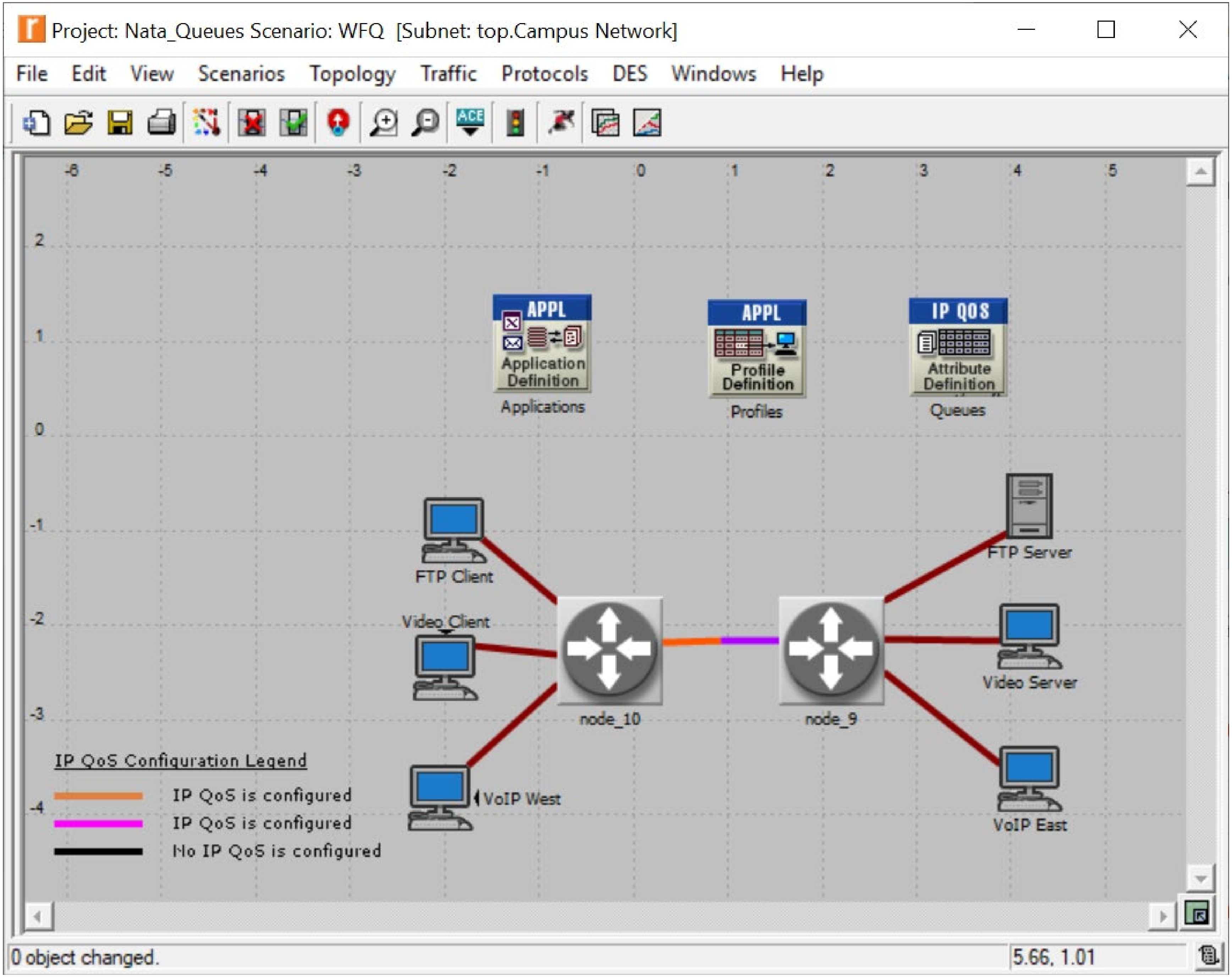
Network Layer:
Queueing Disciplines

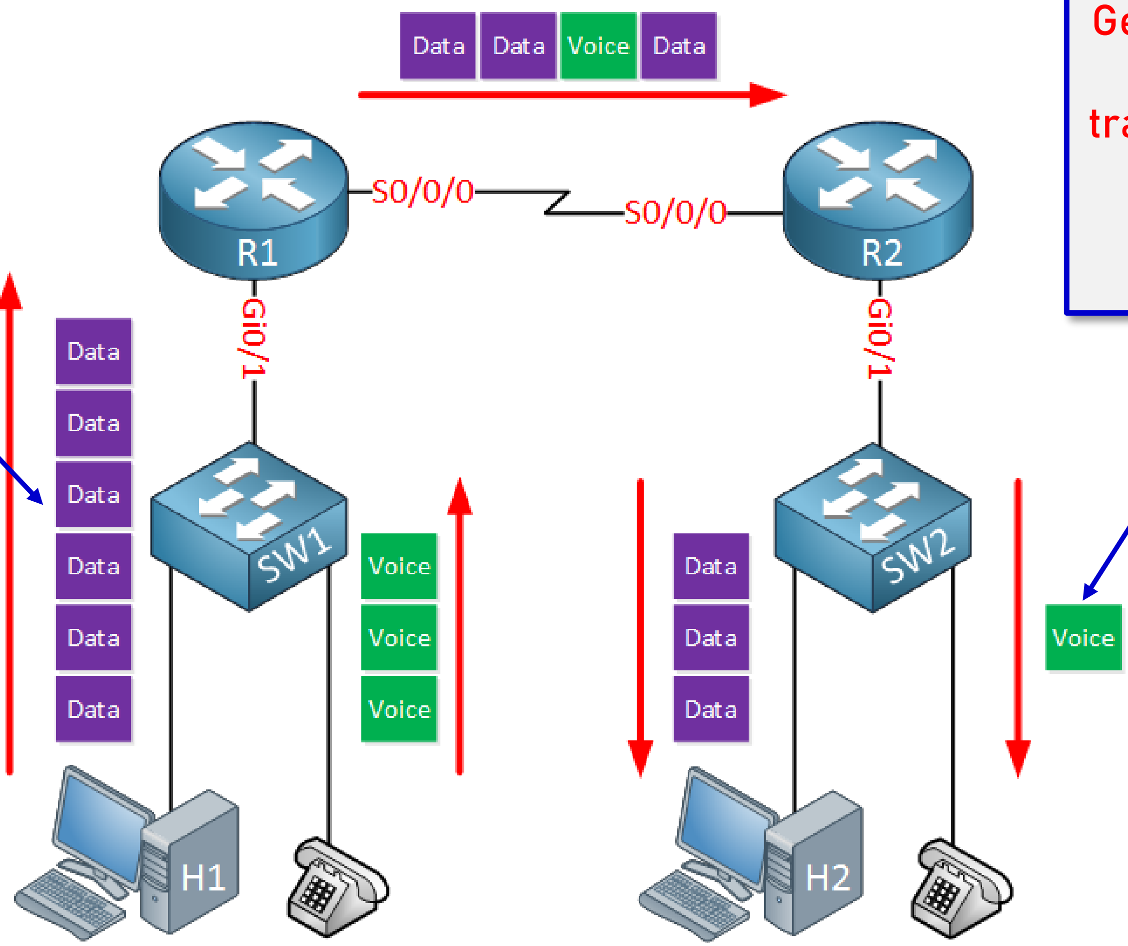
(LAB 1 Related Material)


Project: Nata_Queues Scenario: WFQ [Subnet: top.Campus Network]
File Edit View Scenarios Topology Traffic Protocols DES Windows Help
APPL
Application Definition
Applications
APPL
Profile Definition
Profiles
IP QOS
Attribute Definition
Queues
FTP Client
Video Client
VoIP West
node_10
node_9
FTP Server
Video Server
VoIP East
IP QoS Configuration Legend
IP QoS is configured
IP QoS is configured
No IP QoS is configured
0 object changed.
5.66, 1.01

characteristics & requirements of data transfer ≠ streamed video ≠ VoIP

May have lots of traffic/packets, but there are no strict time-related constraints.

Generally does not produce much traffic/packets, but has strict time-related constraints.



https://networklessons.com/cisco/ccna-routing-switching-icnd2-200-105/introduction-qos-quality-service

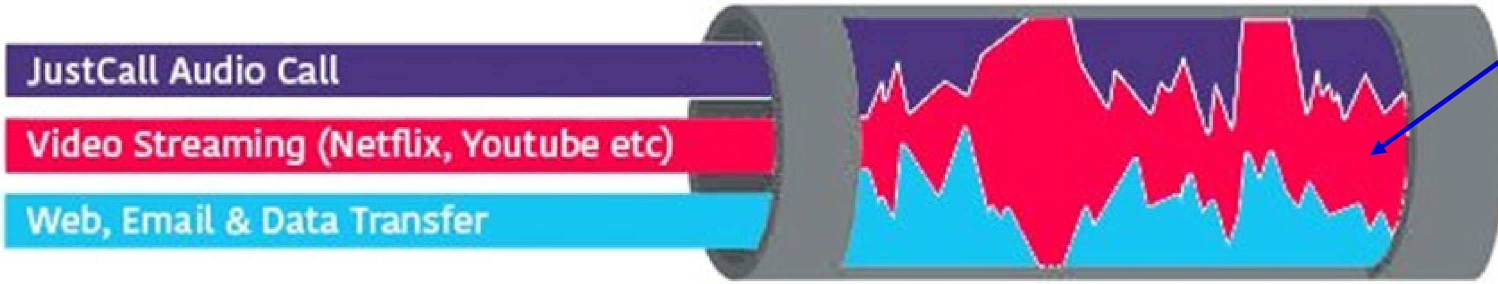
Quality of Service (QoS) requirements of different applications.

Should throughput be high or low??

Application Type	Throughput Demand	Latency Tolerance	Jitter Tolerance	Loss Tolerance
Email	Low	High	High	High
Web browsing	Low	High	High	High
File transfer (FTP)	Low - High	High	High	High
Chat (IM)	Low	Medium	Medium	Medium
Video streaming (e.g. surveillance)	Medium - High	Medium	Medium	Medium
Video on demand (e.g. YouTube / NetFlix)	High	Medium	Medium	Low
Voice over IP / WiFi	Low	Low	Low	Low
Videoconferencing (e.g. Skype, FaceTime)	Medium - High	Low	Low	Low

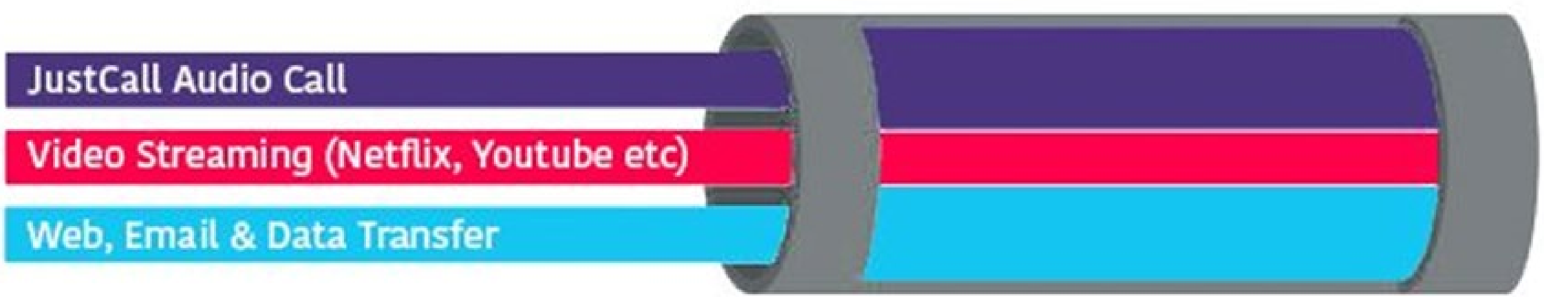
https://blogs.cisco.com/collaboration/connectivity-and-network-service

Sharing of bandwidth/router capacity when there are NO special QoS provisions in place.

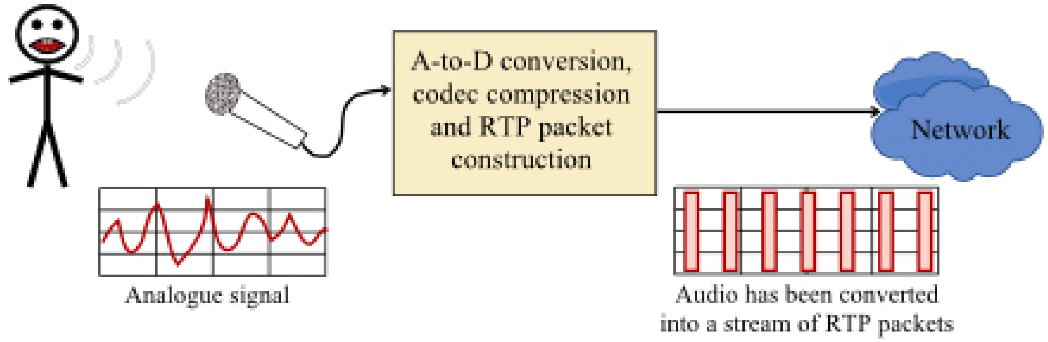


More dominant traffic flows get allocated more bandwidth 'without any regard' for other traffic flows.

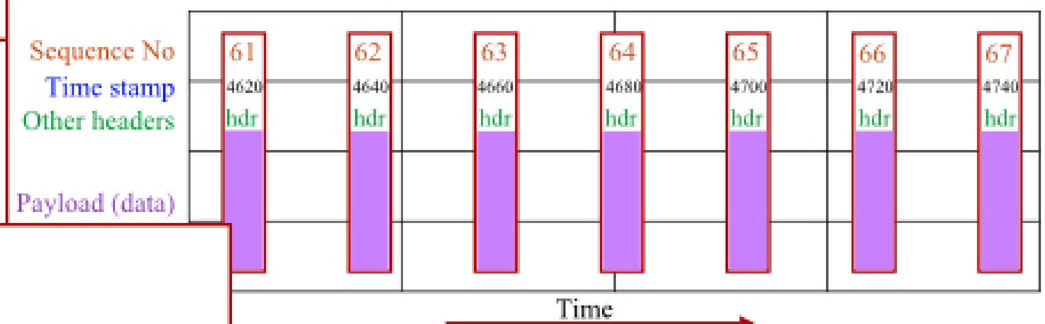
Sharing of bandwidth/router capacity when there are special QoS provisions in place.



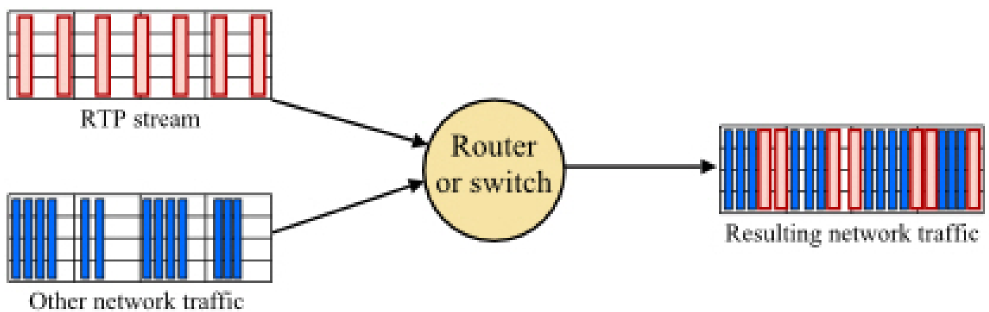
https://justcall.io/blog/bandwidth-requirements-voip-phone-systems.html



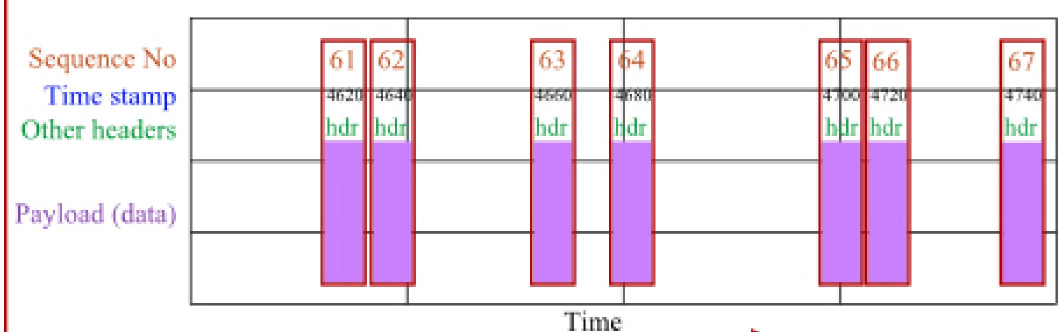
RTP packets generated at the source device



A jitter-free stream of RTP packets



RTP packets have to jostle for position with other network traffic



This sample of the RTP stream has jitter (but has no dropped packets)

https://kb.smartvox.co.uk/voip-sip/rtp-jitter-audio-quality-voip/

Quality of Service

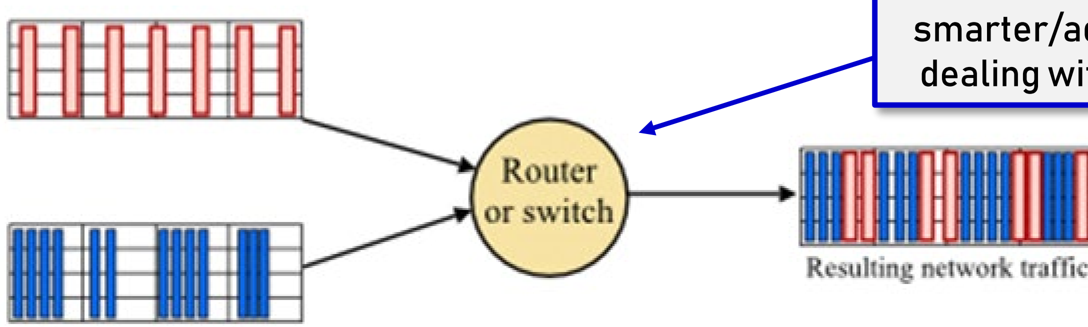
Quality of Service (QoS) – capability of a network to guarantee a certain level of performance to an application, user or data flow

- parameters used to define / measure QoS
  - (1) RELIABILITY - measures packet loss
  - (2) DELAY - measures delay in individual packet transmiss.
  - (3) JITTER - measures variation in packet delay
  - (4) BANDWIDTH - measures overall demand on [bps]

Different applications required different QoS.

What could we do to try to provide each application with the right QoS??

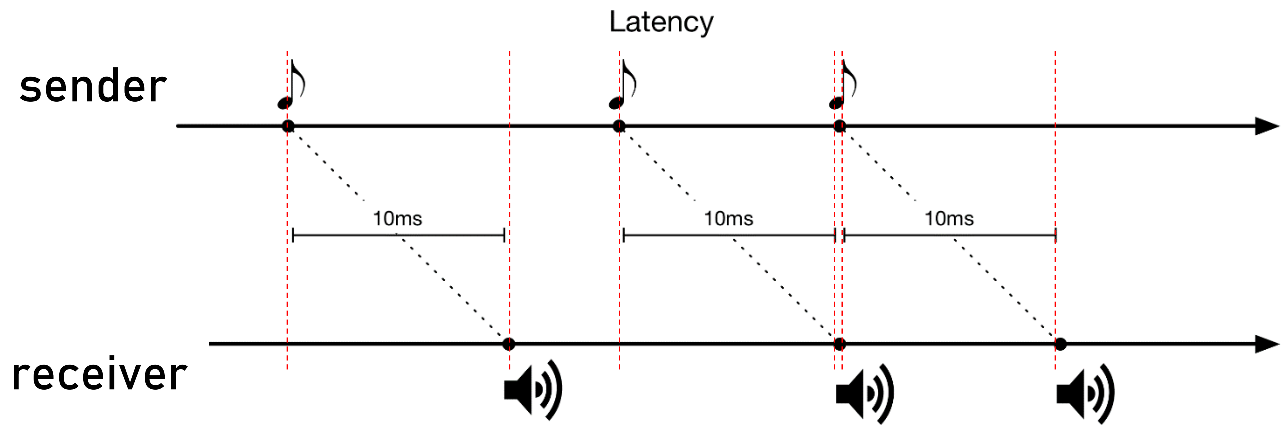
Make routers deploy smarter/advanced techniques of dealing with incoming packets …



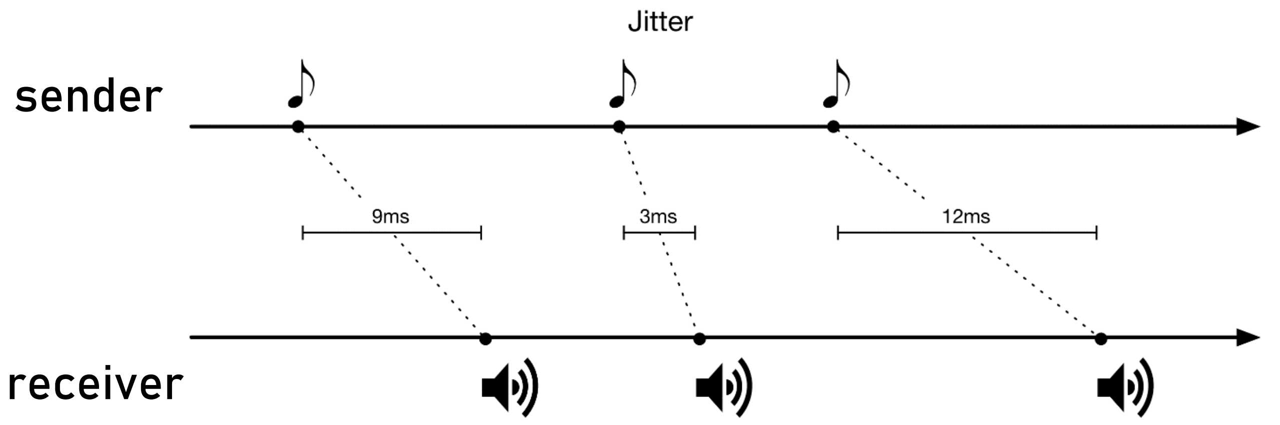
Quality of Service (cont.)

Example **[delay vs. jitter]**

delay ≠ 0
jitter = 0



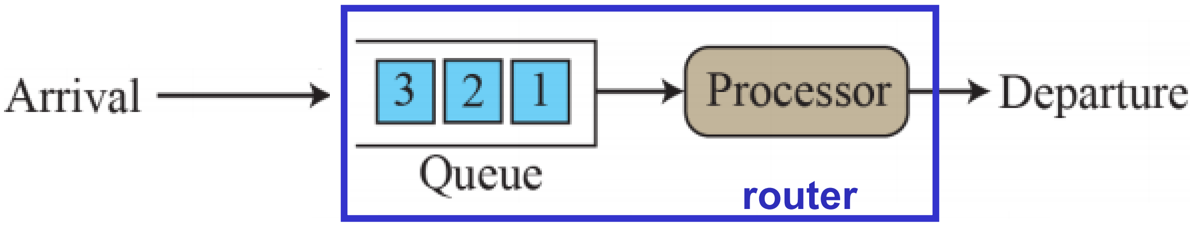
delay ≠ 0
jitter ≠ 0



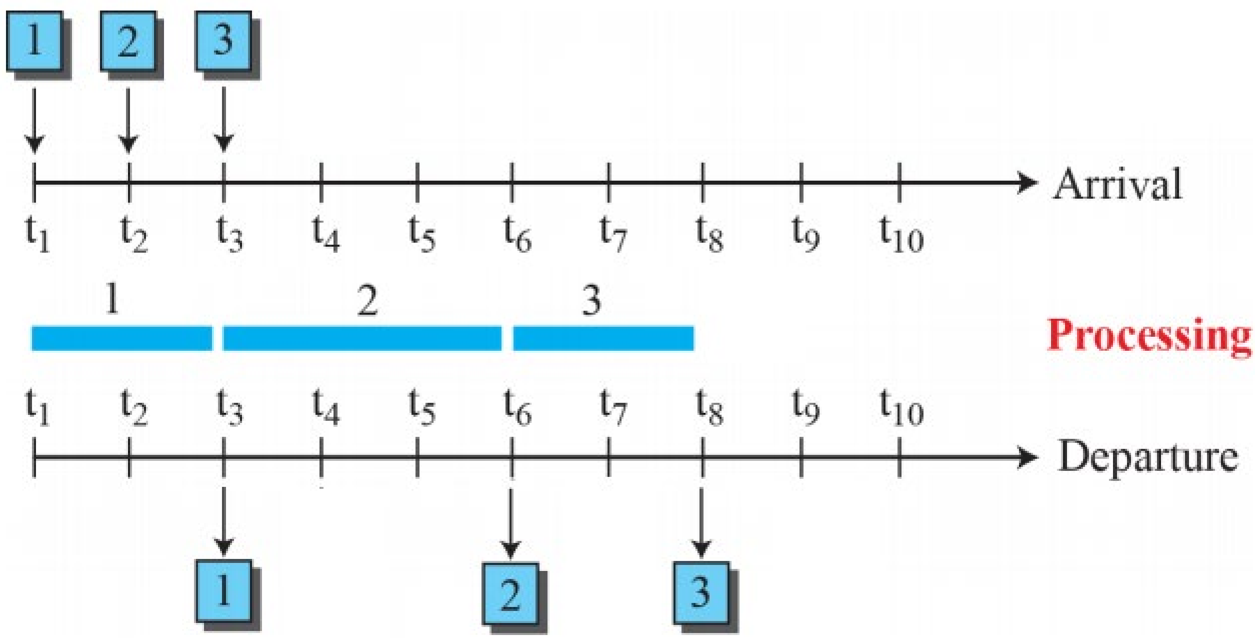
http://expressiveness.org/2012/12/04/midi-jitter

Quality of Service (cont.)

Example [**causes of packet delay & jitter**: **processing** + **queueing delay**]



Required processing time
Packet 1: two time units
Packet 2: three time units
Packet 3: two time units



packet 1:

$d = d_{proc.} = 2$ [tu]

packet 2:

$d = d_{que.} + d_{proc.} =$
$= 1 + 3 = 4$ [tu]

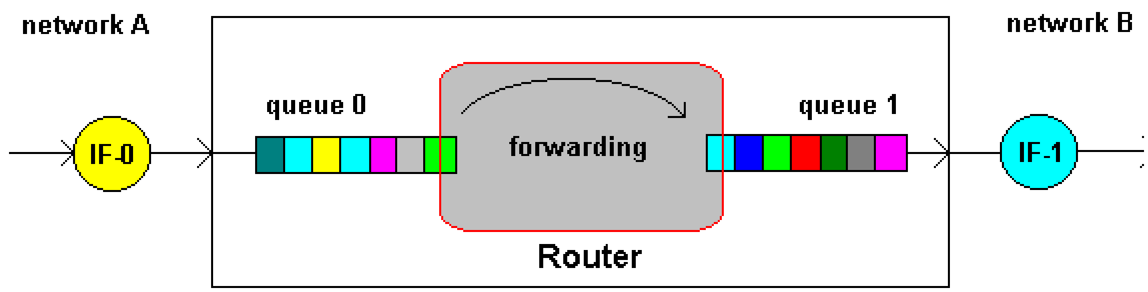
packet 3:

$d = d_{que.} + d_{proc.} =$
$= 3 + 2 = 5$ [tu]

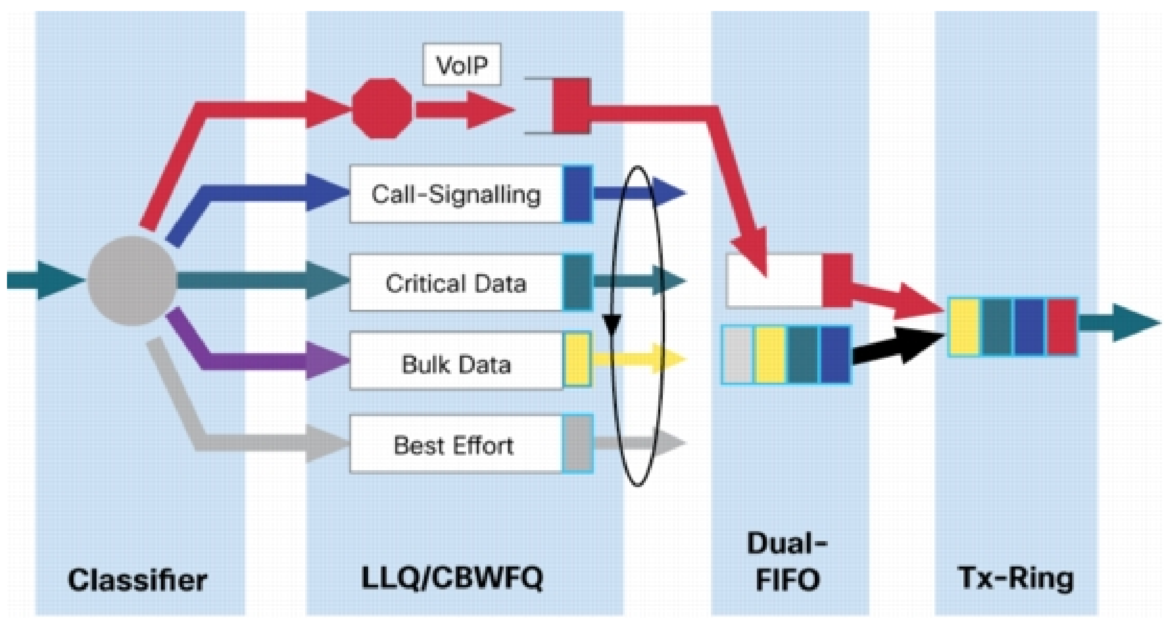
Quality of Service (cont.)

Example **[packet handling at intermediate routers]**

simple packet handing/scheduling
(all packets 'treated' equally, all applications obtain the same QoS)



prioritized packet handing/scheduling
(packets are separated and 'treated' based on their application-type ⇒ different QoS can be obtained)



Packet Scheduling

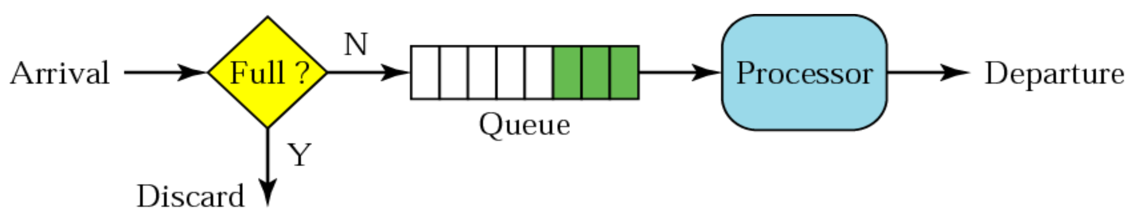
Packet Scheduling to Improve QoS – manner in which queued packets are selected for transmission – has important role in providing QoS

Key element in ensuring priority of multimedia traffic, once packets are properly marked.

- examples:
  (1) FIFO
  (2) Priority Queueing
  (3) Weighted Fair Queueing

FIFO Scheduling – packets are selected for transmission in the same order in which they arrive in the queue

- delays and losses of packets depend on packet interarrival times and packet lengths – no QoS guaranteed!
- packet-discarding policy – used to determine which packets should be dropped if there is not sufficient buffering space
  (a) tail drop: drop last arriving packets
  (b) priority: drop lower-priority packets
  (c) RED – random early detection



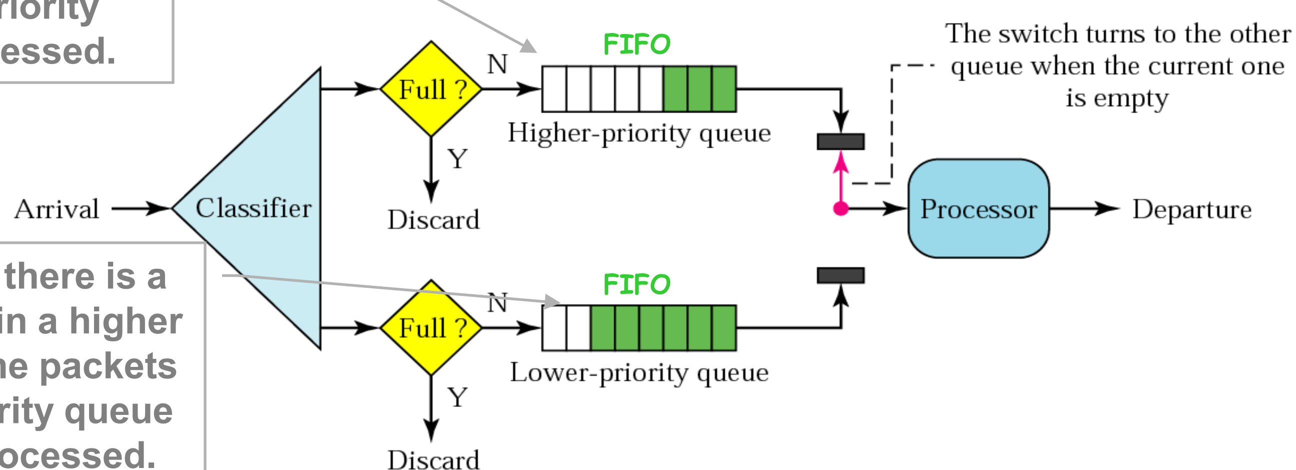
Packet Scheduling (cont.)

Priority Queueing – packets are classified into priority classes – a separate buffer is maintained for each priority class

- assigned priority class may depend on: ToS filed, source or destination IP address, destination port number
- next packet for transmission is selected from the head of the highest priority queue that is not empty – each individual queue is served in a FIFO manner
- advantages: better QoS to high(est) priority traffic than FIFO
- disavantages:
  (a) no QoS guarantees to lower priority traffic (starvation)
  (b) cannot discriminate between users of the same priority – certain users can hog bandwidth with excessive traffic

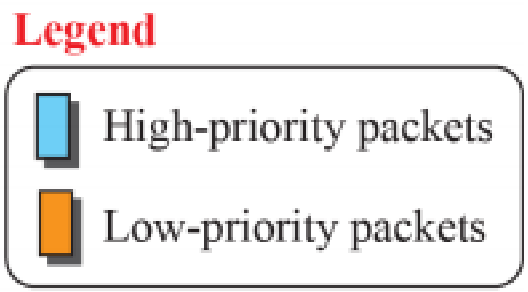
The processor moves to the lower-priority queue only after all packets from higher-priority queue are processed.

STARVATION: if there is a continuous flow in a higher priority queue, the packets in the lower priority queue will never be processed.

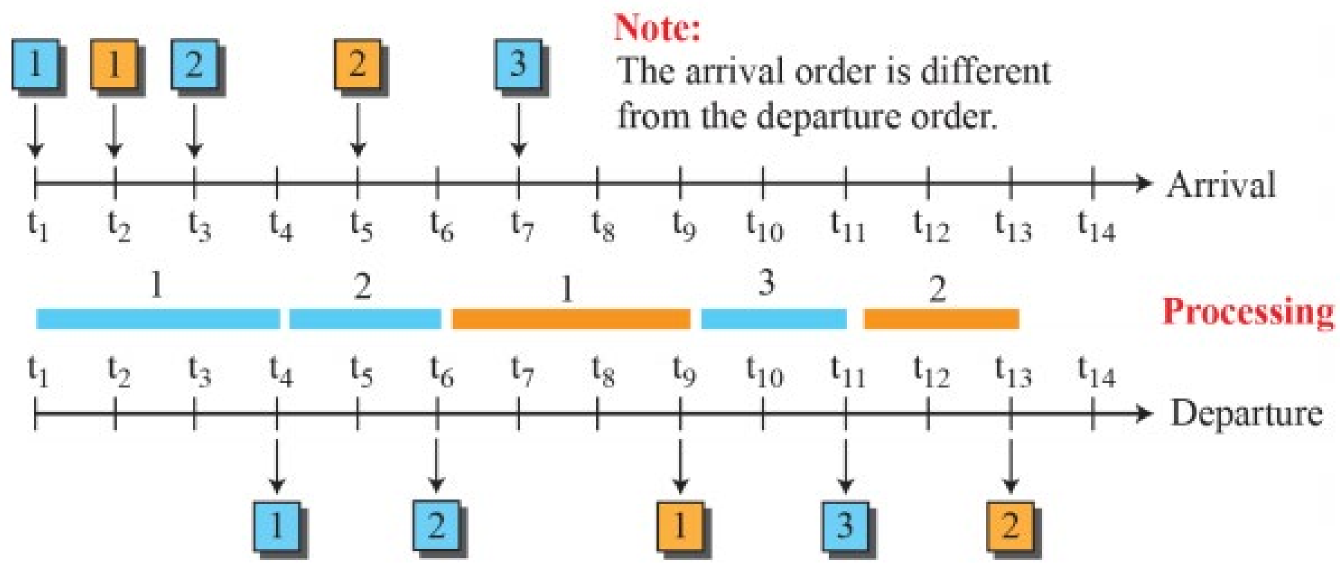


Packet Scheduling (cont.)

Example **[priority queueing]**



required processing time
high-priority P1: 3 time units
high-priority P2: 2 time units
high-priority P3: 2 time units
low-priority P1: 3 time units
low-priority P2: 2 time units



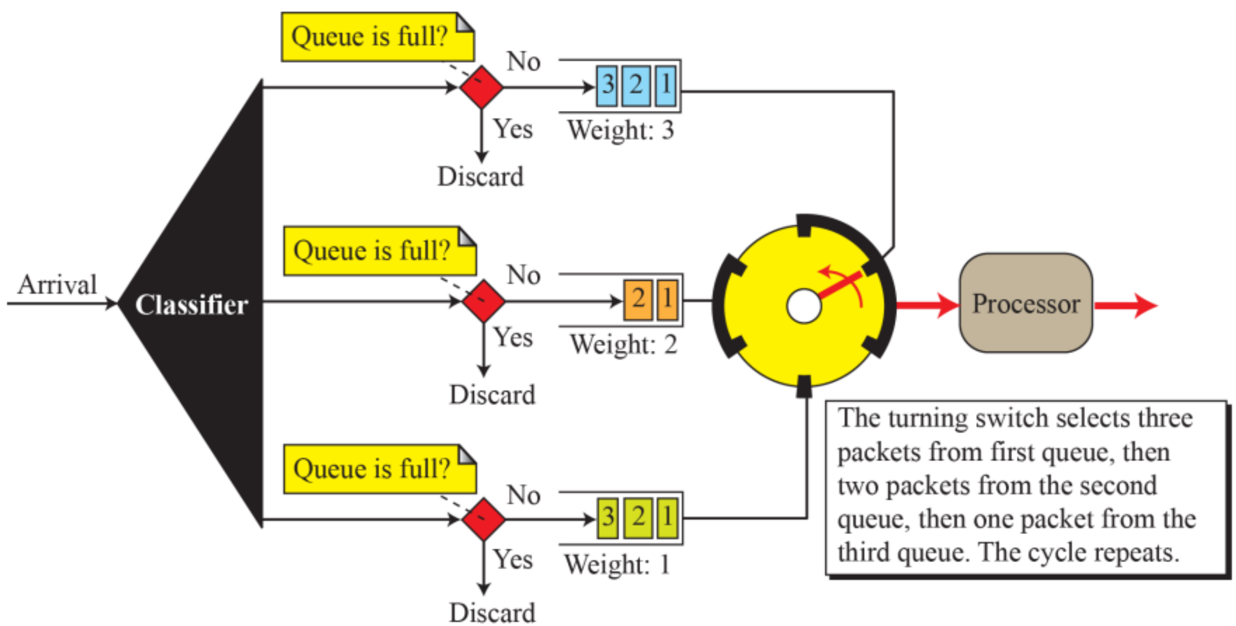
Packet Scheduling (cont.)

Weighted Fair Queueing – packets are classified into priority classes as in priority queueing, however WFQ scheduler serves classes in a circular manner (but not plain round-robin!)

- class(i) receives a weight that determines its relative share of bandwidth (w_i): in any interval of time class(i) is guaranteed to receive <u>at least</u> a fraction of service/bandwidth equal to

$$R(i) = R \cdot \frac{w_i}{\sum w_j}$$

- <u>advantages</u>: QoS guarantees to all priority classes



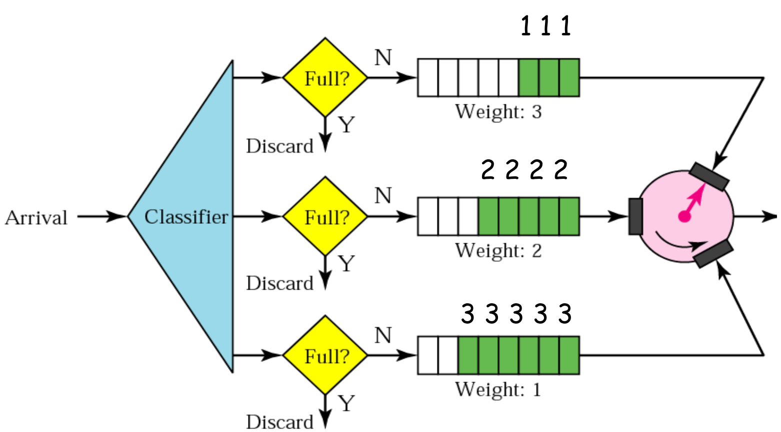
Packet Scheduling (cont.)

Example [WFQ]

Suppose that the WFQ scheduling policy is applied to a buffer that supports 3 classes, and suppose the weights are: 0.5, 0.25, 0.25 for the three classes.

(a) Suppose that each class has a large number of packets in the buffer. In what sequence might the three classes be served in order to achieve the WFQ weights?

(b) Suppose that classes 1 and 2 have a large number of packets in the buffer, and there are no class 3 packets in the buffer. In what sequence might the three classes be served to achieve the WFQ weights?



Solution:

(a) Possibility 1: 1 1 2 3 1 1 2 3 1 1 2 3 …
Possibility 2: 1 2 1 3 1 2 1 3 1 2 1 3 …

(b) The new value of weights for class 1 and class 2 has to satisfy:

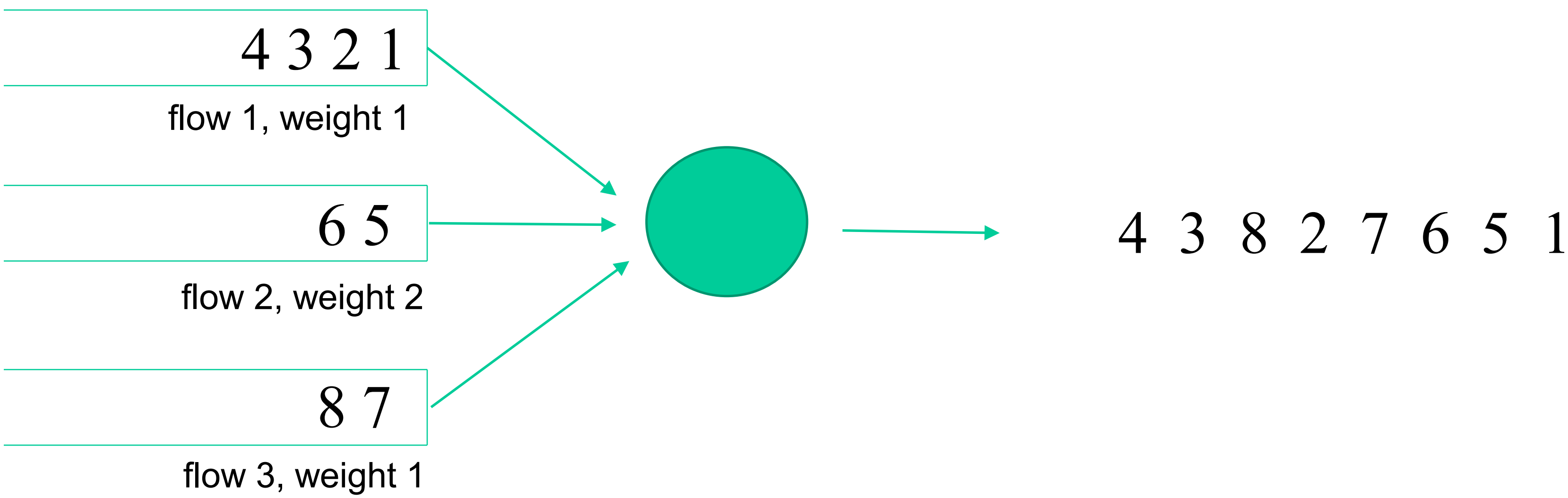
$$\left.\begin{array}{l} w1_{new} = 2^*w2_{new} \\ w1_{new} + w2_{new} = 1 \end{array}\right\} \quad \begin{array}{l} w1_{new} = 2/3 \\ w2_{new} = 1/3 \end{array}$$

Possible solution: 1 1 2 1 1 2 1 1 2

Packet Scheduling (cont.)

Example [WFQ]

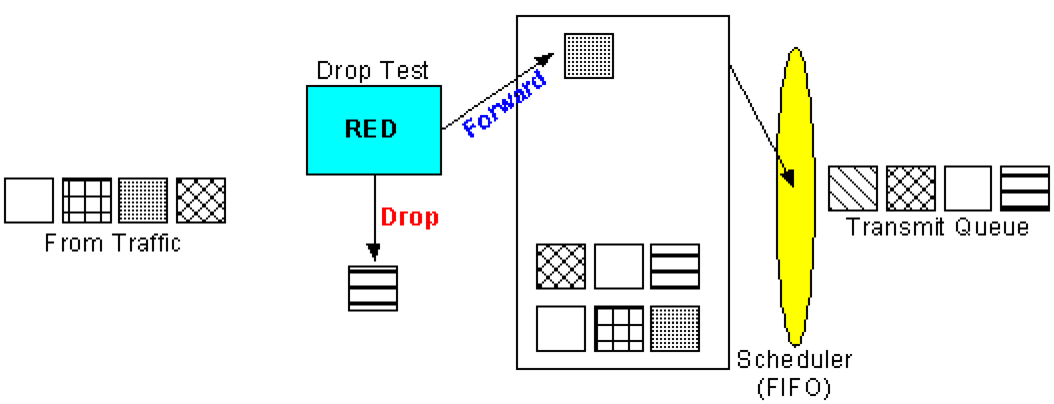
Suppose a router has three input flows and one output. It receives the packets listed in the table below all at about the same time in the order listed during a period in which the output port is busy but all queues are otherwise empty. Give the order in which packets are transmitted, assuming: weighted fair queuing with flow #2 having weight 2, and flows #1,3 with weight 1.

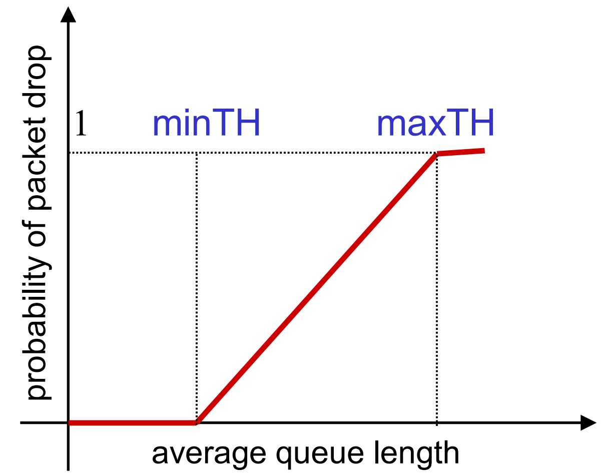


Packet Scheduling (cont.)

Random Early Detection (RED) Algorithm – **proactive packet discard – router discards one or more incoming packets <u>before</u> output buffer gets full ⇒ improved network performance**

- **<u>motivation</u>: TCP connections enter slow start around the same time, and come out of slow start around the same time, causing another cycle of congestion ('global synchronization' phenomenon)**
- **<u>solution</u>: anticipate the onset of congestion and tell one TCP connection at a time to slow down – "apply brakes gradually to gently reduce traffic"**
- **<u>RED idea</u>: packets are randomly discarded with increased probability as the queue size grows**





Packet Scheduling (cont.)

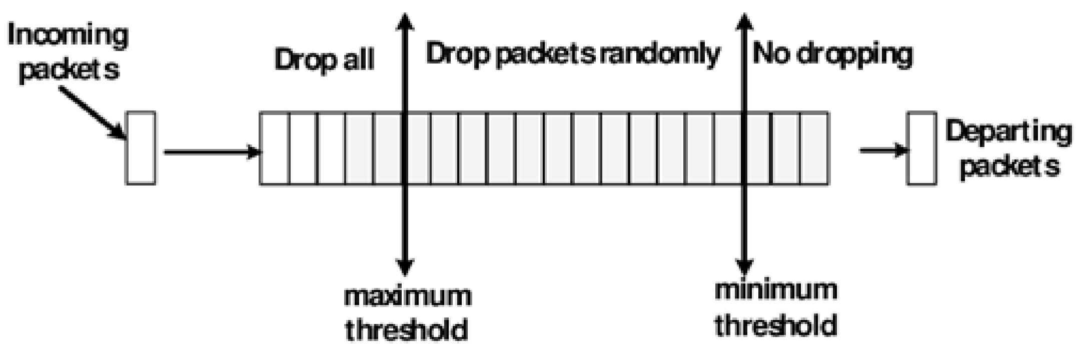
Exponentially weighted average of previous queue lengths.

for each packet arrival
calculate the average queue size avg

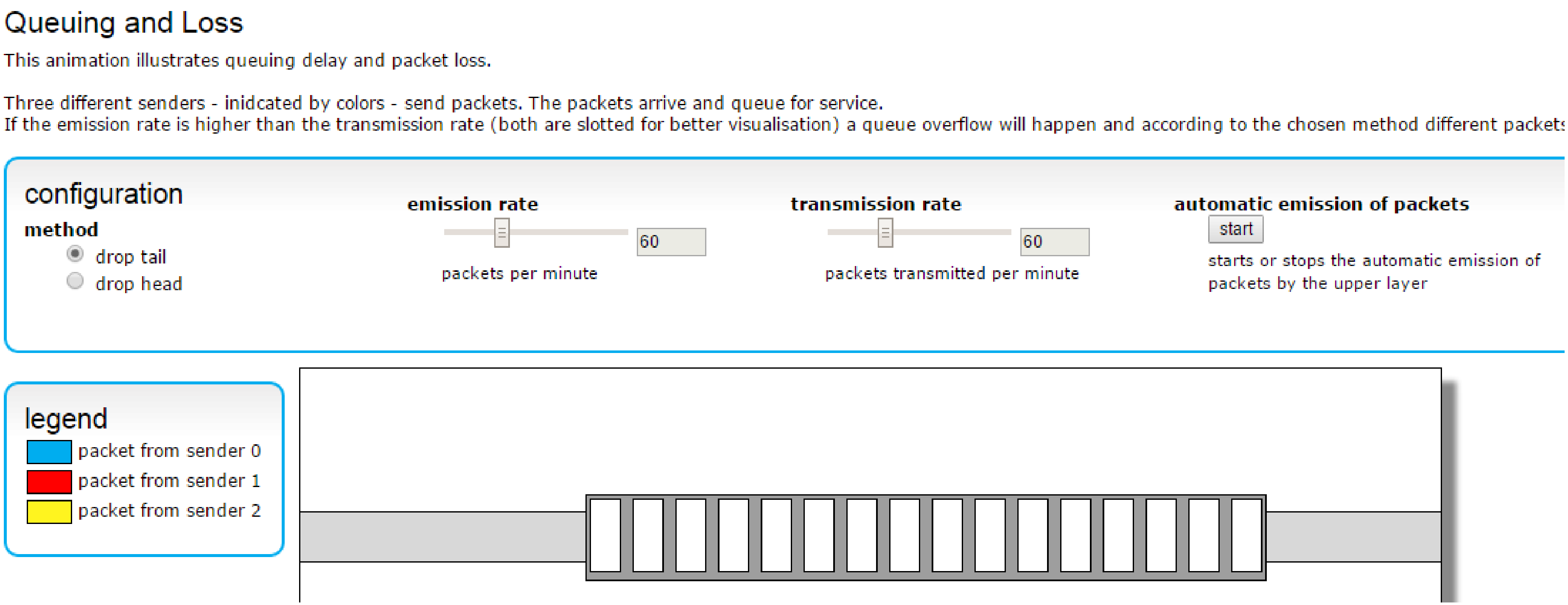
if avg < minTH
    queue packet

else if minTH ≤ avg < maxTH
    calculate probability P_a
    with probability P_a discard packet
    with probability $(1-P_a)$ queue packet

else if avg ≥ maxTH
    discard packet



The method of "early drop" is used to notify the source to reduce its transmission rate before the buffer becomes full.



http://www.ccs-labs.org/teaching/rn/animations/queue/index.htm