Traffic Engineering
EN3370

Chathuranga Weeraddana

February 2021

Lecture 10: Flow Control I

Why Flow Control?

Flow Control?

- the externally offered load can be larger than what can be handled even with optimal routing [1]

[1] See § 5.4 of [Bertsekas-Gallager-1992]: D. Bertsekas and R. Gallager, *Data Networks*, Prentice Hall, Englewood Cliffs, NJ, 2nd Ed., 1992.

Flow Control?

- if no measures to restrict the traffic into the network
  - queue sizes at bottleneck links grows
  - packet delays will increase
  - violation of maximum delay specifications
  - queue sizes grow indefinitely
  - the buffer space at some nodes may be exhausted
  - some of the packets arriving at these nodes will be discarded
  - those are to be retransmitted later
  - thus, waste of communication resources

Flow Control?

- if no measures to restrict the traffic into the network
  - the actual network throughput decreases
  - at the same time, packet delay increases
  - recall: the trade-off curve plotted in your Project 1
  - one has to prevent some of the offered traffic from entering the network
  - this is why flow control

Flow Control Vs Congestion Control

- flow control and congestion control: considered as same meaning

Flow Control Between?

- flow control is between two points, e.g.,
  - between two users at the transport layer
  - between a user and an entry point of network layer
  - between two network layer nodes

Current Approaches

Approaches to Flow Control

- call blocking:
  - when a session requires a minimum guaranteed QoS and no network resources
- packet dropping:
  - when a node with no available buffer space receives a packet
  - when maintaining a fair resource allocation among the users of the system (buffer space is still available)
  - corrective actions for discarded packets: depends on the type of the session
  - different priority levels can be attached with packets $\rightarrow$ packet dropping is controlled

Approaches to Flow Control

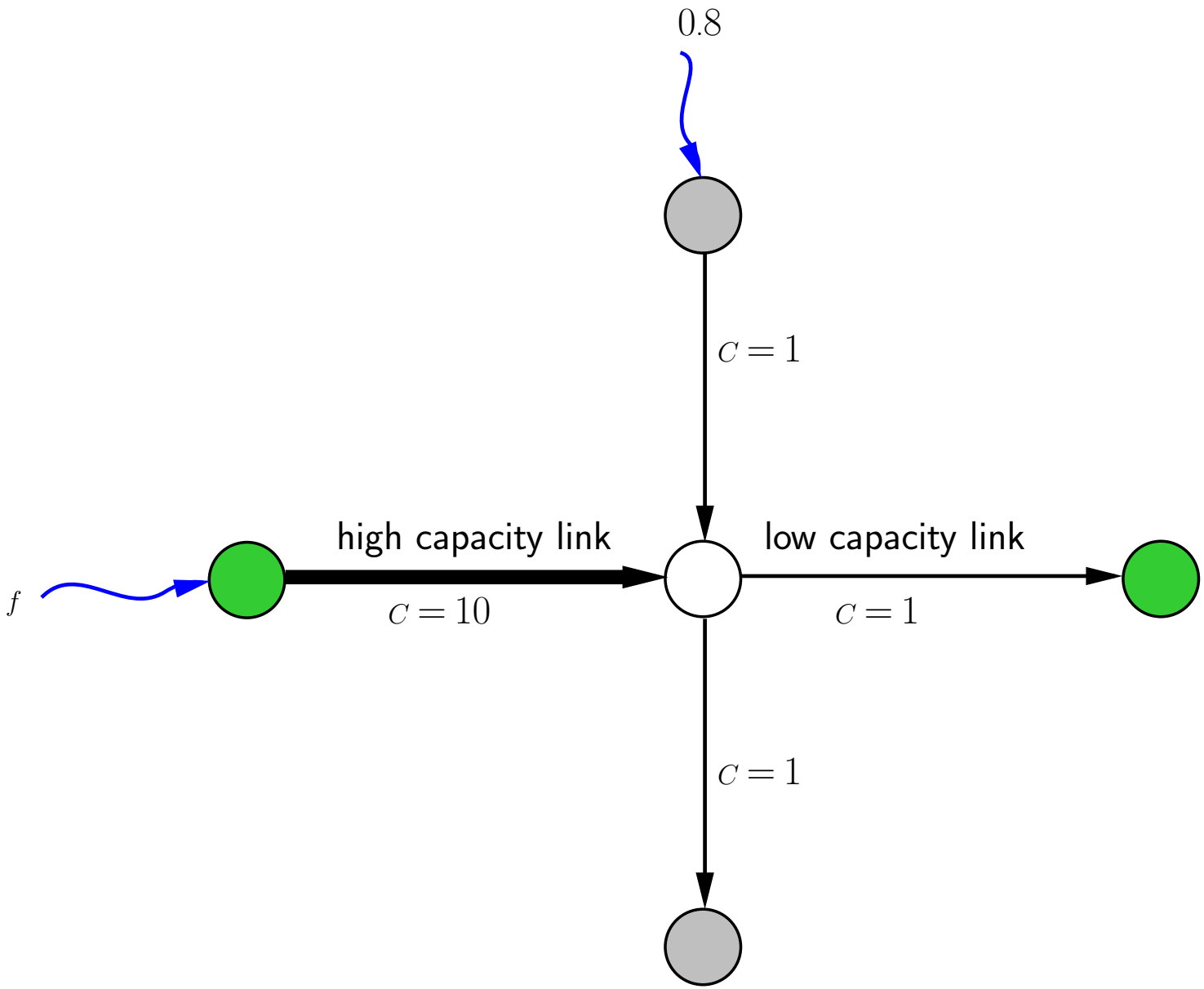
- packet blocking:
  - the network resources up to the place of packet dropping is wasted
  - more meaningful to restrict a session's packets from entering the network if its dropping is anticipated
- packet scheduling:
  - selectively expediting or delaying the transmission of the packets of various sessions

Main Objectives of Flow Control

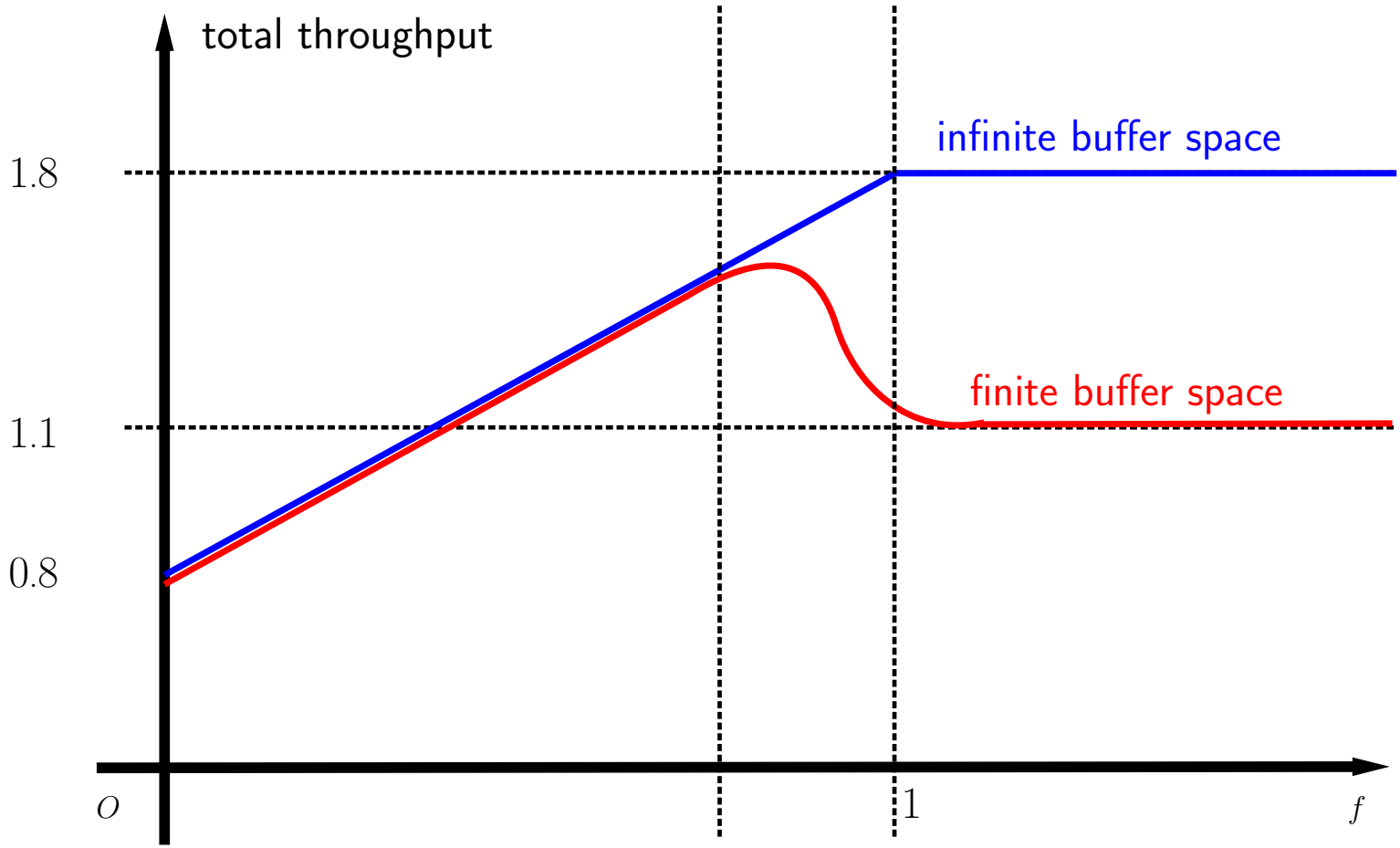
Main Objectives of Flow Control

- case 1: strike a good compromise between throttling sessions and keeping average delay and buffer overflow at a reasonable level
- case 2: maintain fairness between sessions in providing the requisite quality of service

Case 1, an Example



Case 1, an Example



Case 1

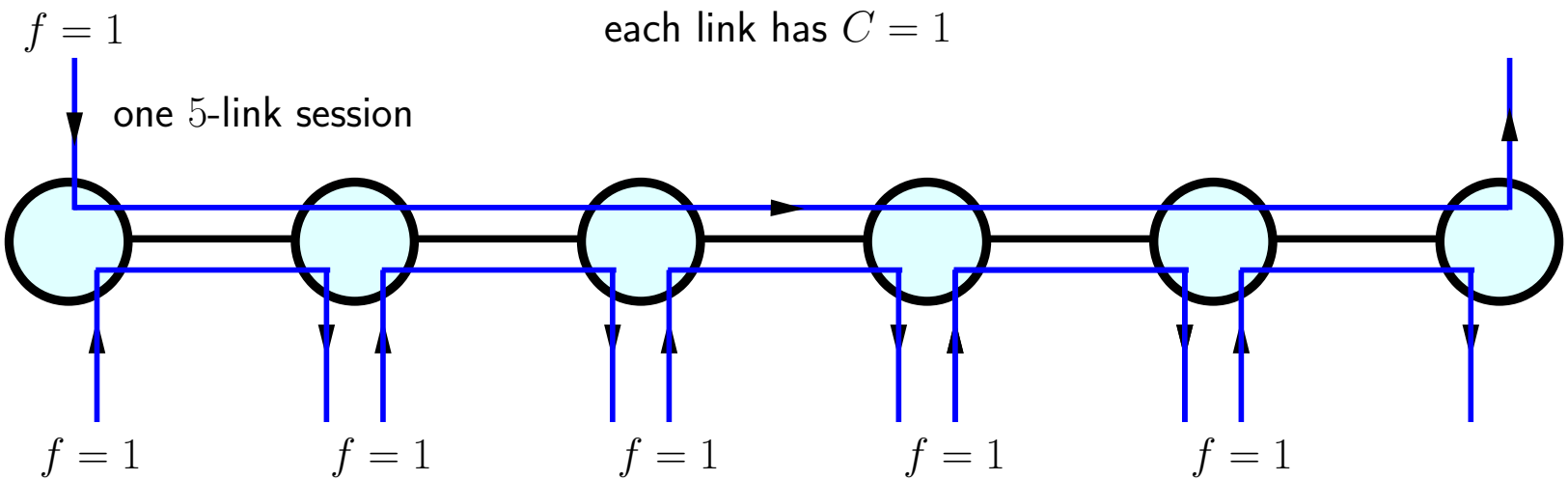
- offered load [2] is large $\rightarrow$ delay and buffer overflow are resolved by lowering the input to the network

[2] throughput = offered load when the traffic load offered by external sites to the network is relatively low. On the other hand, throughput = offered load-rejected load when the traffic load offered by external sites to the network is relatively high.

Case 1

- rough guideline:
  - flow control should not be exercised when network delay is below some requisite level
  - flow control should reject as much offered traffic as necessary to keep delay at the critical level
  - easy to say but not easy to handle: why? neither delay nor throughput can be represented meaningfully by single numbers in a flow control context

Case 2, an Example



five 1-link sessions

Case 2, an Example

- 6 sessions each offering 1 unit/sec of traffic
- 5 links with capacity of 1 unit/sec

Case 2, an Example

- suppose we need to achieve the maximum throughput
- how?
  - admit all the traffic of 1-link sessions but the 5-link session
  - the total throughput will be 5 units/sec

Case 2, an Example

- suppose we need to achieve a 'fair' admission of the sessions, in the sense that rates of each session is equal
- how?
  - admit all the traffic of 1-link sessions only in the first half of the second
  - thus, the rate for a 1-link session is 0.5 unites/sec
  - admit the traffic of 5-link session in the second half
  - thus, the rate for the 5-link session is 0.5 unites/sec
  - the total throughput will be $(0.5 \times 5 + 0.5) = 6/2$ units/sec

Case 2

- the fairness is an important objective that is to be optimized
- that is when offered traffic is to be cut back (an approach to flow control), it should be done fairly
- many different sessions, priorities and QoS $\rightarrow$ the notion of fairness is complicated
- a compromise is required between equal access to a network resource and total throughput, easy to say but not easy to handle

Flow Control Methods

Window Flow Control

Window Flow Control

- there is an upper bound on the number of data units that have been transmitted by `A` and are not yet known by `A` to have been received by `B`
- then the session between Tx `A` and Rx `B` is said to be window flow controlled
- the upper bound is referred to at as the window size or the window

Involved Entities

- nodes `A` and `B` can be:
  - two nodes of the communication subnet
  - user's machine and the entry node of the communication subnet
  - users' machines at the opposite ends of a session
- the data units in a window can be:
  - packets
  - bytes

Basic Mechanism

- `B` notifies the `A` that it has received a data unit by sending a special message to `A`
- the message is called a permit (or acknowledgement or ACK)
- upon receiving a permit, `A` is free to send one more data unit to `B`
- a permit is a form of passport that a data unit must obtain before entering the logical communication channel between `A` and `B`

Realizing Flow Control

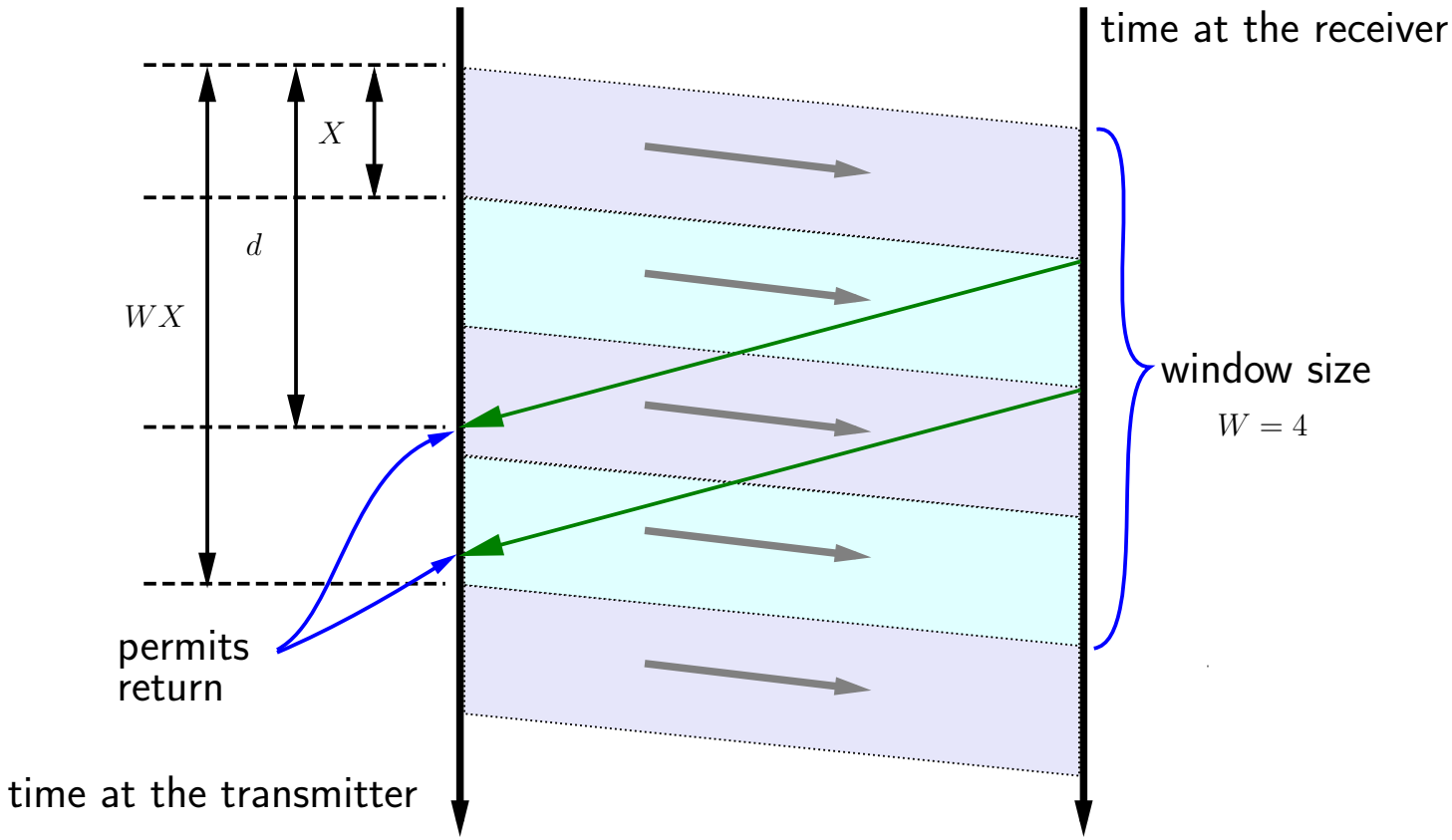
- how flow controlling is realized from the above mechanism?
  - (permits + data) units in transit $\leq$ the window size
  - i.e., a sort of regularization
  - network is congested $\rightarrow$ permits are delayed
  - source must wait $\rightarrow$ network finds time to resolve the ongoing congestion
  - receiver `B` can intentionally slow down the flow into the networkby delaying permits, why?
    - to prevent buffer overflow

Commonly Use Strategies

- two commonly used strategies
  - end-to-end windowing
  - node-by-node windowing

End-to-End Windows

Basic Idea


time at the receiver
X
d
WX
window size
$W = 4$
permits
return
time at the transmitter

Basic Idea

- X : transmission time of a single data unit
- W : window size
- d : round-trip delay

End-to-End Windows, Transmission Rate

- note that in the previous figure we have

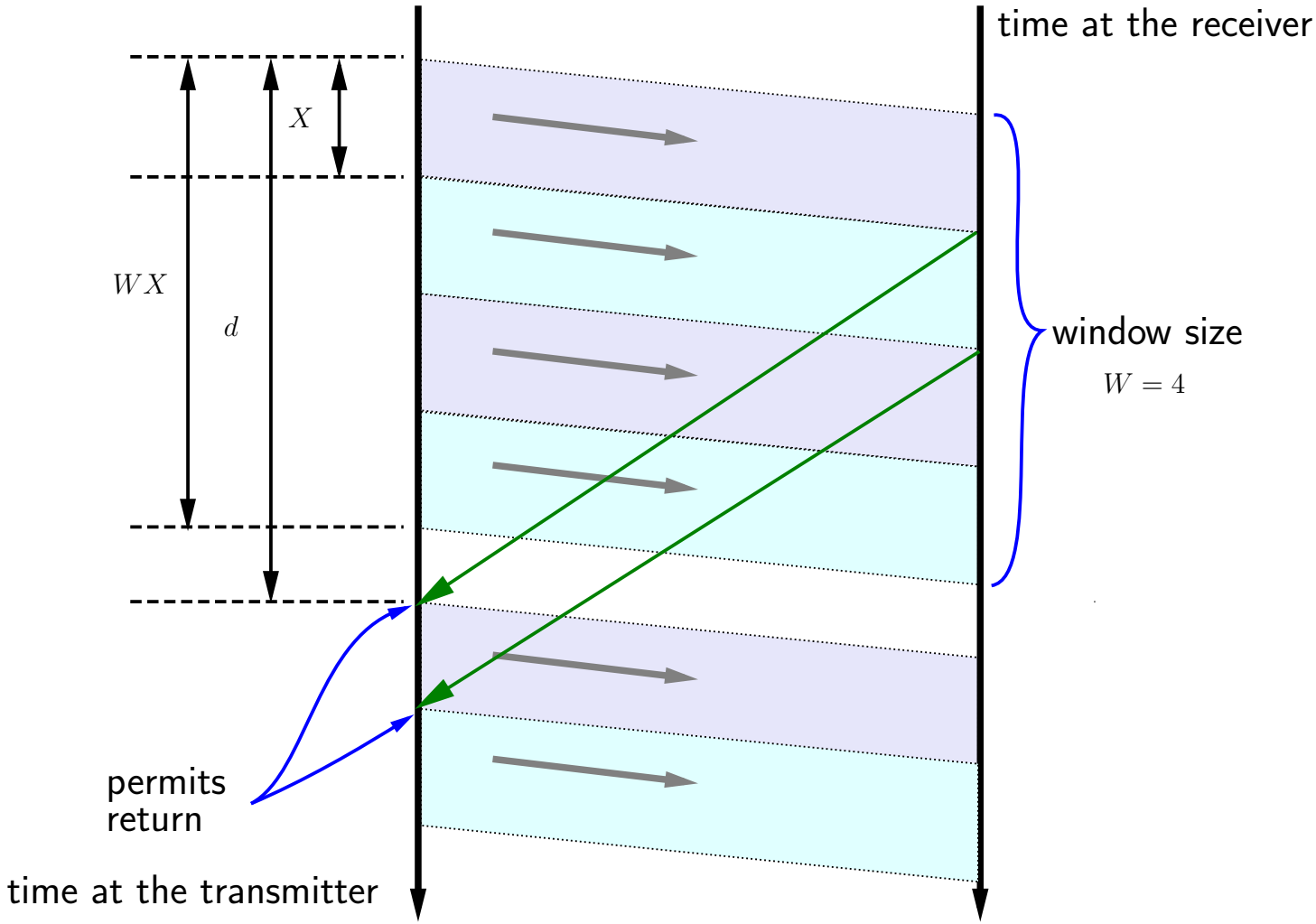
$$d \leq WX$$

- thus the rate r of the transmission is

$$r = \frac{W}{WX} = \frac{1}{X} \quad \text{data units/sec}$$

- flow control is not active!

Basic Idea


time at the receiver
X
WX
d
window size
$W = 4$
permits
return
time at the transmitter

End-to-End Windows, Transmission Rate

- note that in the previous figure we have

$$d > WX$$

- thus the rate r of the transmission is

$$r = \frac{W}{d} \quad \text{data units/sec}$$

- flow control is active!

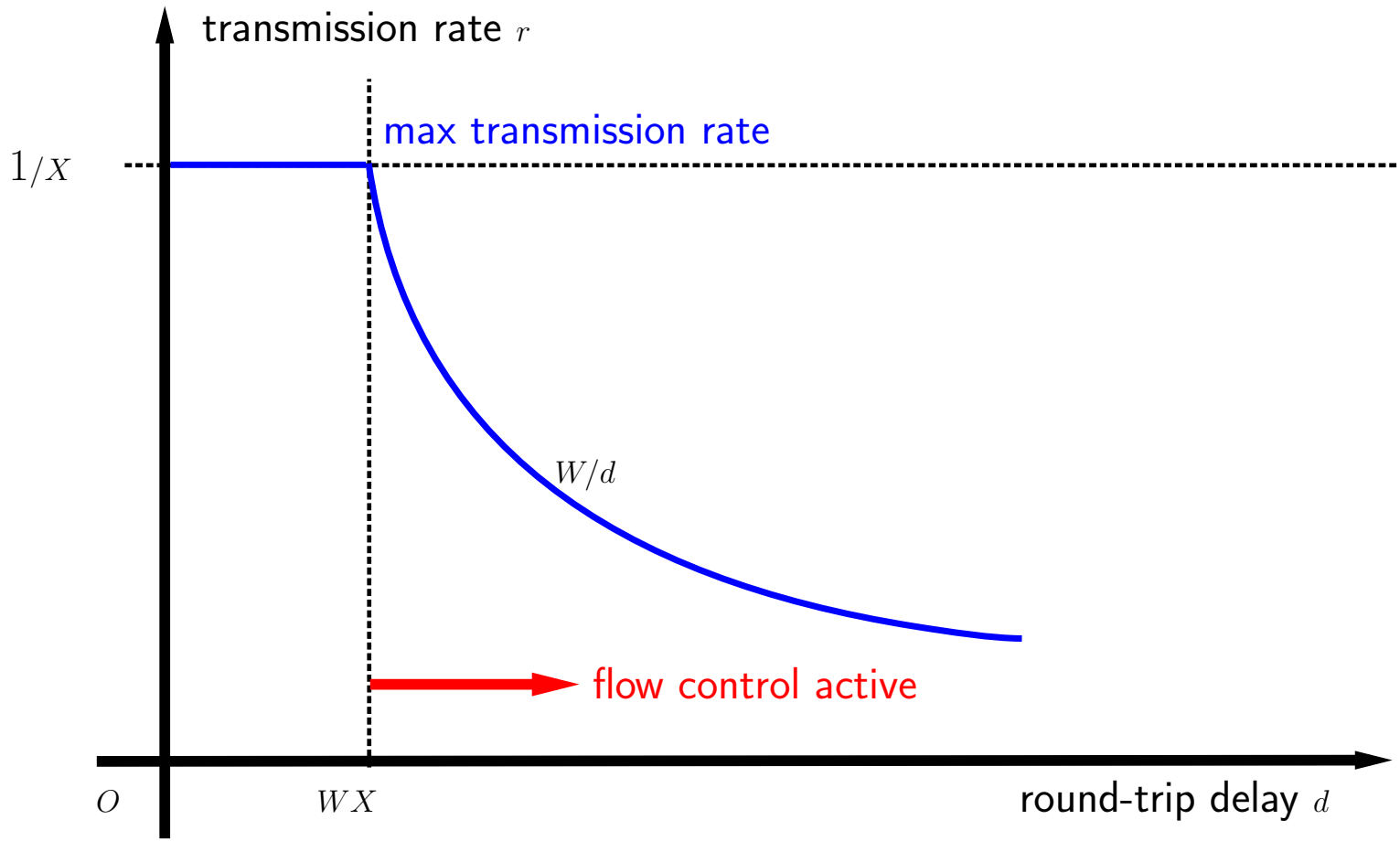
End-to-End Windows, Transmission Rate

- maximum rate corresponding to a round-trip delay d

$$r_{\max} = \min\left\{\frac{W}{d}, \frac{1}{X}\right\} \quad \text{data units/sec}$$

- the source transmission rate is reduced in response to congestion and the attendant large delay
- W is relatively small $\rightarrow$ fast reactions to congestion
- fast reaction + low overhead are major advantages

End-to-End Windows, Transmission Rate



End-to-End Windows, Limitations, with Regard to Case 1

End-to-End Windows, Limitations

- a minimum rate for a session is not guaranteed
  - not suitable for voice, video
- properwindow size and adjusting that size in response to congestion is not easy
  - a basic trade-off exists
  - small window size (avoid congestion) vs large window size (max throughput operation)
  - $W \propto d$ and $W \propto$ (the maximum desired rate)

End-to-End Windows, Limitations

- adequate control of packet delay is not supported, e.g.,
  - n actively flow controlled sessions, window sizes, $W_1, \ldots, W_n$
  - # of (permits+data) units $= \sum_n W_n$
  - # of data units $= \sum_n \beta_n W_n$, $\beta_n \in [0, 1]$
  - average delay T per packet is

$$T = (1/\lambda) \sum_n \beta_n W_n \tag{1}$$

where λ is the throughput

End-to-End Windows, Limitations

- e.g., continues
  - as $n \uparrow$, throughput $\rightarrow$ const
  - on the other hand, roughly the delay $T \propto n$ [see (1)]
  - therefore, as $n \uparrow$, the delay T increases
  - one cannot remedy the situation by simply $\downarrow W_i$ for all i, why?
    - imposed lower bounds on W_i, e.g., what if no lower bounds on W_i in an m-hop session? [note: $d \propto m$]
    - recommended window sizes are typically between m and $3m$

End-to-End Windows, Limitations

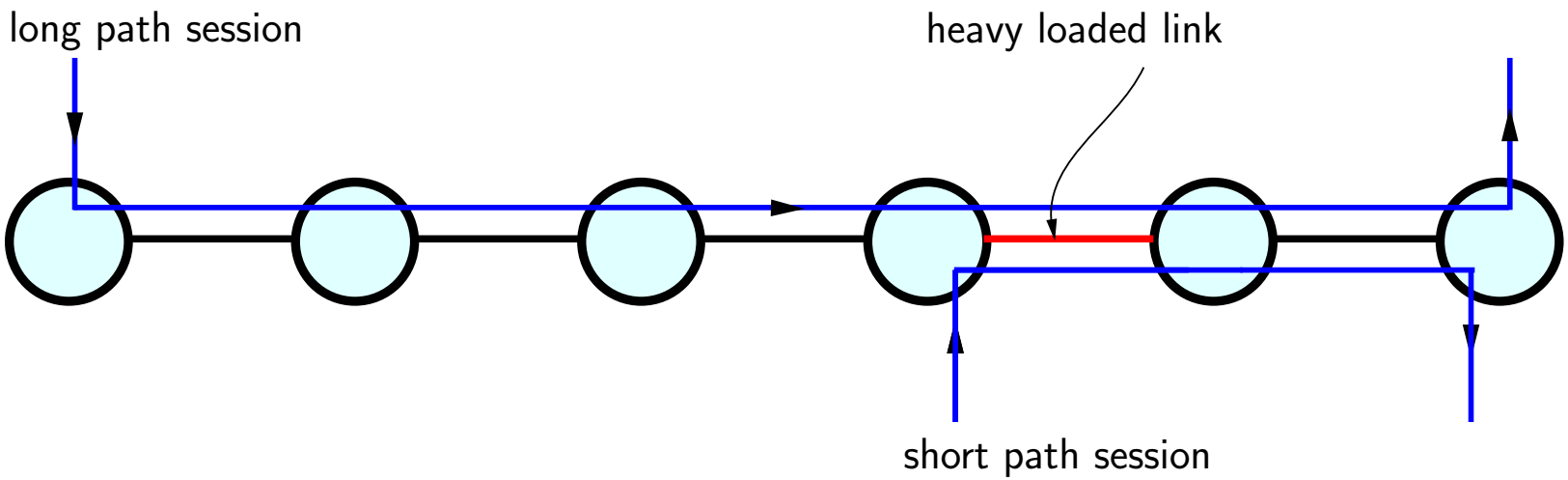
- windows should be used with care when high-speed and large distances are involved
- they require excessive memory
- they respond to congestion relatively slowly when the $d \gg X$
- if a good trade-off between delay and throughput needed $\rightarrow$ dynamic window adjustment (not easy)
  - light load conditions $\rightarrow$ windows should be large (i.e., high throughputs)
  - heavy load conditions $\rightarrow$ windows should shrink (reasonable delays)

End-to-End Windows, Limitations, with Regard to Case 2

End-to-End Windows, Limitations

- recall the window size of a session is recommended to be proportional to the number of links on its path, i.e., $W \propto m$
- as a result, at a heavily loaded link, a long-path session can have many more packets awaiting transmission than short-path sessions
- i.e., a larger throughput to the long-path session
- the notion of fairness is destroyed
- can be improved by changing queuing discipline

Fairness Problem



Node-by-Node Windows for Virtual Circuits

Basic Idea

- there is a separate window for every virtual circuit and pair of adjacent nodes
- flow control is exercised over one link long and many such segments
- thus W is typically $2 - 3$ (packets) for moderate-speed terrestrial links
- for high-speed networks $\rightarrow$ the node-by-node strategy is less attractive

Basic Idea

- the receiver avoids substantial accumulation of packets into its memory
- to this end, it slows down the rate at which it permits are sent
- usually, the receiver returns a permit to the transmitter as soon as it releases [3] a packet from its buffer

[3] e.g., releasing to data-link-control unit.

Basic Idea

- consider the windows along three successive nodes $i-1$, i, and $i+1$
- suppose the W-packet buffer of node i is full
- node i must wait for a permit from node $i+1$
- upon the reception of the permit from node $i+1$, node i releases a packet from its buffer towards node $i+1$
- soon after the release, node i sends a permit to node $i-1$

Basic Idea

- there is an inextricable coupling of successive windows along the path of a virtual circuit
- as a result, the message "there is a congestion" is implicitly directed to the source of the session
- thus, the origin of the virtual circuit (source) is enforced to wait and see, i.e., flow is controlled
- the phenomenon whereby windows progressively fill up from the point of congestion toward the virtual circuit origins is known as backpressure

Distribution of Data Units In Network

- the packets inside the network [4] in transit is roughly mW, assuming an m-link session
- this is comparable to that of a flow controlling via an end-to-end window, e.g.,
  - $W \approx 2m$ in end-to-end window case [see slide 33]
  - with $W = 2$ per link [see slide 37] in node-to-node window case, $mW = 2m$

[4] in the worst case, where the congestion is at the last link.

Distribution of Data Units In Network

- nonetheless, there is a vital difference with regard to the packet distribution
  - uniformly distributed in node-by-node (NN) case
  - concentrated at the congested link in the end-to-end (EE) case
- disparity in packet distribution $\rightarrow$
  - disparity in memory requirements in NN and EE cases
  - disparity in buffer overflows in NN and EE cases

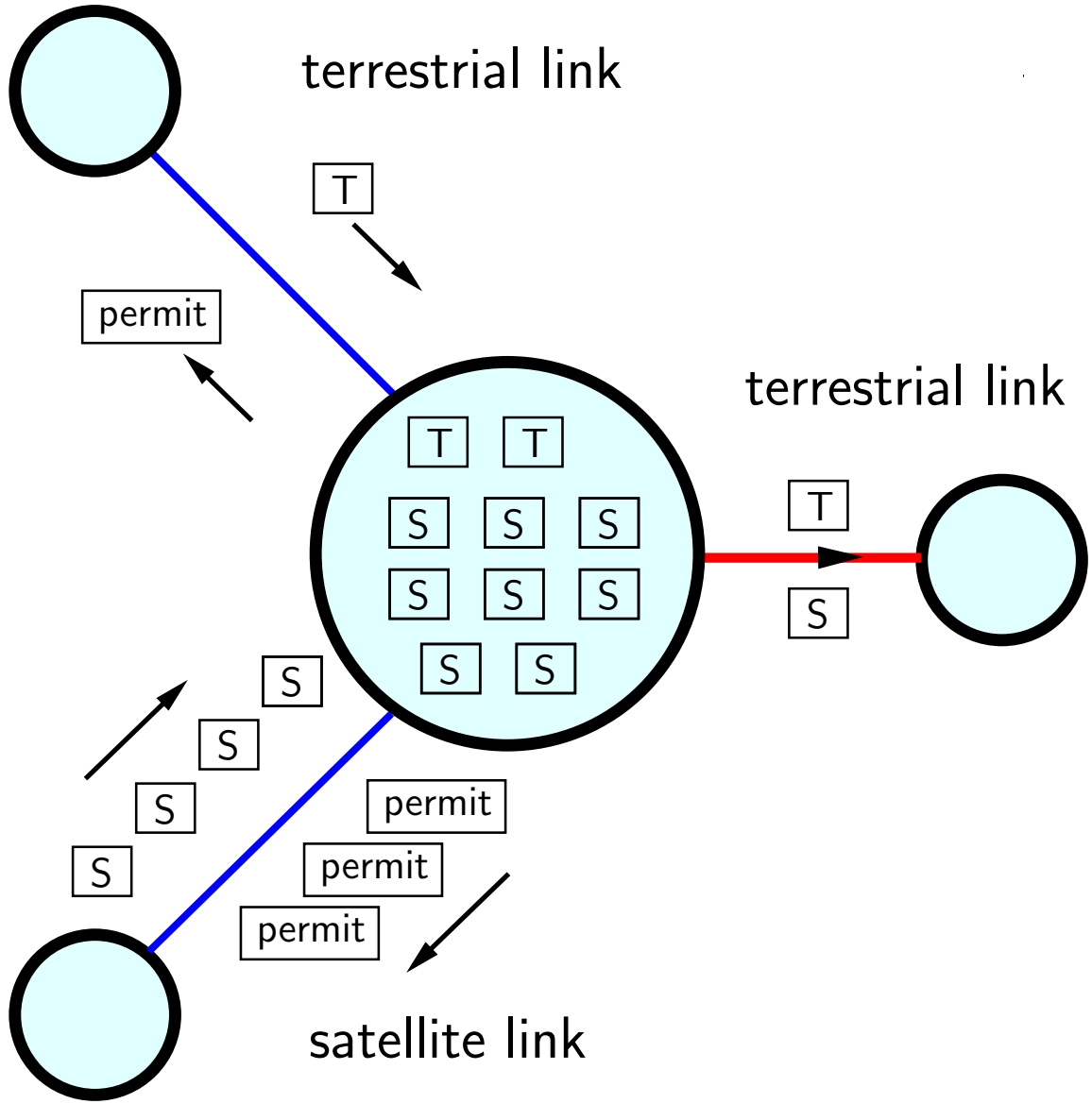
Fairness Managed?

- recall in EE windows, there is a fairness problem, i.e.,
  - large window sessions monopolize a congested link at the expense of small window sessions
- uniform packet distribution in NN $\rightarrow$ fairness problem in EE case is handled to some extent
  - when the window sizes of all virtual circuits are roughly equal

Fairness Managed?

- problem may still remain if links with large disparities in propagation delay are there in circuits
- e.g., problem may start arising at a node serving both
  - virtual circuits with large window size
  - virtual circuits with small window size

Fairness Problem

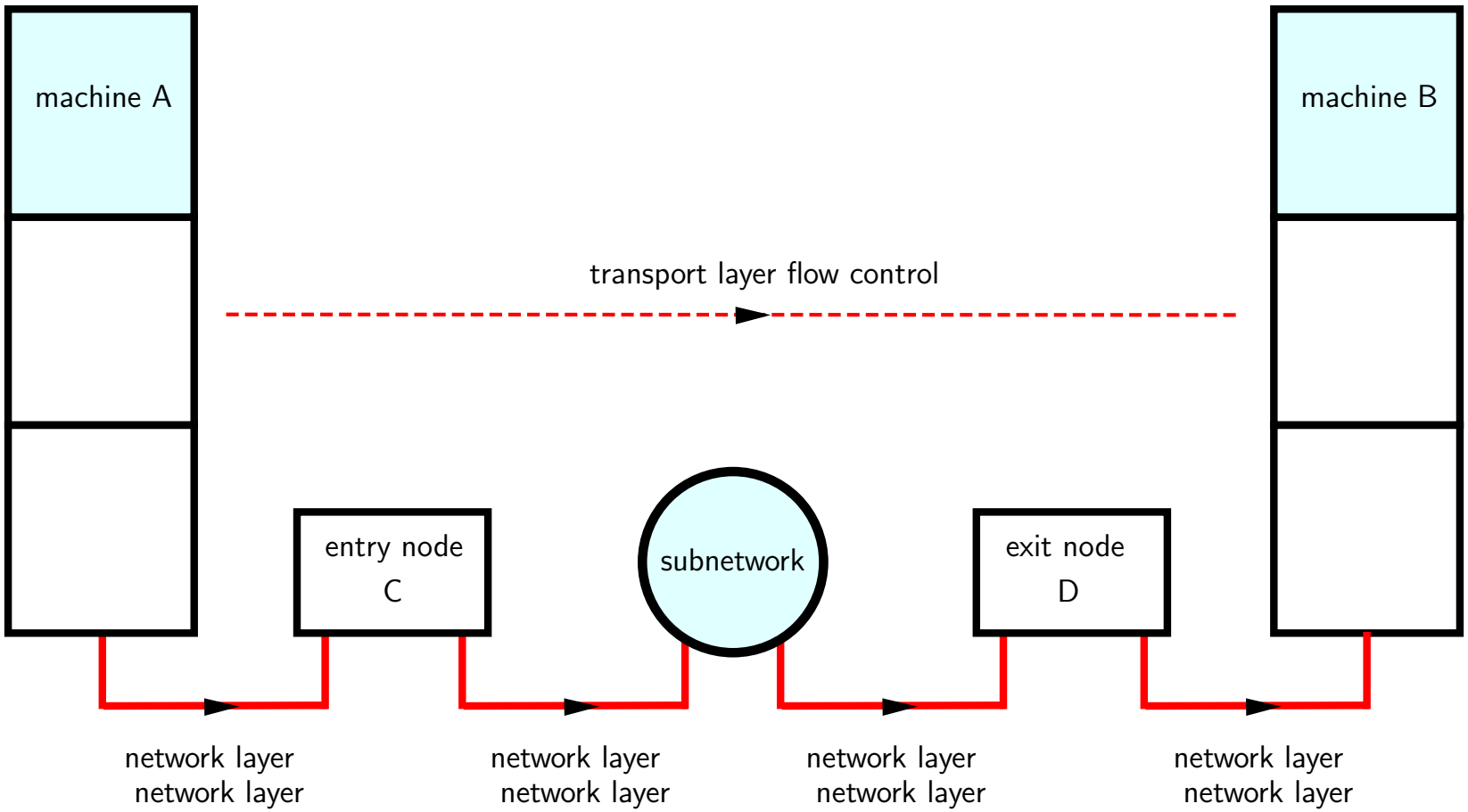


Window Flow Control at Higher Layers

Window Flow Control at Higher Layers

- content of foregoing discussion is applied directly
- possibly communicating across several interconnected networks

Window Flow Control at Higher Layers



Dynamic Window Size Adjustments

Dynamic Window Size Adjustments

- see § 6.2.5 of [Bertsekas-Gallager-1992]