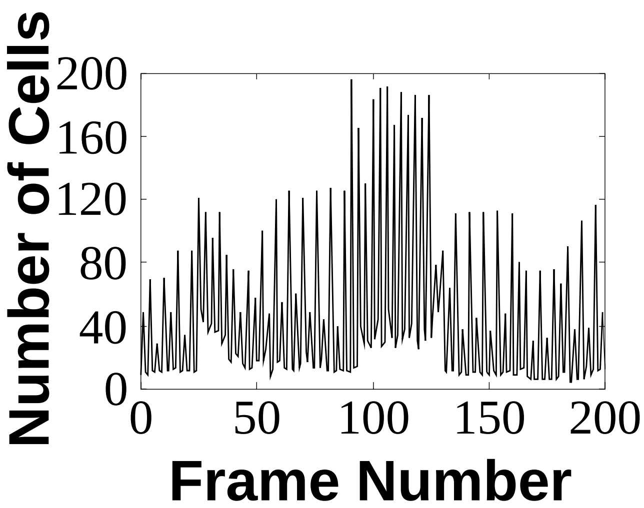


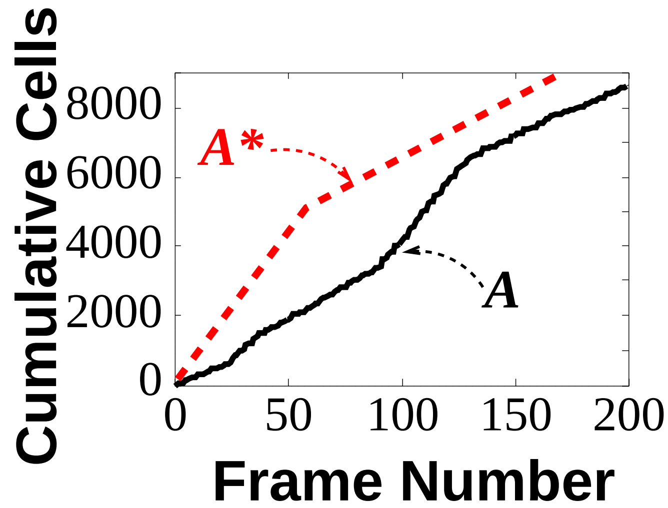
- *Problem:* How do we provide deterministic QoS without peak-rate reservation?

Design Space of a Multimedia Network



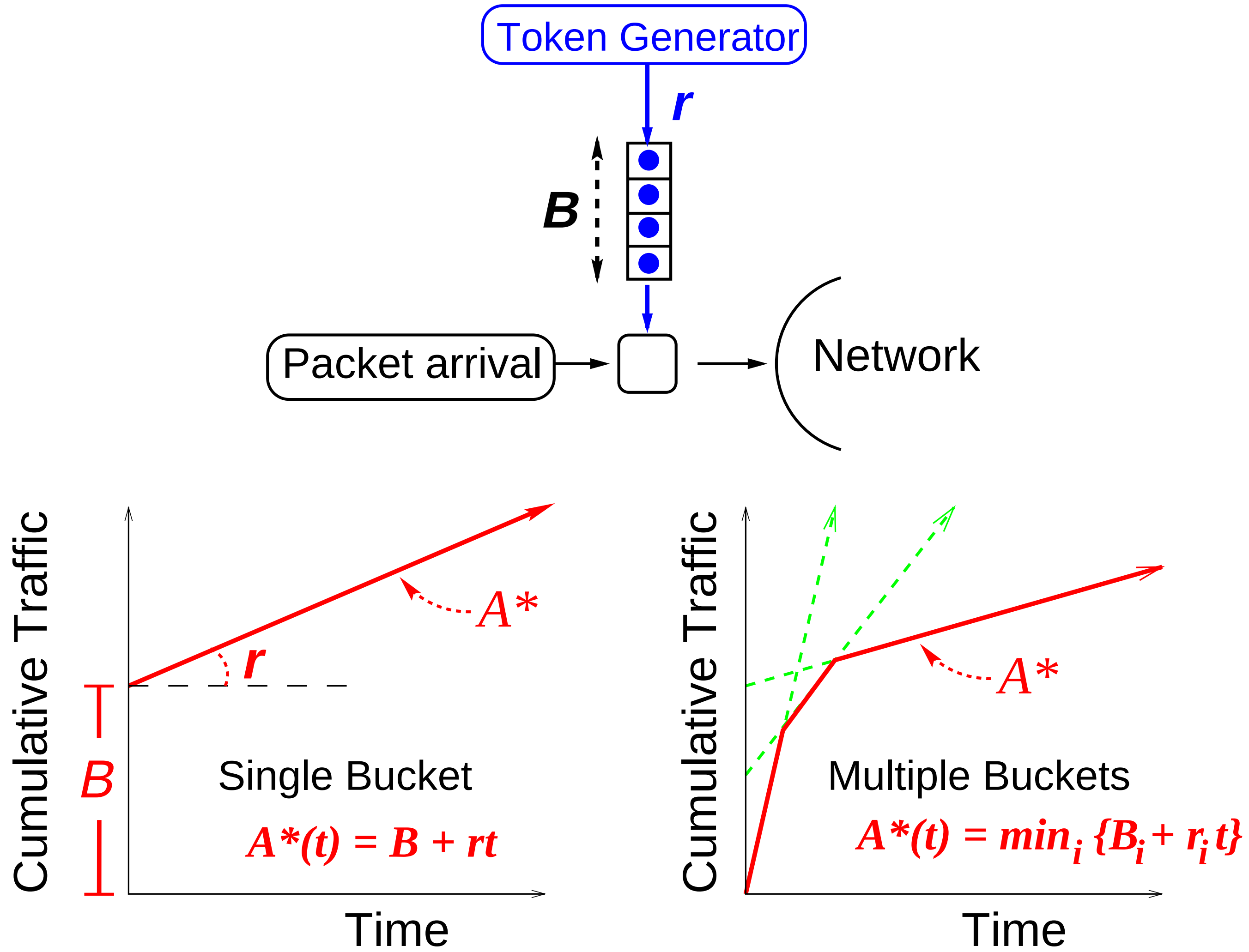
What is Traffic Characterization?





- A *traffic characterization* is a bound for the traffic over any interval.
  - *Time-invariant*: $A^*(t) \geq A[\tau, \tau + t], \quad \forall t, \tau$
  - *Subadditive*: $A^*(t_1 + t_2) \leq A^*(t_1) + A^*(t_2), \quad \forall t_1, t_2$
- Traffic characterization must map to traffic policer.

The "Leaky Bucket" Traffic Characterization

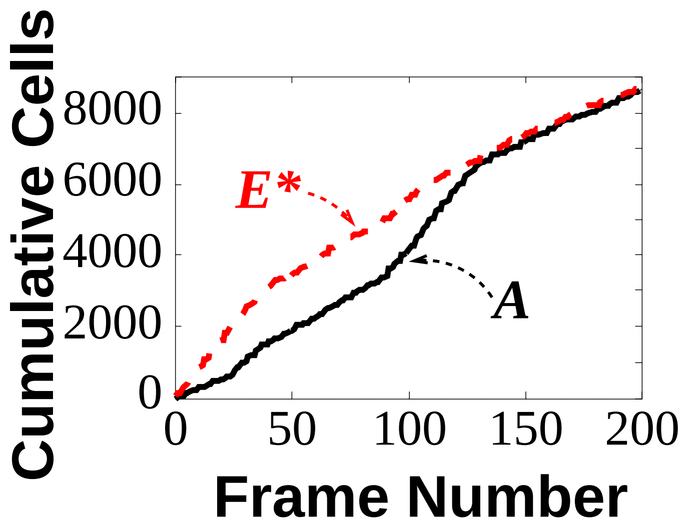


- Used in: ATM, Integrated-services Internet

Traffic Characterization Problem

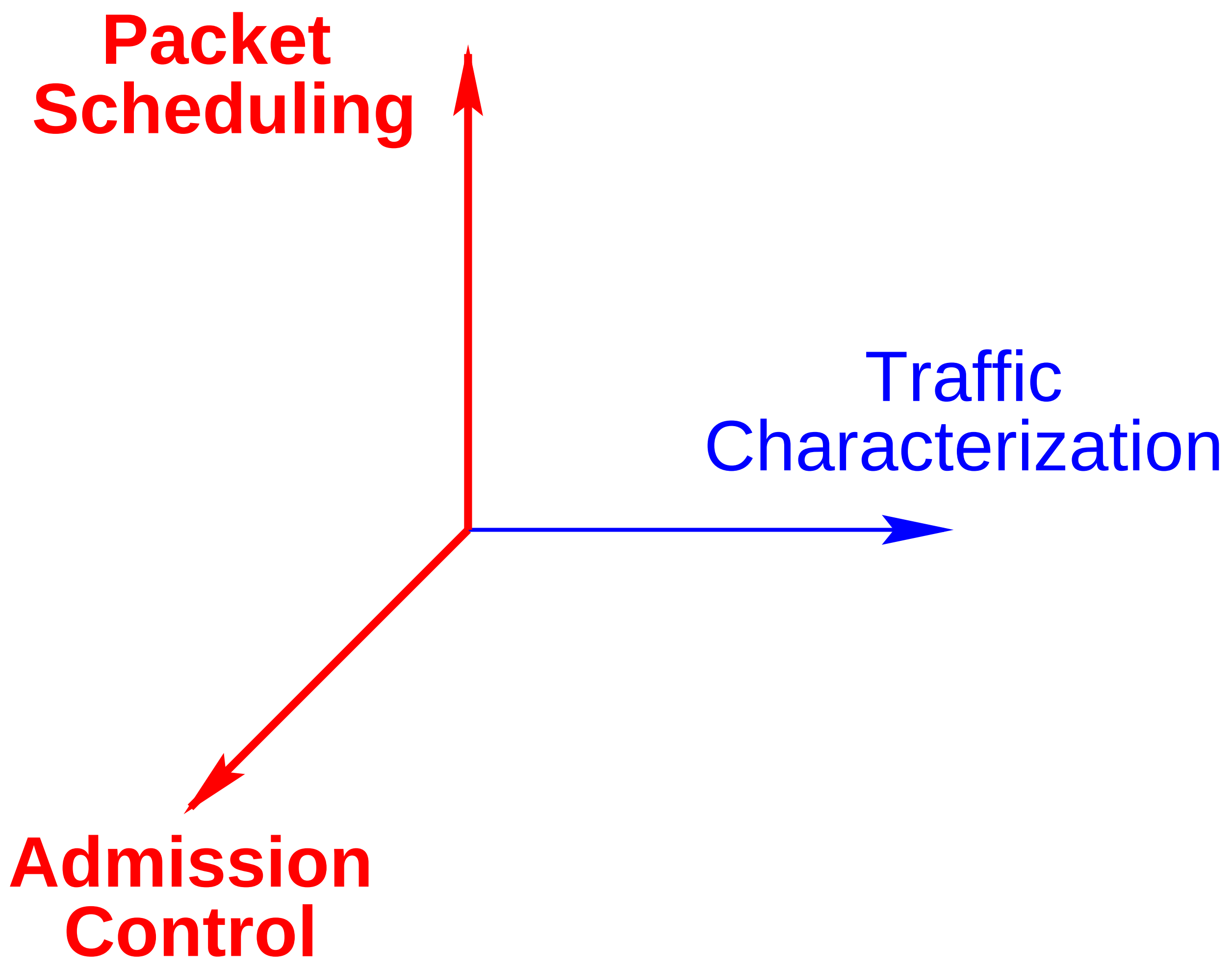
- Given a video sequence, how do I select leaky bucket parameters?
- Previous approaches:
  - *Candidate Sets* (Low and Varaiya 1991).
  - Choose B according to buffer space availability (Pancha and El Zarki 1995).
  - Relative importance of buffer space and bandwidth (Guillemin et. al. 1995).
  - *Empirical envelope* (Wrege, Knightly, Zhang, and Liebeherr 1996).

Empirical Envelope

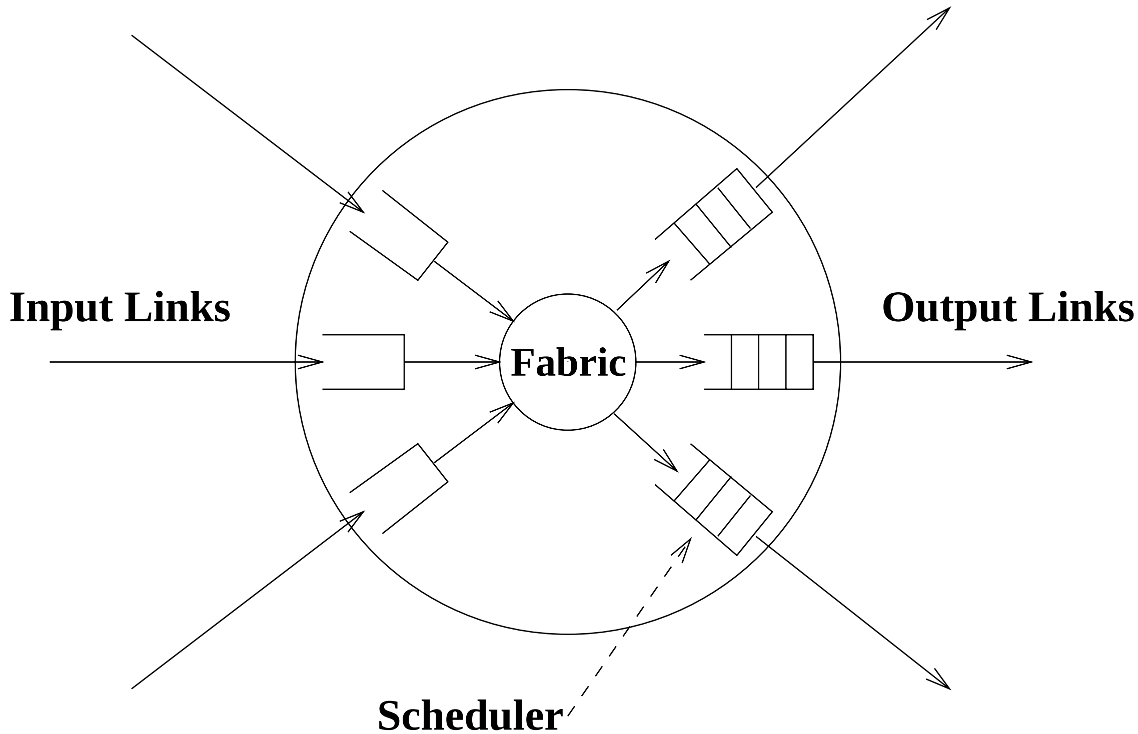


- The best possible characterization for a video source is its *empirical envelope* E^*.
- $E^*(t) := \sup_{\tau \geq 0} A[\tau, \tau + t]$, for all $t \geq 0$.
- *Difficult to police*, *expensive to compute*.

Design Space of a Multimedia Network



Packet Scheduling



- A connection j has a *delay bound* d_j.
- *Packet scheduling discipline* determines *delay*.

Admission Control

Schedulability Condition:

Given a packet scheduler and a set of connections. The connections are said to be **schedulable** if a violation of the delay bounds will never occur.

Schedulability Condition

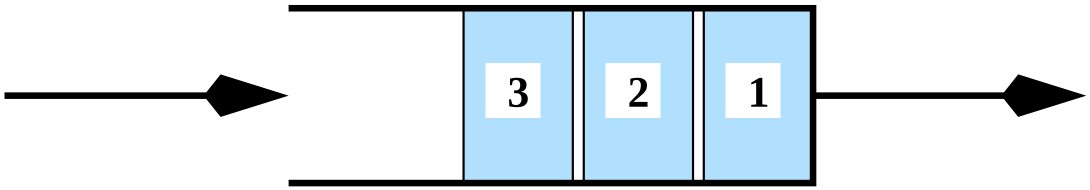
=

Delay Bound Test for Admission Control

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.

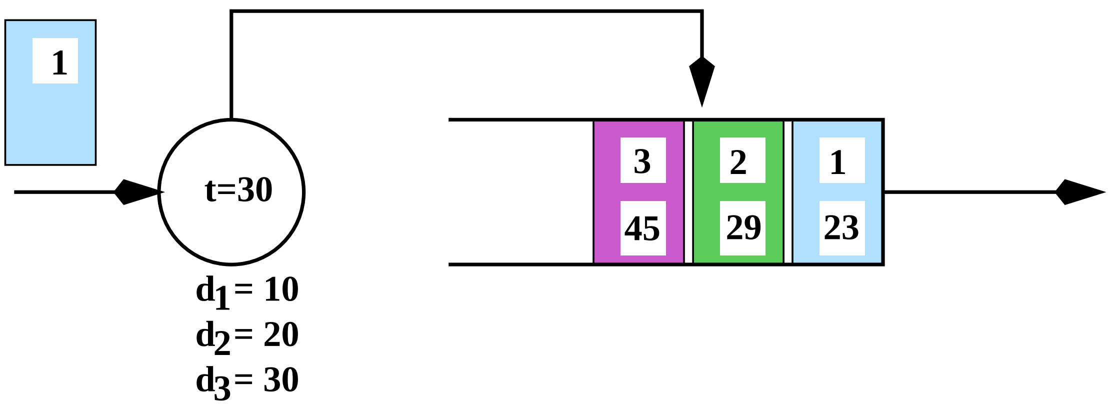
First-Come-First-Served (FCFS)



- Exact Admission Control Test:

$$d \geq \sum_{j \in \mathcal{N}} A_j^*(t) - t + \max_{k \in \mathcal{N}} s_k \qquad t \geq 0$$

Earliest-Deadline-First (EDF)

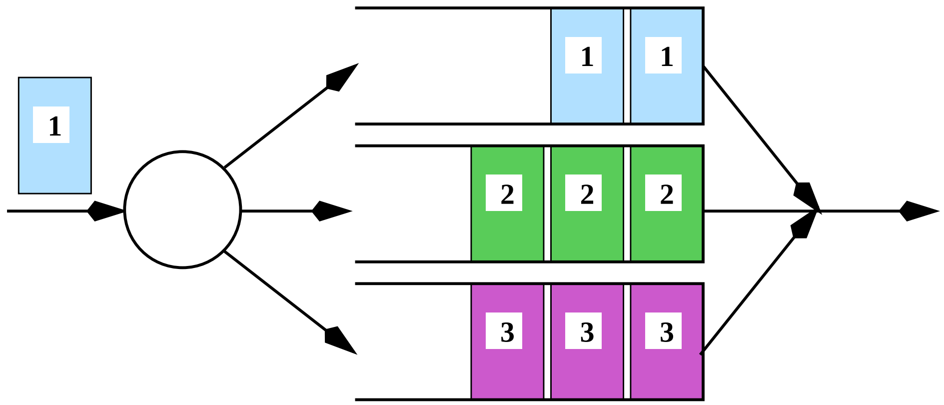


- Exact Admission Control Test (Liebeherr/Wrege/Ferrari):

$$t \geq \sum_{j \in \mathcal{N}} A_j^*(t - d_j) + \max_{k, d_k > t} s_k \qquad t \geq 0$$

where $\max_{k,d_k>t} s_k \equiv 0$ for $t > \max_{k \in \mathcal{N}} d_k$

Static Priority (SP)



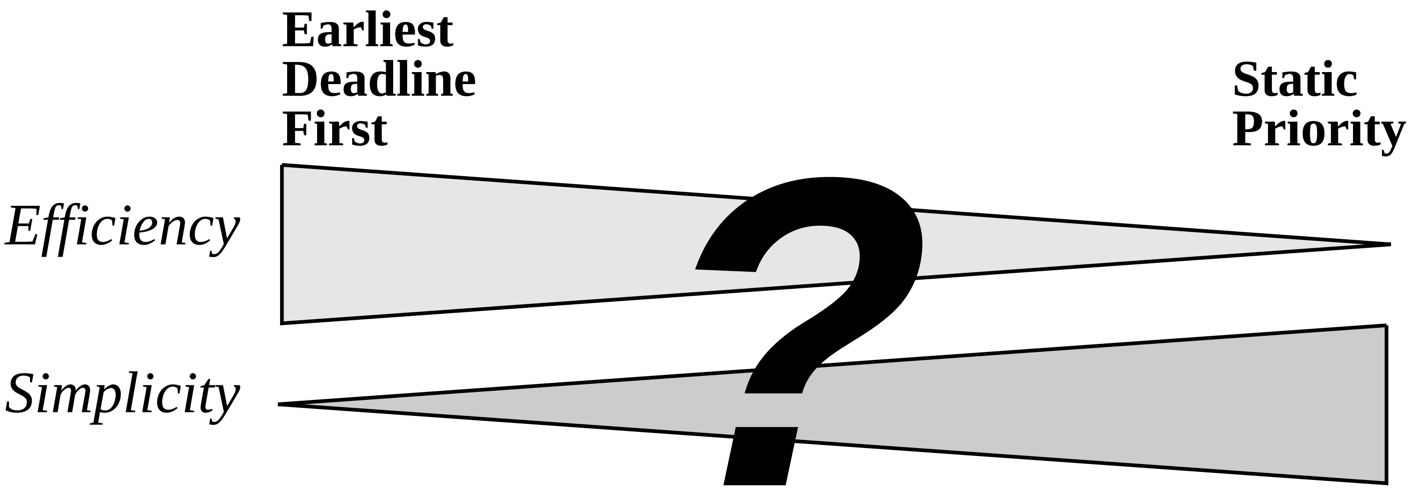
- Exact Admission Control Test (Liebeherr/Wrege/Ferrari):

$$(\exists \tau \leq d_p)$$

$$t + \tau \quad \geq \sum_{j \in \mathcal{C}_p} A_j^*(t) + \sum_{q=1}^{p-1} \sum_{j \in \mathcal{C}_q} A_j^*(t+\tau) + \max_{r>p} s_r$$

for all p, $t \geq 0$

What is a good scheduler?



Approximate EDF with FIFO queues

Approximations that require *no sorting*:

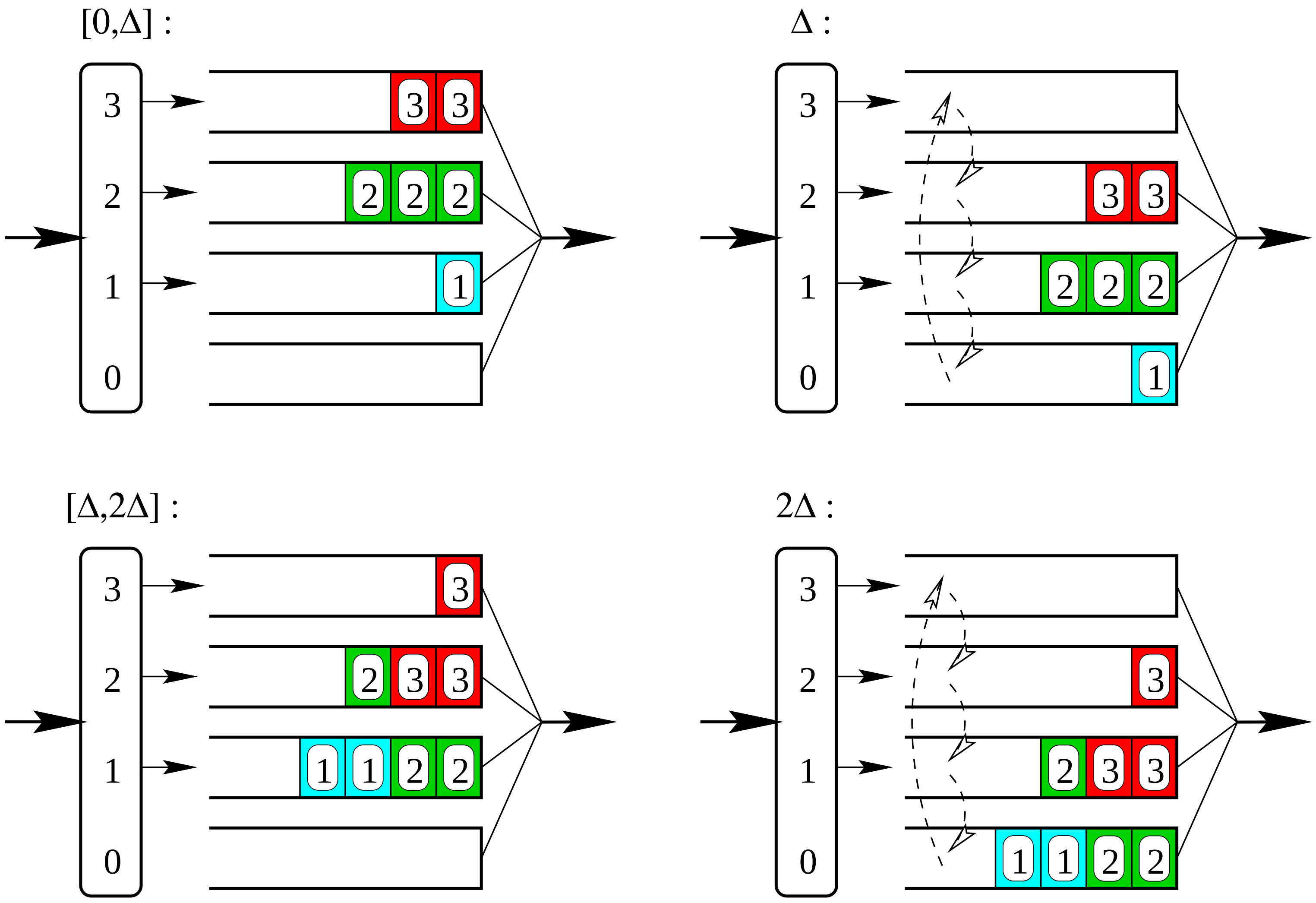
- *HOL-PJ* (Lim/Kobza 1990)
- *Relabeling Architecture* (Peha/Tobagi 1991)
- *Rotating-Priority-Queues* (RPQ) (Liebeherr/Wrege 1994)

Rotating-Priority-Queues (RPQ)

Design Principles:

- P priority sets.
- $P + 1$ FIFO queues with labels.
- Relabel queues every Δ time units.
- One delay bound for each priority set: $d_p = p \cdot \Delta$.

RPQ Scheduler



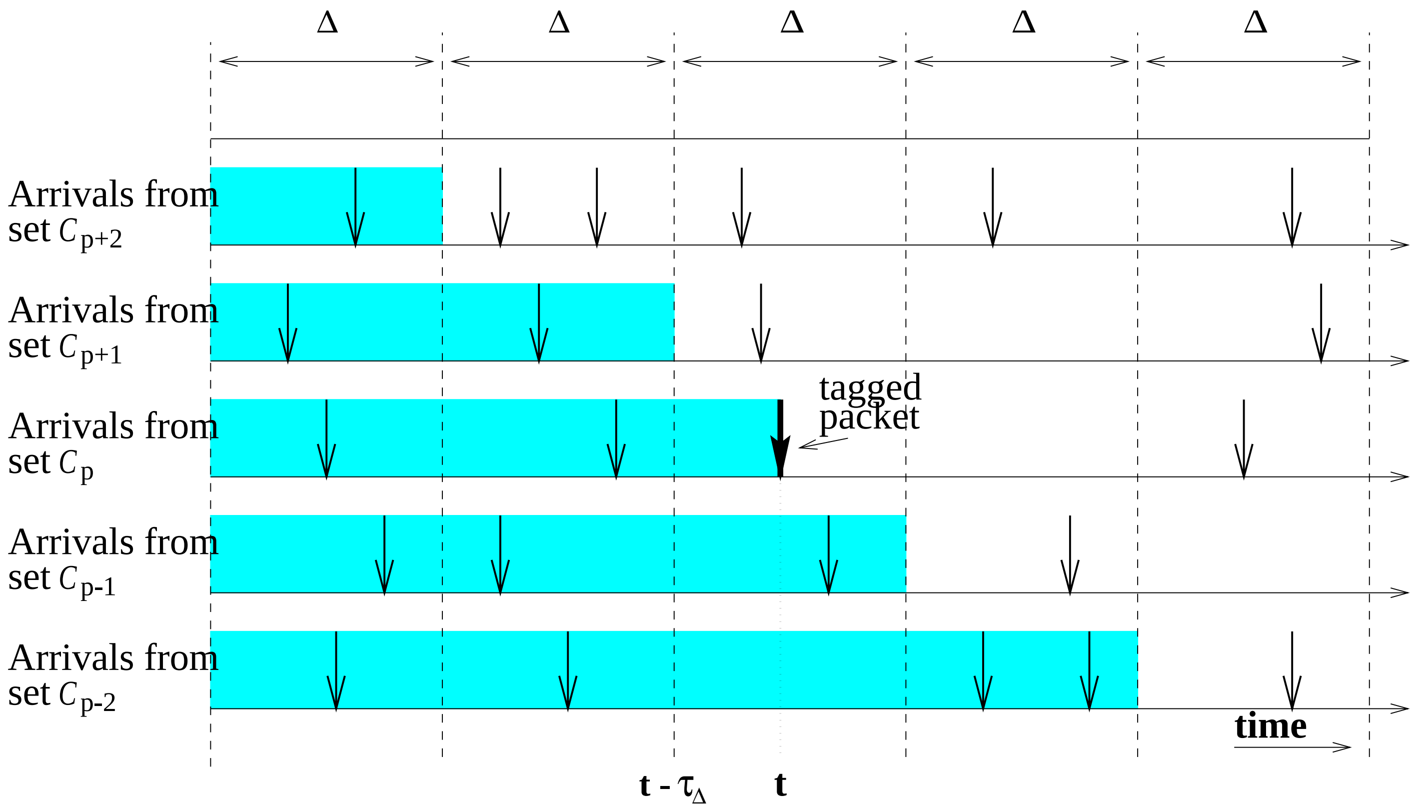
Admission Control Test for RPQ

For all priorities p and all $t \geq d_1$,

$$t \geq \sum_{j \in \mathcal{C}_1} A_j^*(t - d_1) + \sum_{q=2}^{P} \sum_{j \in \mathcal{C}_q} A_j^*(t + \Delta - d_q) + \max_{r, d_r > t+\Delta} s_r^{max}$$

RPQ

- Transmissions *before* a tagged packet.

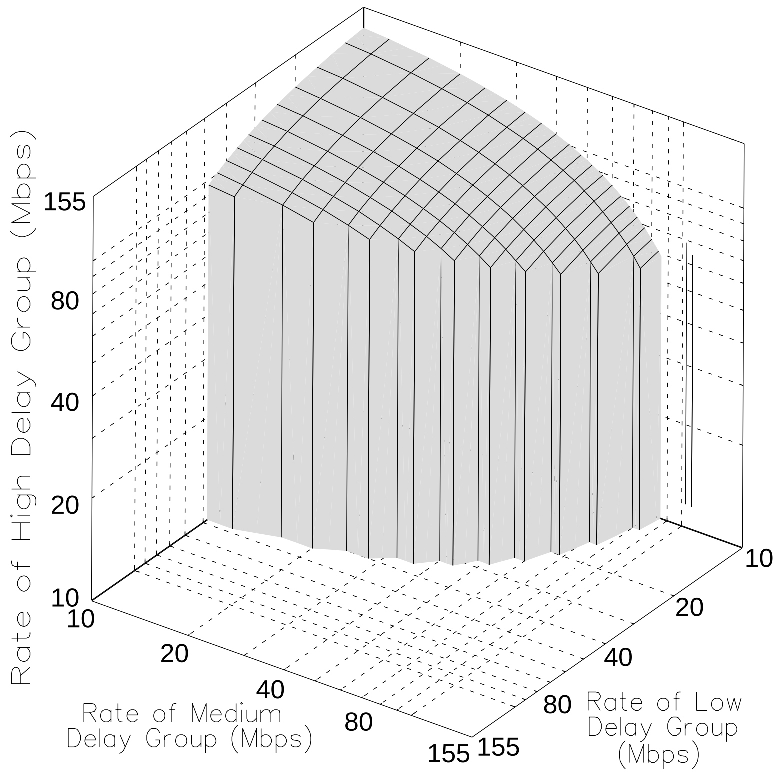


Experimental Setup

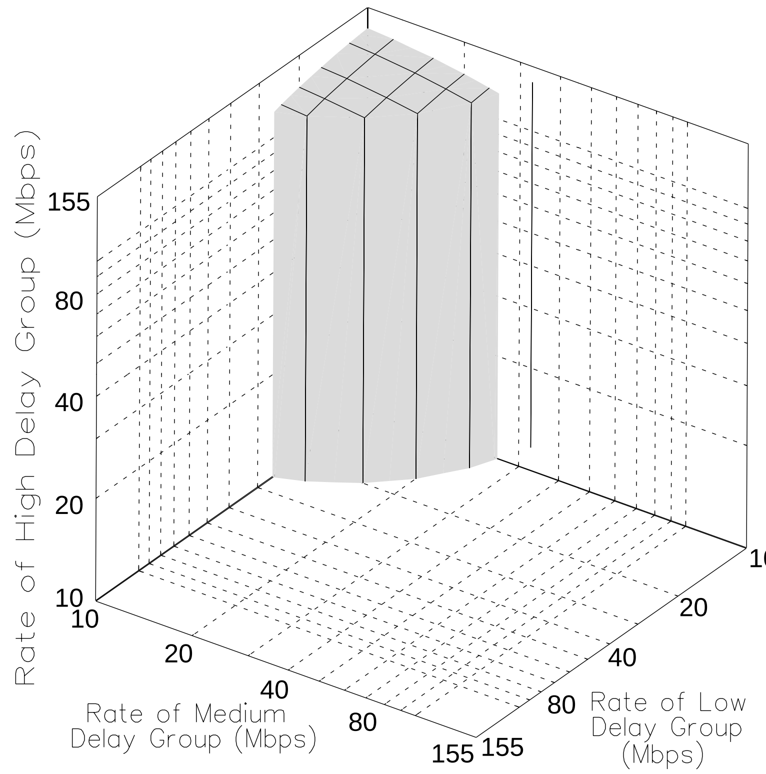
- Single 155 Mbps switch.
- Three connection groups *Low, Medium, High Delay*.

	Index j	Delay Bound d_j	Burst Size B_j	Rate r_j
Low	1	12 ms	4,000 cells	10-155 Mbps
Medium	2	24 ms	2,000 cells	10-155 Mbps
High	3	36 ms	4,000 cells	10-155 Mbps

Evaluation

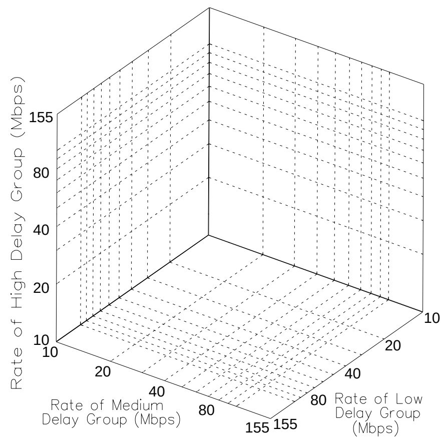


EDF

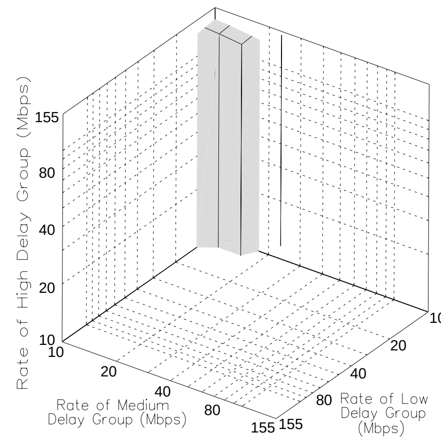


SP

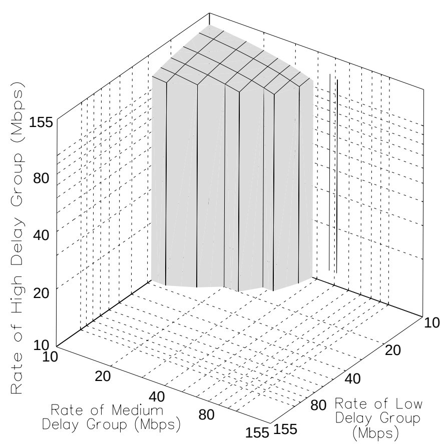
Evaluation of RPQ



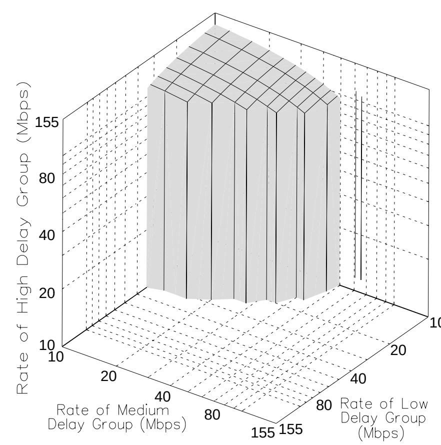
RPQ ($\Delta = 12ms$; 6 FIFOs)



RPQ ($\Delta = 6ms$; 12 FIFOs)



RPQ ($\Delta = 4ms$; 18 FIFOs)



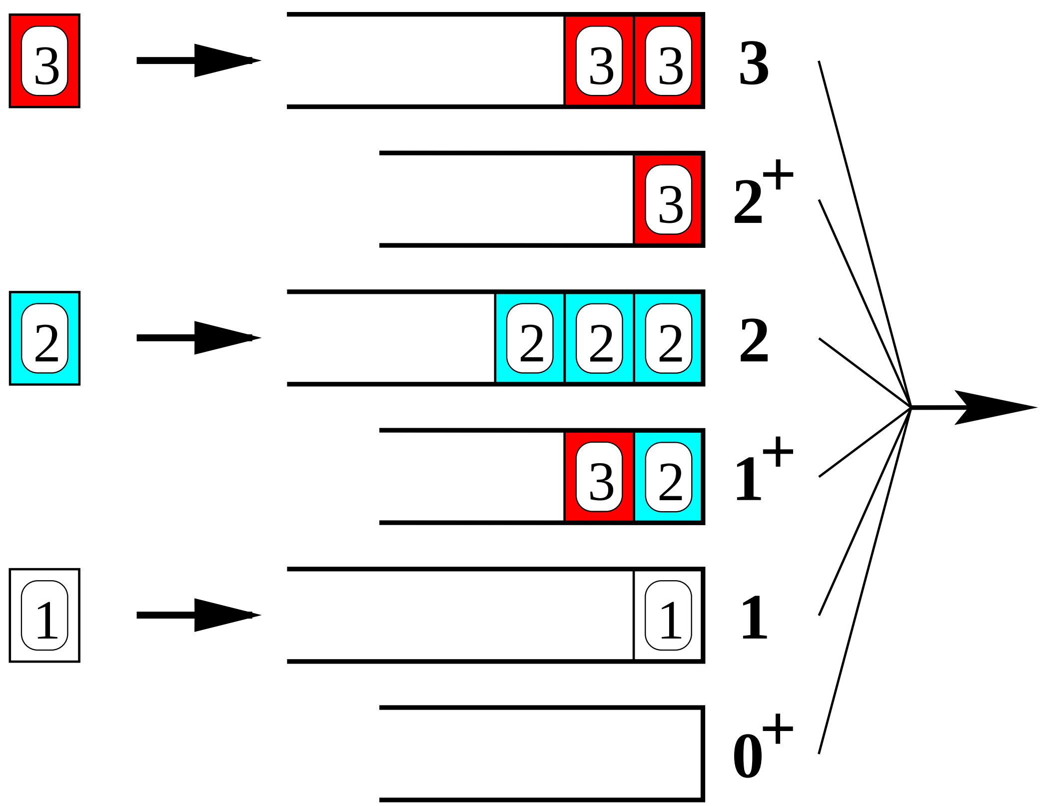
RPQ ($\Delta = 3ms$; 24 FIFOs)

Rotating-Priority-Queues$^+$ (RPQ$^+$)

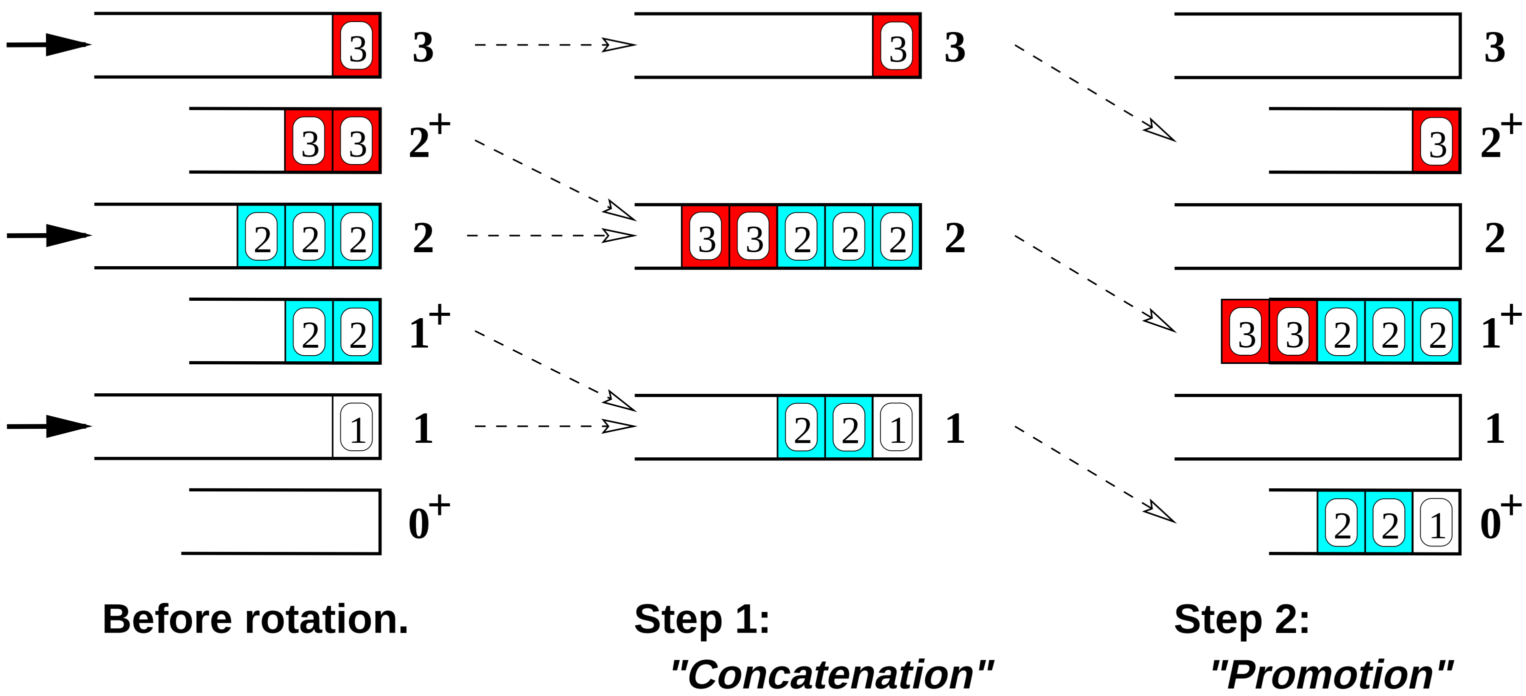
Design Principles:

- P priority sets.
- $2P$ FIFO queues with labels.
- Relabel queues every Δ time units.
- One delay bound for each priority set: $d_p = p \cdot \Delta$.

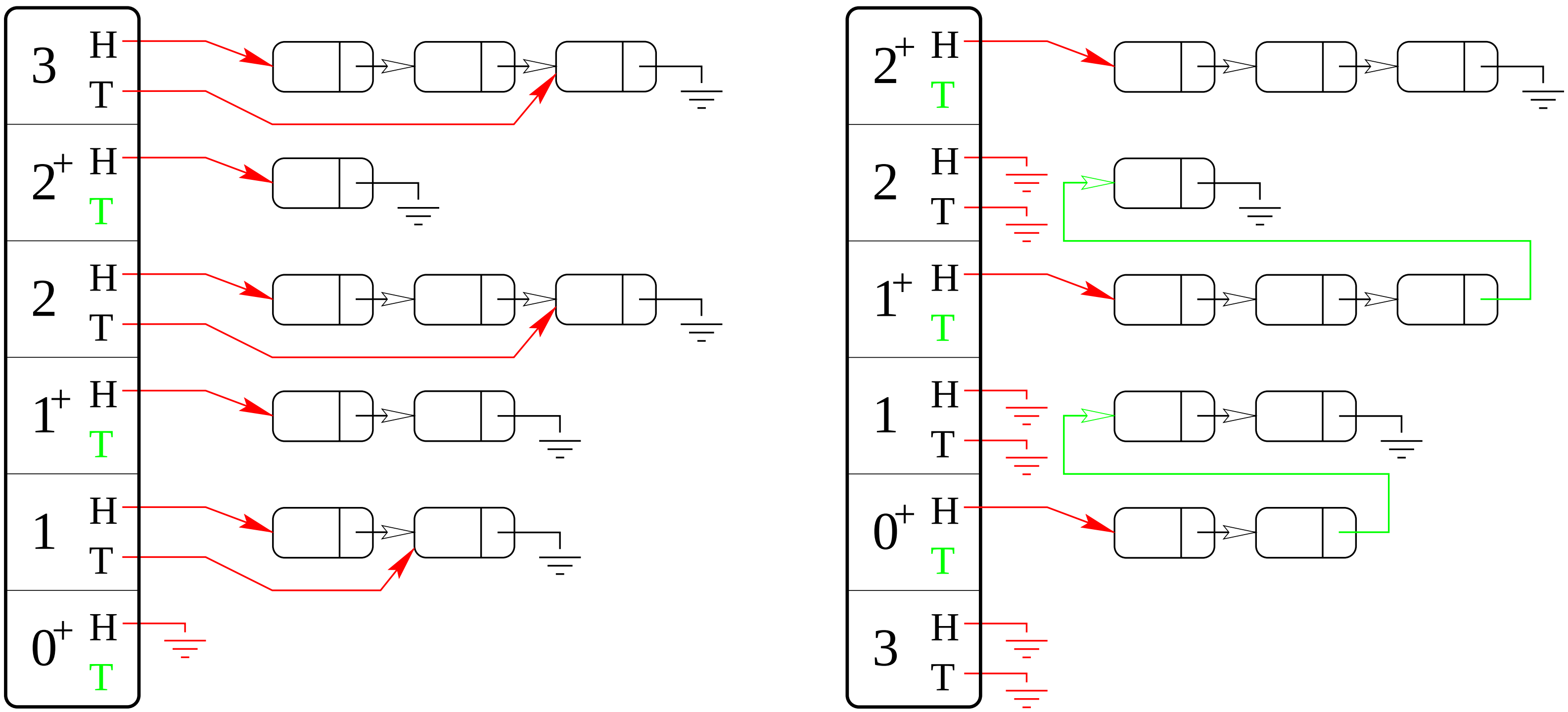
RPQ$^+$ Scheduler



RPQ$^+$ Queue Rotation



Implementating RPQ$^+$ in Shared Memory



- No movement of packets.
- Operations independent of queued packets.

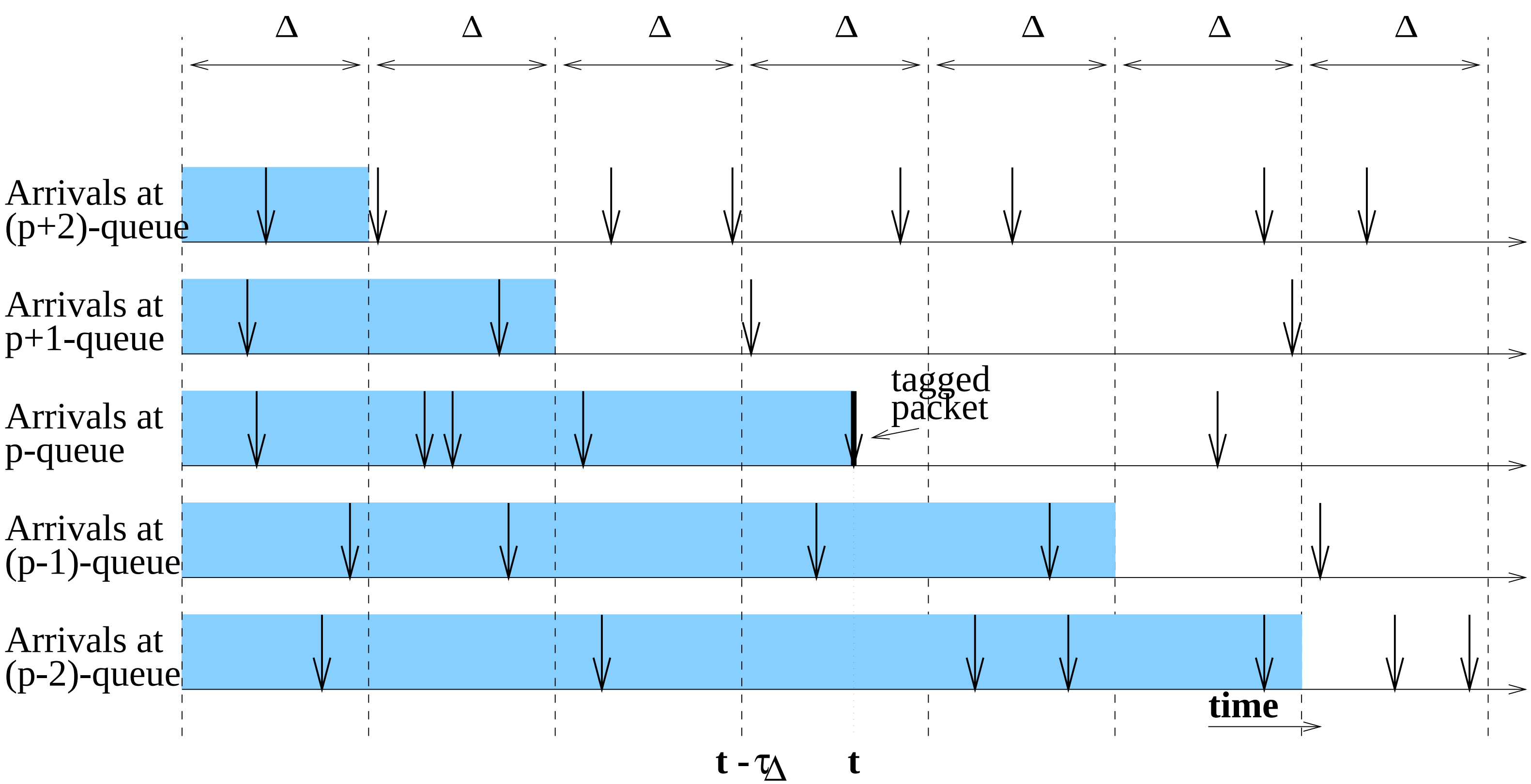
Admission Control Test for RPQ$^+$

For all priorities p and all $t \geq d_p$,

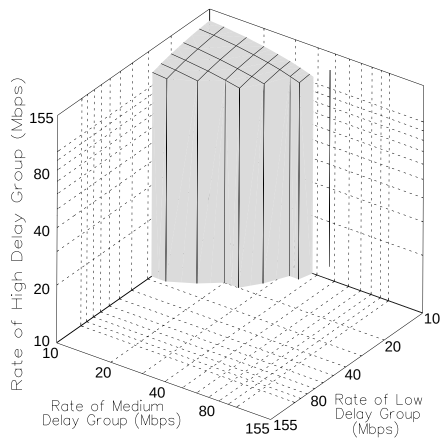
$$t \geq \sum_{q=1}^{p-1} \sum_{j \in \mathcal{C}_q} A_j^*(t - d_q + \Delta) + \sum_{q=p}^{P} \sum_{j \in \mathcal{C}_q} A_j^*(t - d_q) + \max_{r, d_r > t} s_r^{max}$$

RPQ$^+$

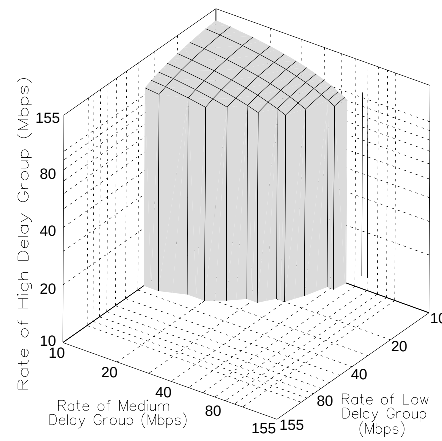
- Transmissions *before* a tagged packet.



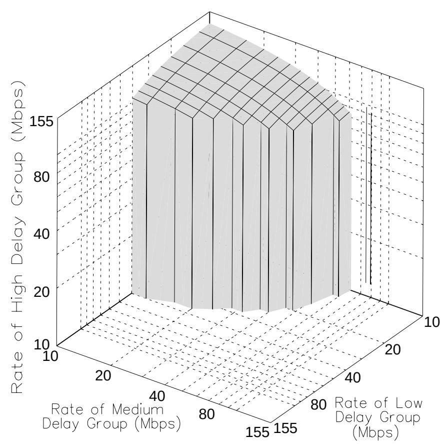
Evaluation of RPQ$^+$



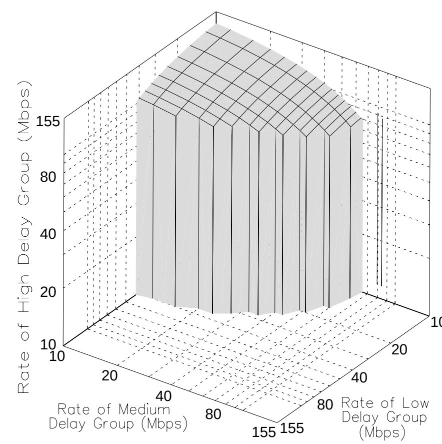
RPQ$^+$ ($\Delta = 12ms$; 6 FIFOs)



RPQ$^+$ ($\Delta = 6ms$; 12 FIFOs)



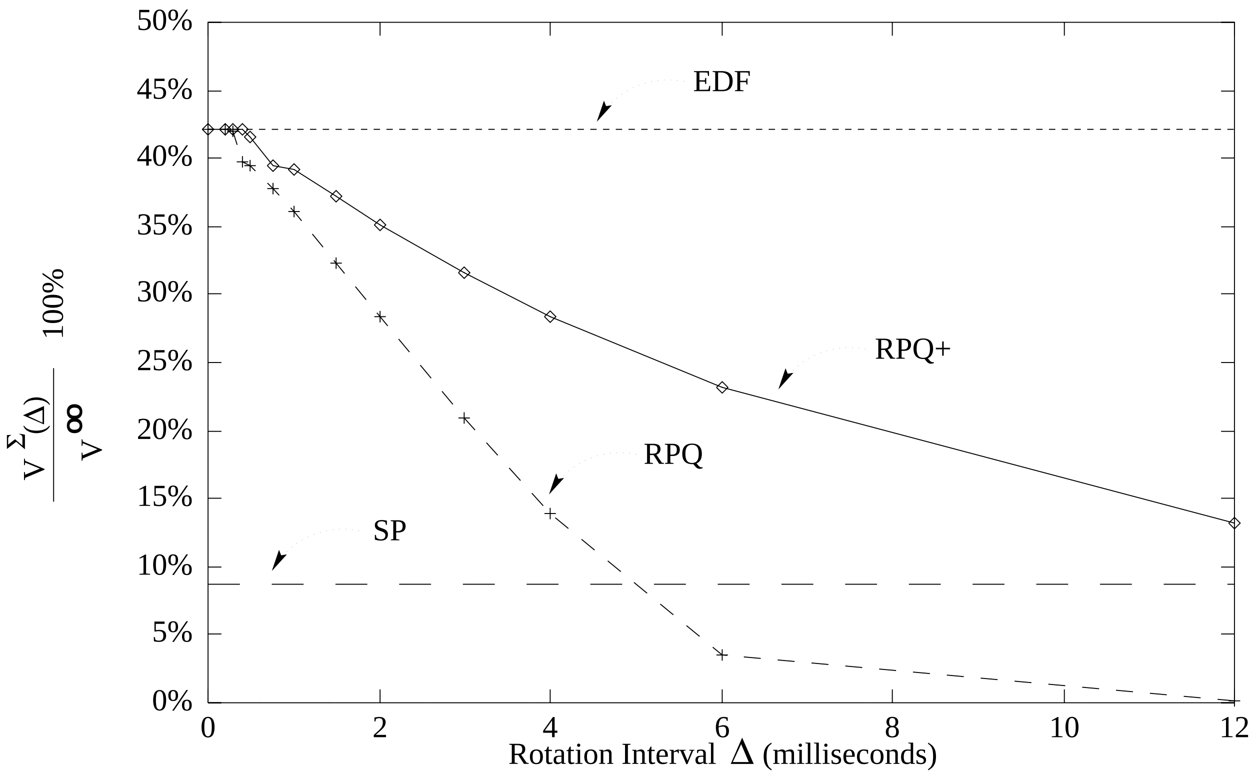
RPQ$^+$ ($\Delta = 4ms$; 18 FIFOs)



RPQ$^+$ ($\Delta = 3ms$; 24 FIFOs)

Summary of Evaluation

- Compare volume of the schedulable regions:
  $\frac{V^{\Sigma}(\Delta)}{V^{\infty}} \cdot 100\%$



Conclusions

- Approximate EDF with rotating FIFO queues.
- Simple solution (RPQ) can be worse than SP.
- RPQ^+ is "between" SP and EDF.
- Reading:

  `IEEE/ACM Transactions on Neworking`, June 1996.

  `IEEE/ACM Transactions on Neworking`, December 1996.

  `Proc. IEEE Infocom '96`, San Francisco, March 1996.

  `Proc. IEEE Infocom '97`, Kobe, April 1997.