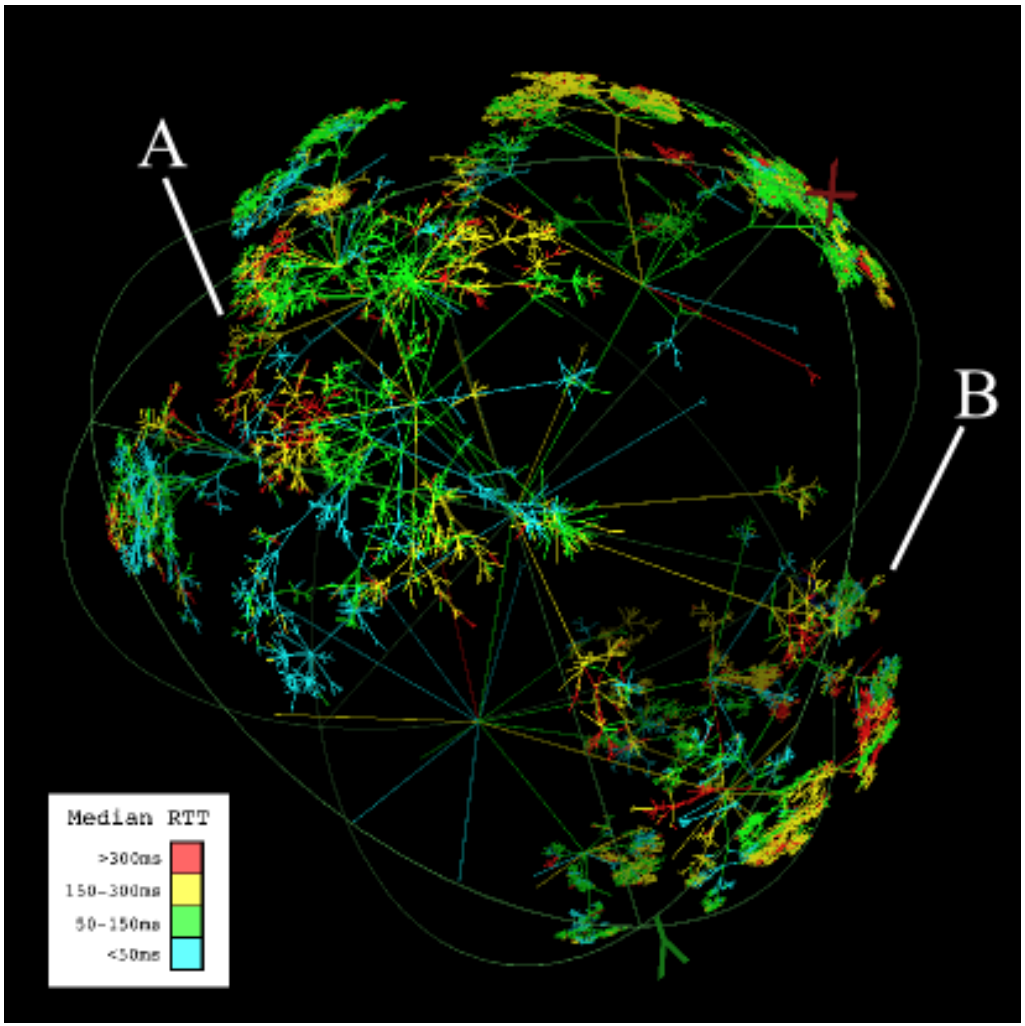
Delay, Loss, and Throughput

Daniel Zappala

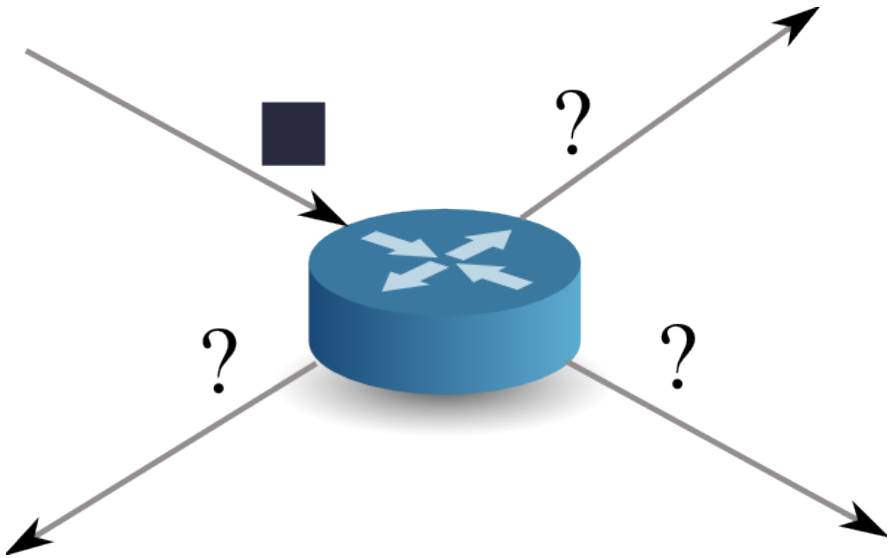
CS 460 Computer Networking
Brigham Young University

Delay



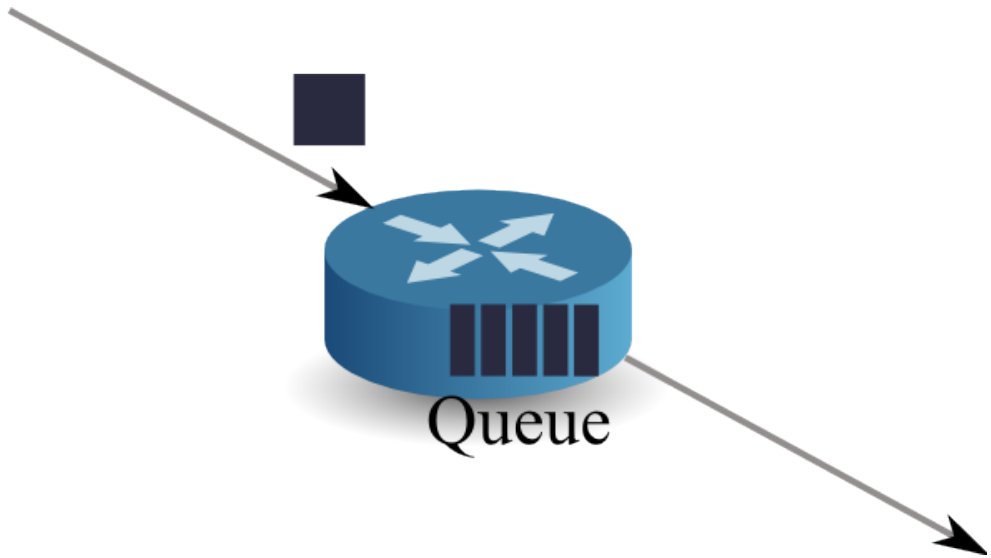
- How fast will your data get from point A to point B?

Processing Delay



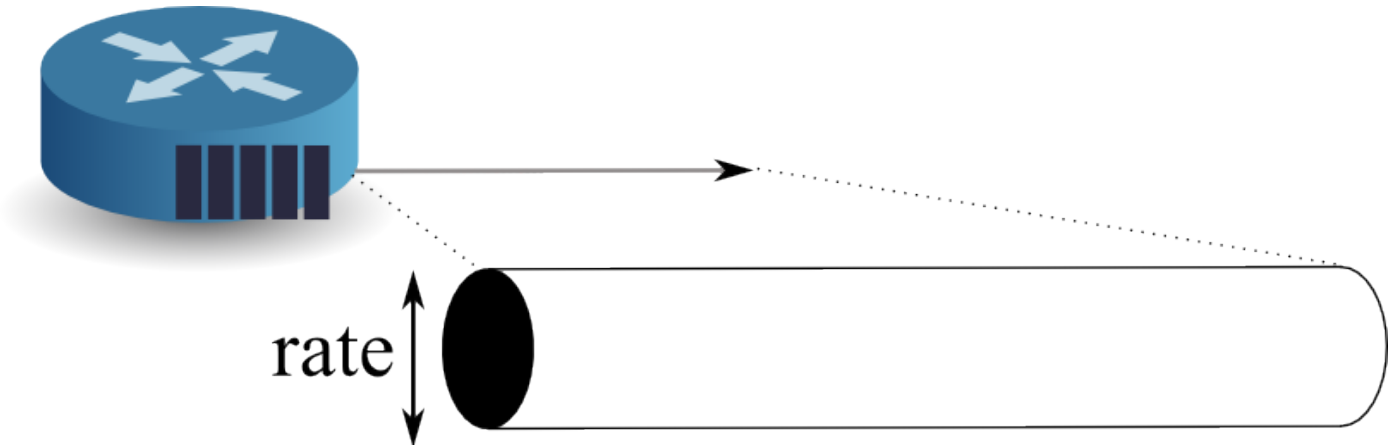
- the time it takes a node to check a message for errors, decide where it goes next, and queue it for delivery

Queueing Delay



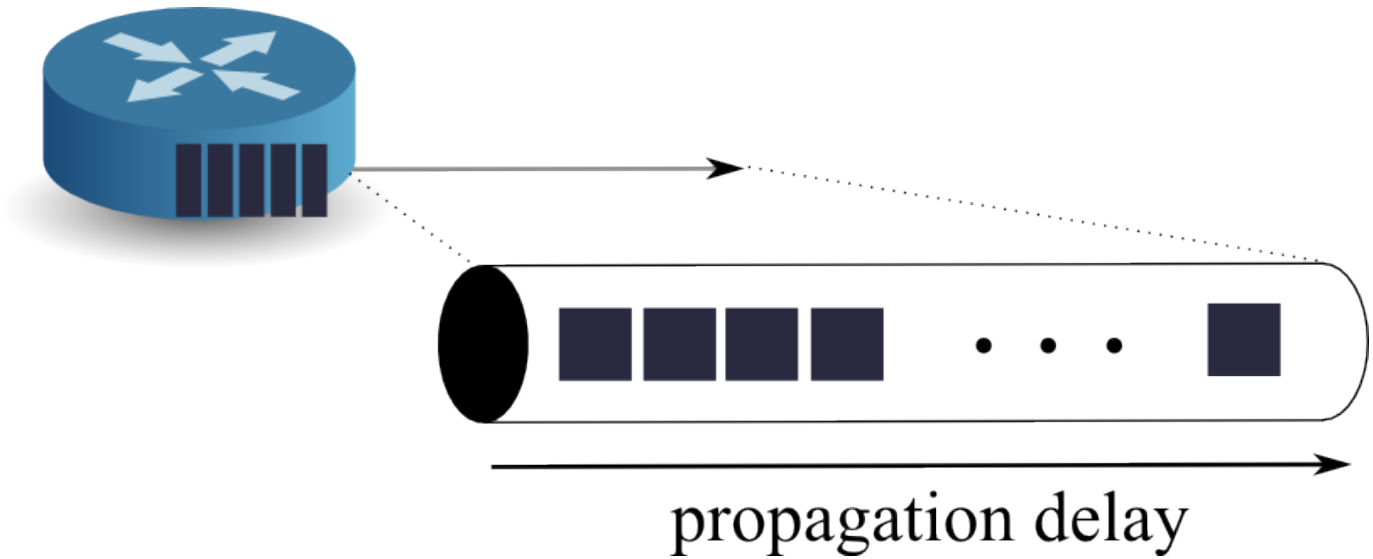
- the time a message spends waiting in a queue before being sent on a link
- most network queues are FIFO

Transmission Delay



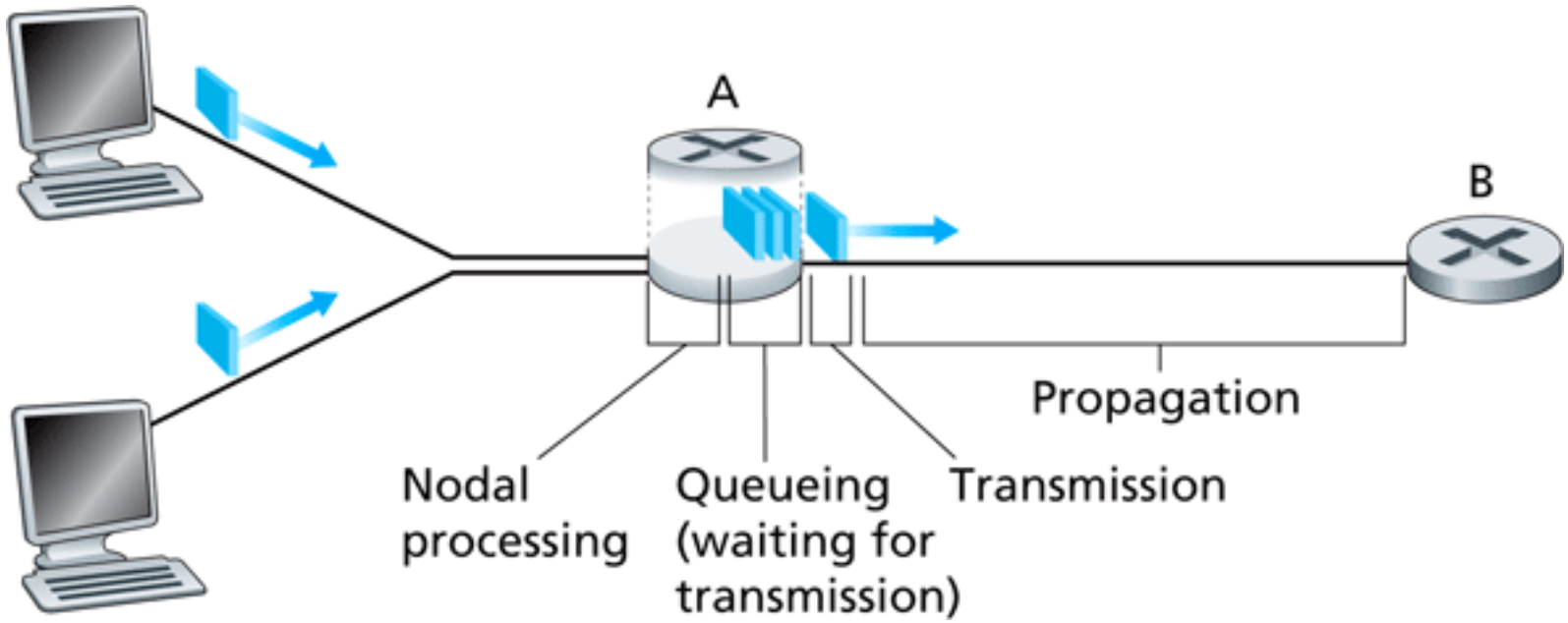
- the time it takes for the source to send a message on the link
- given by L/R
  - L = length of the message in bits
  - R = rate of the link, in bits/second

Propagation Delay



- the time it takes for the message to travel down the link
- limited by the speed of light
- speed of electricity in cable: $s = 2 * 10^8 m/s$
- propagation delay $= d/s$, d is the length of the link in meters

Sources of Delay



Queueing Theory

Observations and Questions

- transmission delay and propagation delay are fixed, depend on the type of link
- processing delay depends on CPU speed, CPU load, complexity of table lookup
- queueing delay depends on load
  - how often do packets arrive?
  - how big are they?
  - how do we model queueing delay?

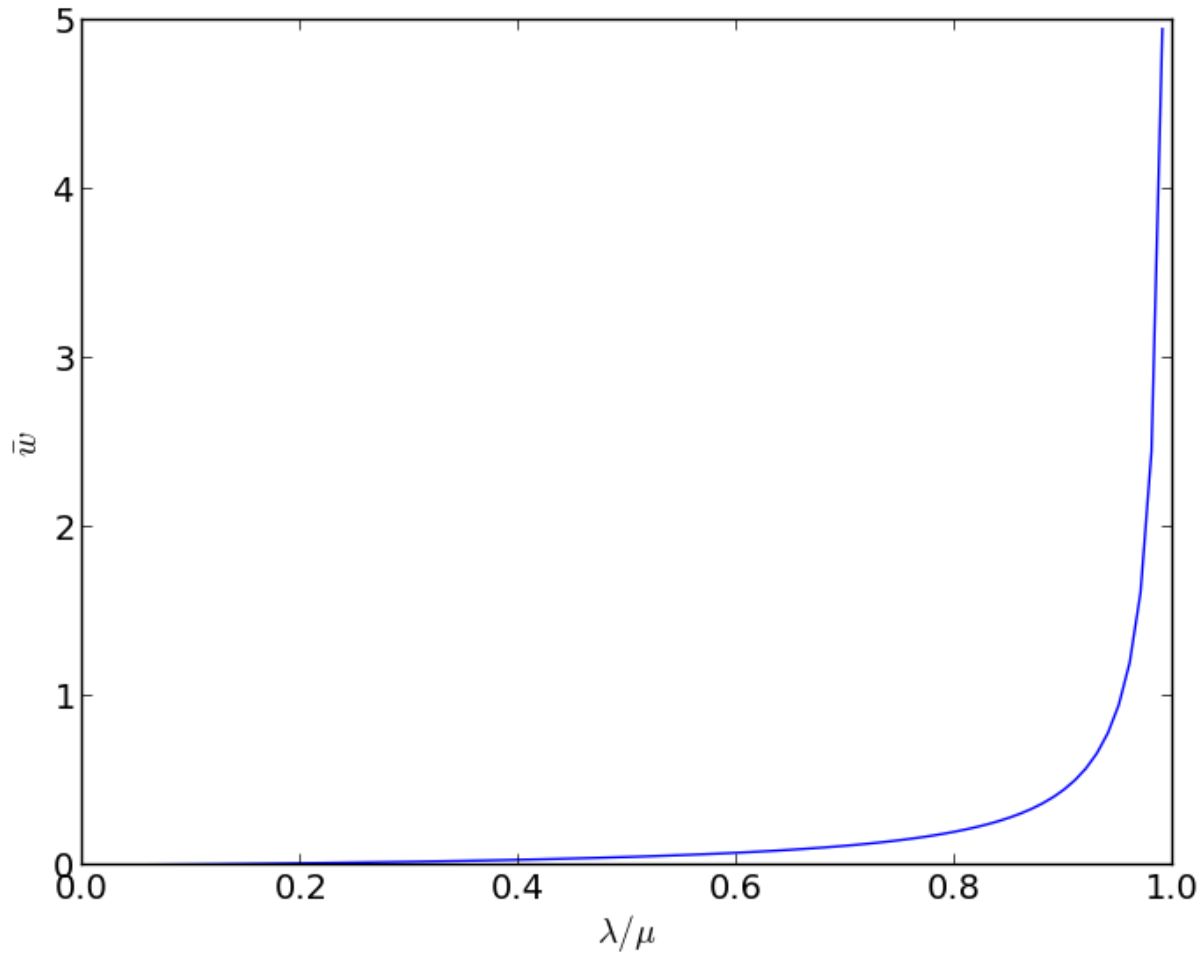
Basic Definitions

- arrival process: probability distribution to model arriving packets
- service process: probability distribution to model service time, equivalent to probability distribution for packet size
- number of servers: how many links are sending to the destination, one in the simplest case

M/D/1 Queue

- definition
  - M = exponential arrival rate (Poisson process)
  - D = deterministic service rate
  - 1 = 1 server (or 1 link to the destination)
  - **unlimited queue size**
- parameters
  - λ = mean arrival rate (packets per second)
  - $\mu = R/L$ = mean service rate (packets per second)
  - $\rho = \lambda/\mu$
- forumulas
  - average queue length, $\bar{Q} = \frac{\rho^2}{2(1-\rho)}$
  - average wait time in queue, $\bar{w} = \frac{1}{2\mu}(\frac{\rho}{1-\rho})$
  - average time in system, $\bar{t} = \frac{1}{2\mu}(\frac{2-\rho}{1-\rho})$

M/D/1 Queueing Delay



- as ρ approaches 1, delay becomes infinite

Packet Loss

- what causes packet loss?
  - queue overflow
  - what causes the queue to overflow?
  - packets arriving faster than they can be serviced
  - $\rho > 1$
- can packet loss happen with $\rho < 1$?

Calculating Delay

Single Queue

- What if you want to send a message between two directly-connected machines?

Calculating Delay over a Single Queue

- ignore processing delay (very short in modern routers)
- problem
  - 2 machines, joined by 1 link
  - R = 10 Mbps, d = 1000 m
  - L = 10,000 bits, divided into 10 packets of 1,000 bits
  - D = total delay from first machine to second machine = ?
- solution
  - transmission delay = $L/R = 100\mu s$
  - propagation delay = $d/s = 5\mu s$
  - one packet takes $L/R + d/s = 105\mu s$

Calculating Delay over a Single Queue

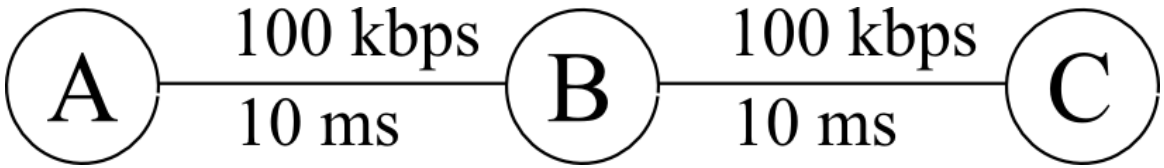
- ignore processing delay (very short in modern routers)
- problem
  - 2 machines, joined by 1 link
  - R = 10 Mbps, d = 1000 m
  - L = 10,000 bits, divided into 10 packets of 1,000 bits
  - D = total delay from first machine to second machine = ?
- solution
  - transmission delay = $L/R = 100\mu s$
  - propagation delay = $d/s = 5\mu s$
  - one packet takes $L/R + d/s = 105\mu s$
  - 10 packets take $10 * (L/R + d/s) = 1050\mu s = 1.05ms$?

Parallelism

- parallelism in propagation delay
  - transmission delay $= L/R = 100\mu s$
  - propagation delay $= d/s = 5\mu s$
  - propagation delay so small that packet 1 arrives at 2nd machine before it is finished being transmitted at first machine
  - as last part of a packet travels down the link, the next packet can be sent on the link
- $D = 10 * L/R + d/s = 1005\mu = 1.005ms$

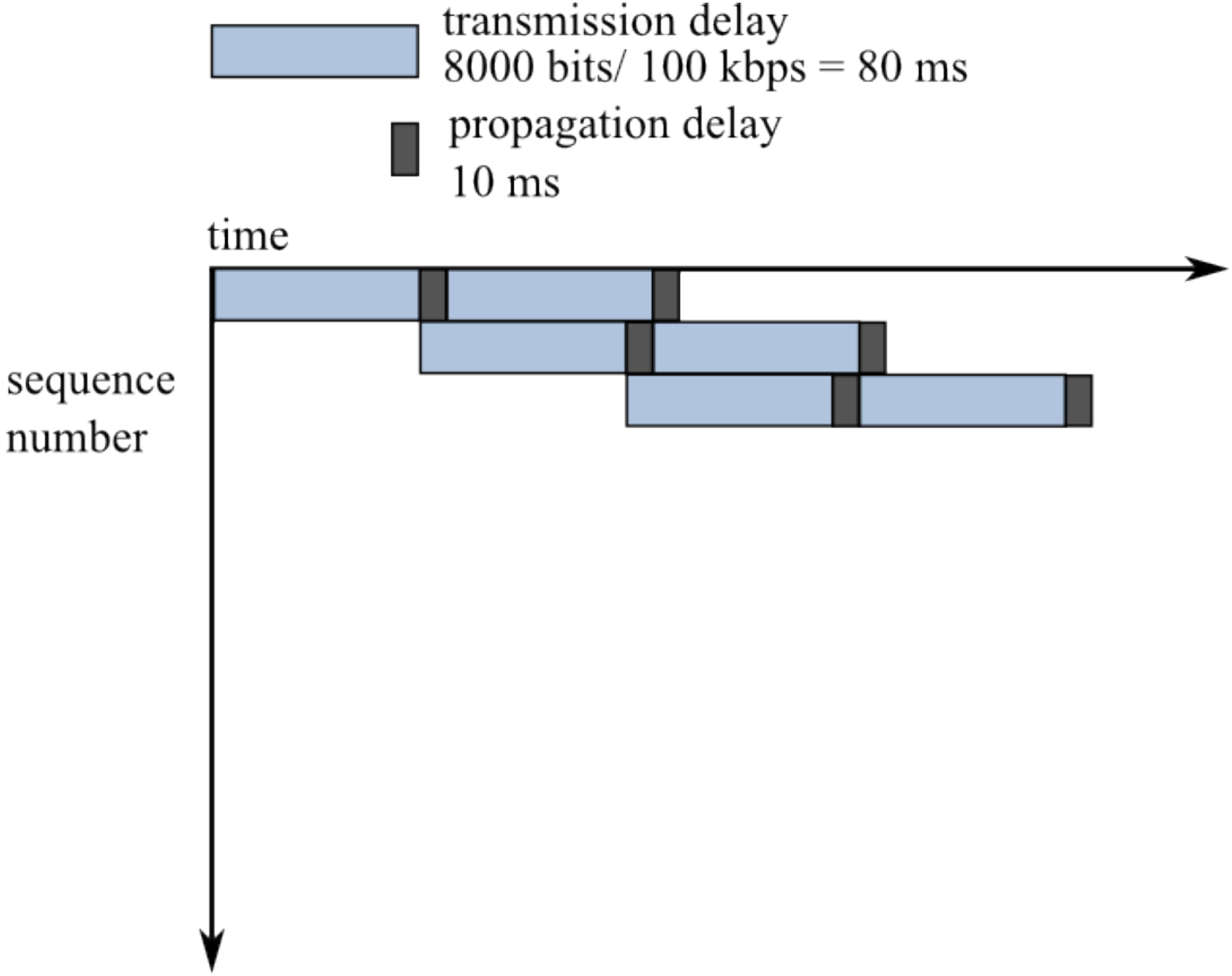
What about queueing delay?

Calculating Delay over Multiple Queues



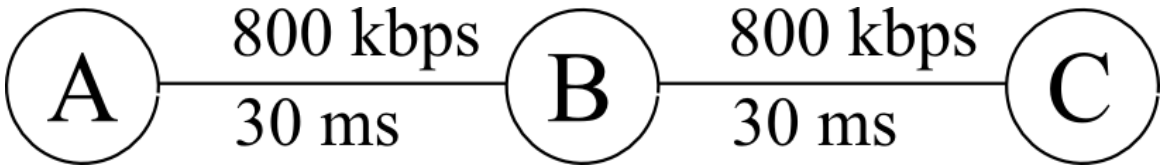
- A transmits a 1 MB file to C, using 1 kB packets
- how long will it take?
- is it bandwidth or delay constrained?

Calculating Delay over Multiple Queues



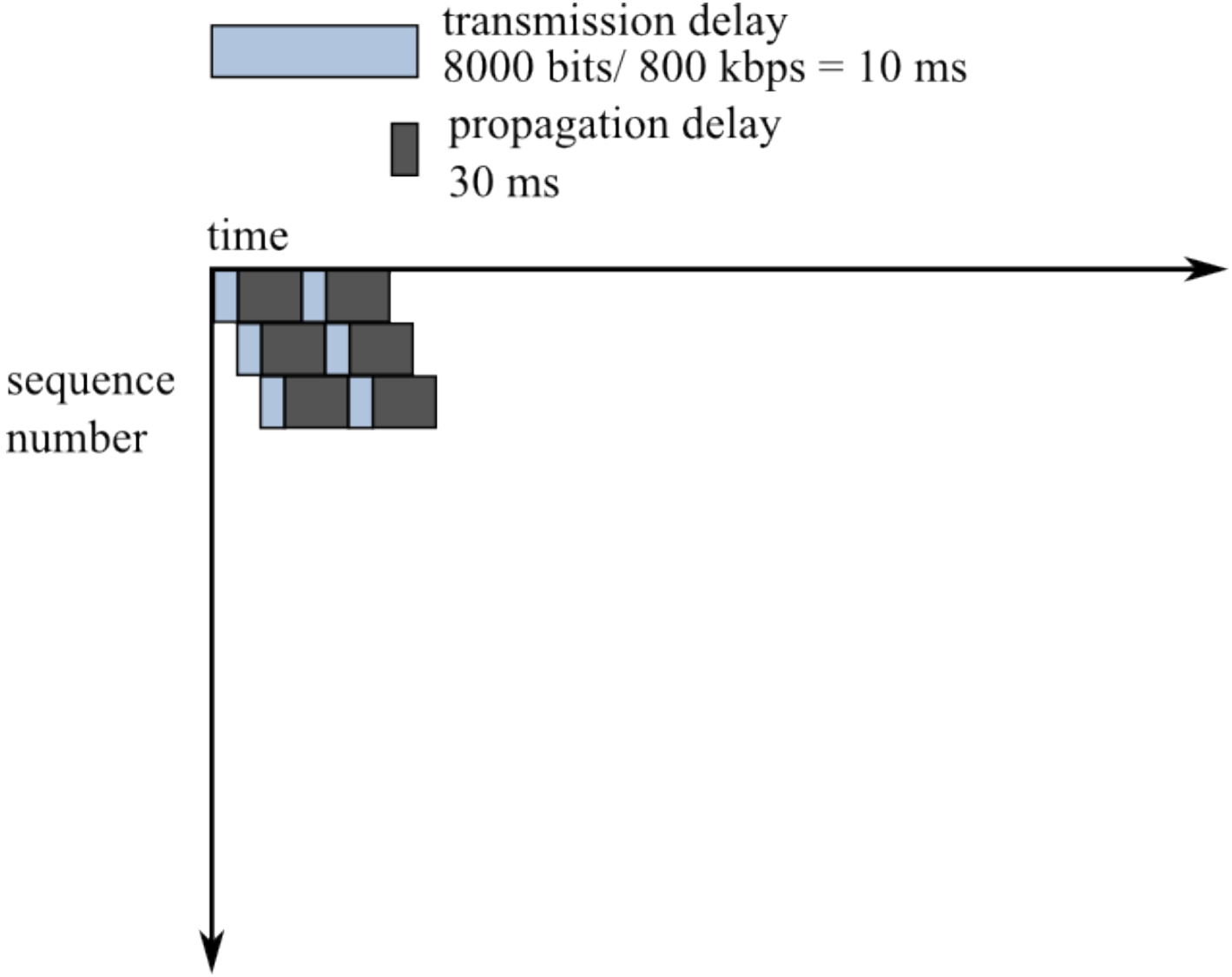
- $90ms + 1000 * 80ms + 10ms = 80.1seconds$

Calculating Delay over Multiple Queues



- A transmits a 1 MB file to C, using 1 kB packets
- how long will it take?
- is it bandwidth or delay constrained?

Calculating Delay over Multiple Queues

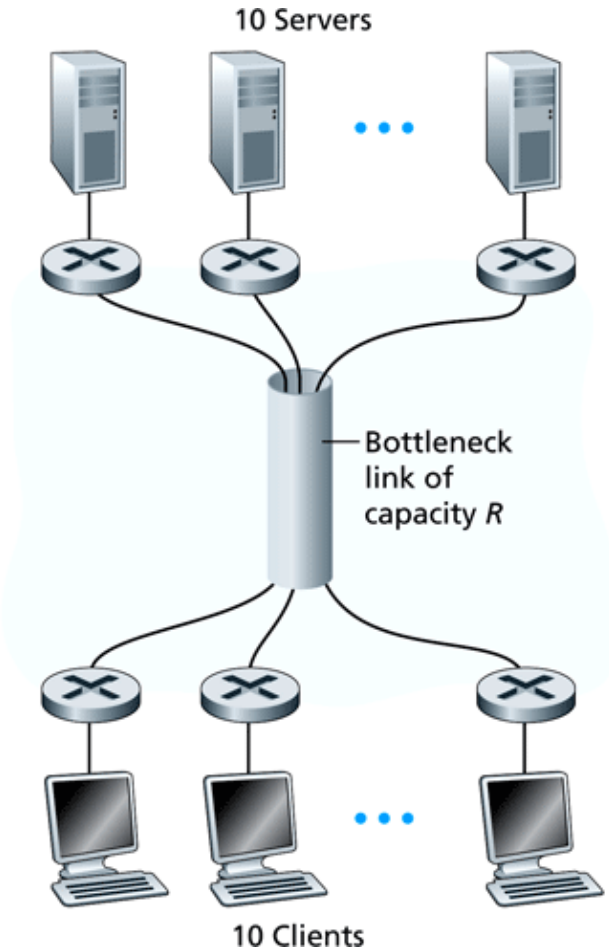


- $40ms + 1000 * 10ms + 30ms = 10.07seconds$

Throughput

Delay Approximation using Throughput

- where is the bottleneck?
  - at the edges?
  - in the core?
- delay = file size ÷ bottleneck rate



Measurement Tools

ping

- measures the round-trip time and records any packet loss
- how it works
  - sends ICMP echo request packets to the target host
  - listens for ICMP echo response replies
- *demonstration*

traceroute

- records the route a packet will take and the delay to each hop
- how it works
  - sends a series of UDP packets to a machine on port 33434
  - the first three packets use a TTL of 1, the next three use a TTL of 2, etc.
  - each hop that sees an expired TTL will send an ICMP time exceeded message
- *demonstration*
- tcptraceroute uses TCP packets and open ports to get around traceroute blocking