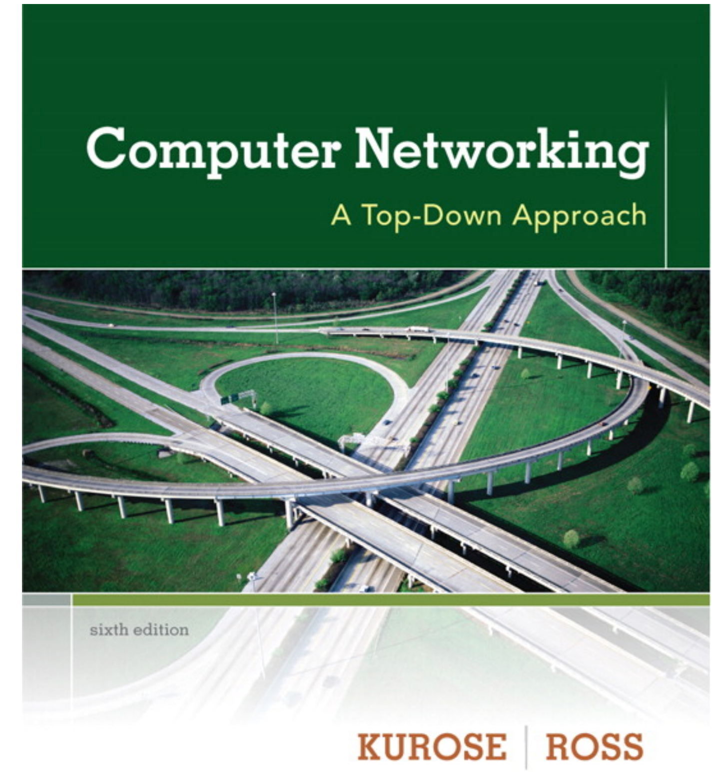
CN-Advanced
L38, L39

Network Support for Multimedia

Dr. Ram P Rustagi
rprustagi@ksit.edu.in
http://www.rprustagi.com
https://www.youtube.com/rprustagi

Acknowledgements

Chapter 7
Multimedia Networking



A note on the use of these ppt slides:
We're making these slides freely available to all (faculty, students, readers). They're in PowerPoint form so you see the animations; and can add, modify, and delete slides (including this one) and slide content to suit your needs. They obviously represent a lot of work on our part. In return for use, we only ask the following:

- If you use these slides (e.g., in a class) that you mention their source (after all, we'd like people to use our book!)
- If you post any slides on a www site, that you note that they are adapted from (or perhaps identical to) our slides, and note our copyright of this material.

Thanks and enjoy! JFK/KWR

Computer Networking: A Top Down Approach
6th edition
Jim Kurose, Keith Ross
Addison-Wesley
March 2012

Multimedia Networking

- Mechanisms used by application to improve performance
  - Pre-fetching and buffering
  - Adapting media quality to available bandwidth
  - Adaptive playout
  - Loss mitigation techniques
- System level approaches to delivery multimedia contents
  - CDNs
  - P2P networking
- Underlying assumption
  - Network (Internet) provides Best effort delivery
- Can network provide a better service?
  - Yes, but at increased complexity.
    - Makes deployment challenges, take off is slow

Network Support for Multimedia

Network Approach	Granularity	Guarantee	Mechanism	Comple xity	Deplo yed?
Making best of best effort service	All traffic treated equally	None or soft	Applicationl evel support, CDNs, overlays	Minimal	Every where
Differenti ated Service (IP ToS)	Traffic class	None or soft	Packet marking, scheduling, policing	Medium	Some
Per connecti on QoS	Per connection Flow	Soft or hard once flow is admitted Bandwidth reserved on each link	Pkt marking, scheduling, policing, call admission, requires resource reservation	High	Little

Dimensioning best effort networks

- Multimedia content delivery requirements
  - low delay, jitter, and pkt loss
- *Approach:* enough link capacity to avoid congestion
  - multimedia traffic flows without delay or loss
  - low complexity of network mechanisms
  - high bandwidth costs
- Challenges:
  - *Network dimensioning:* how much bandwidth ?
  - *Bandwidth provisioning (Estimating traffic demand):*
    - Determine how much bandwidth is "enough"
  - Cost of enough bandwidth?

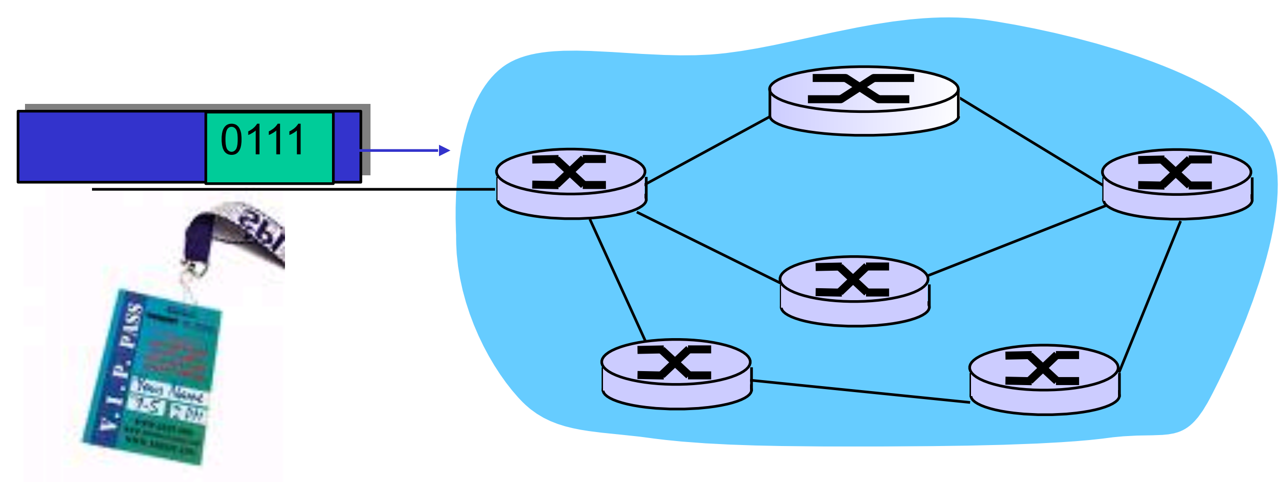
Issues with Best Effort Service

- Issues in predicting application level performance
- *Model of traffic demand between network endpoints*
  - Models at call levels (user arriving, starting applications)
  - Models at packet level (traffic generated at application level)
- *Well defined performance requirements*
  - Probability of end to end delay exceeding tolerable limit
- *Model for minimum cost bandwidth allocation*
  - *Design of optimization techniques for min cost bandwidth*
- *Economic and organization issues*
  - *Links between two end points owned different ISPs*
  - *Native ISP can possibly dimension the network*
    - *How to make other ISPs cooperate*

Providing multiple classes of service

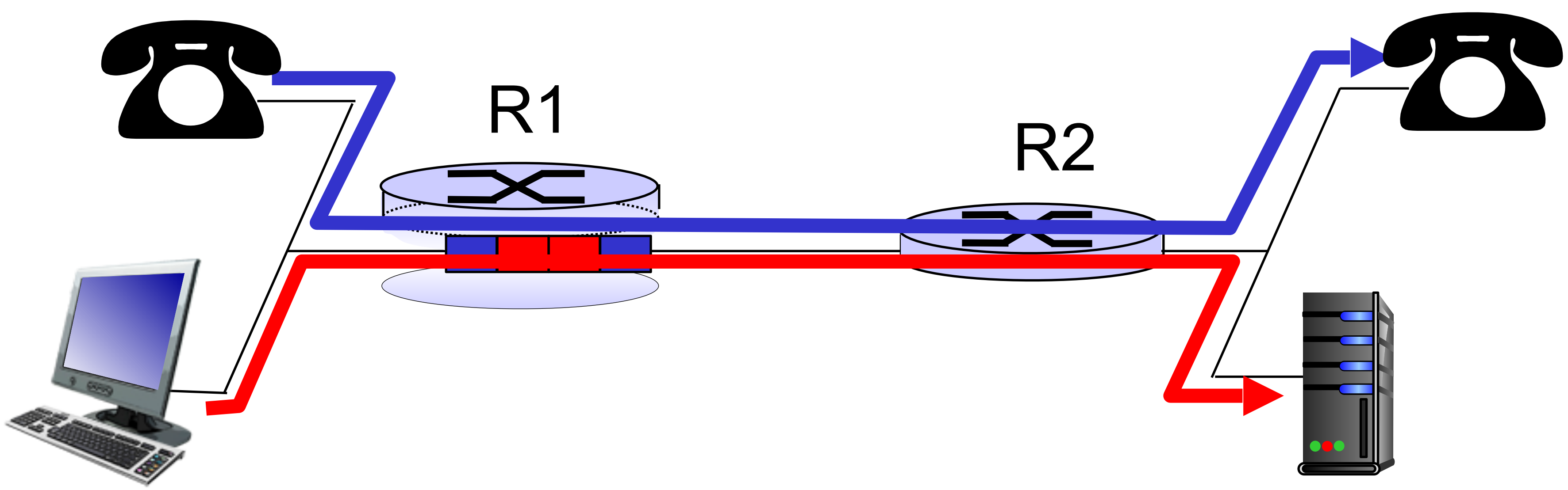
- Till now: making use of the best of best effort service
  - one-size fits all service model
- Alternative: multiple classes of service
  - Partition traffic into classes
  - Network treats different classes of traffic differently (analogy: VIP service versus regular service)
  - All pkts belonging to a class receive same service
    - e.g. all business class passengers get same service

- Granularity: differential service among multiple classes, not among individual connections
- History: ToS bits



Scenario 1: Mixed HTTP and VoIP

- Example: 1Mbps VoIP, HTTP share 1.5 Mbps link.
  - HTTP bursts can congest router, cause audio loss
  - Want to give priority to audio over HTTP
  - Q:How to differentiate audio pkts from HTTP?

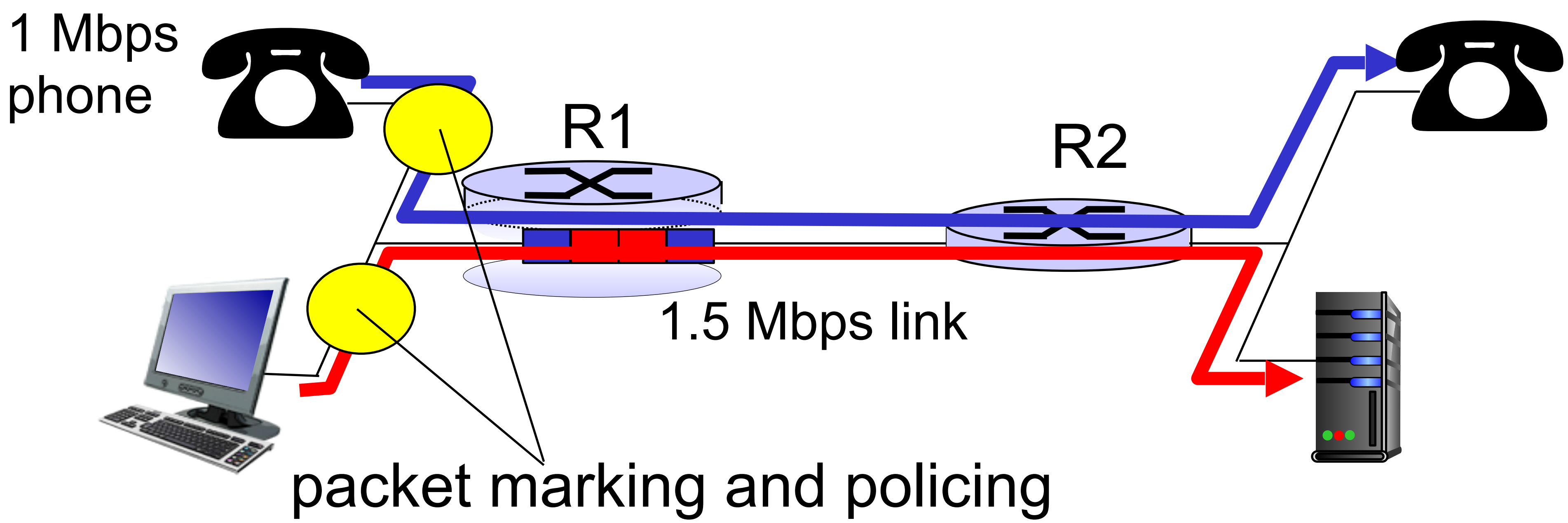


Principle 1

Packet marking needed
- Router distinguish between different classes;
- New router policy to treat packets accordingly

Principles for QOS guarantees (more)

- What if applications misbehave
  - VoIP sends higher than declared rate
  - HTTP packets may starve
- Policing: force source adherence to bandwidth allocations
  - *Marking*, *policing* at network edge
  - Policing: pkt delay, pkt drop (implement leaky bucket)

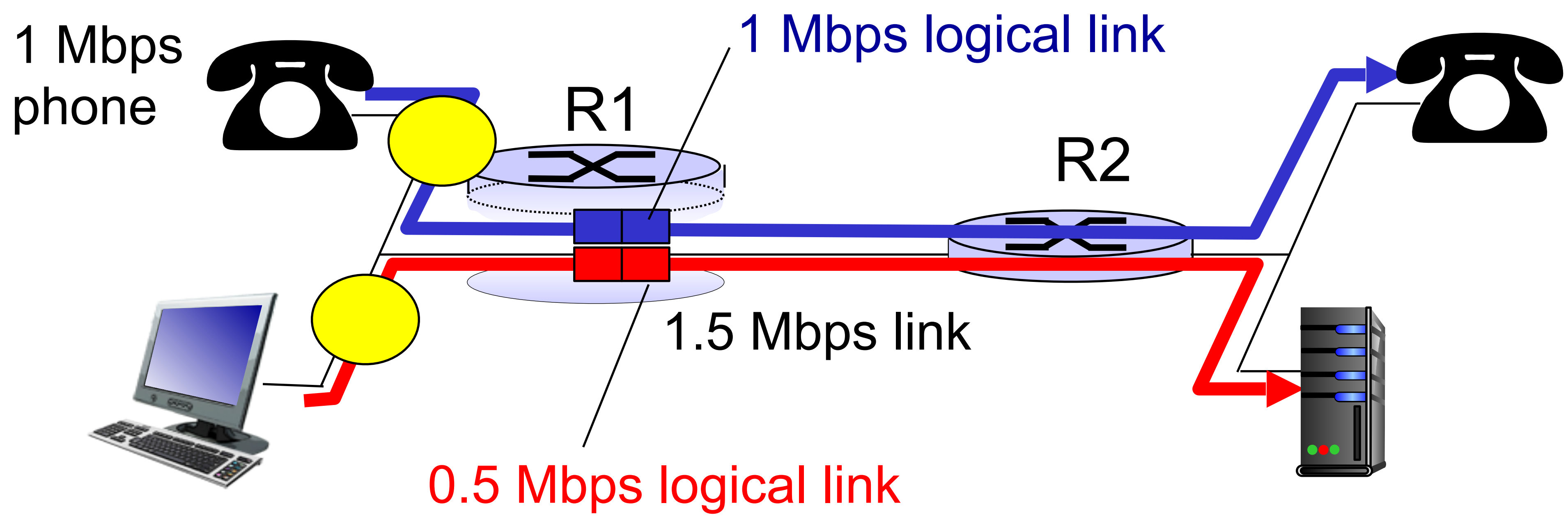


Principle 2

provide traffic isolation for one class from others

Principles for QOS guarantees (more)

- Scheduling policy:
  - Allocating *fixed* (non-sharable) bandwidth to flow:
  - Possible *inefficient* use of bandwidth
    - e.g. when no audio traffic

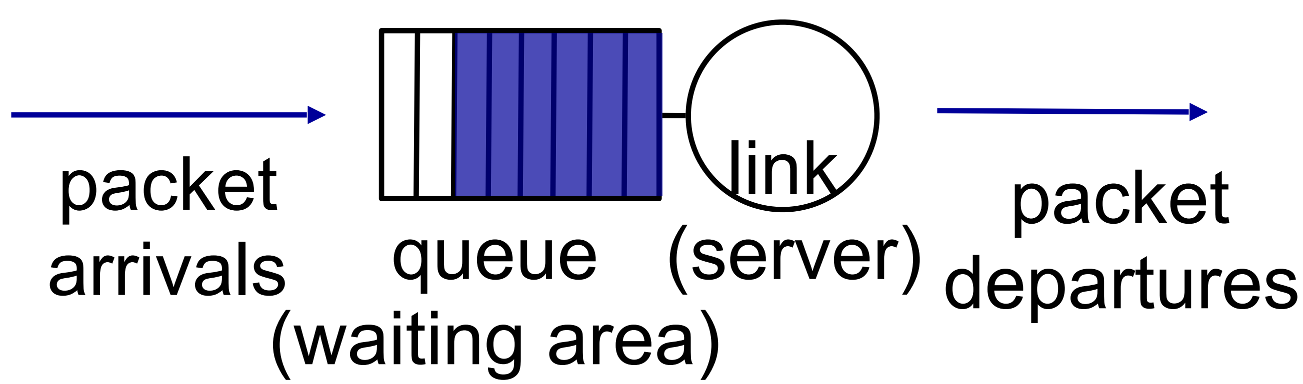


Principle 3

While providing isolation,
use resources as efficiently as possible

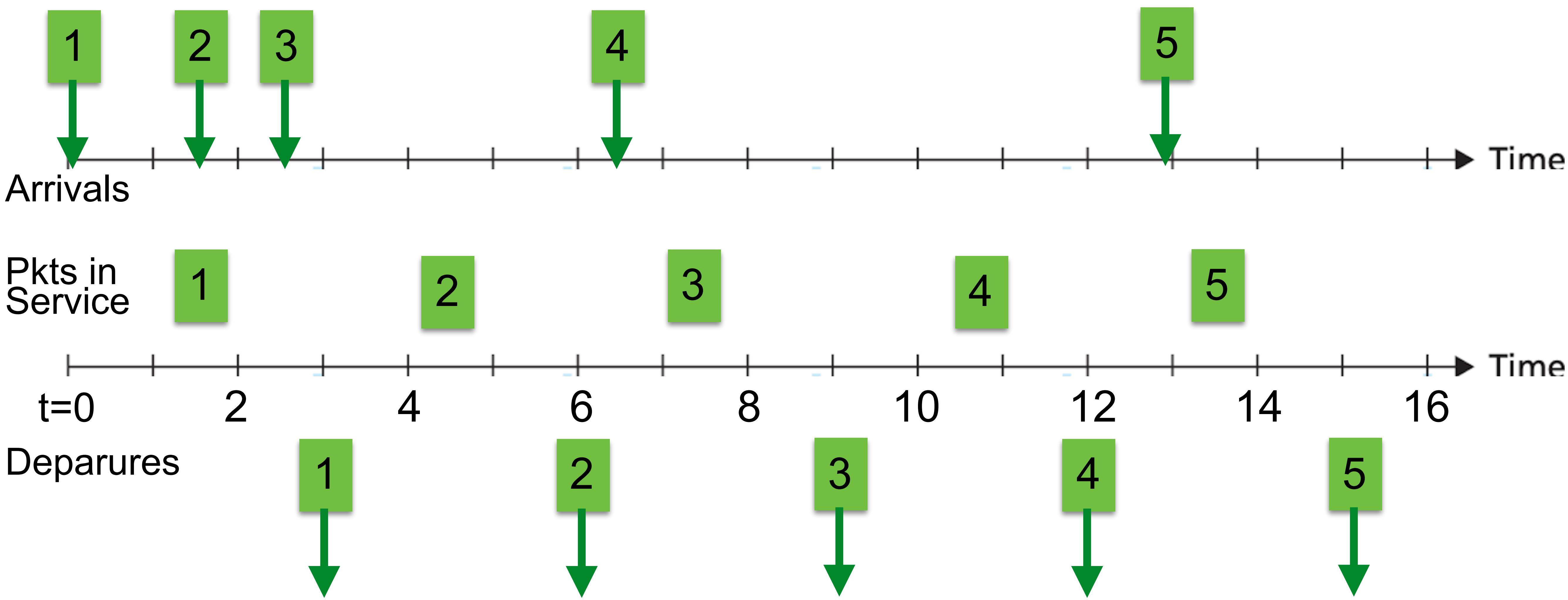
Scheduling and Policing Mechanisms

- *Scheduling:* choose next packet to send on link
- *FIFO scheduling (or FCFS):*
  - Real-world example?
  - *Discard policy:* if packet arrives to full queue: who to discard?
    - *Tail drop:* drop arriving packet
    - *Priority:* drop/remove on priority basis
    - *Random:* drop/remove randomly



FIFO Scheduling

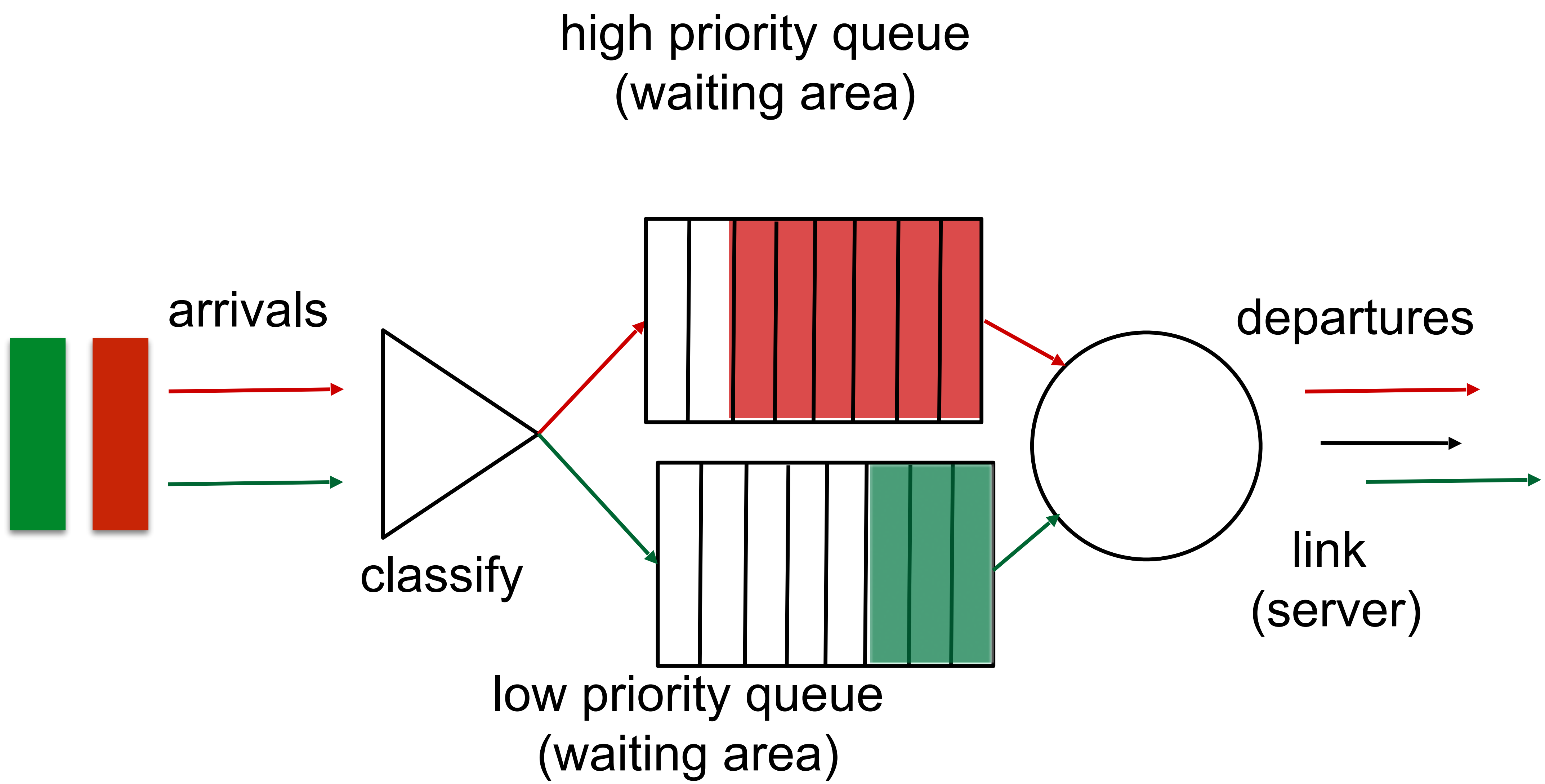
- FIFO Queue in operation



Scheduling policies: priority

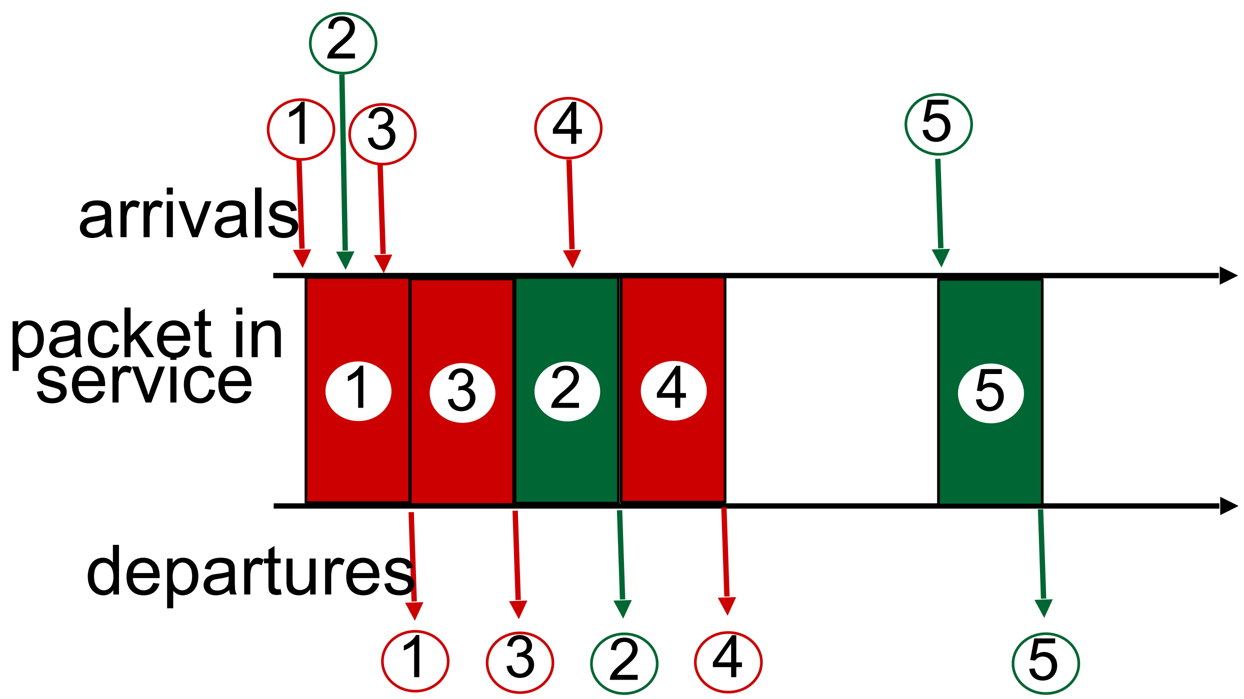
- *priority scheduling:*
  - At output links, packets are prioritized
  - Packets priority class may depents upon
    - packet marking in packet header, or
    - source and/or destination address
  - Each priority class has its own queue
    - Multiple *classes*, with different priorities
  - *H*ighest priority queued packet is sent first
- Real world example?

Scheduling policies: priority



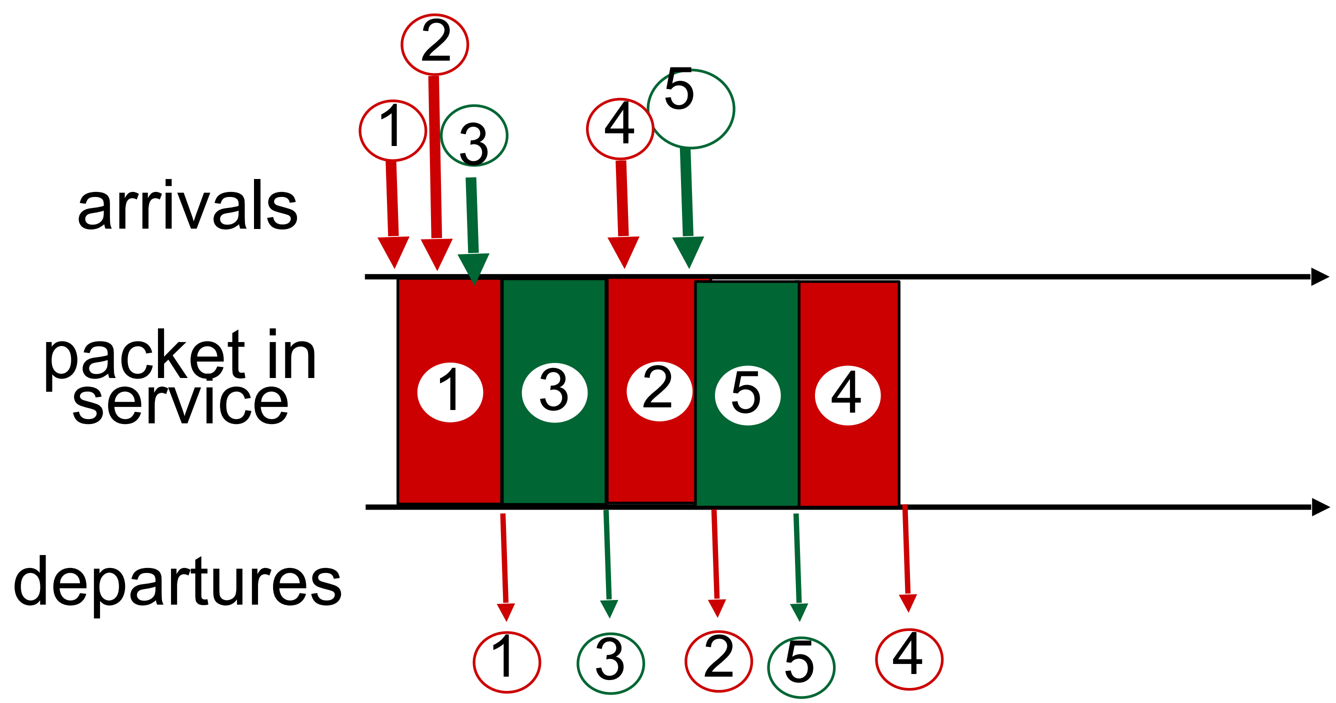
Scheduling policies: priority

- Priority Queue with 2 priority classes



Scheduling policies: still more

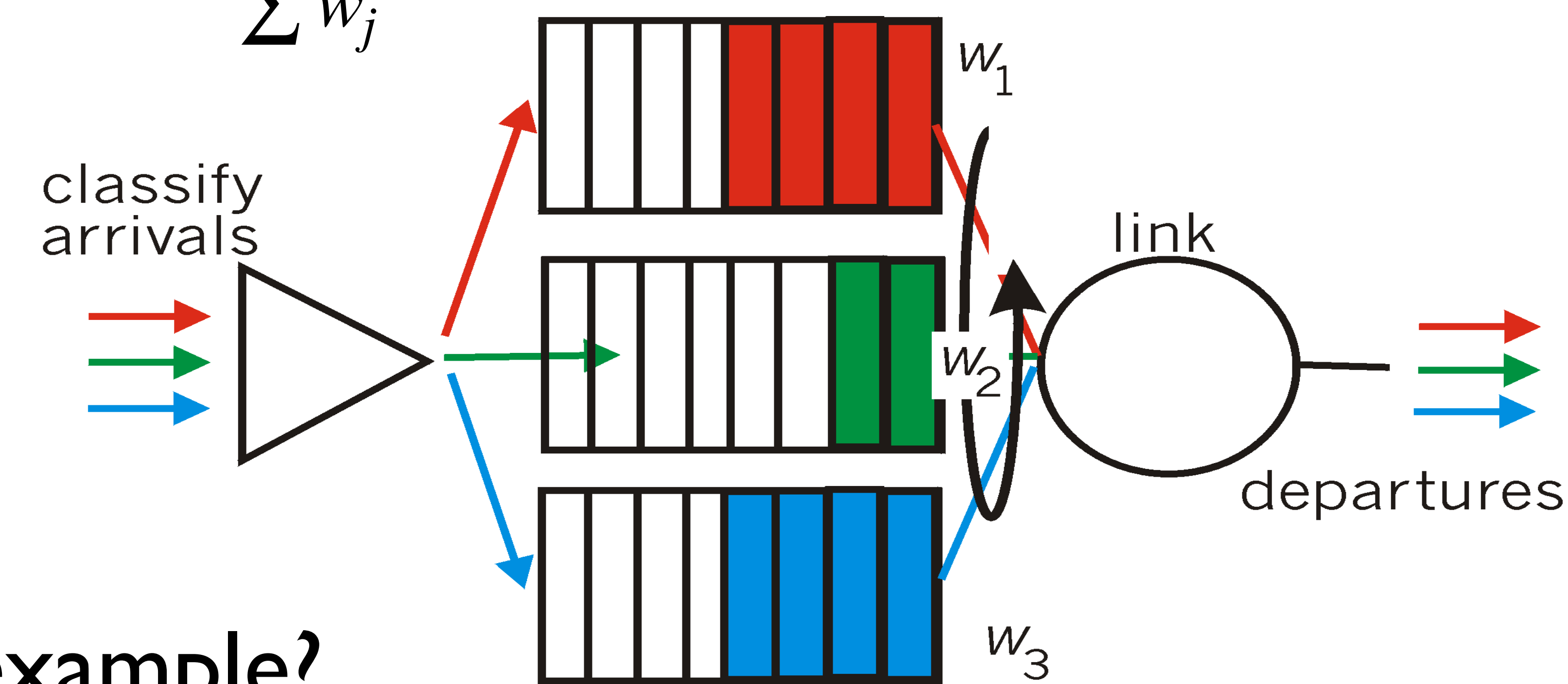
- *Round Robin (RR) scheduling:*
- Multiple classes
- Cyclically scan class queues, sending one complete packet from each class (if available)
- Real world example?



- Work conserving round robin
  - if no packet in current class, then take from next class

Scheduling policies: still more

- *Weighted Fair Queuing (WFQ):*
- Generalized form of Round Robin
- Each class gets weighted amount of service in each cycle
  - Consider class `i` is assigned weight w_i, then
  - Class `i` gets service for a fraction of $\frac{w_i}{\sum w_j}$
- If link bandwidth is R, then class i is guaranteed to get bandwidth of $R * \frac{w_i}{\sum w_j}$



- Real-world example?

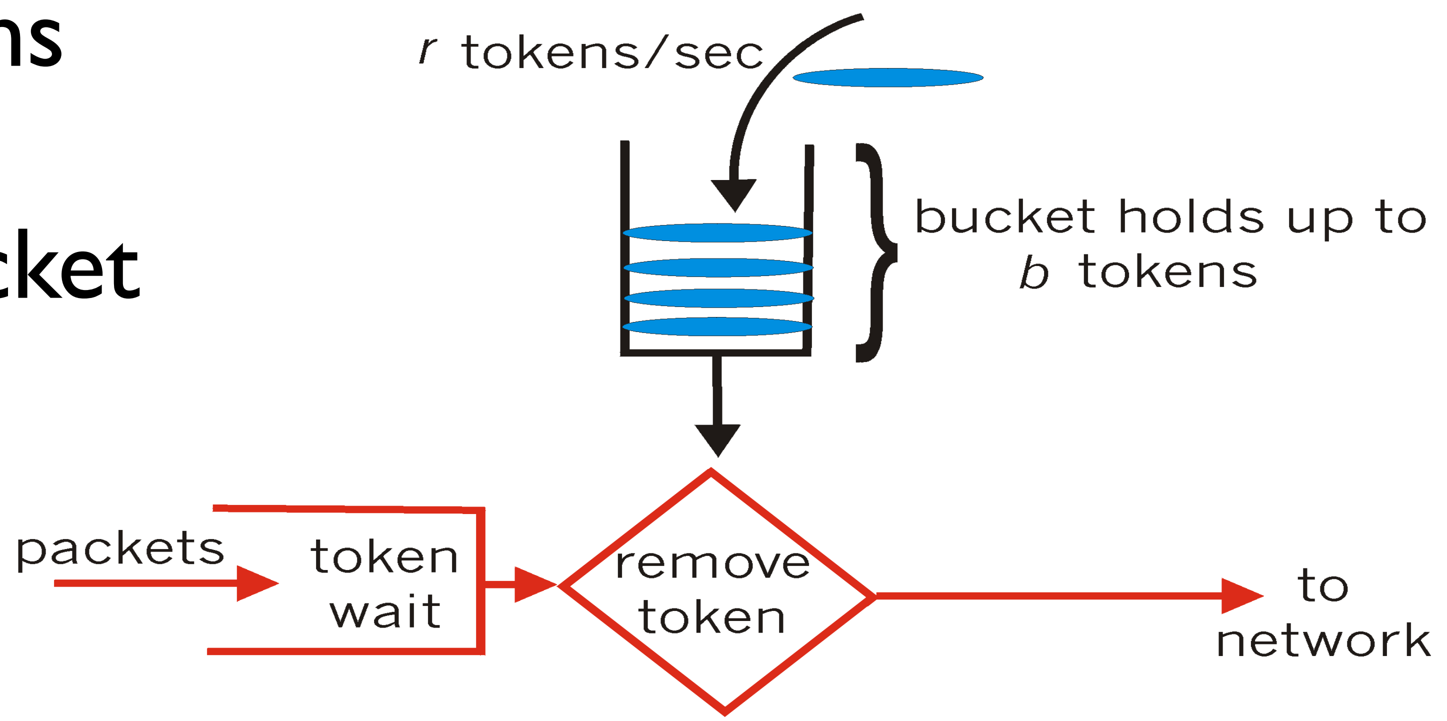
Policing mechanisms

- *Goal:* limit traffic to not exceed declared parameters
- Three common-used criteria:
- *(long term) average rate:* how many pkts can be sent per unit time (in the long run)
  - Crucial question: what is the interval length:
  - e.g. 100 packets per sec or 6000 packets per min have same average!
- *Peak rate:* Average rate of 6000 pkts per min
  - with some peak rate e.g. 1500 pkts in a second
  - peak rate of 100 pkts/sec wouldn't allow 1500pkts/sec
- *(max.) burst size:* max number of pkts sent consecutively
  - in a short period (with no intervening idle)

Policing mechanisms: implementation

Leaky (or token) bucket: limit input to specified *burst size* and *average rate*

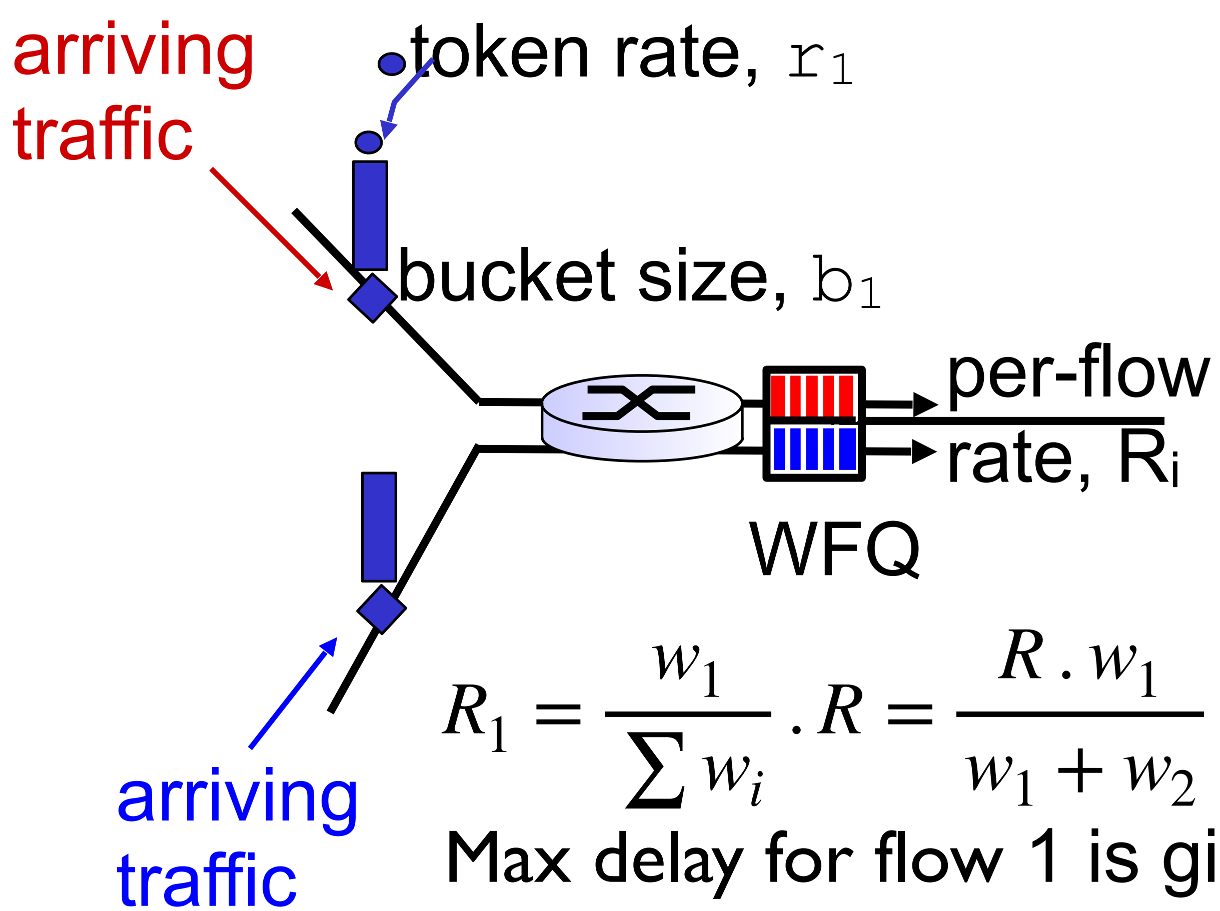
- Bucket can hold b tokens
- Tokens generation rate
  - r *token/sec* unless bucket full

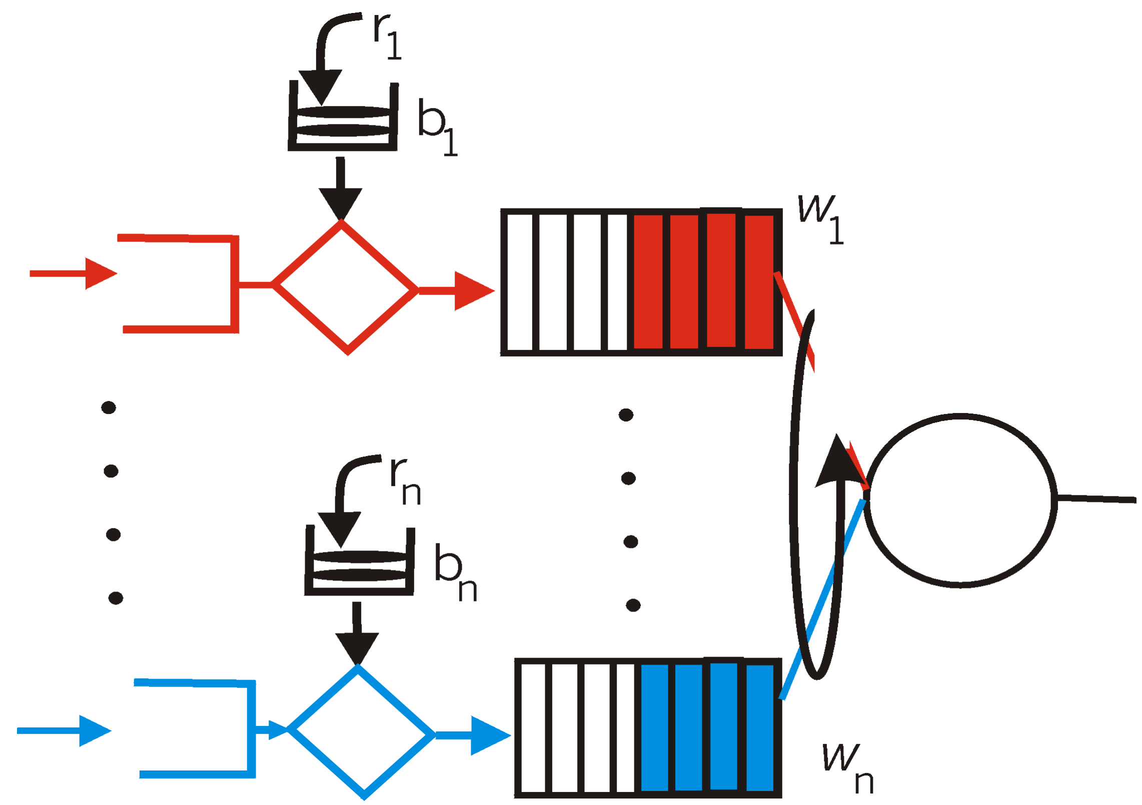


- *Over interval of length t:*
  - *number of packets admitted <= (r.t + b)*
- Token generation rate determines long term average
- Bucket size b limits the burst rate
- Two leaky buckets in series implements flow's peak rate

Policing and QoS guarantees

- Token bucket, WFQ combine to provide guaranteed upper bound on delay, i.e., *QoS guarantee!*
- Consider router's output link multiplexes N queues
  - Each policed by bucket size b_i and token rate r_i
  - with WFQ weight of w_i

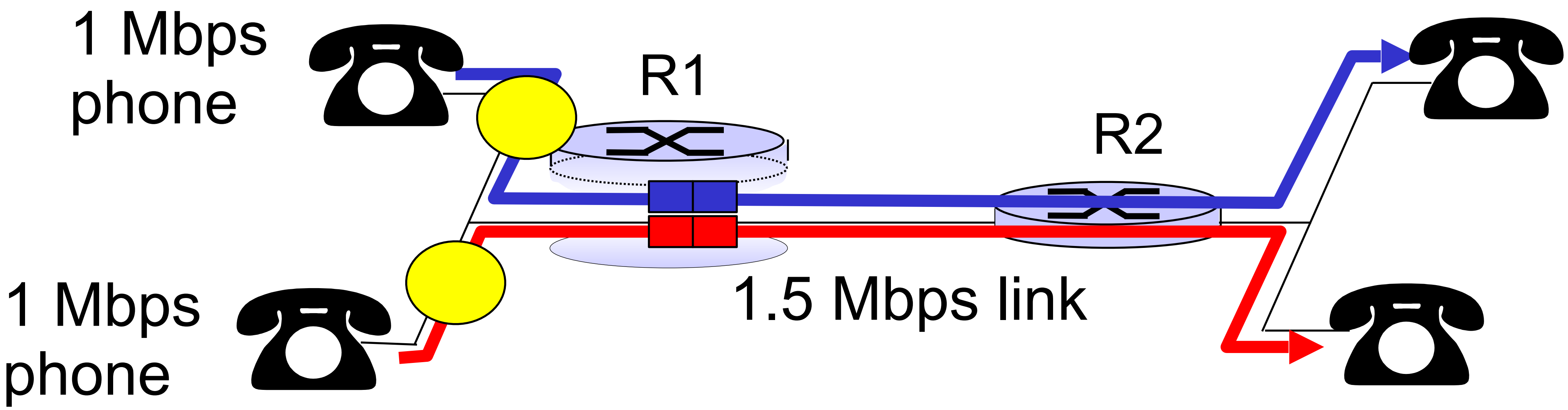




$$R_1 = \frac{w_1}{\sum w_i} . R = \frac{R . w_1}{w_1 + w_2}$$

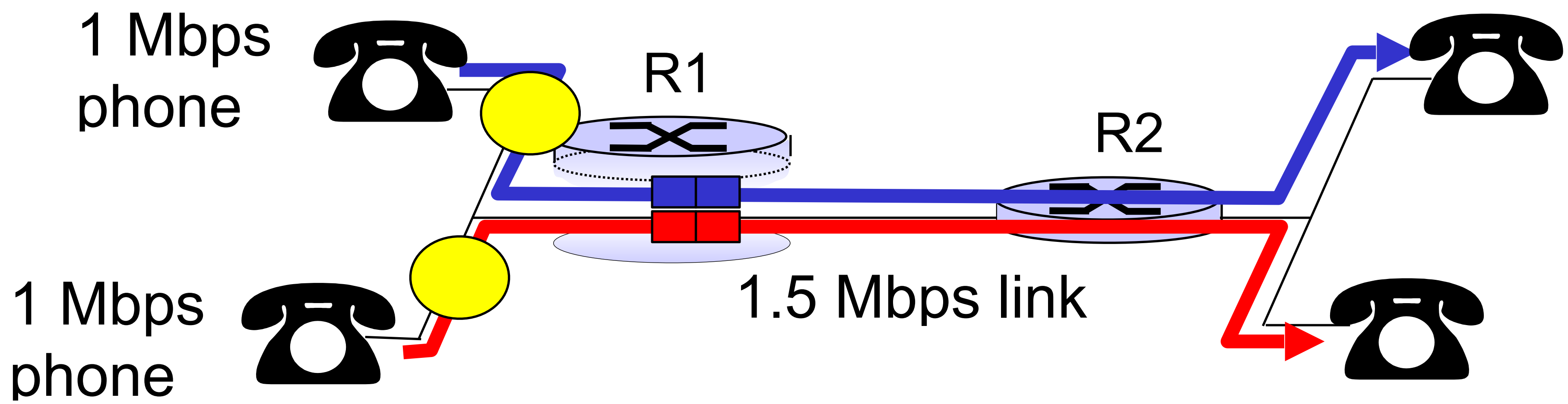
Max delay for flow 1 is given by $D_1 = b_1 / R_1$
$= b_1 / (R . w_1 / \Sigma w_i)$

Per-connection QoS guarantees



- Conside two VoIP calls, where each needs 1 Mbps
- The bandwidth of common link is 1.5 Mbps
- What happens with packet marking and policing work?
  - Each call loses 25% of packet, poor quality for both
  - neither one gets a good service
- Solution: block one of the two calls.
- *Basic fact of life:* can not support traffic demands beyond link capacity

Per-connection QoS guarantees



Principle 4

call admission: flow declares its needs, network may block call (e.g., busy signal) if it cannot meet needs
Real life example: Telephone network

- Result of admitting a flow without required QoS support?
  - Network resources are used
  - Does not provide useful activity to user
  - Possible impact other users as well.

Per-connection QoS guarantees

- How to admit a flow with required QoS support?
  - Requires **Call Admission Process**
  - If call can' be admitted, it is blocked (or delayed)
  - Example: Busy signal in a telephone network
- Network mechanisms for call admission process
  - Resourse reservation
  - Call Admision
  - Call setup signaling

Network Mechanism for Qos Support

- Resource Reservation
  - A process of allocating(resources) to the call.
    - Resurces: link bandwidth, buffers, (compute power?)
  - Guarantees resources to meet desired QoS for a call
  - Call is guaranteed availability of reserved resources
    - Throughout the entire duration of the call
  - If `x` Mbps is reserved,
    - Call is guaranteed to get up to `x` Mbps.
    - Call will see delay free and loss free performance
  - What happens if call try to use >`x` Mbps

Network Mechanism for Qos Support

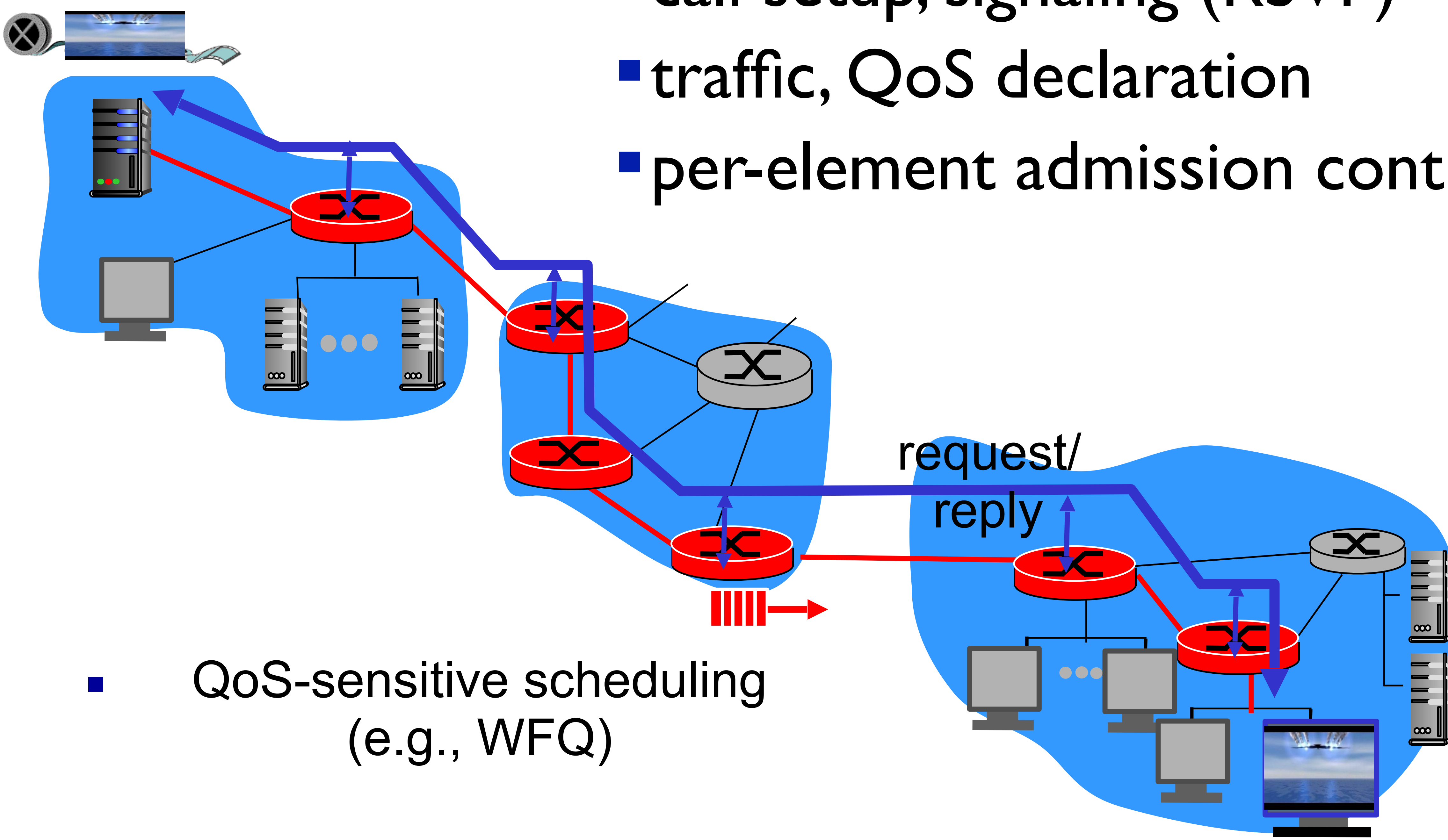
- Call admission
  - How does network know what resources a call needs?
  - A call has to request and reserve resources
    - Example: Dialing a number in telephone network
    - call is blocked (busy signal) when circuits unavailable
  - A Call sends traffic only after call admission is complete
  - Router (in the path) must ensure during call admission
    - Should not allocate more bandwidth than available
    - Typically, a call reserve only a fraction of link'w BW
    - Router allocates link bandwidth to multiple calls.
  - Hard service guarantee
    - sum of allocated BW <= link capacity

Network Mechanism for Qos Support

- Call setup signaling
  - End to end QoS: each router needs to reserve resources
  - Each router determines availability of local resources
    - After accouting for resources already reserved
- Need of a signaliing protocol (or call setup protocol)
  - Per hop allocation of local resources
  - Overall end to end decision
    - call has been (un-) able to reserve resources
- Solution: RSVP protocol (RFC 2210)
  - Resource Reservation Protocol
  - ATM Networks: Q2931b protocol

QoS guarantee scenario

- *resource reservation*
  - call setup, signaling (RSVP)
  - traffic, QoS declaration
  - per-element admission control



- QoS-sensitive scheduling (e.g., WFQ)

Status: Per Connection QoS Support

- Very limited (or little) deployment
- Possible reasons:
  - Simple application level mechanisms with network dimensioning provides good enough service
  - High complexity of managing a network in per conn. QoS
  - High cost makes it infeasible for ISP
    - Revenues from customers doesn’t justify the cost

Summary

- Best effort networks
- Multiple classes of service
- Scheduling policies
  - Round Robin
  - WFQ
- Policy mechanisms
  - Leaky bucket implementation
- Per Connection QoS
  - Resource reservation
  - Call admision
  - Call setup

VTU Questions (2015 scheme)

- July 2019 - Q9
  - Bring out leaky bucket mechanism for traffic policing (7 marks)
  - Classify the multimedia network applications (3 marks)
  - Describe the link scheduling mechanisms (6 marks)
- July 2019 - Q10
  - List the categories of streaming stored video. Explain any one of them (8 marks)
  - Explain the working of CDN (8 marks)
- Jan 2019 - Q9
  - With diagram, explain naive architecture for audio/video streaming (8)
  - Explain audio compression in internet (8 marks)
- Jan 2019 - Q10
  - With diagram, explain interaction between client and server using RTSP (8 marks)
  - Explain how streaming from streaming server to media player is done.

VTU Questions (2015 scheme)

- July 2018 - Q9
  - Brief out 3 broad categories of multimedia network applications (8)
  - Discuss the followings: (8 marks)

    (i)Adaptive streaming (ii) DASH
- July 2018 - Q10
  - With general format, explain the various fields of RTP (8 marks)
  - Explain the working procedure of leaky bucket algorithm (8 marks)
- Jan 2018 - Q9
  - Elaborate the features of streaming stored video (3 marks)
  - With a neat diagram, explain CDN operation (8 marks)
  - Summarize the limitation of best effort IP service (5 marks)
- Jan 2018 - Q10
  - Explain the diffserve internet architecture (5 marks)
  - Describe the leaky bucket policing mechanism (6 marks)
  - Discuss the round-robin and WFQ scheduling mechanism (5 marks)