CSc 461/561
Multimedia Systems
Multimedia QoS

Jianping Pan

Spring 2015

MM application requirements

- The amount of multimedia data is huge
  - many need certain (minimum) bandwidth
  - some can tolerate packet loss to a certain extent
- Multimedia applications often interactive
  - many have upper bound on end-to-end delay
  - some are sensitive to delay variance (jitter)
- Multimedia may involve multiple endpoints
  - some need multicast, session management

MM QoS metrics: example

- Multimedia QoS: timeliness
  - bounded end-to-end delay/jitter
  - interactivity, smooth playback
- Network QoS: delay
  - processing delay (packetization, protocol, etc)
  - transmission delay (bandwidth constraints)
  - propagation delay (distance constraints)
  - queuing delay (store-and-forward, contention)

Internet (QoS) status

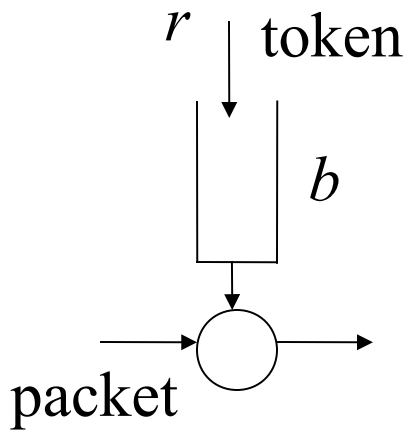
- Best effort services
  - no (session) admission control
  - no resource reservation
- Drop-tail router queues
  - no packet classification
  - no packet scheduling
- Good for *elastic* applications
  - email, FTP, web, etc

Internet QoS goals

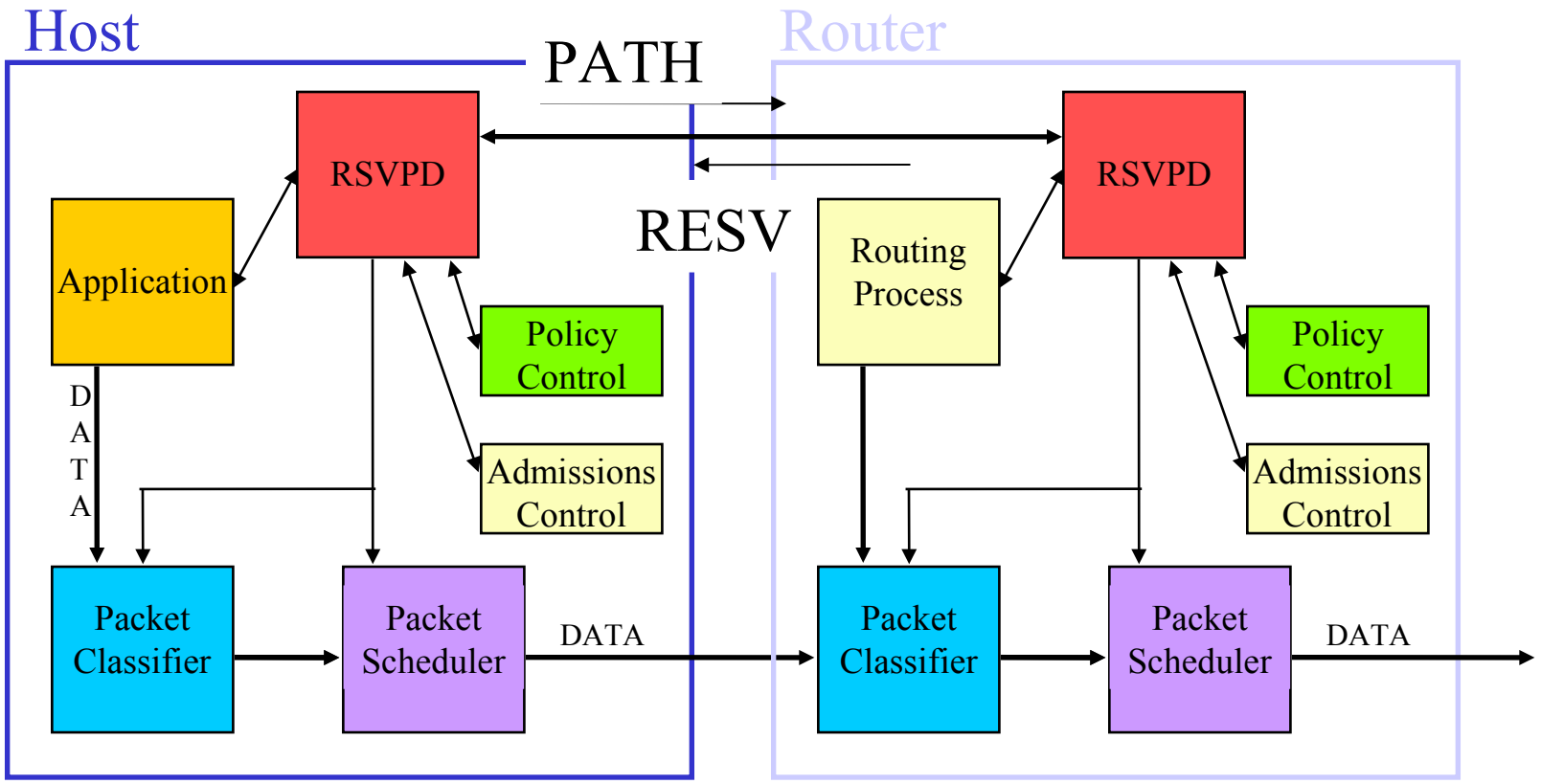
- RSVP
  - resource reservation
- Integrated Services
  - fine-granularity QoS
- Differentiated Services
  - coarse-granularity QoS
- Other approaches

RSVP design guidelines

- The ReSerVation Protocol
  - upon unicast or multicast routing
  - receiver-oriented
  - soft state
- Per-flow resource reservation
  - flow definition
  - traffic specification
    - leaky token bucket



RSVP in action



RSVP messages

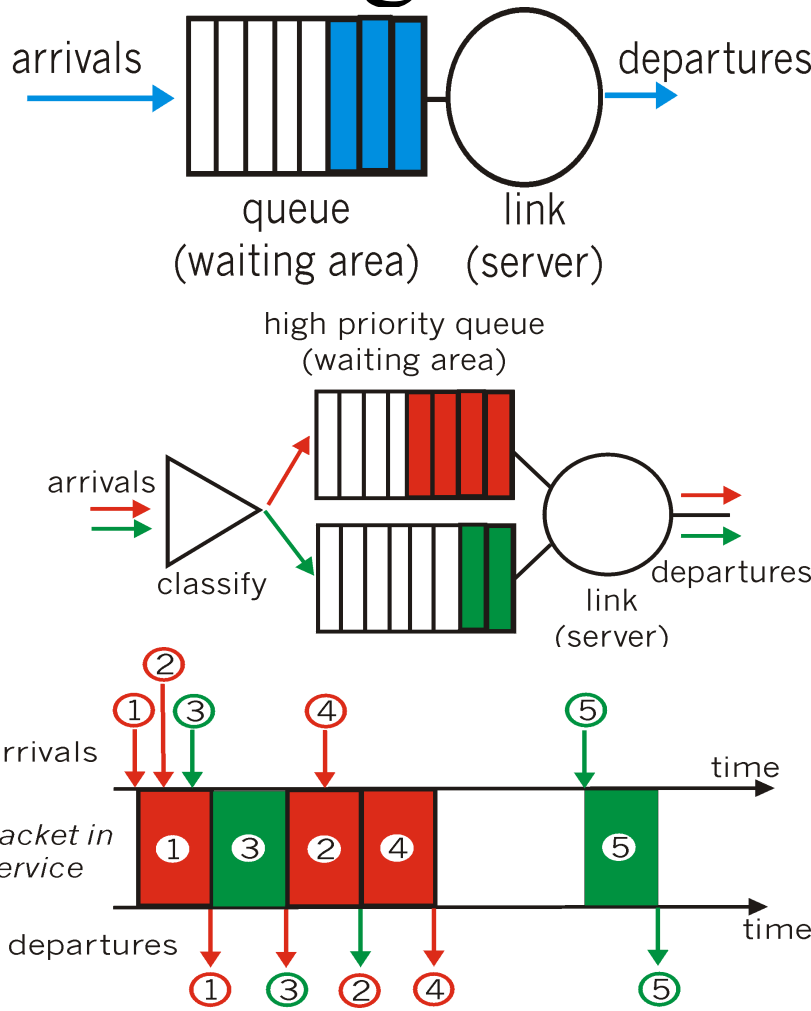
- PATH
  - sender => routers => receiver(s)
  - traffic specification (T-spec)
    - token rate, bucket size, peak rate, etc
- RESV
  - receiver => routers => sender; reserve PATH
  - reservation specification (R-spec): e.g., rate
  - filters: wildcard, shared, explicit

Integrated Services

- Guaranteed service
  - emulating dedicated virtual circuit
  - for *hard* real-time applications
  - bounded queuing delay, admission control
- Controlled-load service
  - equivalent to best-effort in unloaded condition
  - for *adaptive* real-time applications
  - e.g., measurement-based admission control

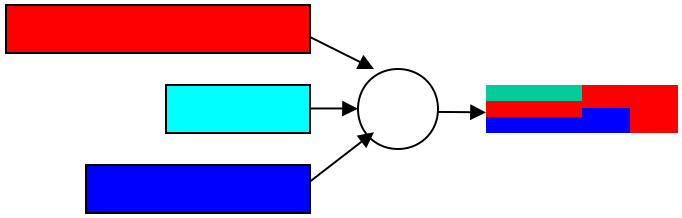
Queuing and scheduling

- First-Come-First-Serve
  - when queue is full
    - drop tail
    - drop head
    - random drop
- Priority
- Round robin
  - weighted round robin



General processor sharing

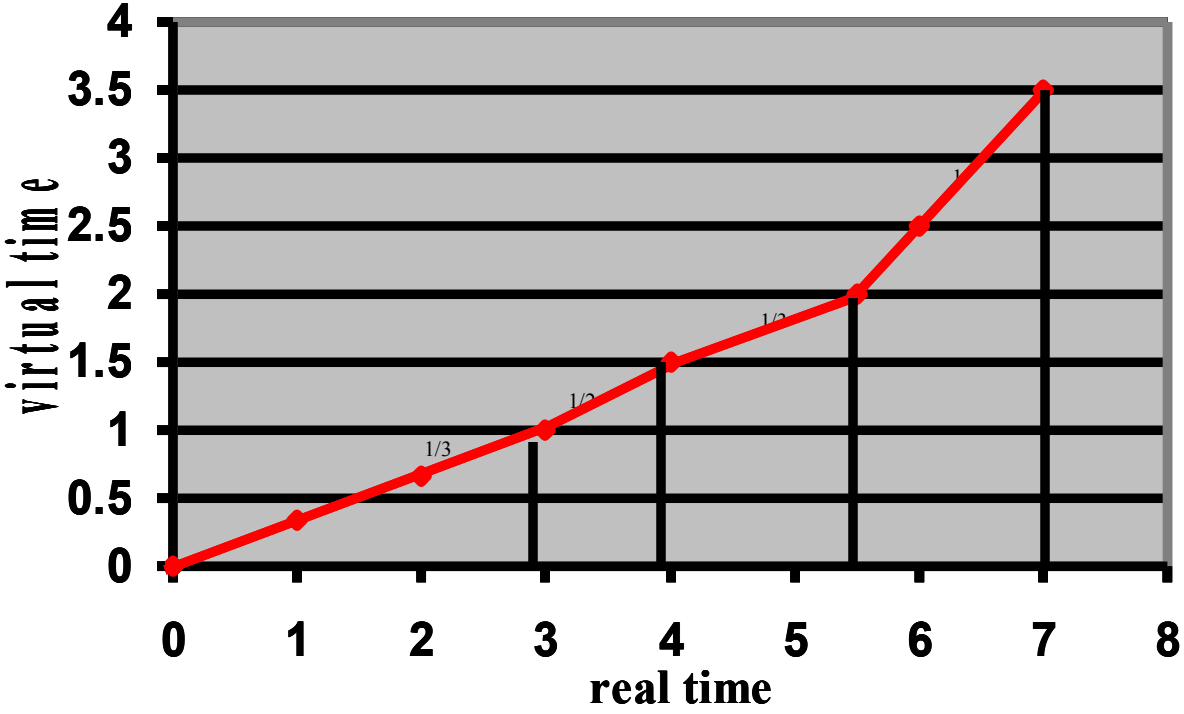
- GPS: ideal case
  - treat data as fluid
  - serve active flows simultaneously
    - fair allocation among active flows
  - not realizable
    - network handles packets
    - packets of different size
    - arriving at different time

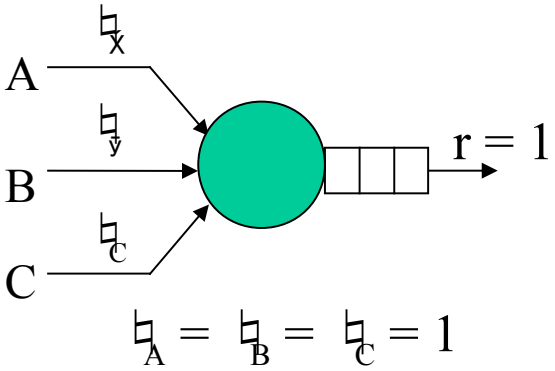


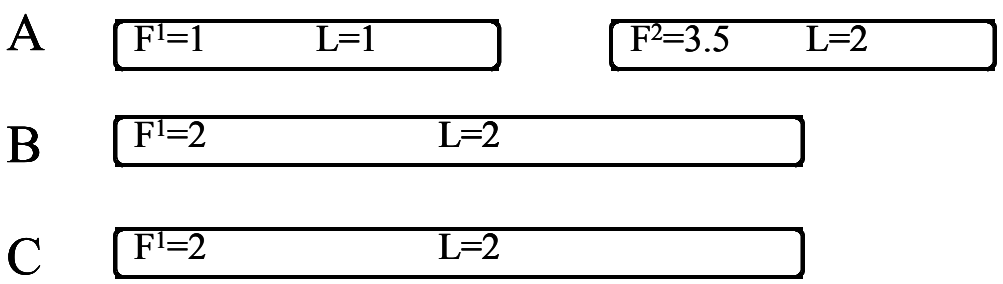
Weighted fair queuing

- Packetized GPS: approximation
  - serve packet by packet
  - pick the first one to *finish* in GPS
    - virtual time: round number (variable in real time)
    - assume active flows served one bit in each round
    - update finish round number upon packet arrival
- WFQ: weighted fairness
  - bounded difference with GPS

WFQ example







t=0: Packets of sizes 1,2,2 arrive at connections A, B, C.

t=4: Packet of size 2 arrives at connection A

Differentiated Services

- Issues with IntServ
  - scalability: per-flow
  - flexibility: two QoS classes
- DiffServ: new approach by IETF
  - more work at edge routers
    - classifying, metering, marking, shaping
  - less work at core routers
    - forwarding based on DSCP and PHB

Per-Hop-Behavior

- Expedited Forwarding (EF)
  - guarantee a minimum rate for EF
  - admitted based on peak rate
  - virtual circuit like
- Assured Forwarding (AF)
  - service classes (e.g., gold, silver, bronze)
  - a few drop priorities in each class

This lecture

- Multimedia QoS
  - QoS metrics in different layers
  - IntServ and DiffServ
  - RSVP
  - Queuing and scheduling
- Explore further
  - GPS bounds
  - WFQ and variants (SCFQ, WF^2Q, etc)

Next lecture

- Multimedia congestion control