Module M5

CPSC 317

October 26, 2022

Learning Goals

A. Basic Definitions

- Define the terms bandwidth, latency, throughput, goodput, and jitter.
- Explain the difference between bandwidth, latency, throughput and goodput
- Define the terms "bits on the wire", and "packets in flight" and be able to apply the definitions .
- Perform computations with respect to bandwidth, throughput, goodput, and latency
- Calculate the protocol overheads with respect to performance (goodput vs bandwidth)
- Describe how jitter is introduced into a network

B. Delay

- List and define the types of delay and how they contribute to delay.
- Calculate the end-to-end delay in a network.
- Perform bottleneck analysis on a path
- Compute traffic intensity and relate traffic intensity to queuing delays.
- Calculate link utilization
- Use the formula of Average Delay = S/(1-U) where U is the network utilization and S is the average service time for a single packet. This formula is only relevant for randomly arriving packets.

BANDWIDTH and LATENCY

BW -> rate (bps = bits/second)
Lat -> time (delay)

Kilobit = kb = 10^{3} bits,

but for bytes,

Kilobyte = kB = 2^{10} bytes

Bandwidth and Latency

- Bandwidth, in networks the term refers to a rate (bits/sec). The maximum rate at which bits are transferred in the communication medium

- Latency, the time it takes for the transfer of some communication unit, from one end to the other end.

Measurements

- Rates: kbps (1000 bits per second), Mbps, Gbps, Tbps
- Data sizes: 1Kilobyte (1024 bytes, 8x1024 bits), Megabyte, Gigabyte, Terabyte

We will always give rates in bps (bits per second).

https://www.lifewire.com/bits-per-second-kbps-mbps-gbps-818122

Latency (or delay)

- The delay from when information is sent until it is received
- The “thing sent” can be bits, bytes, packets, etc. as long as it is used consistently
- Examples
  - Packets: start of sending a packet until it is completely received
  - Bit: start of sending the encoded signal until the signal is decoded at the receiver.
  - Bytes: start of sending the byte until it is completely received
- RTT – round trip time is the latency from sending until the response is received
  - RTT is what we get with ping, traceroute etc.

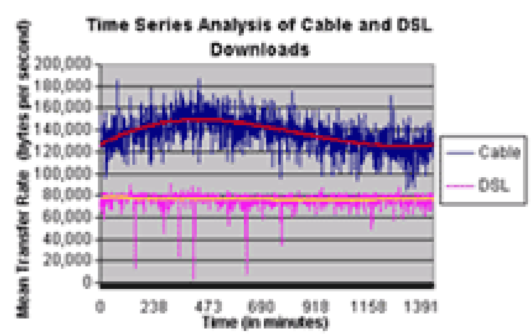
Bandwidth vs Latency

Is Internet speed the same as bandwidth?

Shaw versus Telus (old document by Shaw)

In October 2001 Shaw contracted an organization called Edaptivity to conduct independent speed testing analysis of Shaw High Speed Internet and a DSL service (Telus Corp.) in Calgary. The findings from this independent testing prove that on average, Shaw High Speed Internet provides a faster download experience than that of the local DSL Provider (Telus Corp.) in Calgary.

An evaluation of comprehensive data collected over a 45 day period from 21 residential test sites across Calgary, Alberta establishes that there is sufficient evidence to conclude that **mean file transfer rates** on the Shaw High-Speed Cable Modem network are **faster than** Telus Velocity DSL.



Bandwidth vs Latency (Shaw – old article)

We're building Western Canada's only fibre optic internet network to bring you the fastest, most reliable internet technology.[1]

With the Broadband 100 **Internet** plan, everyone can download HD videos, play online games and transfer data on multiple devices at **ultra-fast** speeds.

"**Internet speed**" refers to how quickly information can be transmitted back and forth from the Internet to your computer. Internet speeds are measured in Mbps (Megabits per second).

- Download speed is how quickly information is transmitted from the Internet to your computer. When it comes to reading, playing games, viewing video and listening to streaming music on the web, this is the key number
- Upload speed is how quickly you can send information from your computer to the Internet

https://support.shaw.ca/t5/internet-articles/about-internet-speeds/ta-p/5648

Bandwidth vs Latency (Telus)

What can affect your Internet speed (TELUS)

Many factors influence the actual Internet speed you experience at any given time, including:

- The speed of the website from which you're downloading information
- The number of people trying to download the same information as you from the same website
- The number of applications or programmes you are running on your computer
- The age and condition of wiring inside your home or apartment
- The distance between the TELUS high speed gateway and your device
- Whether you're connected to your gateway directly or wirelessly
- Your computer's age

https://www.telus.com/en/bc/support/article/understanding-internet-speed

Vehicle Analogy

Sports car may be faster but only carries 2 units (bits). So, using a bus may have more bandwidth because it carries more bits over time

Capacity is a constraint
Think of it like the number of lanes available

Latency versus bandwidth

If you are doing light web browsing and want the websites to be snappy, you want low latency. Bandwidth is only important to a certain extent.

If you are downloading a large game from Steam, you want your bandwidth to be as high as possible. Latency is not really a factor in this case.

Bandwidth and Station-wagons

Never underestimate the bandwidth of a station wagon full of tapes hurtling down the highway.
–Andrew Tanenbaum, 1981

Cisco **estimates** that total internet traffic currently averages 167 terabits per second. FedEx has a fleet of 654 aircraft with a **lift capacity of 26.5 million pounds daily**. A solid-state laptop drive weighs **about 78 grams** and can hold up to a terabyte.
That means FedEx is capable of transferring 150 exabytes of data per day, or 14 petabits per second—almost a hundred times the current throughput of the internet.

https://what-if.xkcd.com/31/

Latency – speed of light

“It’s the Latency Stupid”

http://www.stuartcheshire.org/rants/latency.html

- Fact One: Making more bandwidth is easy.
- Fact Two: Once you have bad latency you're stuck with it.
- Fact Three: Current consumer devices have appallingly bad latency.
- Fact Four: Making limited bandwidth go further is easy.
- This is not to say bandwidth is unimportant: Bandwidth Still Matters

Ping Statistics

- Let's collect some data using ping
- ping -c 10 (windows: ping -n 10)

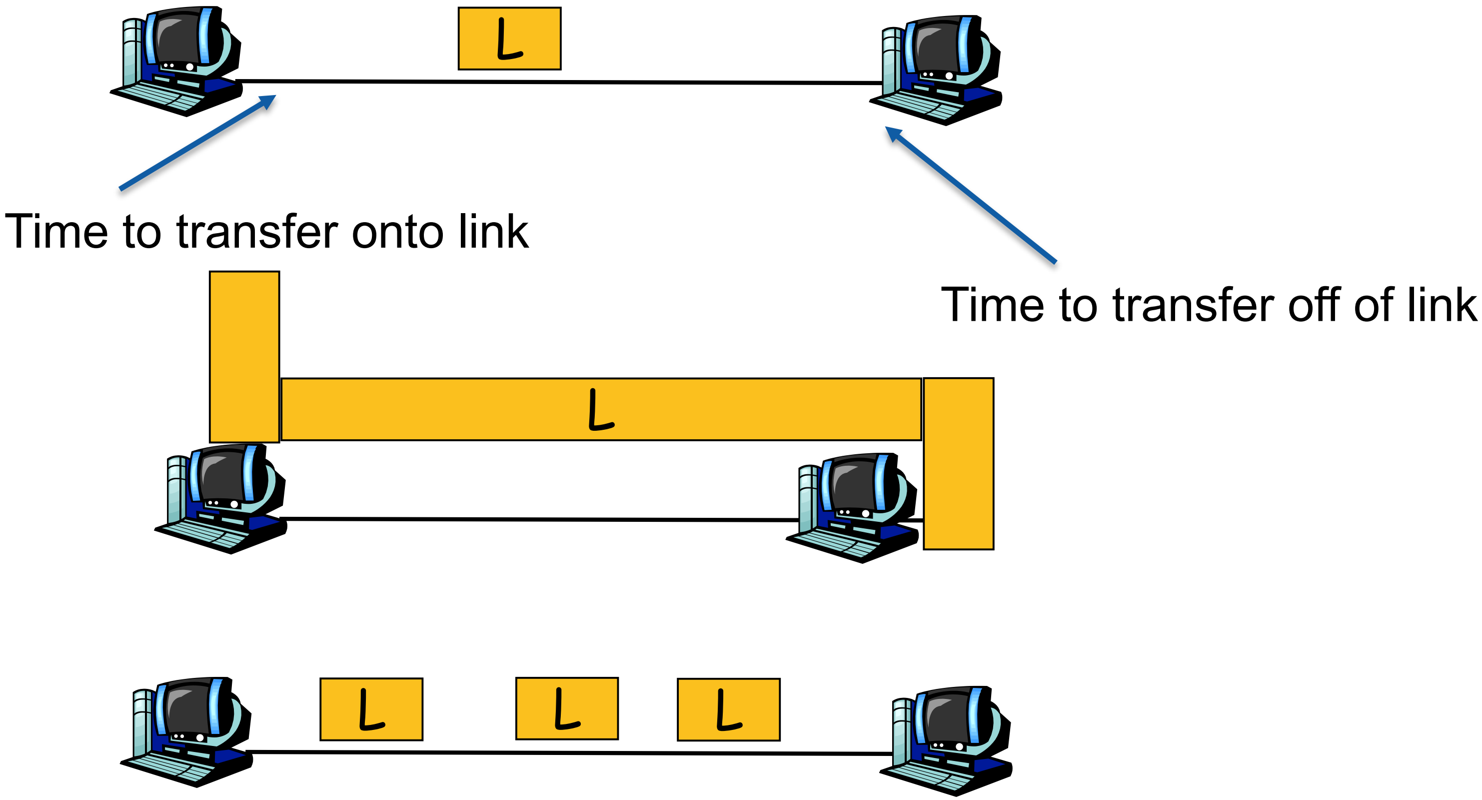
```
$ ping -c 10 www.cs.ubc.ca
PING www.cs.ubc.ca (142.103.6.5) 56(84) bytes of data.
64 bytes from www.cs.ubc.ca (142.103.6.5): icmp_seq=1 ttl=62 time=1.05 ms
64 bytes from www.cs.ubc.ca (142.103.6.5): icmp_seq=2 ttl=62 time=1.22 ms
64 bytes from www.cs.ubc.ca (142.103.6.5): icmp_seq=3 ttl=62 time=1.03 ms
64 bytes from www.cs.ubc.ca (142.103.6.5): icmp_seq=4 ttl=62 time=0.776 ms
64 bytes from www.cs.ubc.ca (142.103.6.5): icmp_seq=5 ttl=62 time=1.05 ms
64 bytes from www.cs.ubc.ca (142.103.6.5): icmp_seq=6 ttl=62 time=1.06 ms
64 bytes from www.cs.ubc.ca (142.103.6.5): icmp_seq=7 ttl=62 time=0.907 ms
64 bytes from www.cs.ubc.ca (142.103.6.5): icmp_seq=8 ttl=62 time=1.09 ms
64 bytes from www.cs.ubc.ca (142.103.6.5): icmp_seq=9 ttl=62 time=1.05 ms
64 bytes from www.cs.ubc.ca (142.103.6.5): icmp_seq=10 ttl=62 time=1.06 ms

--- www.cs.ubc.ca ping statistics ---
10 packets transmitted, 10 received, 0% packet loss, time 9007ms
rtt min/avg/max/mdev = 0.776/1.032/1.221/0.113 ms
```

Jitter

- Observe that not all the ping times are the same
- This variation is called jitter
- What causes jitter?

On the wire…



Number of bits in flight – bits on the wire

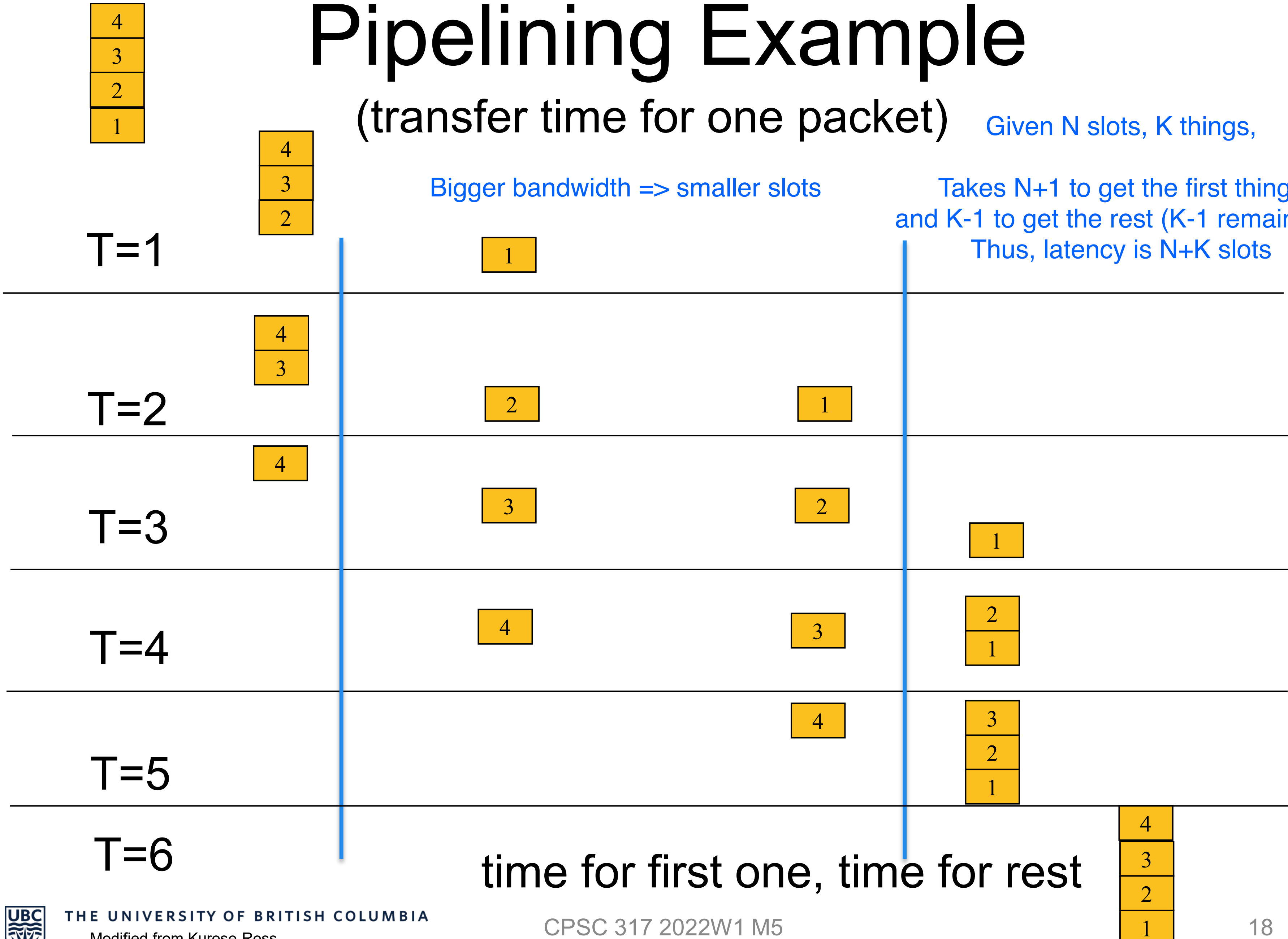
Pipelining Example

(transfer time for one packet)

Given N slots, K things,

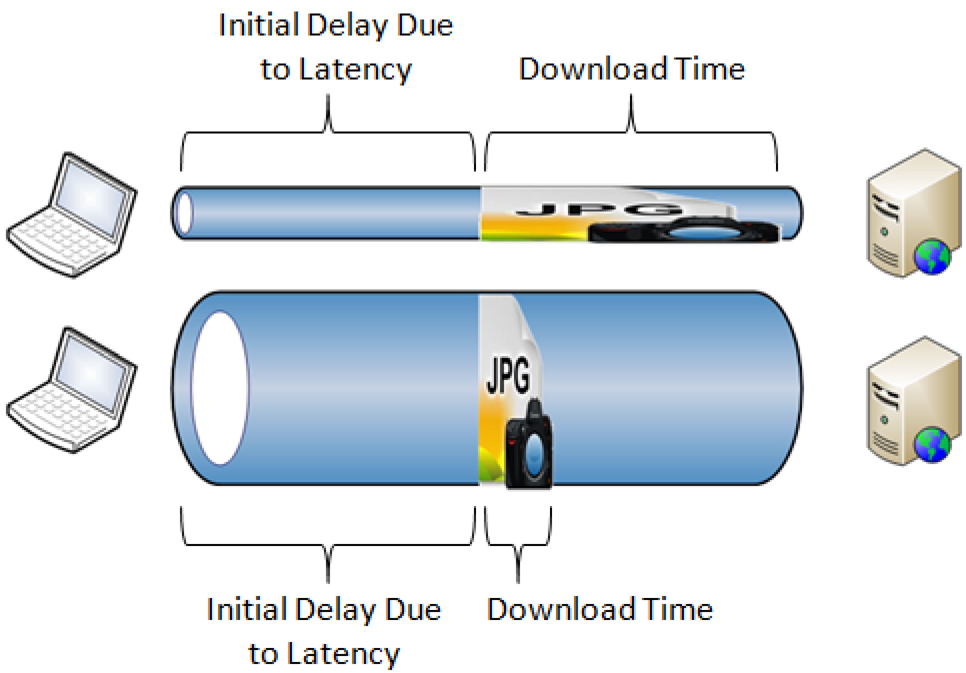
Bigger bandwidth => smaller slots

Takes N+1 to get the first things and K-1 to get the rest (K-1 remaining). Thus, latency is N+K slots



time for first one, time for rest

Pipe Analogy



1/BW
1 Gbps => 1 nanasec
1 Mbps => 1 microsec
1 kbps => 1 ms

Example:
Let 1/BW = 0.01 micro second = 10ns
Suppose signals travel at 2x10^{8} m/s

(2x10^{8}) x (10x10^{-9}) = 2 m

Data Sizes versus Data Rates

- Data sizes measured using binary
  - 1MB 2^20 bytes, 2^20 x 8 bits

- Data rates measured in base 10
  - 1Mbps 10^6 bits/sec,
    - 1MB per second 10^6 x 8

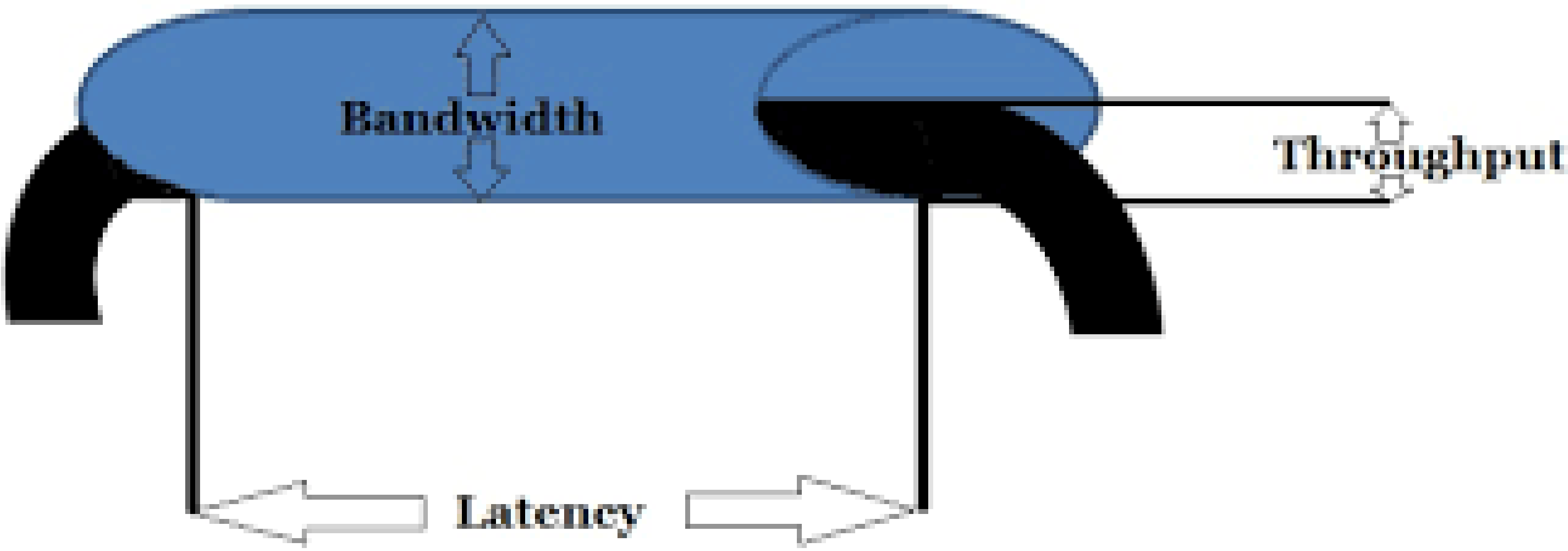
Back-of-the-envelope

2^10 – 10^3   2.4% error
2^20 – 10^6   4.85765% error
2^30 – 10^9   7.374%… error

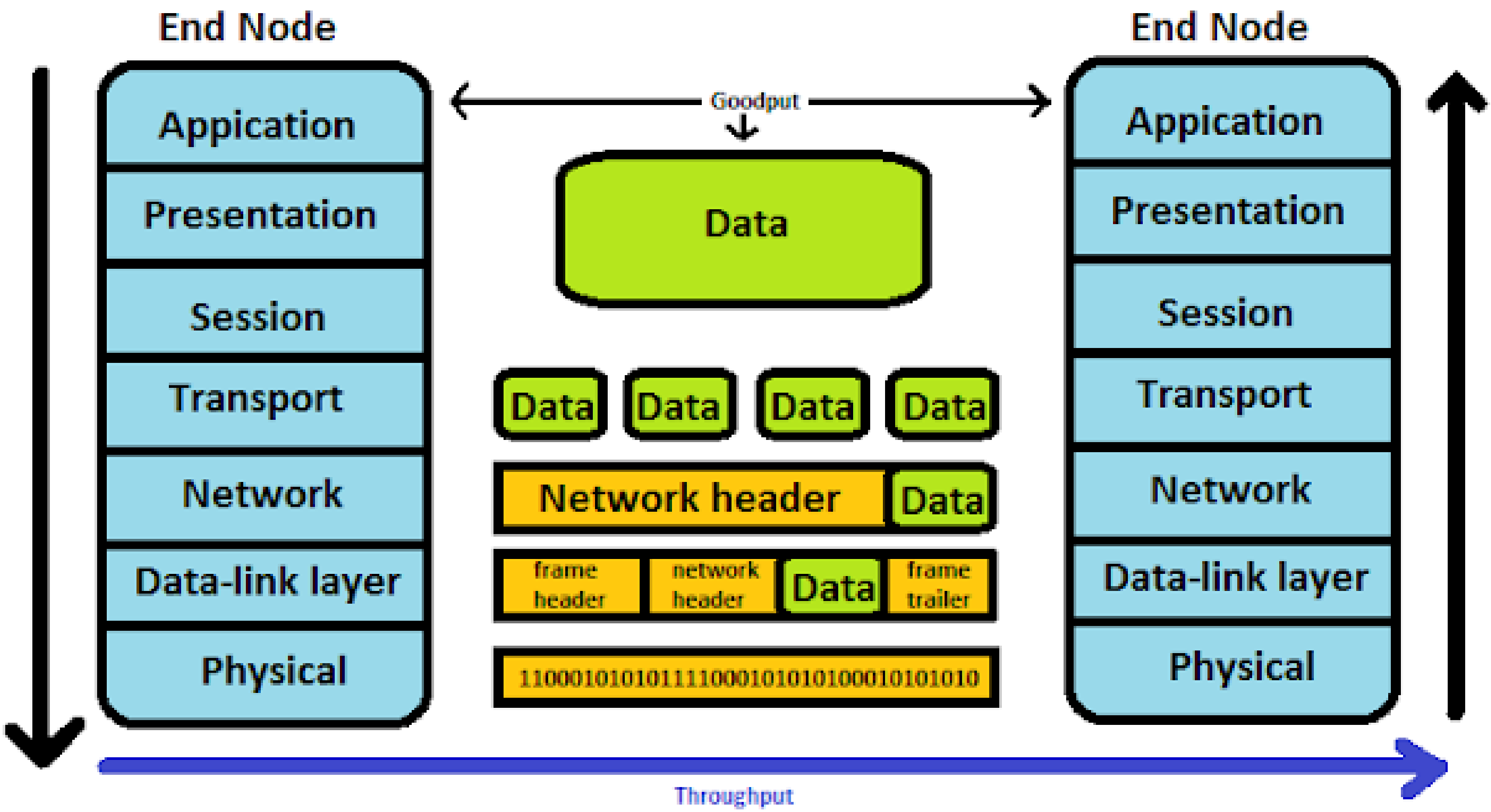
Bandwidth, throughput, goodput

- These are all rates (e.g. bits per second)
- Bandwidth
- Throughput
- Goodput

Throughput

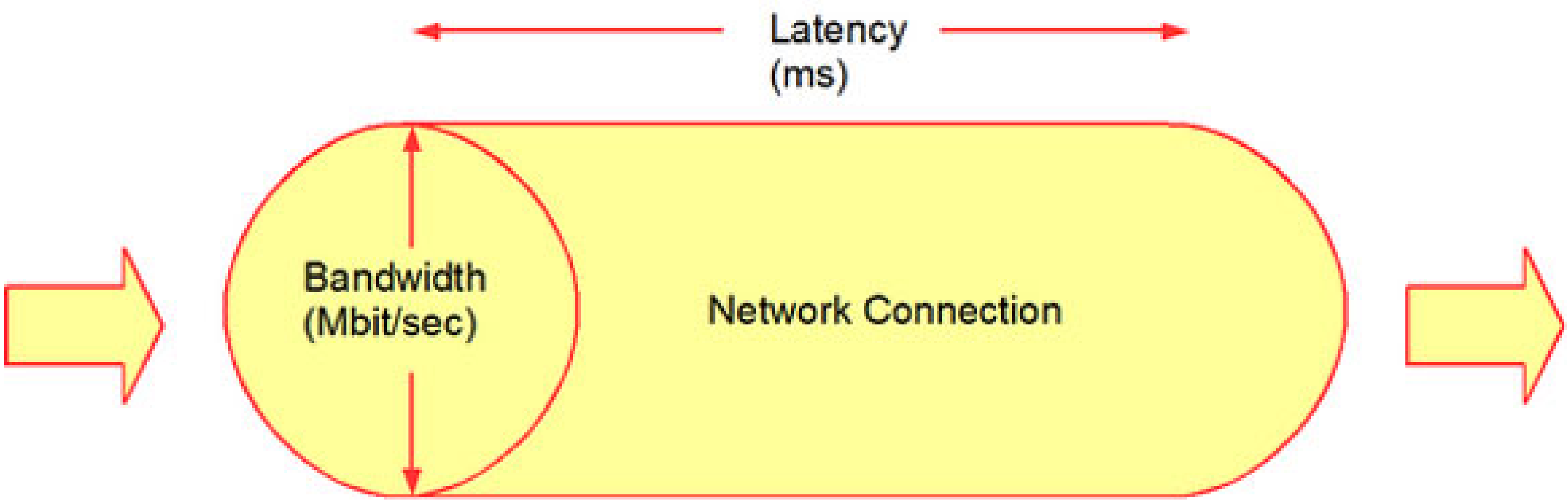


Goodput



End Node
Appication
Presentation
Session
Transport
Network
Data-link layer
Physical
Goodput
Data
Data
Data
Data
Data
Network header
Data
frame header
network header
Data
frame trailer
11000101010111100010101010001010101010
End Node
Appication
Presentation
Session
Transport
Network
Data-link layer
Physical
Throughput

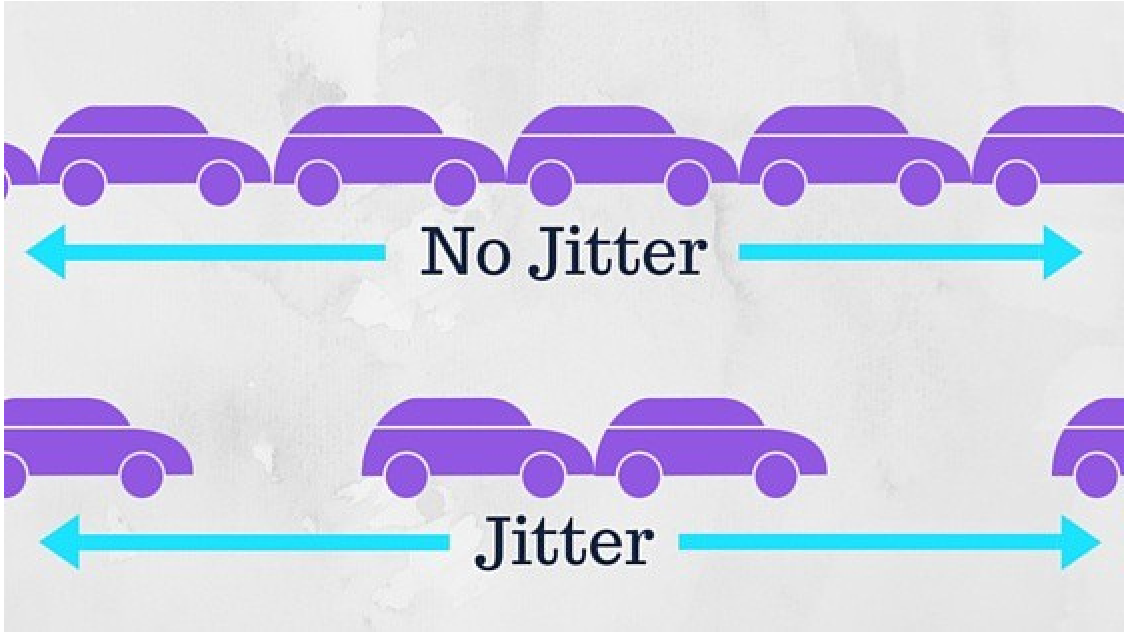
Bandwidth x Latency (BW x Lat)



filling the pipe

Roadway Analogy

Jitter (inter-packet variance)



DELAY

http://www.ccs-labs.org/teaching/rn/animations/propagation/

Types of Delay

1. Processing delay – examine packet to decide where to direct packet
2. Queuing delay – waiting time to get access to the link the packet needs to be sent on
3. Transmission delay – time to actually write the packet onto the "wire"
4. Propagation delay – time spent to move the bit from the source to destination on the transmission medium

Nodal delay

$$\boxed{d_{\text{nodal}} = d_{\text{proc}} + d_{\text{queue}} + d_{\text{trans}} + d_{\text{prop}}}$$

- d_{proc} = processing delay
  - typically a few microsecs or less
- d_{queue} = queuing delay
  - depends on congestion
- d_{trans} = transmission delay 1/BW
  - = L/R, significant for low-speed links
- d_{prop} = propagation delay
  - a few microsecs to hundreds of msecs

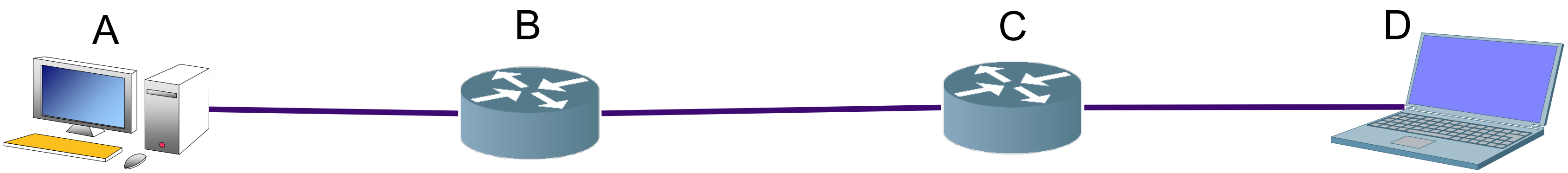
End-to-end Delay

- Sum of the following delay components at every step of the way

1. Processing delay
2. Queuing delay
3. Transmission delay
4. Propagation delay

Store and Forward Links

End-to-end Delay

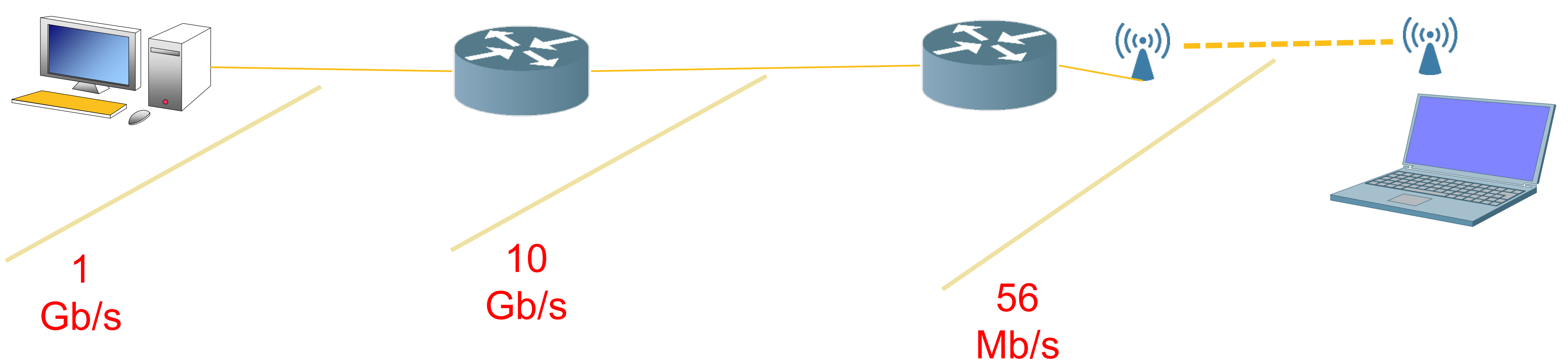


Node	Processing Delay (ms)	Queuing Delay (ms)	TX Delay (ms)	Propagation Delay (ms)	Total Nodal Delay (ms)
A	4	10	15	34	63
B	4	12	11	50	77
C	3	12	15	25	55
D	5	4	-	-	9

Total Delay =

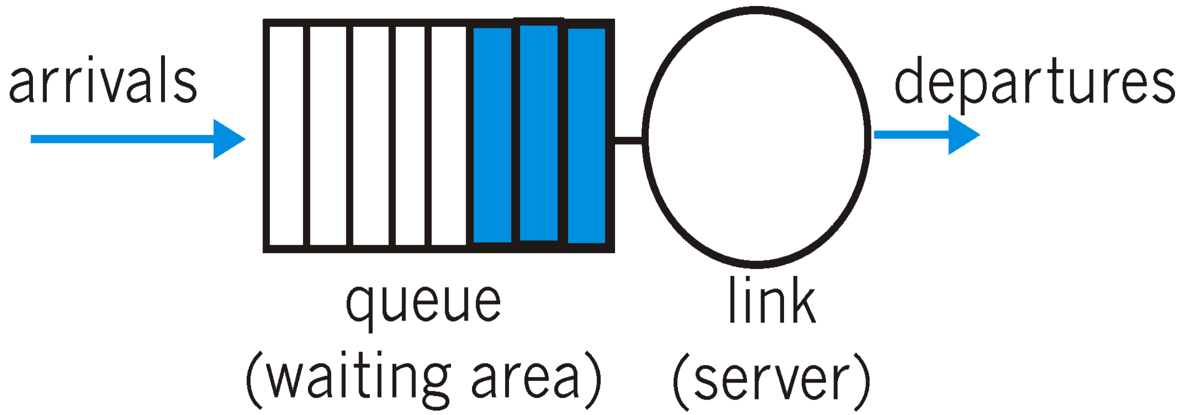
Bottlenecks

- What is the maximum throughput possible (bandwidth) between two nodes connected by a network?



TRAFFIC INTENSITY

Traffic Intensity



Arrival Rate of 100 packets/second

Transfer time of 1 millisecond/packet

Transfer time of 10 millisecond/packet

Transfer time of 100 millisecond/packet

Traffic Intensity Formula

$$\frac{\text{Number coming into system}}{\text{Number leaving the system}}$$

Arrivals

L Average packet size in bits

a Average number of packets arriving per second

Departures

R

The transmission rate, or rate at which bits leave the system.

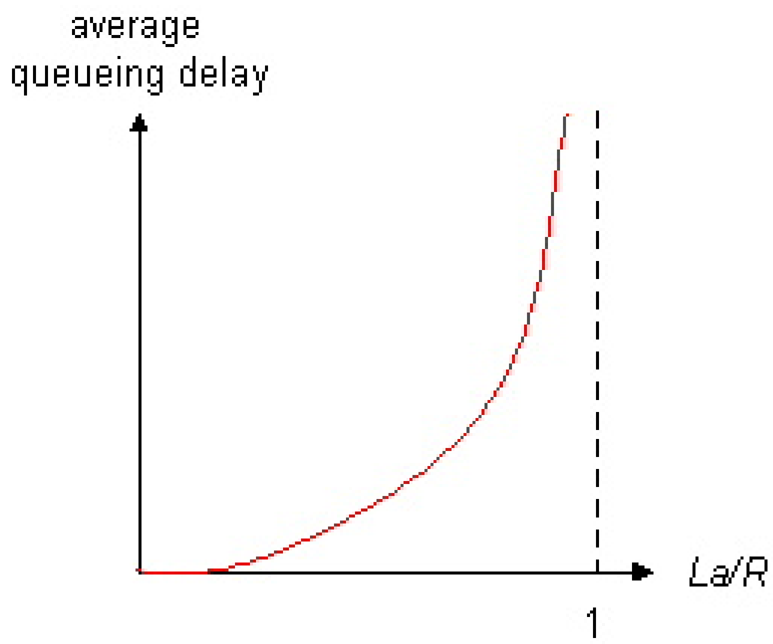
Traffic Intensity and Delay

- La/R << 1

- La/R < 1

- La/R >= 1

Summary of Queueing delay



- R=link bandwidth (bps)
- L=packet length (bits)
- a=average packet arrival rate

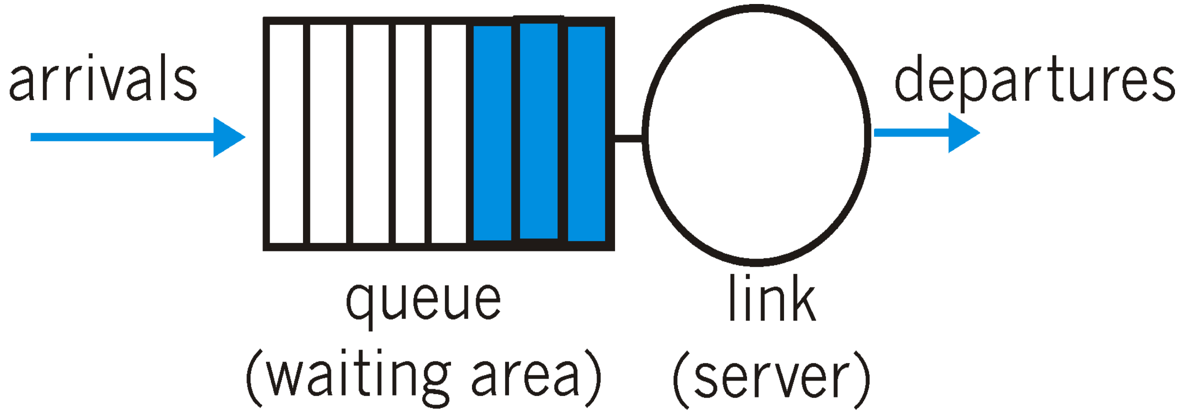
traffic intensity = La/R

- La/R ~ 0: average queueing delay small
- La/R -> 1: delays become large
- La/R > 1: more "work" arriving than can be serviced, average delay infinite!

Why do we care about traffic intensity?

- Helps us understand how busy a link is
- There is a relationship between a link's “busyness” and queuing delay experienced for that link
  - Queuing delay is how long a packet has to wait.
  - Imagine a bunch of packets arriving with some random spacing
  - The first packet goes straight through, but subsequent packets have to wait for the packets already there to leave.
  - The busier the link the higher the probability that there are one or more packets already there.

Average Delay (entire time)



Average Delay = S/(1-U)

U is network utilization (percent busy)

S is the service time (transfer time)

Packet loss

- queue (aka buffer) preceding link in buffer has finite capacity
- when packet arrives to full queue, packet is dropped (aka lost)
- lost packet may be retransmitted by previous node, by source end system, or not retransmitted at all

Delay vs Utilization