Lesson 9: Resource allocation

Van-Linh Nguyen

Fall 2024

國立中正大學
Cybersecurity Lab

Outline

- Resource allocation
    1. Network resource resolution
    2. Queuing Disciplines

Let us play a game!

- We have a bridge with max. load of 40kg

- A group of 7 pigs: 15kg, 25kg, 5kg, 10kg, 25kg, 35kg, 15kg
  - ✓ S: 35, 25, 25, 15, 15, 10, 5
  - ✓ 35-A1:

    For loop for all set S and check A1 + i = 40; stop, get i

    remove i from the the the set S for checking;

    35+ 5:

    - 25, 25, 15, 15, 10

    25 + 15:

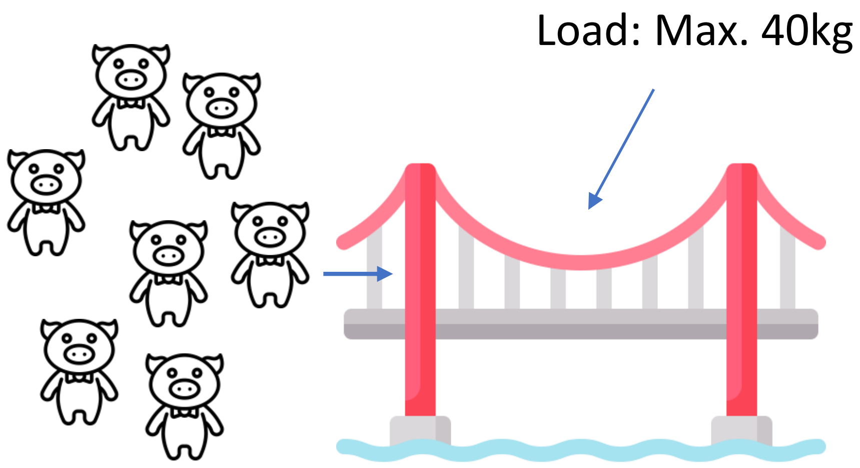
    - 25, 15, 10

    25+15

    - 10:

  → Four rounds + 2mins to be done: 35+5, 25+15, 25+15, 10
- One round to go through takes 30 seconds

- Find a way to allow the pigs to go through the bridge in the fastest time without breaking the bridge!

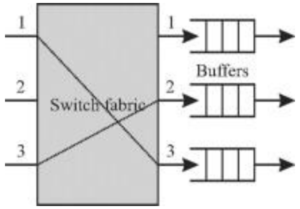


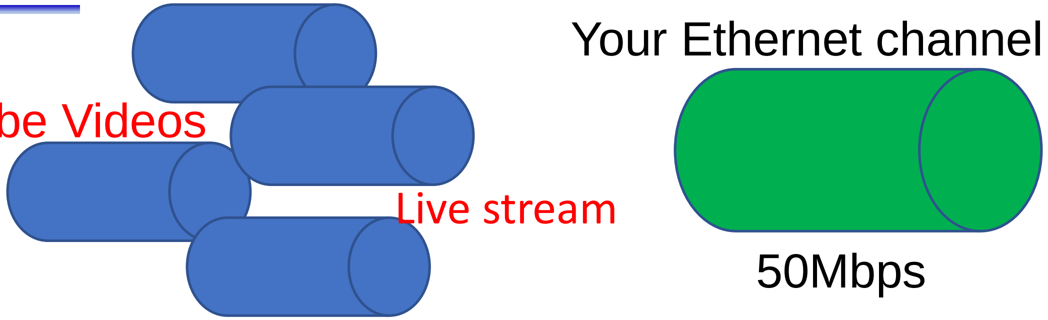
國立中正大學 Cybersecurity Lab

Ask a question

Congestion Control and Resource Allocation

- Resources
  - Bandwidth of the links
  - Buffers at the routers and switches

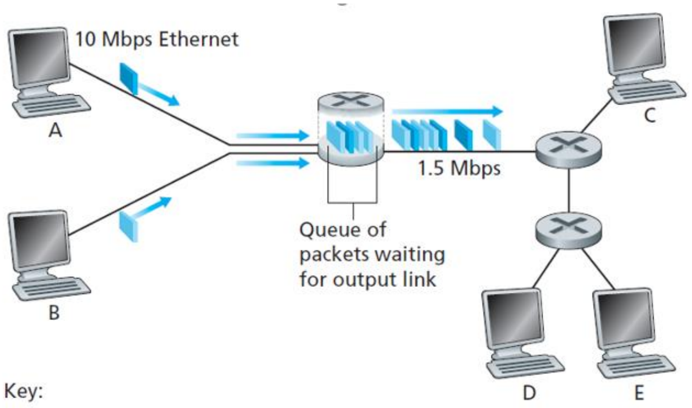


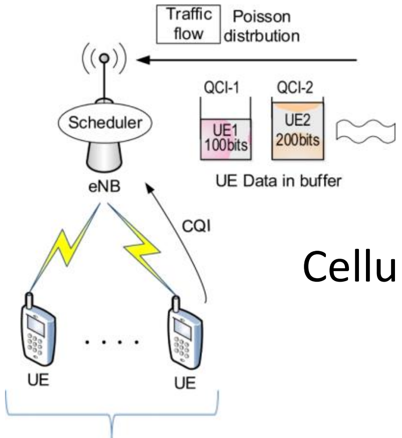


How can the requests use the limited resource channel?

- Packets contend at a router for the use of a link, with each contending packet placed in a queue waiting for its turn to be transmitted over the link

Ethernet





Example of where to require resource allocation

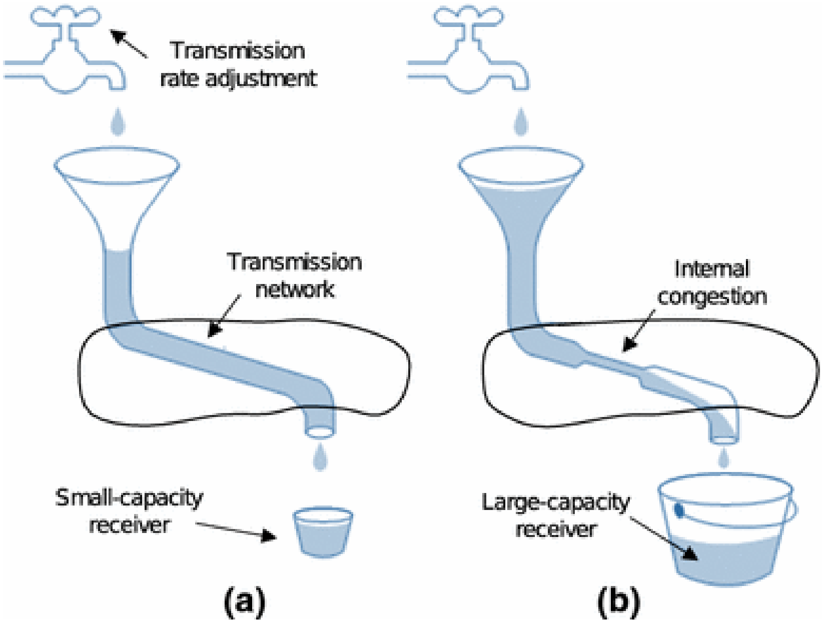
Cellular networks

國立中正大學 Cybersecurity Lab

Congestion Control and Resource Allocation

Ask a question

- When too many packets are contending for the same link
  - The queue overflows
  - Packets get dropped
    - Network is congested!
- Network should provide a congestion control mechanism to deal with such a situation



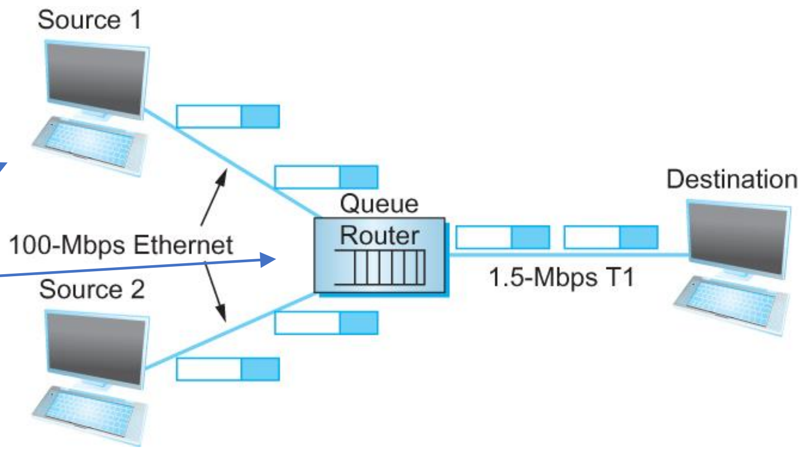
Ask a question

Congestion Control and Resource Allocation

- Congestion control and Resource Allocation
  - Two sides of the same coin

- If the network takes active role in allocating resources
  - The congestion may be avoided
  - No need for congestion control

Control the sending traffic
and buffer at the router



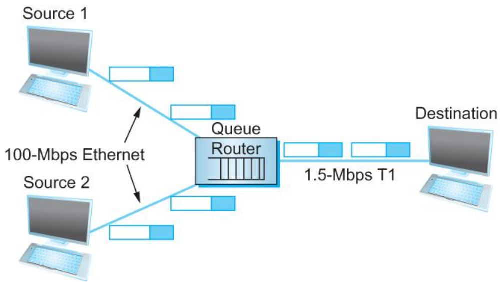
Congestion Control and Resource Allocation

- Allocating resources with any precision is difficult
  - Resources are distributed throughout the network

- We can always let the sources send as much data as they want
  - Then recover from the congestion when it occurs
  - Easier approach but it can be disruptive because many packets many be discarded by the network before congestions can be controlled

Congestion Control and Resource Allocation

- Congestion control and resource allocations involve both hosts and network elements such as routers

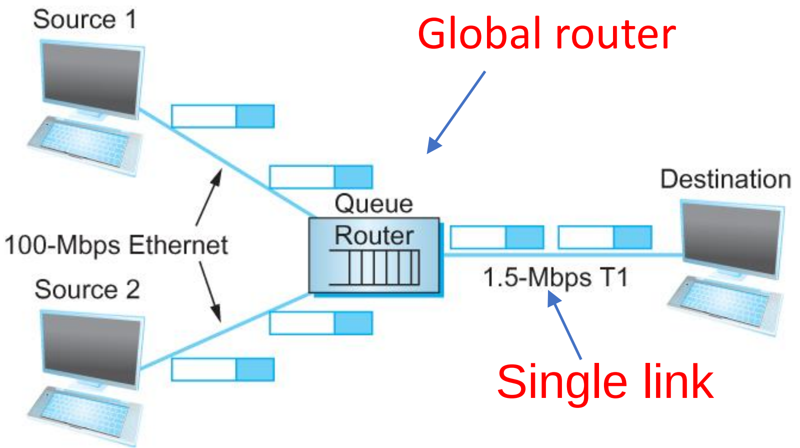


- **In network elements**
  - Various queuing disciplines can be used to control the order in which packets get transmitted and which packets get dropped
- **At the hosts' end**
  - The congestion control mechanism paces how fast sources are allowed to send packets

Ask a question

Resource Allocation

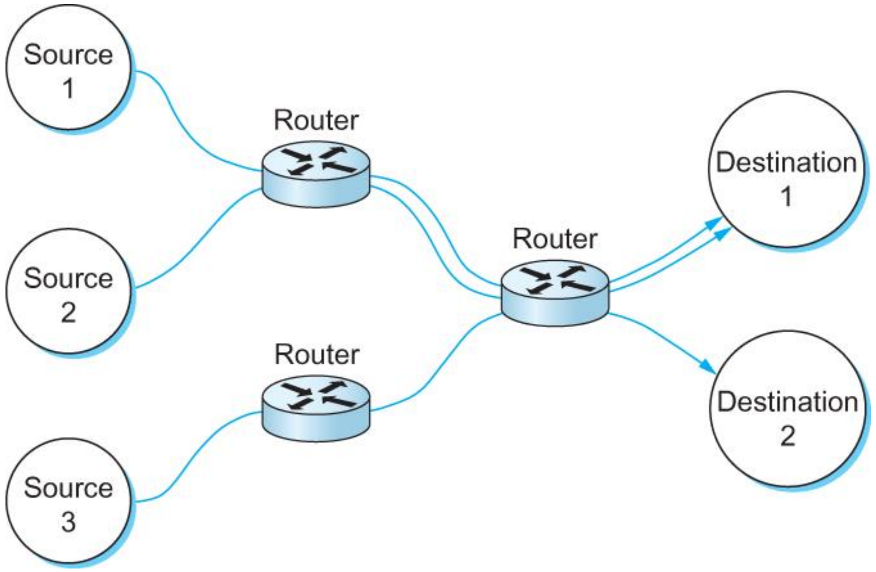
- Assume the network includes multiple links and switches (or routers).
- The links to connect a global router and connect to the destination via a single link



A potential bottleneck router.

Resource Allocation

- In a routing network, data from multiple sources can pass through a set of routers to the destination
- The datagrams are certainly switched/forwarded independently, but it is usually the case that a stream of datagrams between a particular pair of hosts flows through a particular set of routers



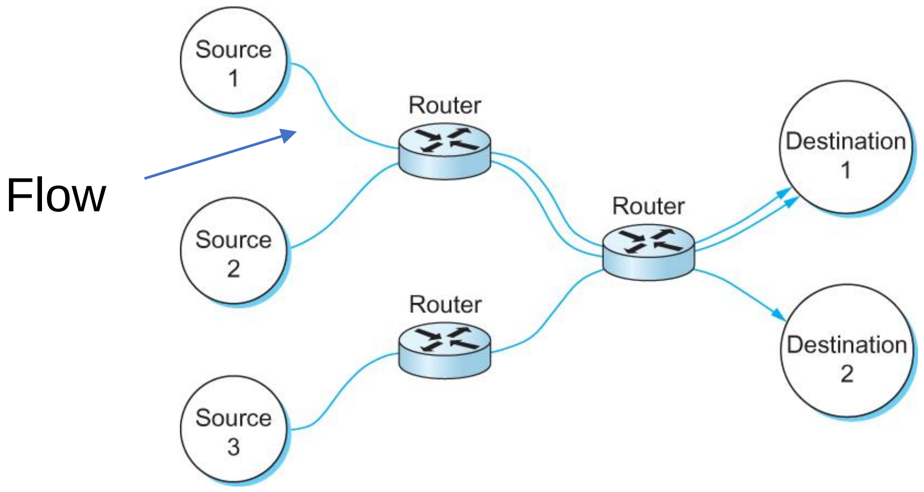
Multiple flows passing through a set of routers

Ask a question

Resource Allocation

- Flow: data stream between host-to-host (i.e., have the same source/destination host addresses) or process-to-process (i.e., have the same source/destination host/port pairs).

- Flow definition is very similar to link but a flow highlights data streams/packets



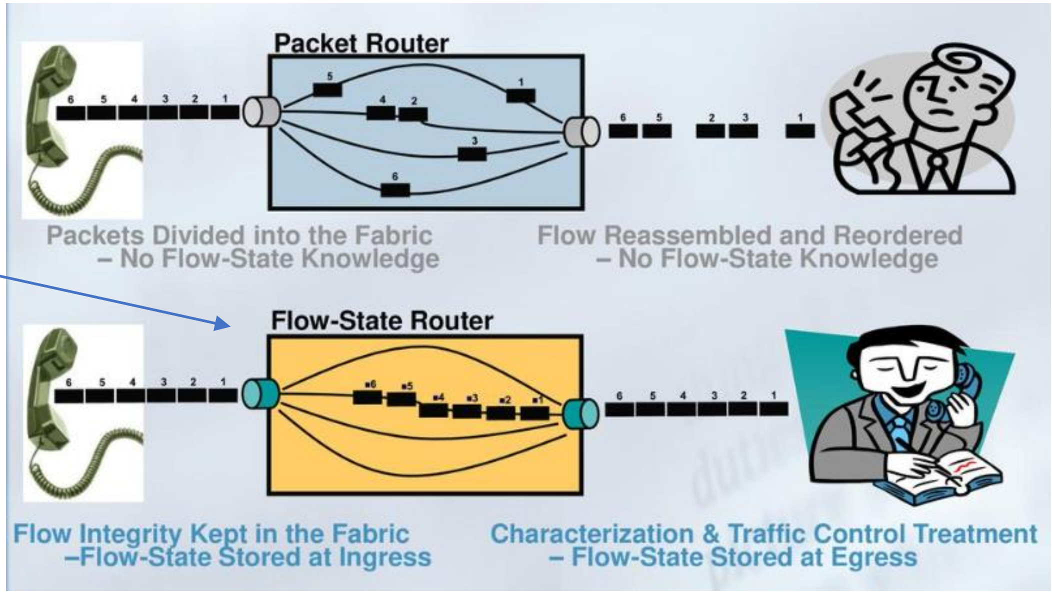
Multiple flows passing through a set of routers

Ask a question

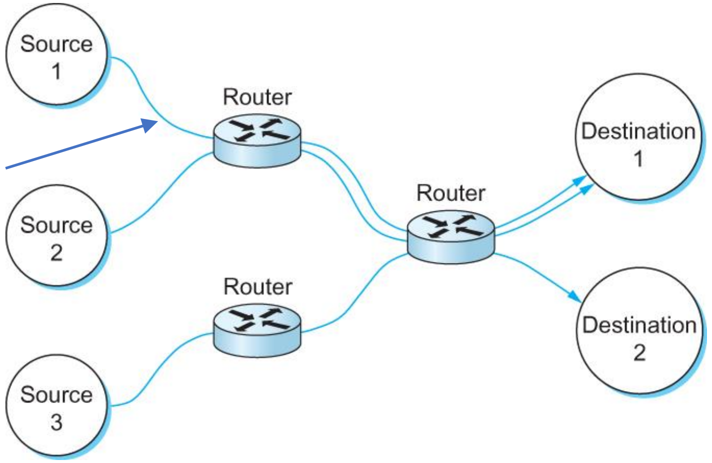
Resource Allocation

- To manage the state of each transferring flow, a router defines some state information for each flow, information that can be used to make resource allocation decisions about the packets that belong to the flow
- Soft state: Connectionless flows or maintains no state at the routers (packet router)
- Hard state: Connection-oriented network that maintains hard state at the routers (flow-state router)

Routers use this state to handle the packets (belong To a flow)



Flow



Multiple flows passing through a set of routers

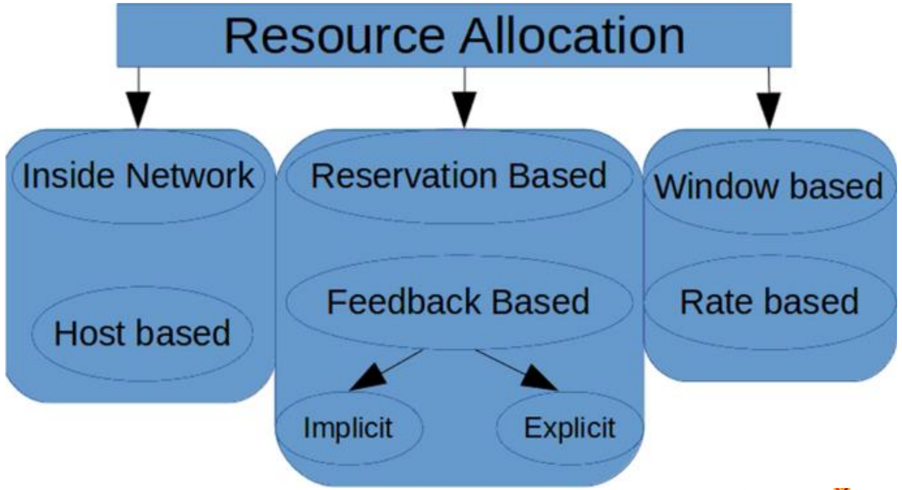
Ask a question

Resource Allocation

- **Router-centric**: each router takes responsibility for
  - ✓ Deciding when packets are forwarded
  - ✓ Selecting which packets are to dropped
  - ✓ Informing the hosts how many packets they are allowed to send

- **Host-centric**:
  - ✓ Each observe the network conditions (e.g., how many packets they are successfully getting through the network) and adjust their behavior accordingly.

- These two groups are not mutually exclusive.

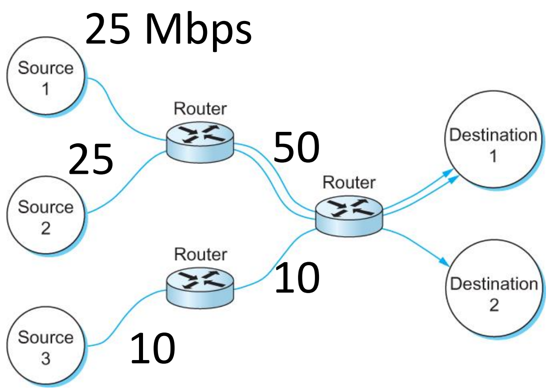


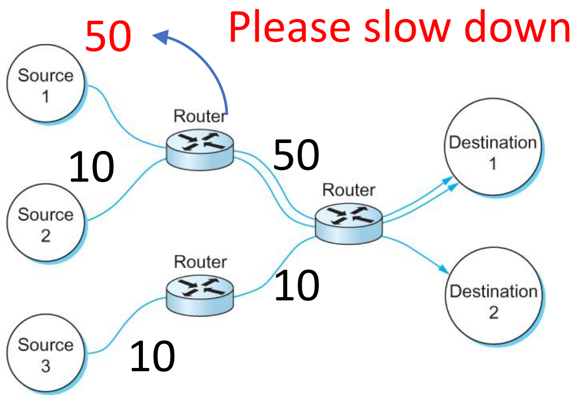
Ask a question

Resource Allocation

- **Reservation-based** : Each router then allocates enough resources (**buffers** and/or **percentage of the link's bandwidth**) to satisfy the requests from the end hosts
  - ✓ If the request cannot be satisfied at some router, because doing so would overcommit its resources, then the router rejects the reservation

- **Feedback-based approach** :
  - ✓ The end hosts begin sending data without first reserving any capacity and then adjust their sending rate according to the feedback they receive.
  - ✓ The feedback can either be *explicit (i.e., a congested router sends a "please slow down" message to the host)* or it can be *implicit (i.e., the end host adjusts its sending rate* according to the externally observable behavior of the network, such as packet losses).

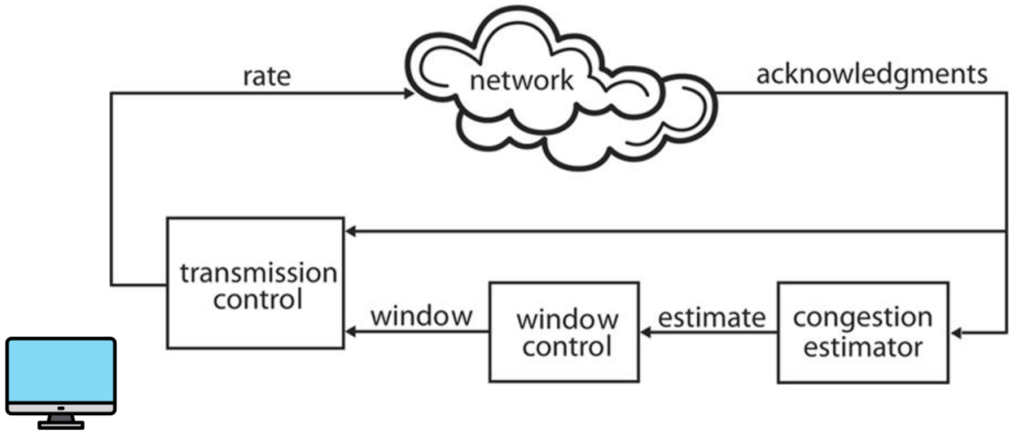




Ask a question

Resource Allocation

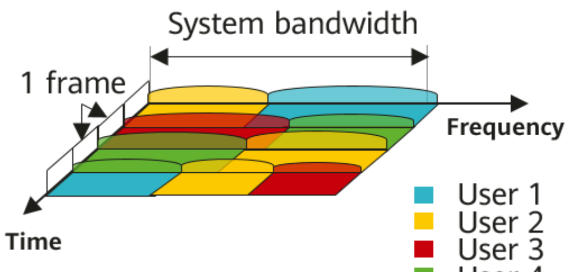
- **Window-based**: Window advertisement is used within the network to reserve buffer space



- **Rate-based**: Control sender’s behavior using a rate, how many bit per second the receiver or network is able to absorb.

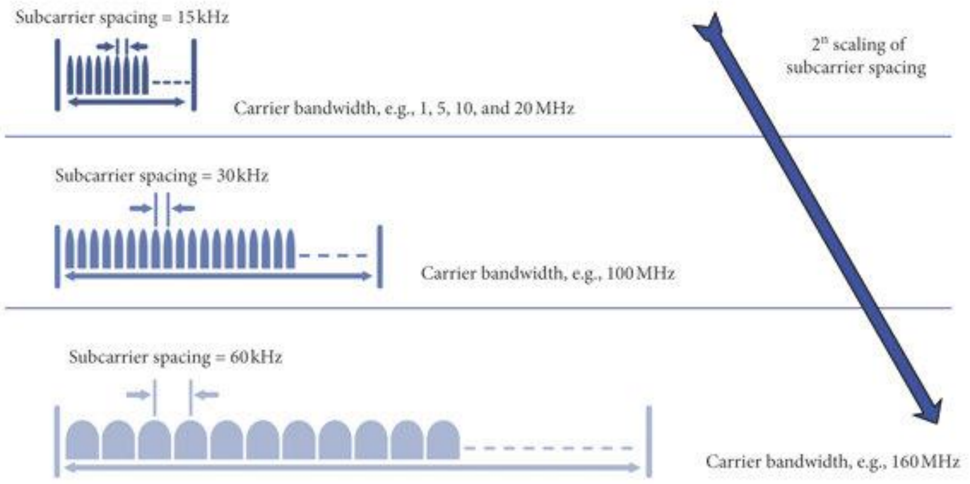
Resource Allocation in WiFi 6

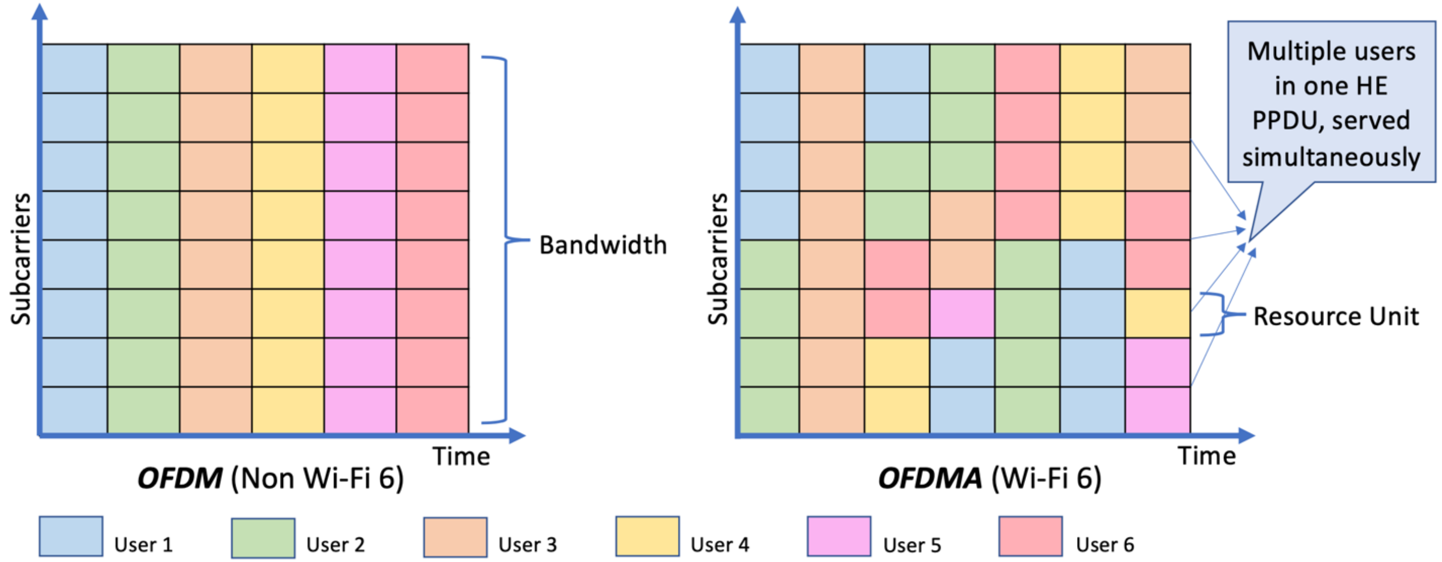
- WiFi 6 uses orthogonal frequency-division multiple access (OFDMA) modulation
- Multiple users can share channel resources



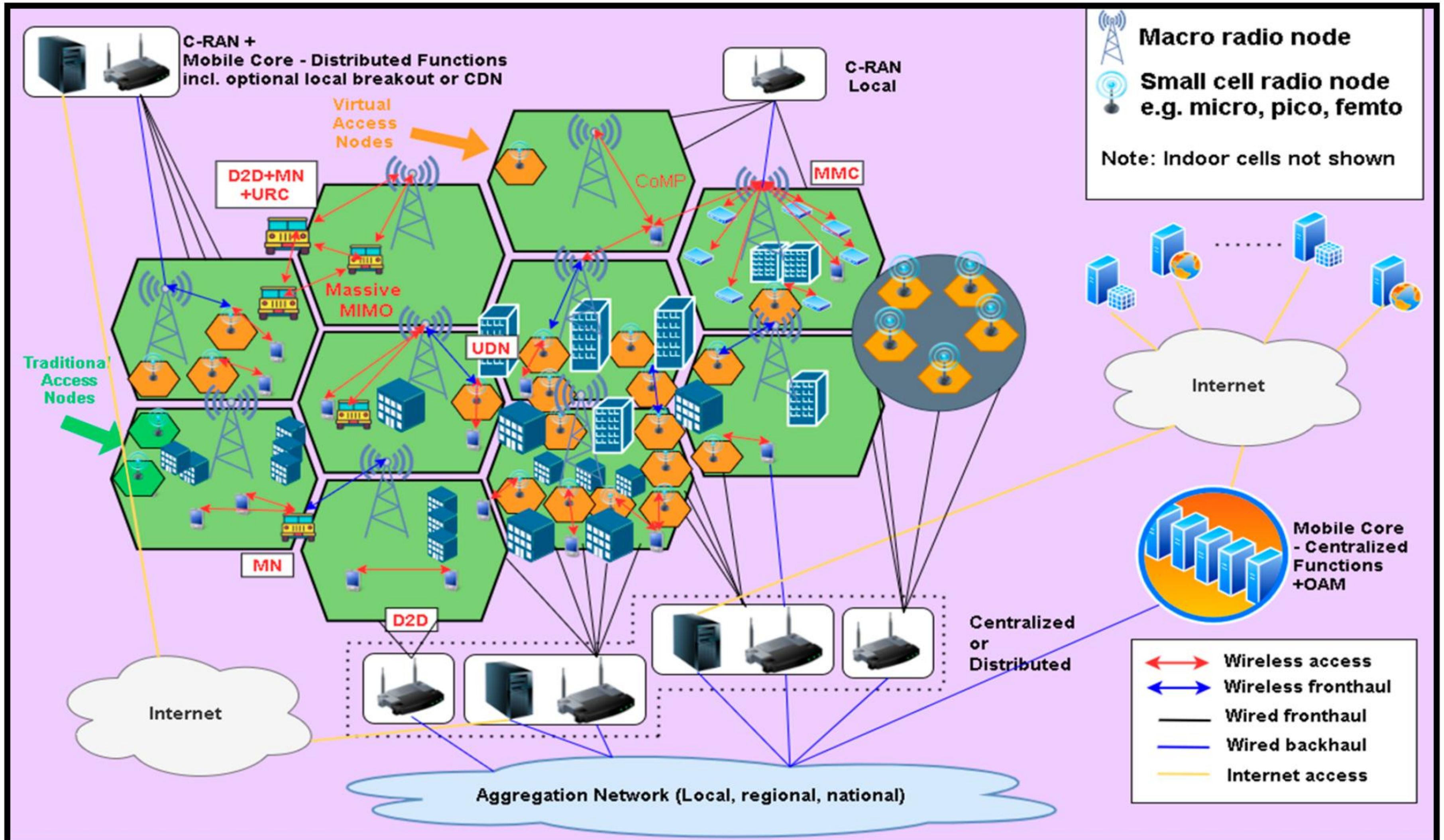
Wi-Fi 6
OFDMA
(Multiple users share channel resources.)

ASUS WiFi



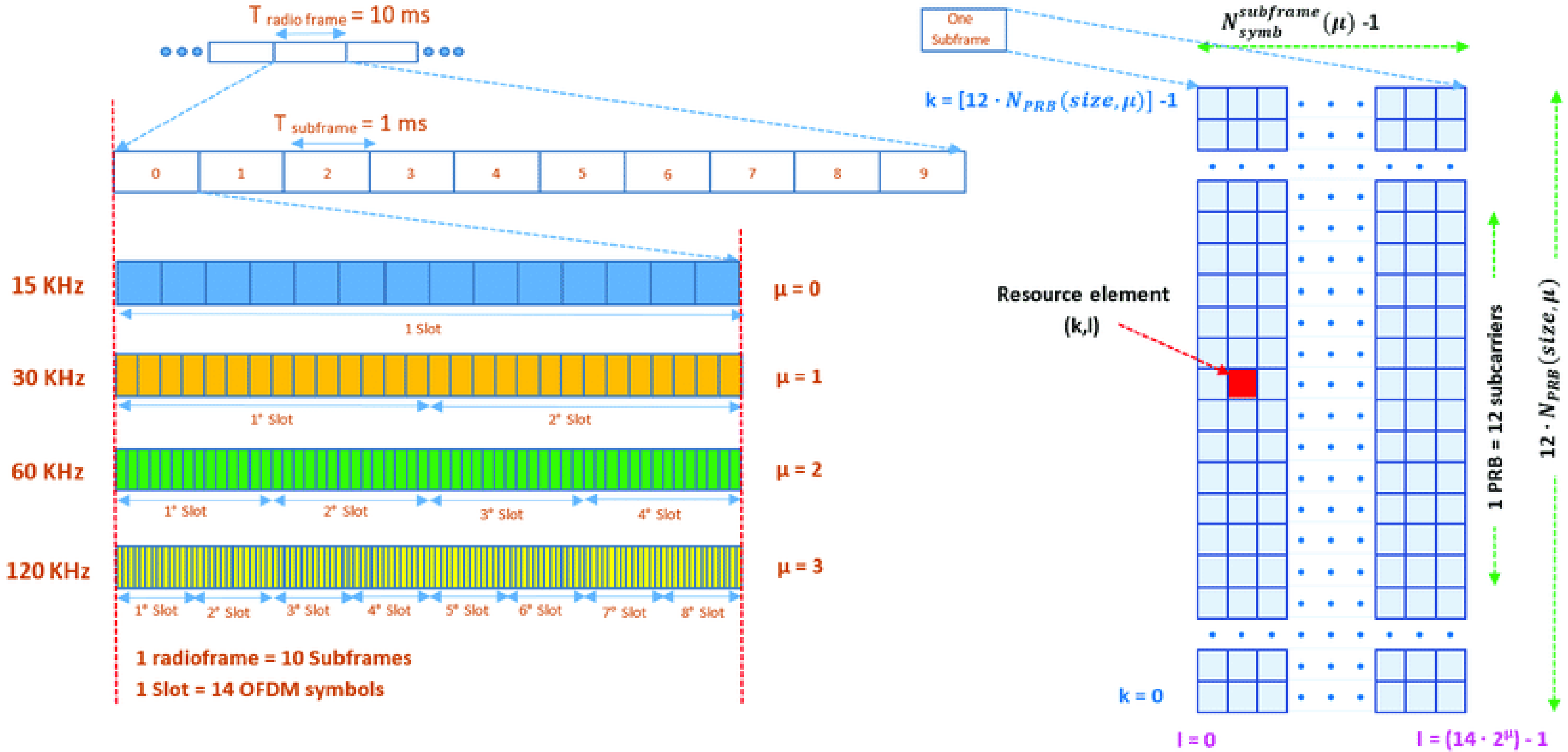


Resource Allocation in Cellular Networks



Resource Allocation in 5G

Ask a question



Ask a question

Resource allocation measurement

- **Two** principal metrics: **Throughput** and **Delay**
- **Effective** resource allocation: Much throughput but little delay
- **Throughput**: allow as many packets into the network as possible, so as to drive the utilization of all the links up to 100%
  - ✓ Avoid the possibility of a link becoming idle
  - ✓ Increase the number of packets in the network also increases the length of the queues at each router
  - ✓ Longer queues, in turn, mean packets are delayed longer in the network

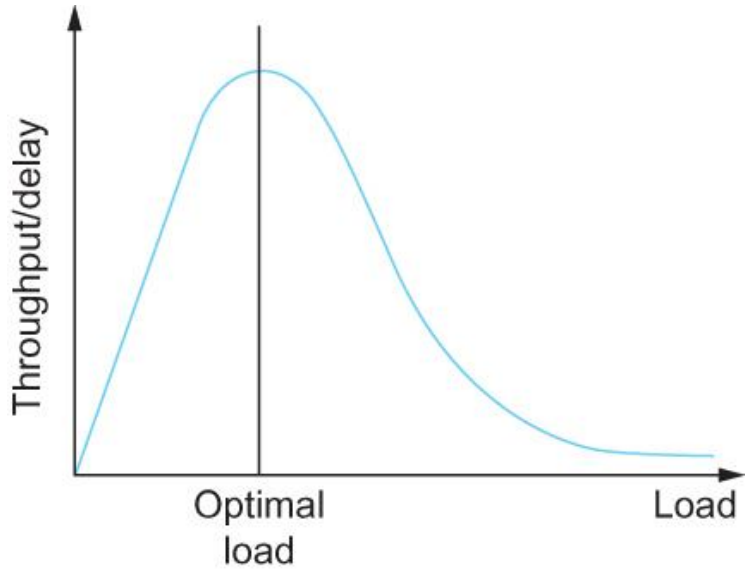
Effective Resource allocation

- To show the relationship of two metrics in evaluating resource allocation algorithms, we can use *the power of the network*

$$Power = \frac{Throughput}{Delay}$$

$$\text{Power} = \frac{10}{2} > \frac{10}{5}$$

$$\text{Power} = \frac{10}{2} > \frac{6}{2}$$



Ratio of throughput to delay as a function of load

Ask a question

Fair Resource allocation

- The effective utilization of network resources is not the only criterion for judging a resource allocation scheme
- We must also consider the fairness: the quality of making judgments that are free from discrimination

Upload Youtube Videos

Which action is prioritized?

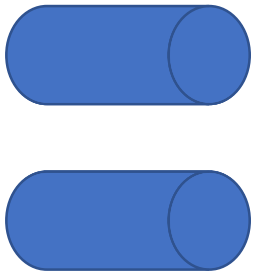
Fairness: The same priority

Download 2GB Game

Fair Resource allocation

- However, a **reservation-based** resource allocation scheme provides an explicit way to create controlled unfairness
  - ✓ **Reservation-based:** Each router then allocates enough resources (**buffers** and/or **percentage of the link's bandwidth**) to satisfy the requests from the end hosts

Video: 2Mbps



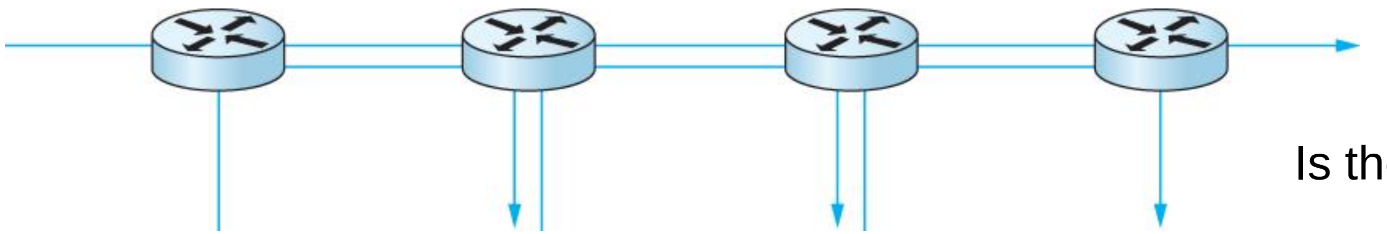
Download game: 200Kbps

Prioritize video upload

There is no fairness in **reservation-based** resource allocation

Fair Resource allocation

- **Fairness setting**: when several flows share a particular link, we would like for each flow to receive an equal share of the bandwidth
- Fairness: a *fair share of bandwidth means an equal share of bandwidth*
- *However,* equal share may not equate to fair shares.



One four-hop flow competing with three one-hop flows

Is the length of the paths fair?

Ask a question

Fair Resource allocation

- Fair implies **equal** and that **all paths are of equal length**

- **Raj Jain** proposed a metric that can be used to quantify the fairness of a congestion-control mechanism.
  - ✓ Jain's fairness index: Given a set of flow throughputs (x_1, x_2, . . . , x_n) (measured in consistent units such as bits/second), the following function assigns a fairness index to the flows:

$$f(x_1, x_2, \ldots, x_n) = \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n \sum_{i=1}^{n} x_i^2}$$

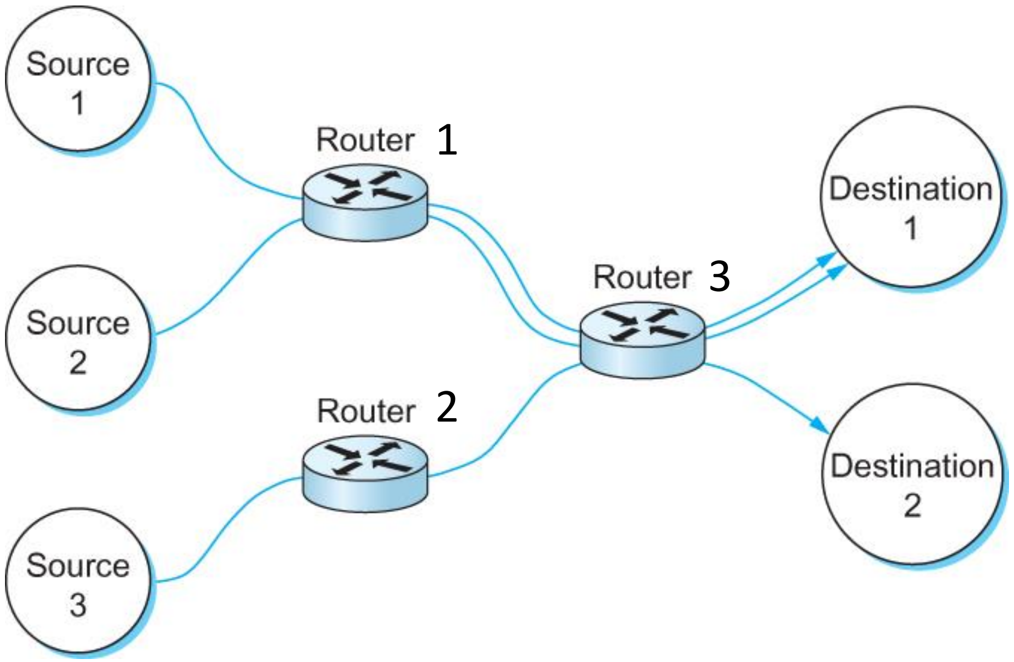
The fairness index always results in a number between 0 and 1, with 1 representing greatest fairness

Let us see an example!

- **Destination 1**: The allocated throughputs of flows from sources (1,2,3) to destination 1 at Router 3 is 10bps, 20bps,15 bps
- Is there a fairness for the network flow to Destination 1?

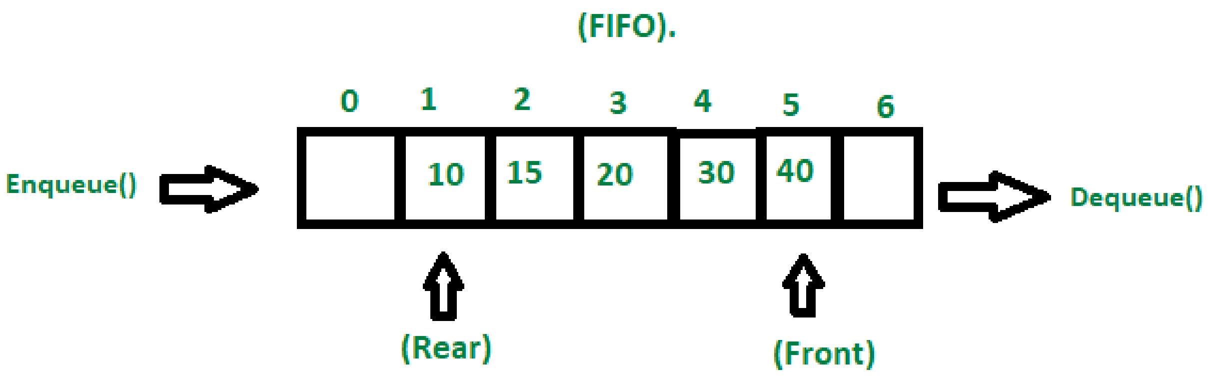
- **Destination 2**: The allocated throughputs of flows from sources (1,2,3) to destination 2 at Router 3 is 3bps, 2bps, 3bps

- Is there a fairness for the network flow to Destination 2?



How to schedule the resources: Queue

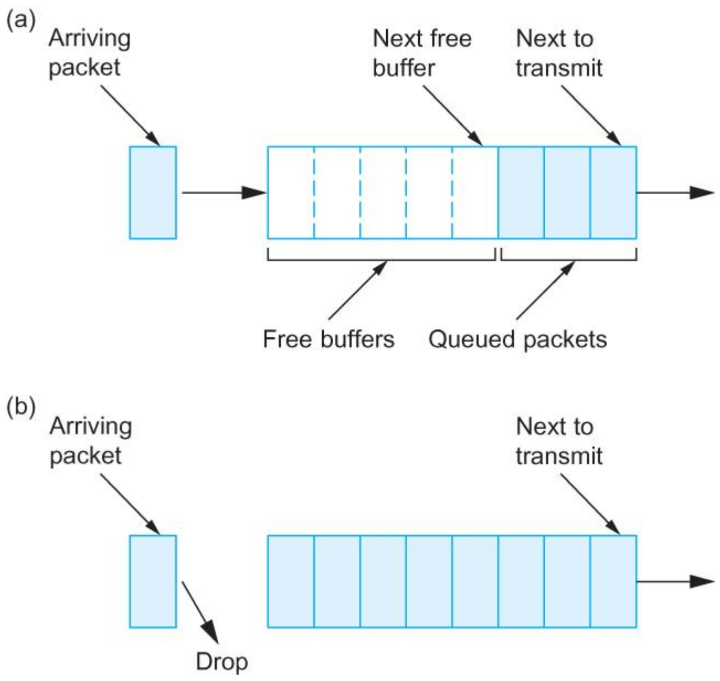
- **FIFO queuing**: first-come-first-served (FCFS) queuing
  - ✓ The first packet that arrives at a router is the first packet to be transmitted
  - ✓ Given that the amount of buffer space at each router is finite, if a packet arrives and the queue (buffer space) is full, then the router discards that packet
  - ✓ The discard is done without regard to which flow the packet belongs to or how important the packet is: *tail drop*



Ask a question

Queuing Disciplines

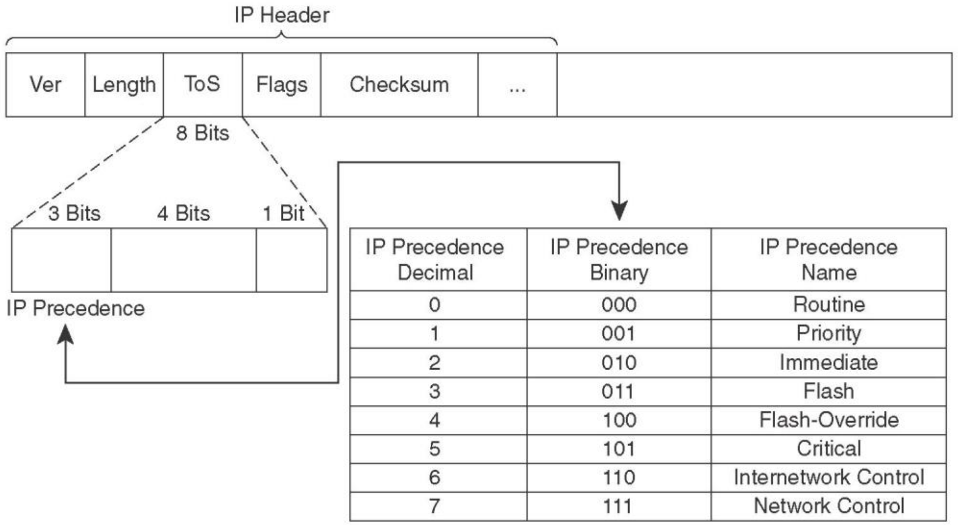
- Tail drop and FIFO are two *separable ideas*:
  - ✓ FIFO is a *scheduling discipline*—it determines the order in which packets are transmitted.
  - ✓ Tail drop is a *drop policy—it determines which packets get dropped*



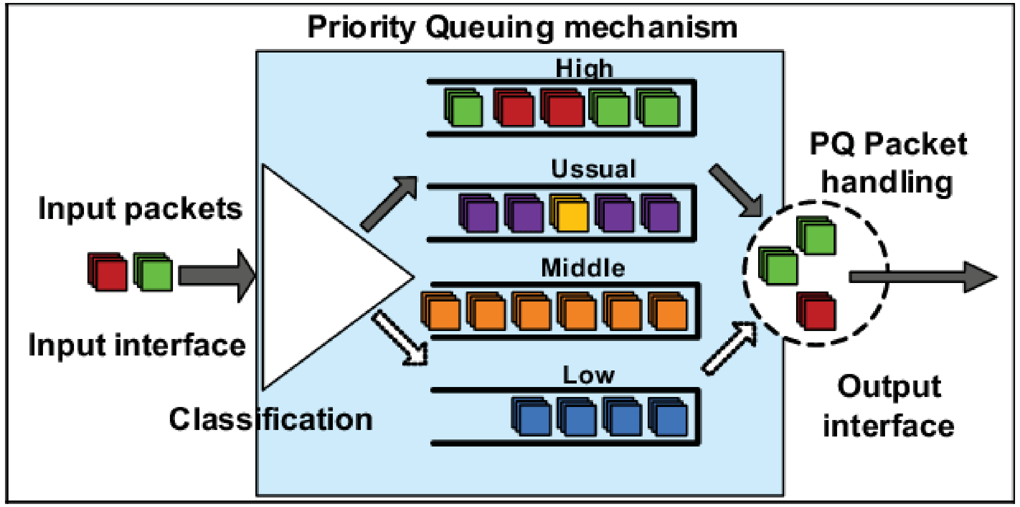
(a) FIFO queuing;
(b) tail drop at a FIFO queue.

Queuing Disciplines

- A simple variation on basic FIFO queuing is priority queuing.
  - ✓ The idea is to mark each packet with a priority; the mark could be carried, for example, in the IP header.

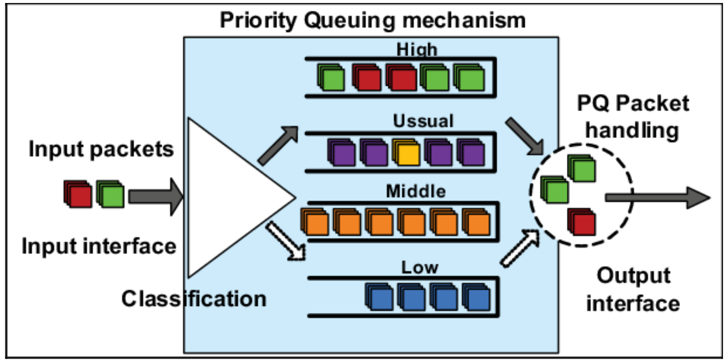


IP Precedence Decimal	IP Precedence Binary	IP Precedence Name
0	000	Routine
1	001	Priority
2	010	Immediate
3	011	Flash
4	100	Flash-Override
5	101	Critical
6	110	Internetwork Control
7	111	Network Control



Priority queuing

- The routers implement multiple FIFO queues, one for each priority class (high, usual, middle, low)
- The router always transmits packets out of the highest-priority queue if that queue is nonempty before moving on to the next priority queue.
- Within each priority, packets are still managed in a FIFO manner.

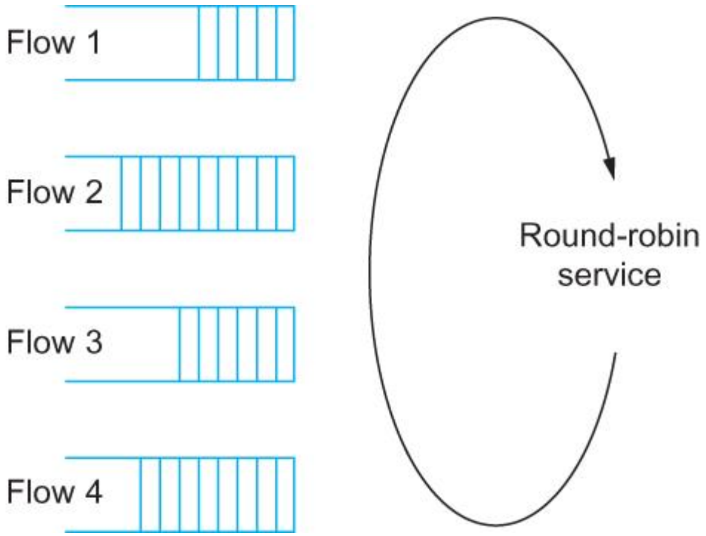


Ask a question

Fair Queuing

- The main problem with FIFO queuing
  - ✓ Does not discriminate between different traffic sources
  - ✓ Does not separate packets according to the flow to which they belong.

- Fair queuing (FQ) is a solution:
  - ✓ Maintain a separate queue for each flow currently being handled by the router.
  - ✓ The router services these queues in a sort of round-robin

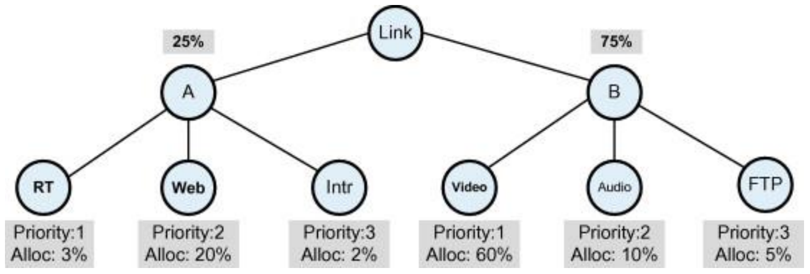


Round-robin service of four flows at a router

Fair Queuing

- Problems of fair queuing
  - ✓ Packets being processed at a router are not necessarily the same length.

- To truly allocate the bandwidth of the outgoing link in a fair manner, it is necessary to take packet length into consideration.
  - ✓ A router is managing two flows:
    - + First flow: 1000-byte packets
    - + Second flow: 500-byte packets



Simple round-robin mechanism:

- ✓ The first flow two thirds of the link's bandwidth
- ✓ The second flow only one-third of the link's bandwidth.

Bit-by-Bit Fair Queuing

Ask a question

- To understand the algorithm for approximating bit-by-bit round robin, consider the behavior of a single flow
- For this flow, let
  - P_i : denote the length of packet i
  - S_i: time when the router starts to transmit packet i
  - F_i: time when router finishes transmitting packet i
  - Clearly, $F_i = S_i + P_i$

Ask a question

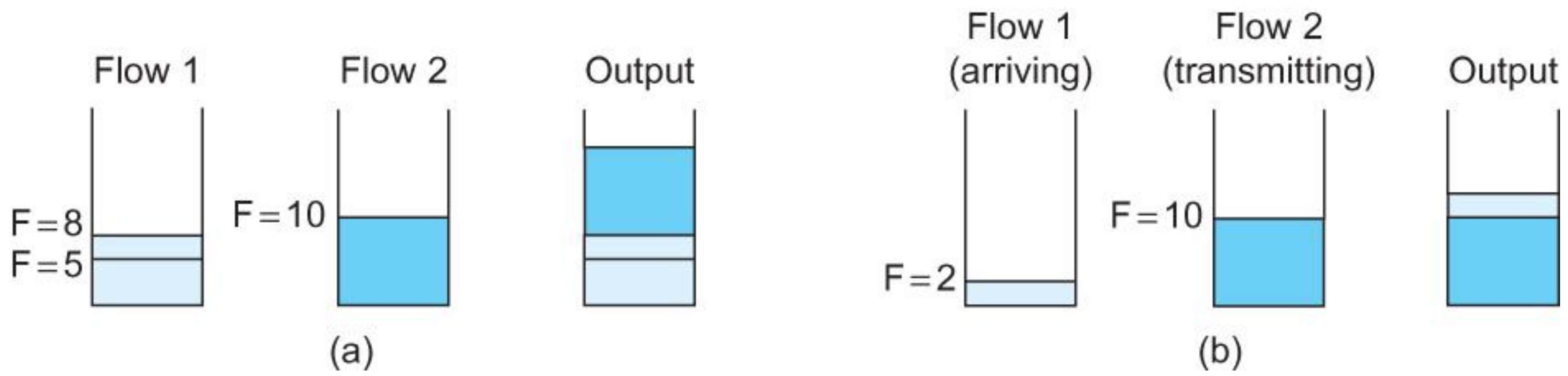
Bit-by-Bit Fair Queuing

- When do we start transmitting packet *i*?
  - Depends on whether packet i arrived before or after the router finishes transmitting packet *i*-1 for the flow
- Let A_i denote the time that packet *i* arrives at the router
- Then $S_i = \max(F_{i-1}, A_i)$
- $F_i = \max(F_{i-1}, A_i) + P_i$
- For every flow, we calculate F_i for each packet that arrives using our formula
  - We then treat all the F_i as timestamps
  - Next packet to transmit is always the packet that has the lowest timestamp
    - The packet that should finish transmission before all others

Queuing Disciplines

Ask a question

- Fair Queuing



Example of fair queuing in action: (a) packets with earlier finishing times are sent first; (b) sending of a packet already in progress is completed

Summary

- Resource allocation is critical in shared computer networks
  - ✓ Allocate fair bandwidths for many senders to access the shared channel/link

- Queue: schedule the resources in a queue for better allocation