Introduction to Computer Networks

CS640

Computer Networks: Performance Analysis

https://pages.cs.wisc.edu/~mgliu/CS640/S25/index.html

Ming Liu
mgliu@cs.wisc.edu

Outline

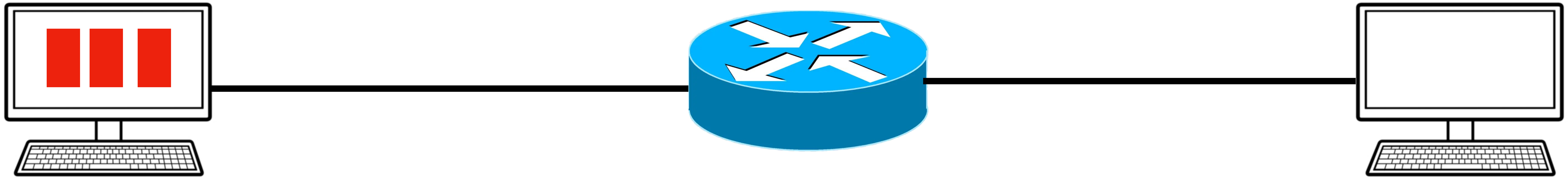
- Last
  - Computer networks: hardware infrastructure
  - Computer networks: software system

- Today
  - Delay
  - Throughput
  - RTT and BDP

- Announcements
  - Lab1 will be released this Thursday
  - No class on Thursday

Packet-Switched Networks

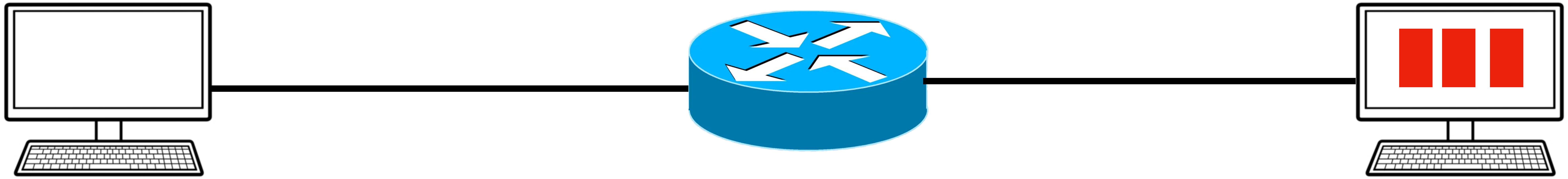
- Packet communication path
  - Host 1 —> Router A —> Router B —> … —> Host 2



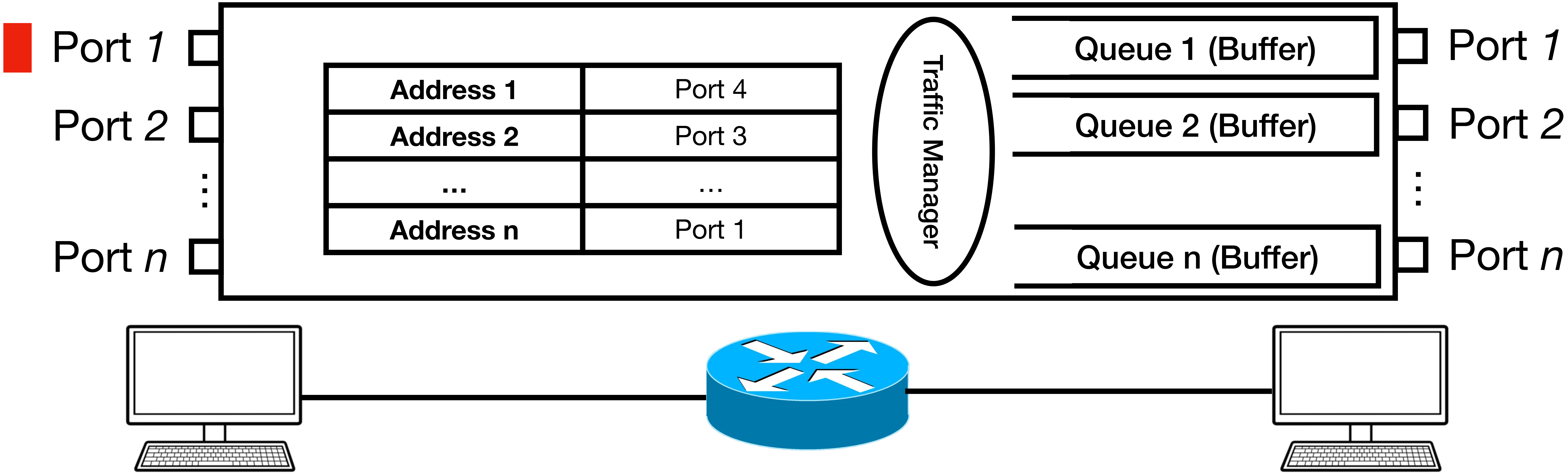
Packet-Switched Networks

- Packet communication path
  - Host 1 —> Router A —> Router B —> … —> Host 2

- Delay
  - The total amount of time transfers N bits across the packet path
  - Four types of delay

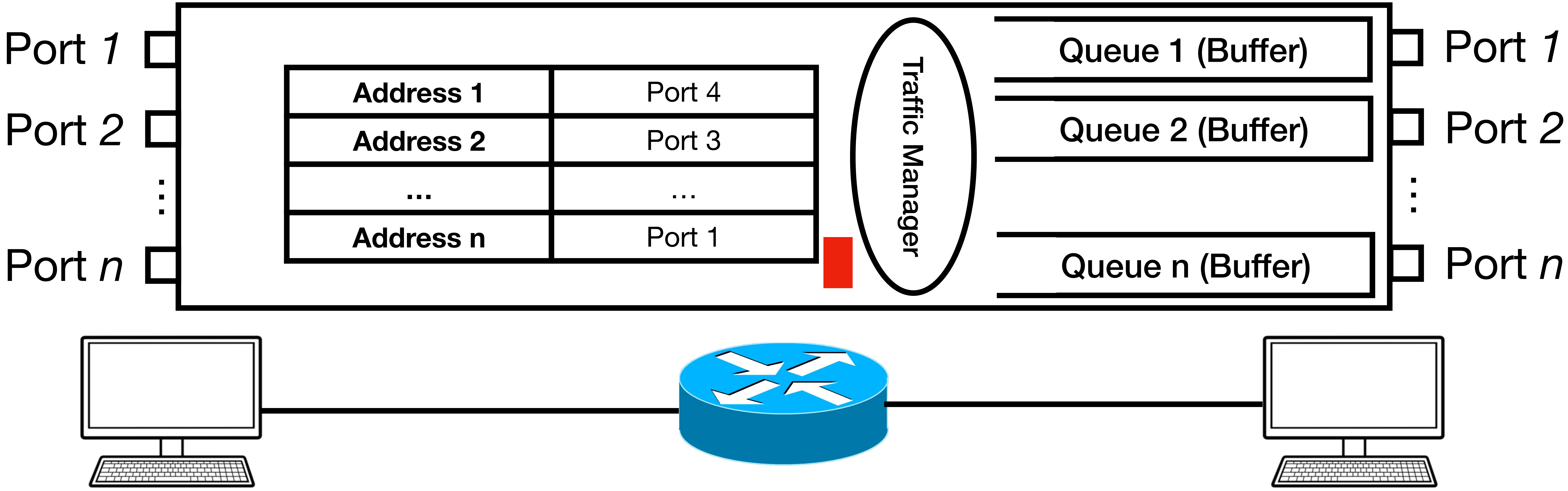


Processing Delay (T_{proc})

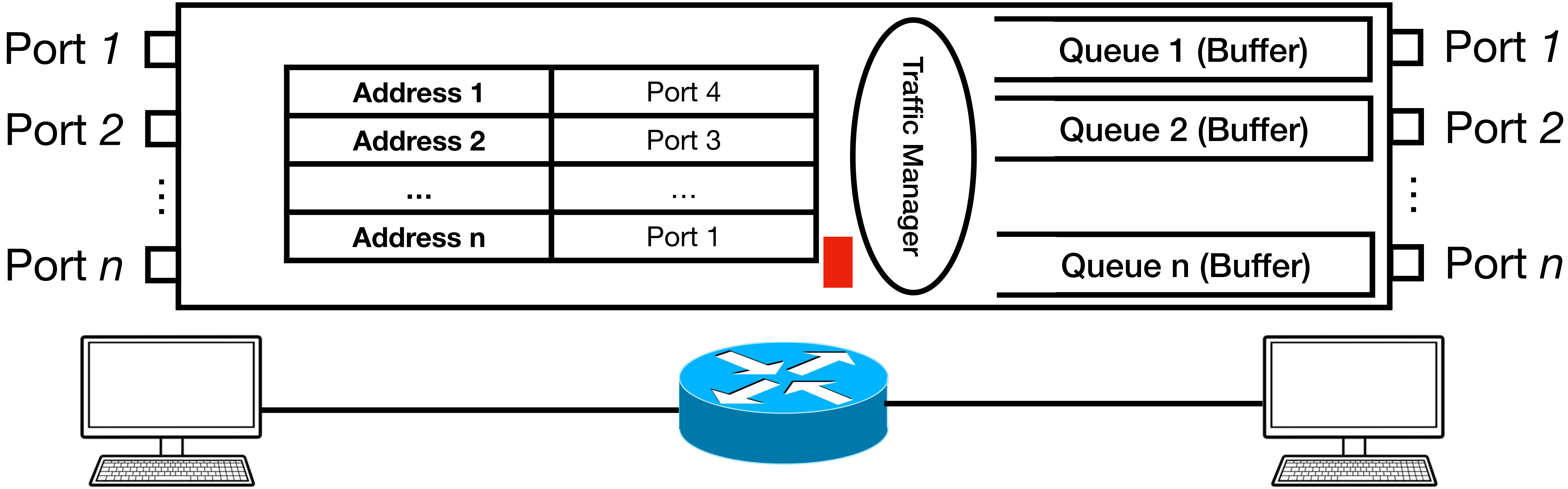


Processing Delay (T_{proc})

- The time required to examine the packet header and determine where to forward
  - Including checking bit-level errors
  - Microsecond

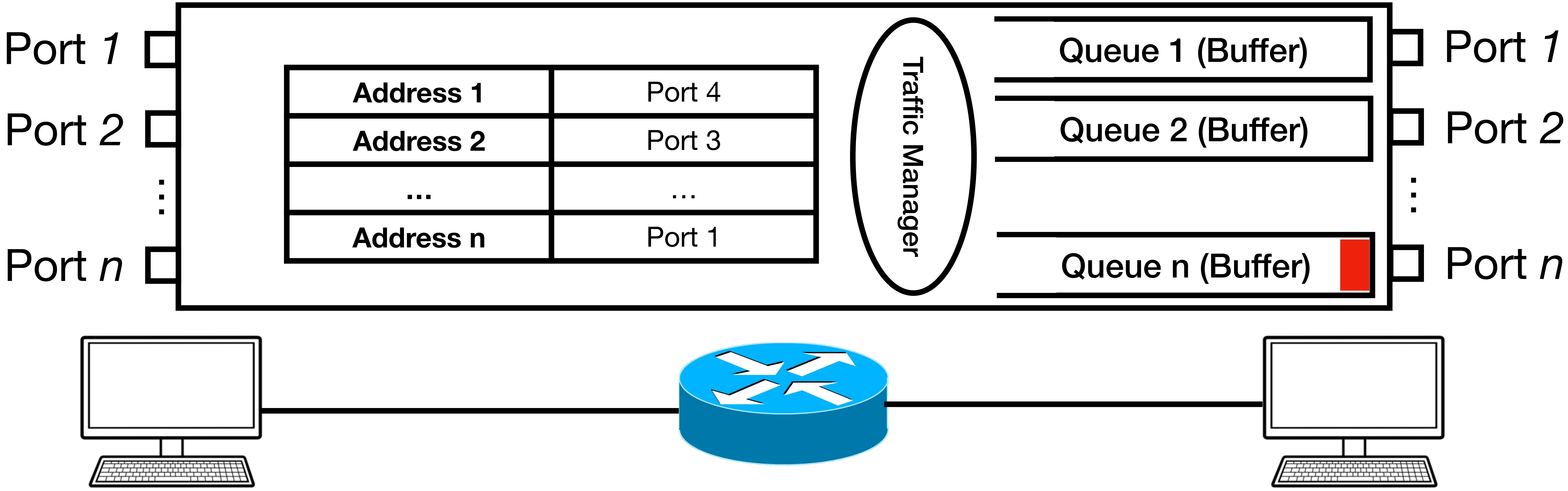


Queueing Delay (T_{queue})

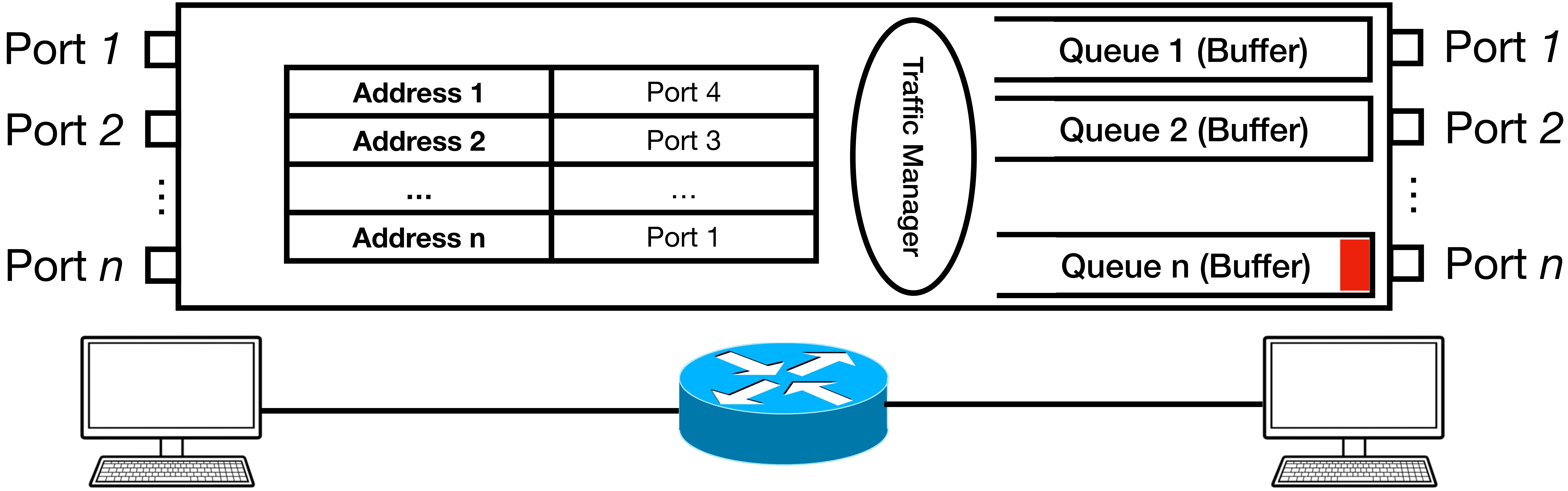


Queueing Delay (T_{queue})

- The time it takes to wait to be transmitted
  - Depend on the number of earlier-arriving packets
  - Microsecond~Millisecond

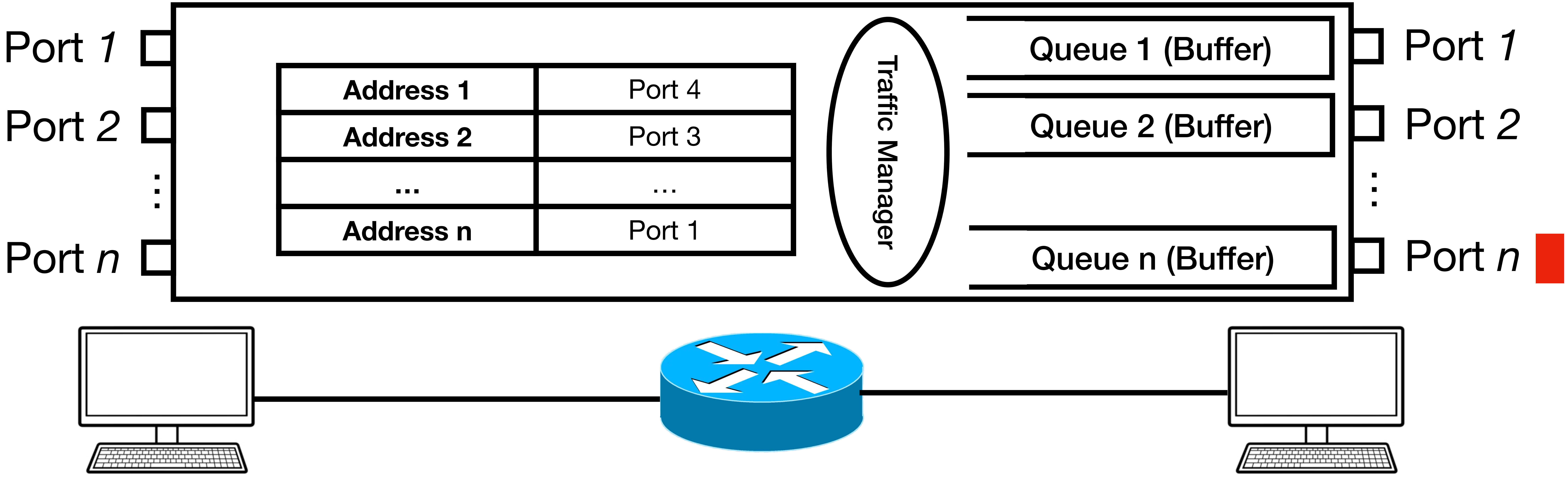


Transmission Delay (T_{trans})



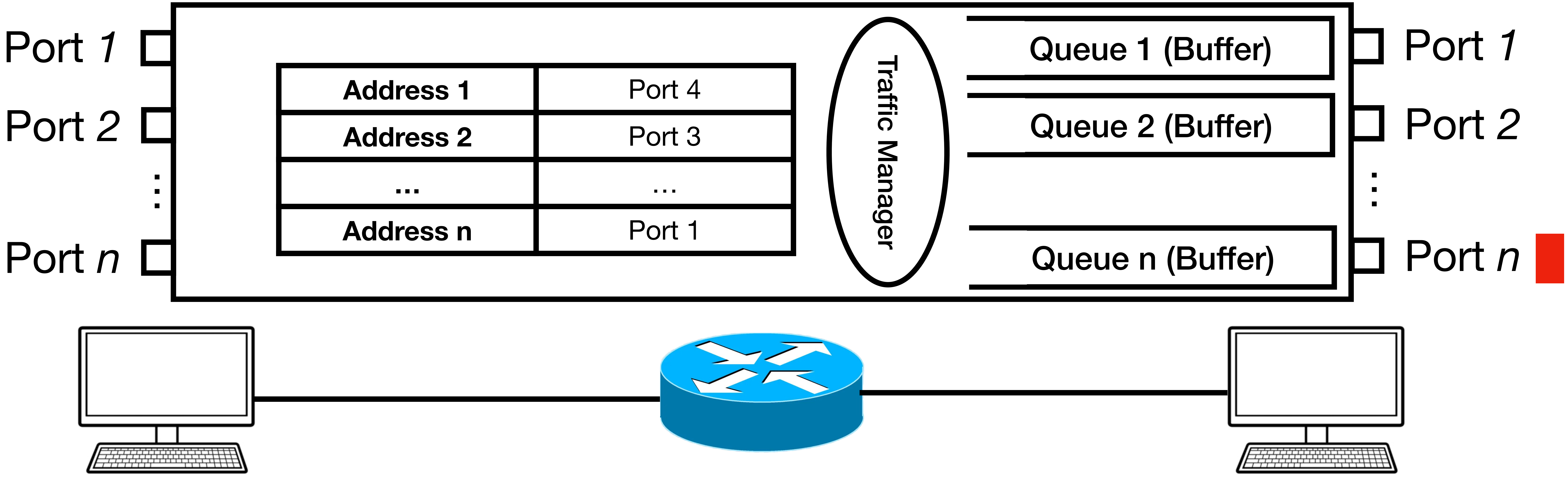
Transmission Delay (T_{trans})

- The time it takes to transmit through the communication ports
  - First-come-first-served manner
  - Depend on the packet length and transmission rate

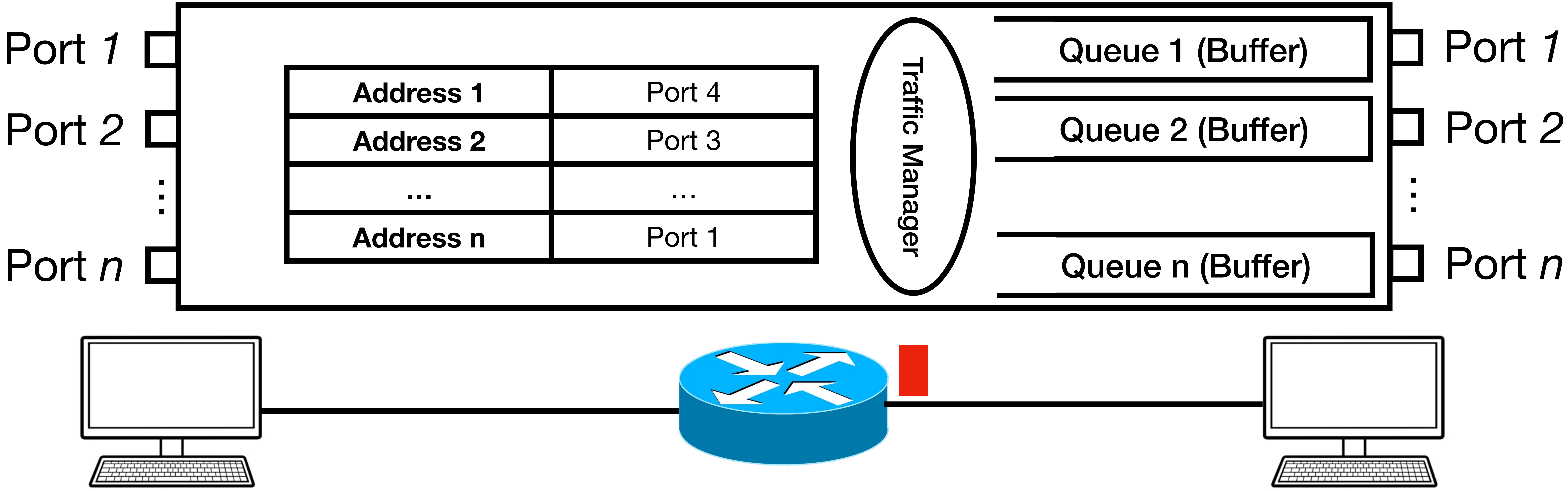


Transmission Delay Calculation

- Packet length (L bits) and Transmission Rate (R bits/sec)
  - Transmission Delay = L/R
  - Suppose L=1KB, R = 10Mbps/100Mbps, what is the transmission delay?

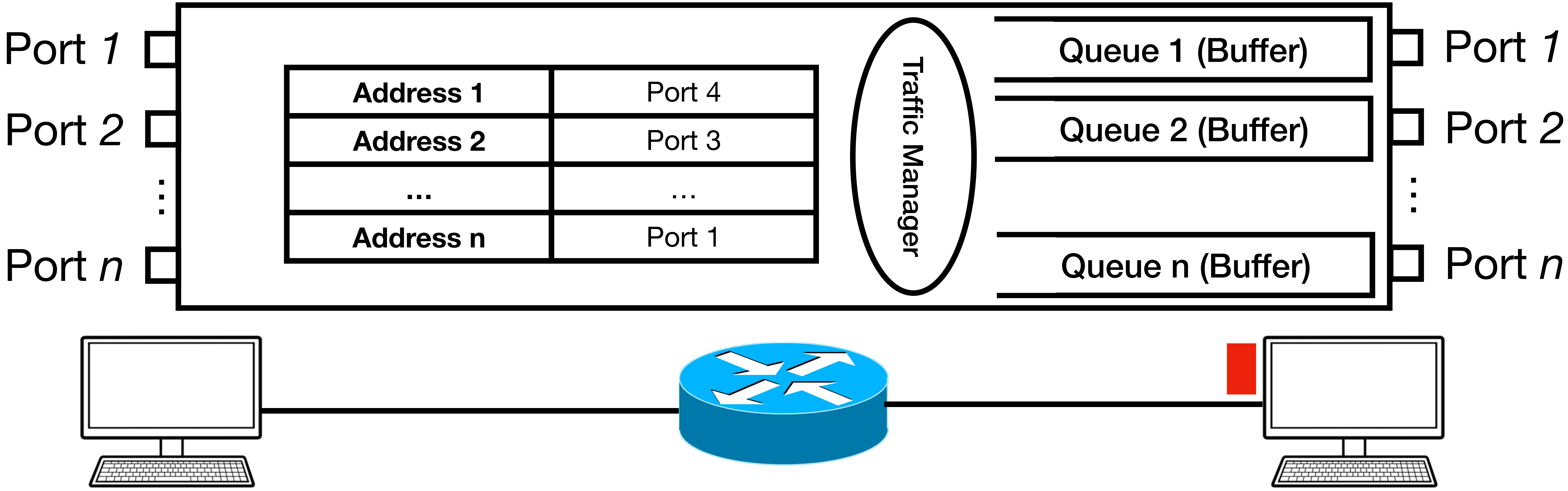


Propagation Delay (T_{prop})



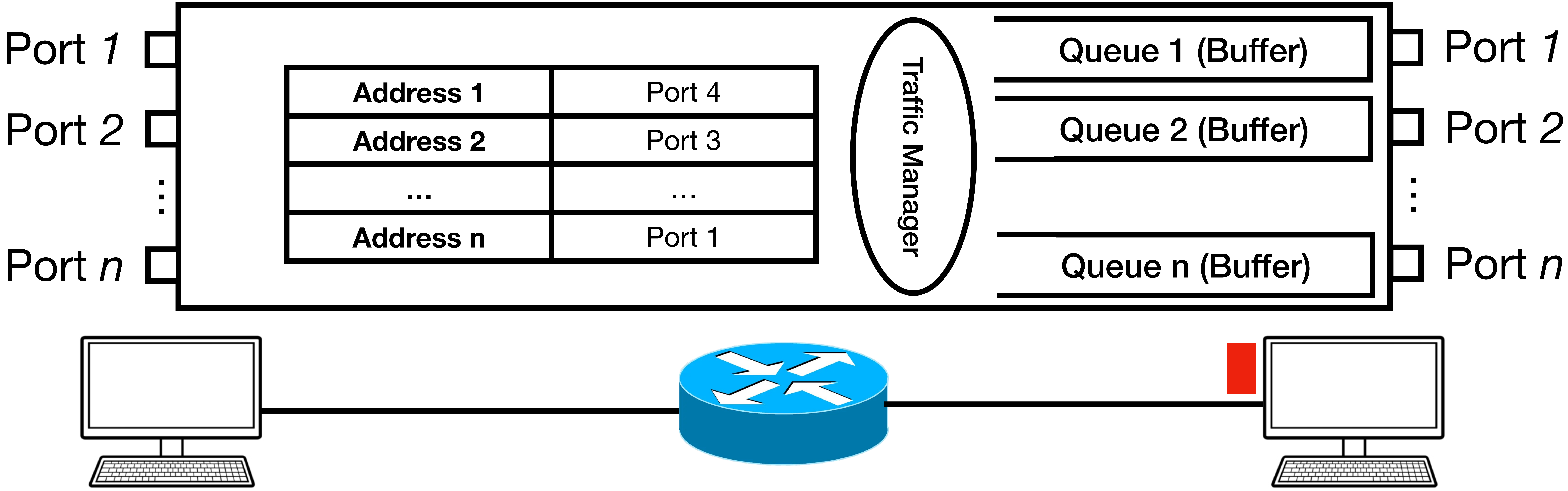
Propagation Delay (T_{prop})

- The time required to propagate over the communication link
  - Depend on the physical media (e.g., fiber optics, copper wire, etc.)
  - 2x10^8 ~ 3x10^8 meters/second



Propagation Delay Calculation

- Distance (d) and propagation speed (s)
  - Propagation delay = d/s
  - Distance matters!



Transmission Delay v.s. Propagation Delay

- Difference
  - Transmission delay: time taken to push out the packet
  - Propagation delay: time taken to traverse the link

Transmission Delay v.s. Propagation Delay

- Difference
  - Transmission delay: time taken to push out the packet
  - Propagation delay: time taken to traverse the link

- Suppose
  - Packet length (L) = 1.5KB and transmission rate (R) = 1Mbps
  - Distance (d) = 3m and propagation speed (s) = 3^8 m/s
  - What is the transmission delay and propagation delay?

Transmission Delay v.s. Propagation Delay

- Difference
  - Transmission delay: time taken to push out the packet
  - Propagation delay: time taken to traverse the link

- Suppose
  - Packet length (L) = 1.5KB and transmission rate (R) = 1Mbps
  - Distance (d) = 3m and propagation speed (s) = 3^8 m/s
  - What is the transmission delay and propagation delay?
  - If we increase the transmission rate to 1Gbps, what happens?

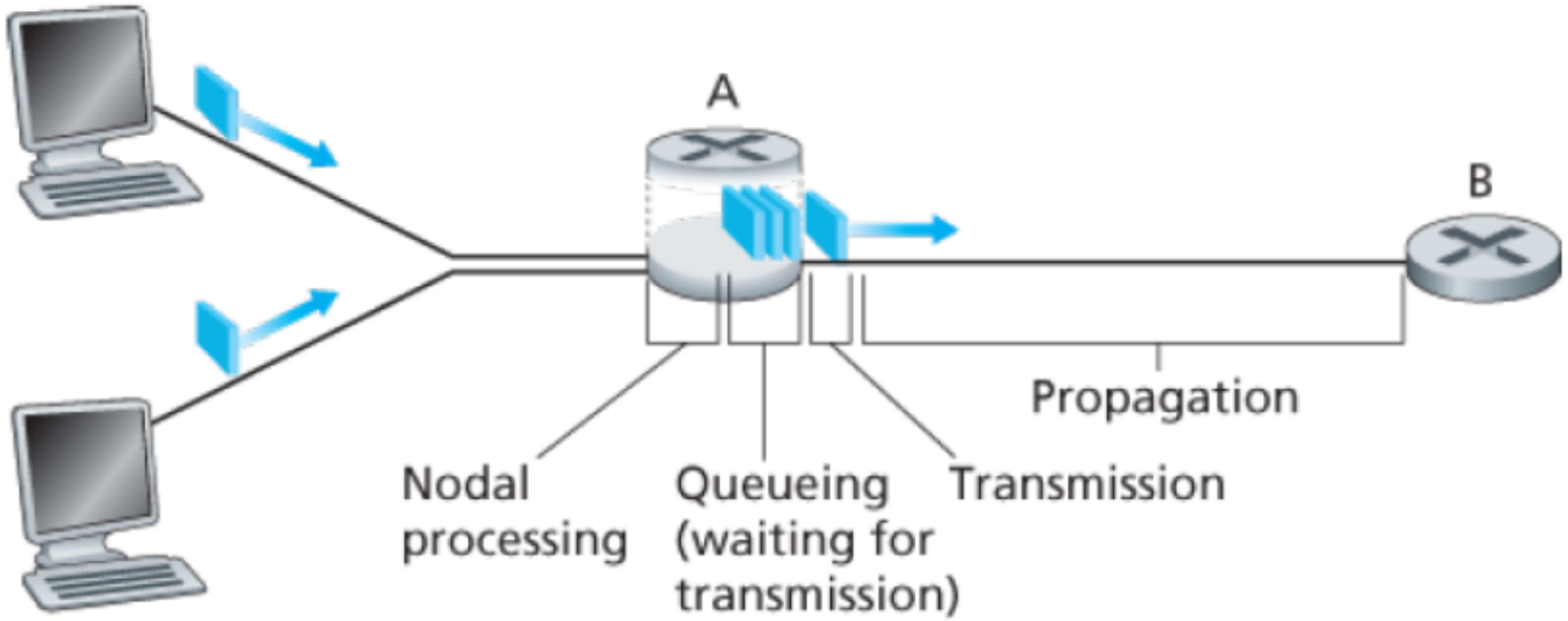
Transmission Delay v.s. Propagation Delay

- Difference
  - Transmission delay: time taken to push out the packet
  - Propagation delay: time taken to traverse the link

- Suppose
  - Packet length (L) = 1.5KB and transmission rate (R) = 1Mbps
  - Distance (d) = 3m and propagation speed (s) = 3^8 m/s
  - What is the transmission delay and propagation delay?
  - If we increase the transmission rate to 1Gbps, what happens?
  - If we increase the distance to 3km, what happens?

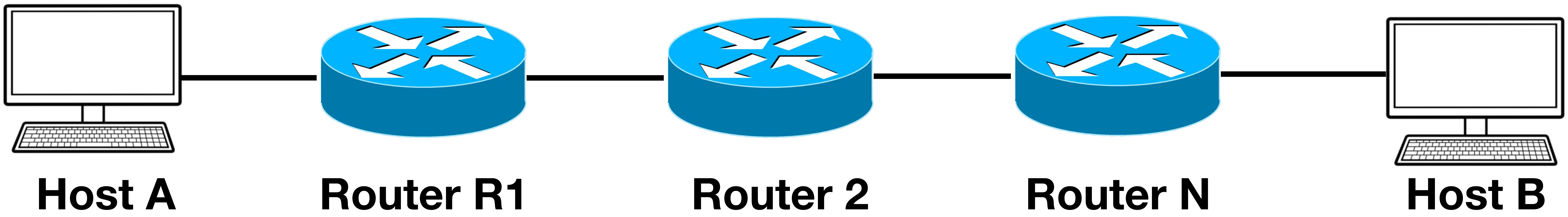
Total (Nodal) Delay

- $T_{total} = T_{proc} + T_{queue} + T_{trans} + T_{prop}$
  - Per-node, also called total nodal delay



End-to-End Delay

- Suppose
  - Host A can send a packet infinitely fast
  - Host B can receive a packet infinitely fast
- What is the end-to-end delay for a packet from host A to host B?
  - $T_{end\text{-}to\text{-}end} = N\,(T_{proc} + T_{queue} + T_{trans} + T_{prop}) + T_{prop}$



Queueing Delay

- The most complicated and interesting one
  - Varying packet to packet and scenario to scenario
- Statistical metrics
  - Average queueing delay
  - Variance of queueing delay
  - Tail queueing delay
  - The probability that the queuing delay exceeds some value

When the Queue is built up?

When the Queue is built up?

- Packet arrival rate **a**
- Unit: packets/second

When the Queue is built up?

- Packet arrival rate **a**
- Unit: packets/second

- Transmission rate **R**
- Unit: bits/second

When the Queue is built up?

- Given time T, incoming traffic load> outgoing traffic load
  - Incoming traffic load = $\sum_{i=1}^{N} pkt\_size_i$
  - Out going traffic load = R * T

- Packet arrival rate **a**
- Unit: packets/second

- Transmission rate **R**
- Unit: bits/second

Traffic Intensity

- We define the ratio (La/R) as the traffic intensity
  - Suppose all packets consist of L bits

Traffic Intensity

- We define the ratio (La/R) as the traffic intensity
  - Suppose all packets consist of L bits

- La/R > 1: queue built up!
  - The average rate at which bits arrive at the queue exceeds the rate at which the bits can be transmitted from the queue

Traffic Intensity

- We define the ratio (La/R) as the traffic intensity
  - Suppose all packets consist of L bits
- La/R > 1: queue built up!
  - The average rate at which bits arrive at the queue exceeds the rate at which the bits can be transmitted from the queue

Design your system so that the traffic intensity is no greater than 1!

Traffic Intensity

- We define the ratio (La/R) as the traffic intensity
  - Suppose all packets consist of L bits

- La/R > 1: queue built up!
  - The average rate at which bits arrive at the queue exceeds the rate at which the bits can be transmitted from the queue

Can we see queueing when La/R <=1?

Traffic Arrival process

- Suppose one packet arrives every L/R seconds

Traffic Arrival process

- Suppose one packet arrives every L/R seconds
  - No queueing

- Suppose N packets arrive every (L/R)N seconds

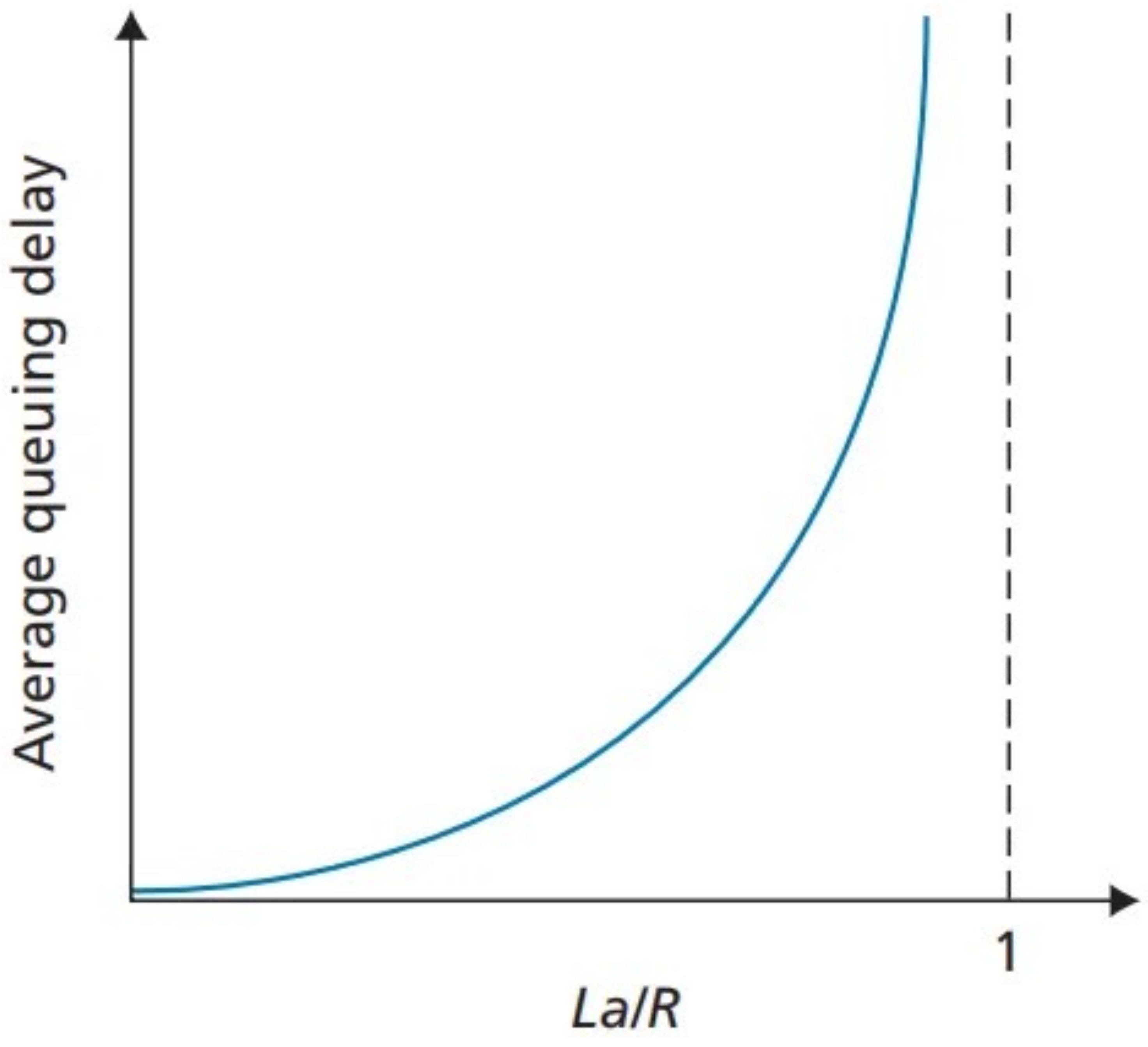
Traffic Arrival process

- Suppose one packet arrives every L/R seconds
  - No queueing

- Suppose N packets arrive every (L/R)N seconds
  - The 1st packet has no queueing delay, $T_{queue} = 0$
  - The 2nd packet has to wait for the 1st one, $T_{queue} = L/R \times 1$
  - The 3nd packet has to wait for the 1st and 2nd ones, $T_{queue} = L/R \times 2$
  - The nth packet has to wait for the (n-1) ones, $T_{queue} = L/R \times (n-1)$

Traffic Intensity Curve

- Close to 1, the average queueing delay increases
  - A small percentage increase causes a significant delay increase
  - In reality, queue is fixed-sized ==> **Packet Loss**

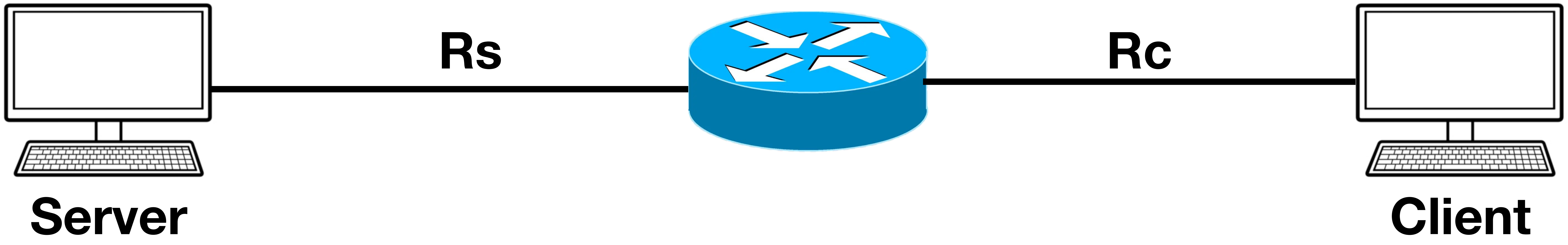


Throughput

- Bandwidth: the number of bits transmitted per second at a communication port and link
  - bps, Kbps, Mbps, Gbps, Tbps, …
- Throughput: the number of bits transmitted from A to B
  - A and B can be host, switch, etc.
  - bps, Kbps, Mbps, Gbps, Tbps, …

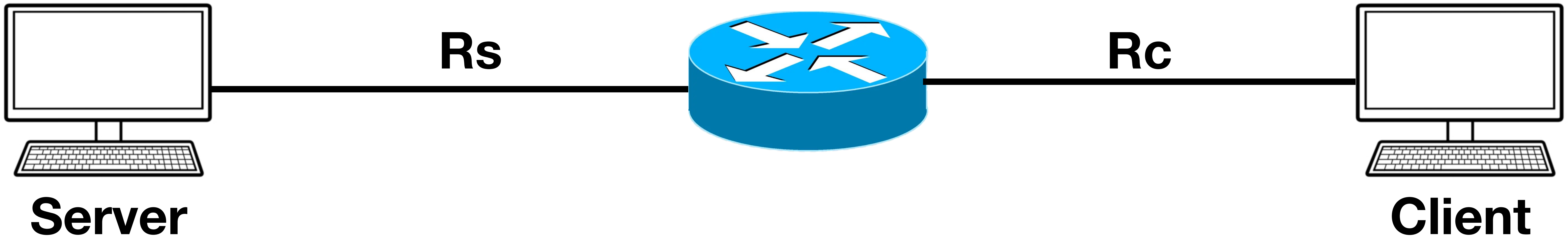
A Simple Throughput Example

- A file transferring from a server to a client



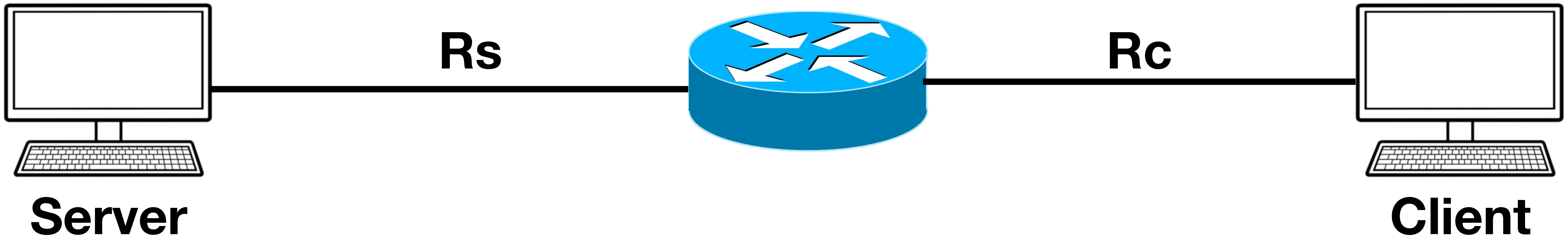
A Simple Throughput Example

- A file transferring from a server to a client
  - If Rs < Rc ==> The client receives the file at Rs
  - If Rs > Rc ==> The client receives the file at Rc, but router is queued



A Simple Throughput Example

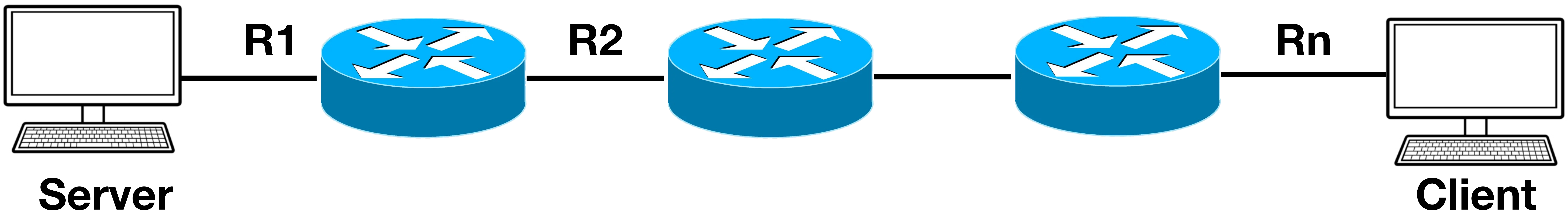
- A file transferring from a server to a client
  - If Rs < Rc ==> The client receives the file at Rs
  - If Rs > Rc ==> The client receives the file at Rc, but router is queued
- Throughput = min {Rs, Rc}
  - Depend on the bottleneck link
- A log file = 32M bits, Rs = 2Mbps, Rc = 1Mbps
  - Transfer time is 32 seconds



Another Throughput Example

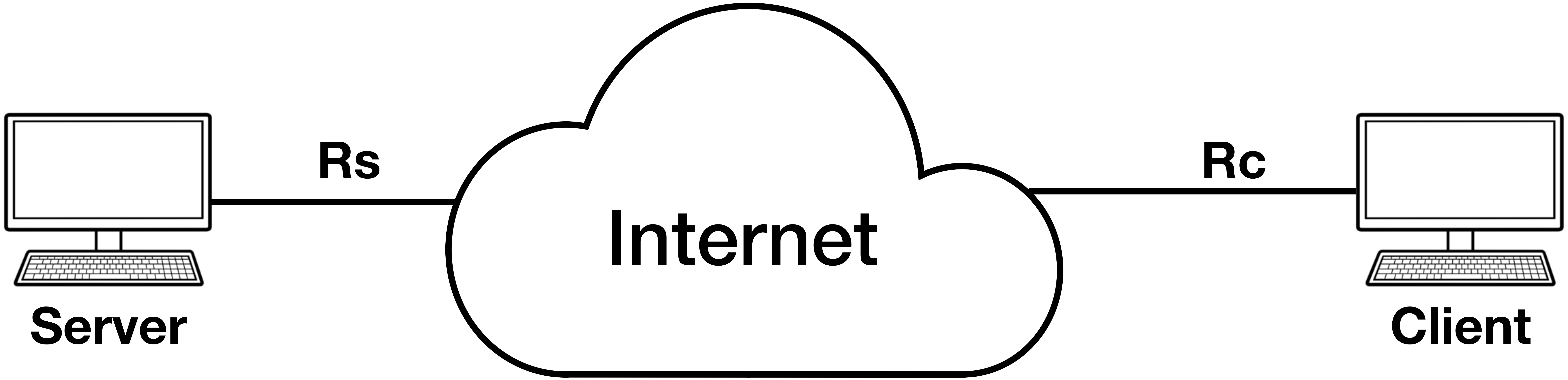
- Communication path: server -> a list of routers -> Client

- Throughput = min {R1, R2, R3, …, Rn}
  - Depend on the bottleneck link



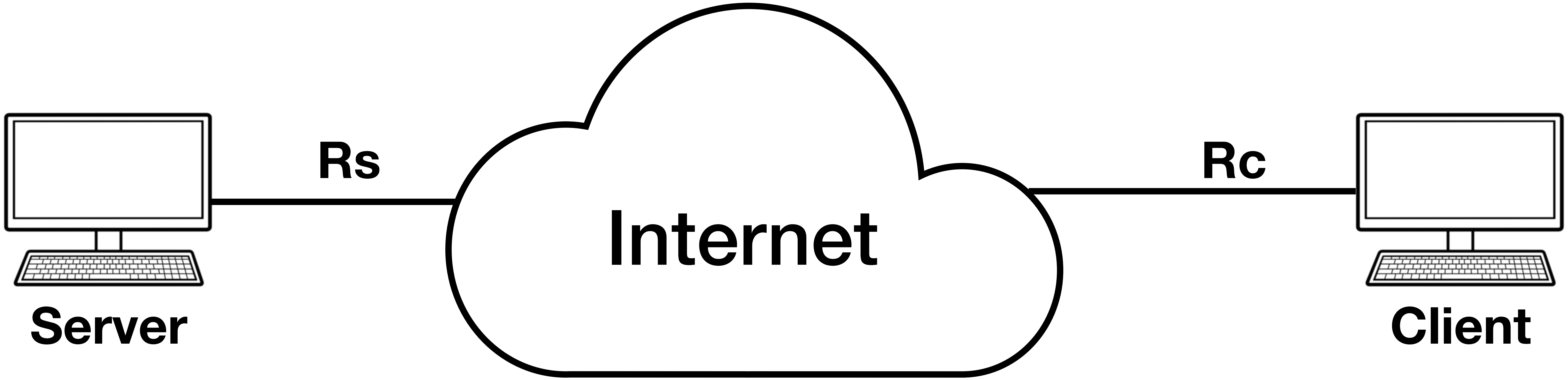
Throughput in a Shared Network

- Impossible to know the communication path details
- What is the throughput to transfer a file from a server to a client?



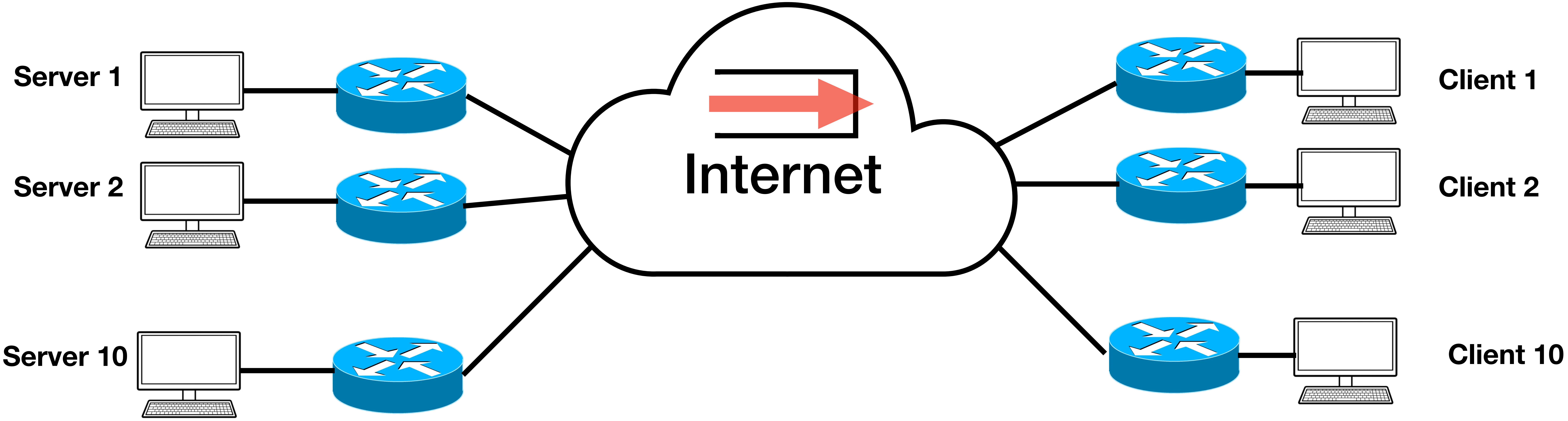
Throughput in a Shared Network

- Impossible to know the communication path details
- What is the throughput to transfer a file from a server to a client?
  - Actual Throughput = File Size / Total Transfer Time
  - Actual Throughput <= min {Rs, Rc}



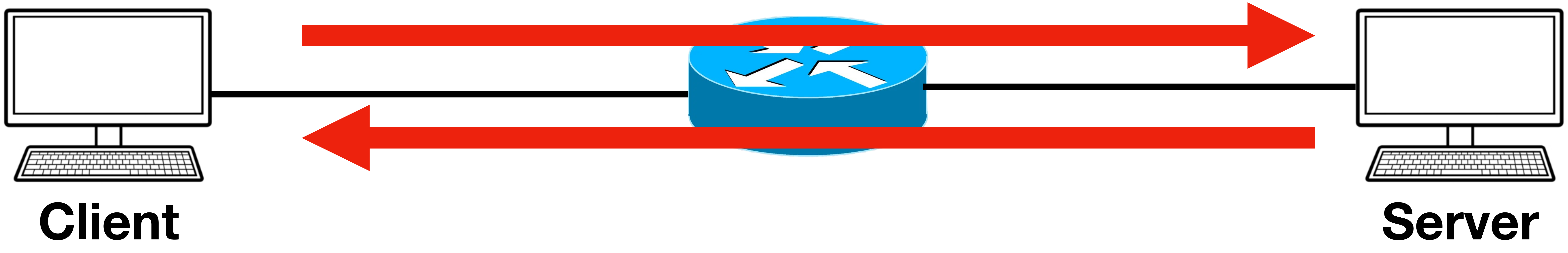
Throughput under Concurrent Transmissions

- Throughput also depends on intervening traffic



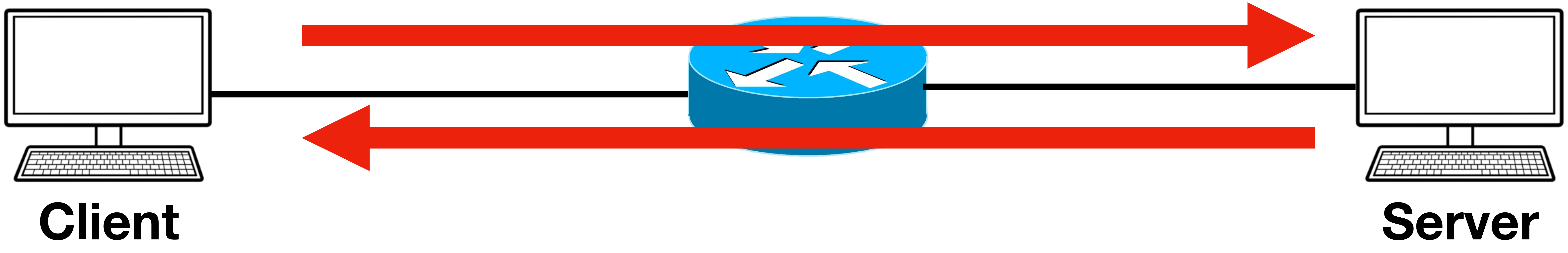
RTT (Round-Trip Time)

- The time it takes to send a request from a client to a server and receive the response back
  - RTT = Total delay (Client -> Server) + Total Delay (Server -> Client)



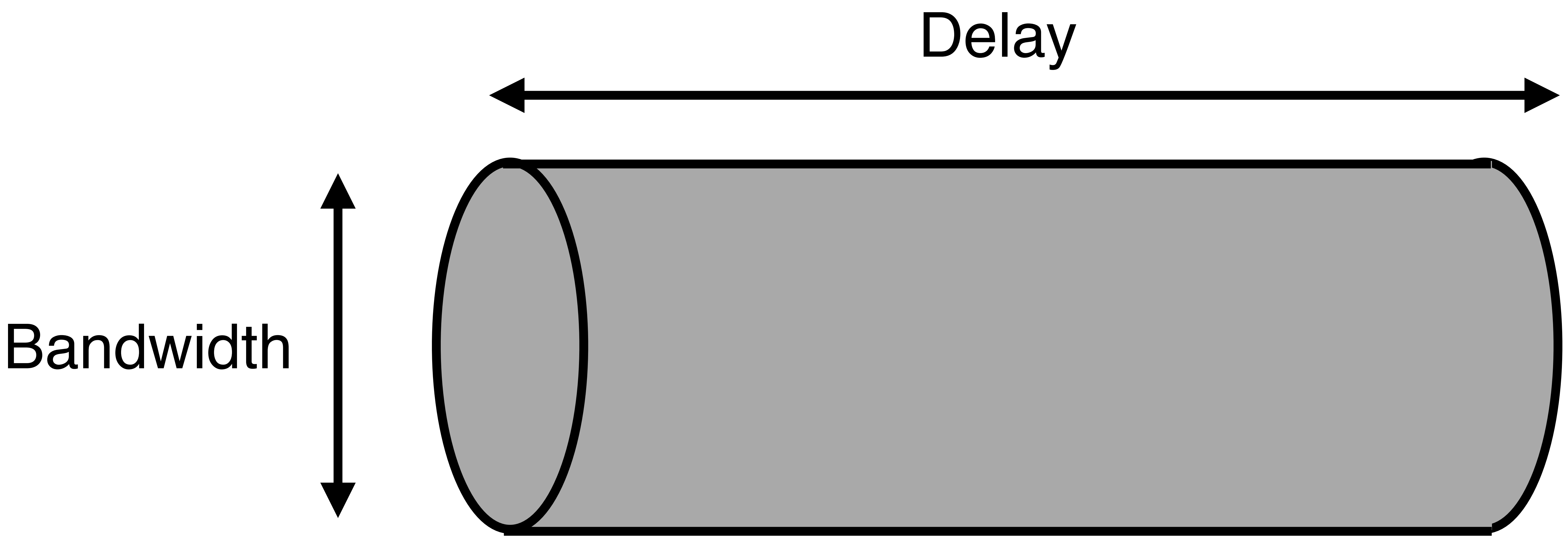
RTT (Round-Trip Time)

- The time it takes to send a request from a client to a server and receive the response back
  - RTT = Total delay (Client -> Server) + Total delay (Server -> Client)
- RTT is application-dependent
  - Web browsing: page download time (time to retrieve the first object)
  - Cloud gaming: interactive latency
  - Video conferencing and streaming: Time to First Frame
  - LLM: Time to first token



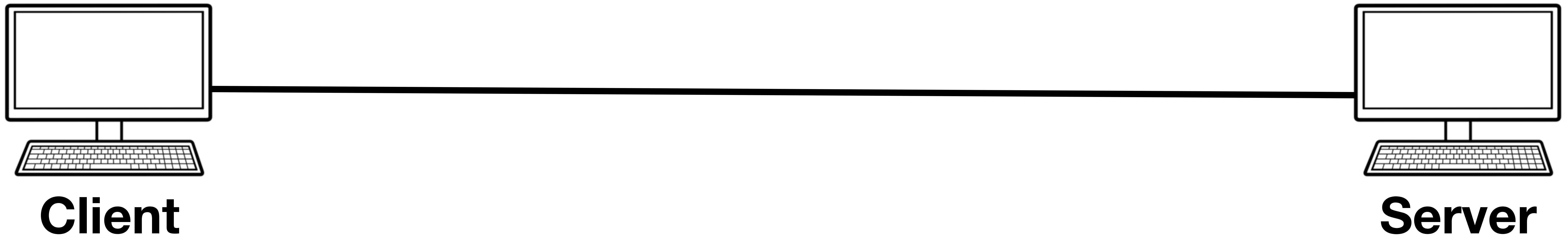
BDP (Bandwidth-Delay Product)

- The volume of a data pipe for one transmission
  - Bandwidth * Delay
  - The number of bits have left the sender and are yet to reach the receiver



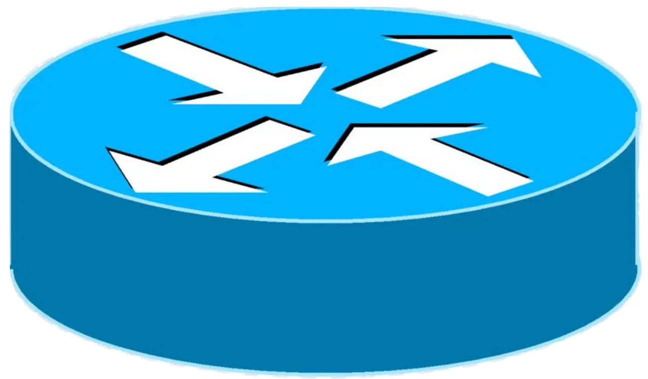
BDP of a Communication Link

- Delay=Propagation delay ==> Link BDP
  - The number of bits over the wire



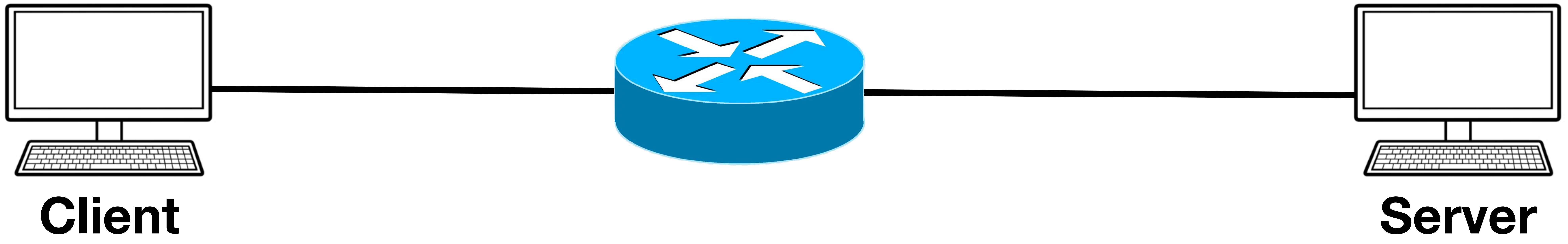
BDP of a Router

- Delay=Propagation delay ==> Link BDP
  - The number of bits over the wire
- Delay=Processing + Queueing + Transmission ==> Router BDP
  - The number of bits a router can hold



BDP of Client —> Server

- Delay=Propagation delay ==> Link BDP
  - The number of bits over the wire
- Delay=Processing + Queueing + Transmission ==> Router BDP
  - The number of bits a router can hold
- Delay=Total delay ==> End-to-End BDP
  - The number of bits that stay in-flight between two hosts



Summary

- Today
  - Computer networks: performance analysis

- Next lecture
  - Physical Layer: Encoding