Network Performance and Design

CSC/ECE 570 Computer Networks
Fall 2024

Agenda

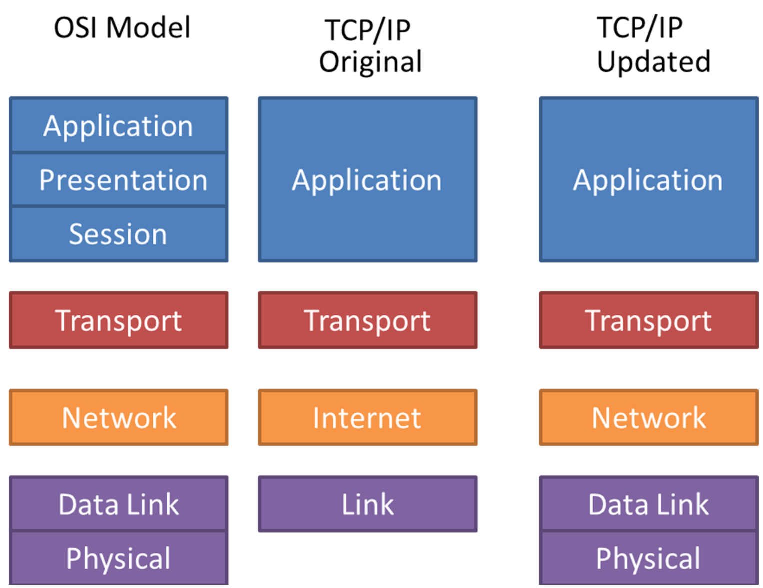
1. Layers
2. Delay, Throughput, Loss
3. Network Design

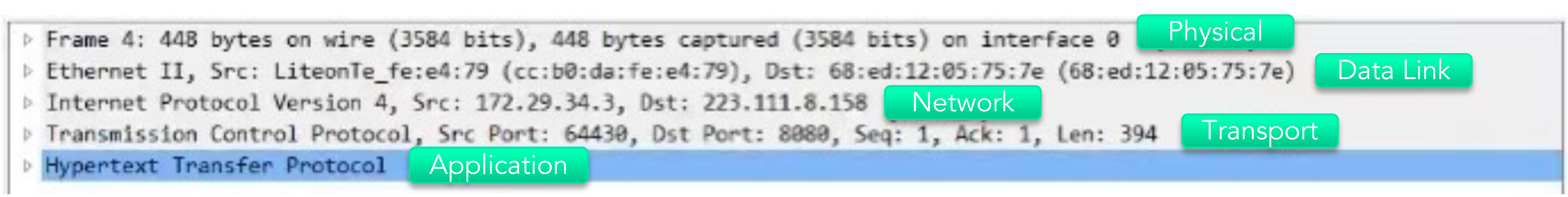
TCP/IP Protocol Suite

- Differences between OSI and TCP/IP?

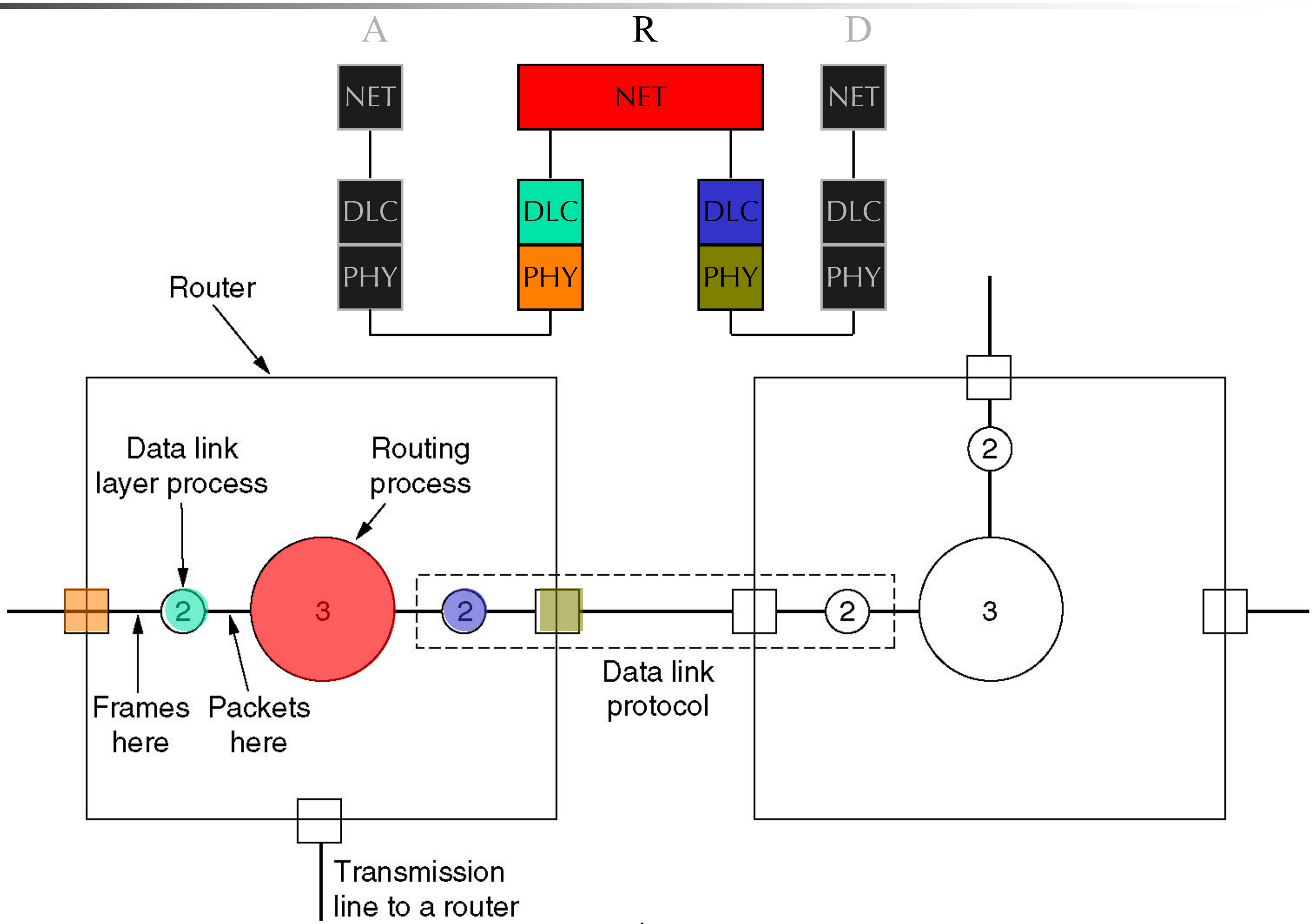
  5 layers:
  - Physical
  - Data link/MAC (ARP, SLIP)
  - Network (IP, ICMP, IGMP)
  - Transport (TCP, UDP)
  - Application

    (http, ftp, telnet, smtp)



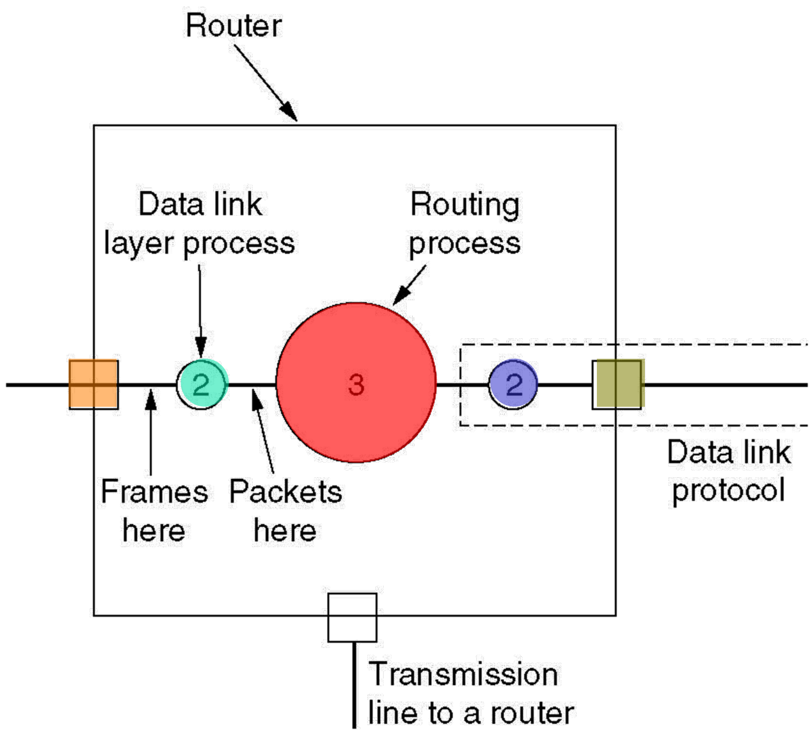


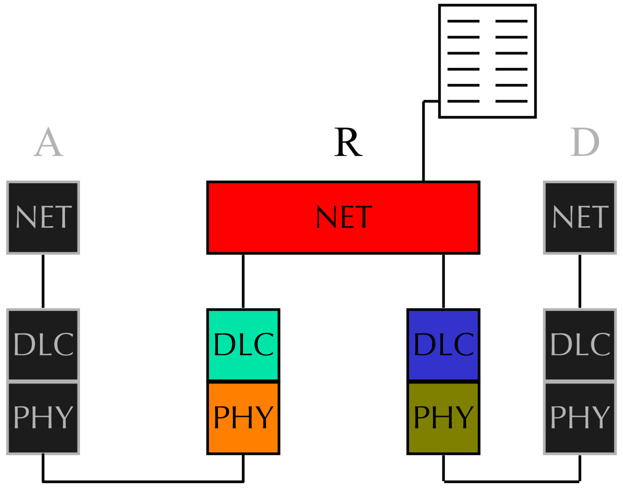
Layers in a Router/Switch



Buffering at L3

- Operation of L3 itself may require buffering data
  - Store-and-forward
- Input-buffer-process-buffer-output cycle
- Discard data → loss





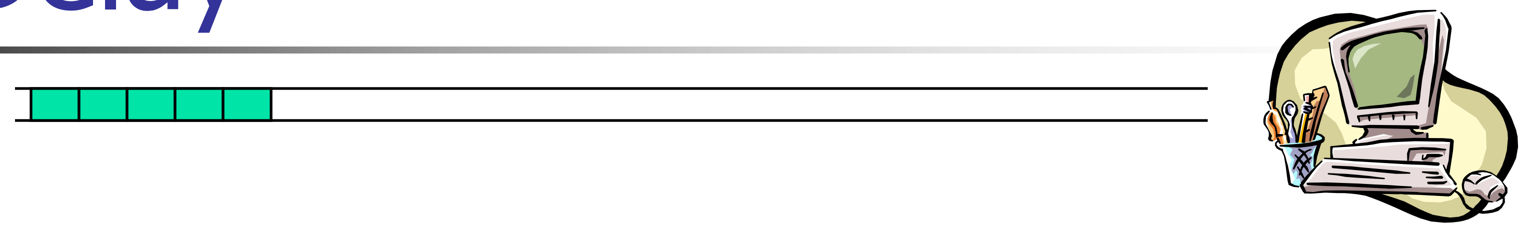
Agenda

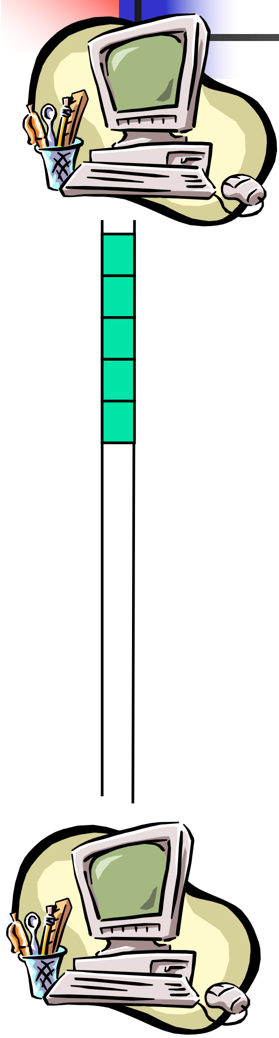
1. Layers in a Router/Switch
2. Delay and Throughput
3. Network Design

Network Performance

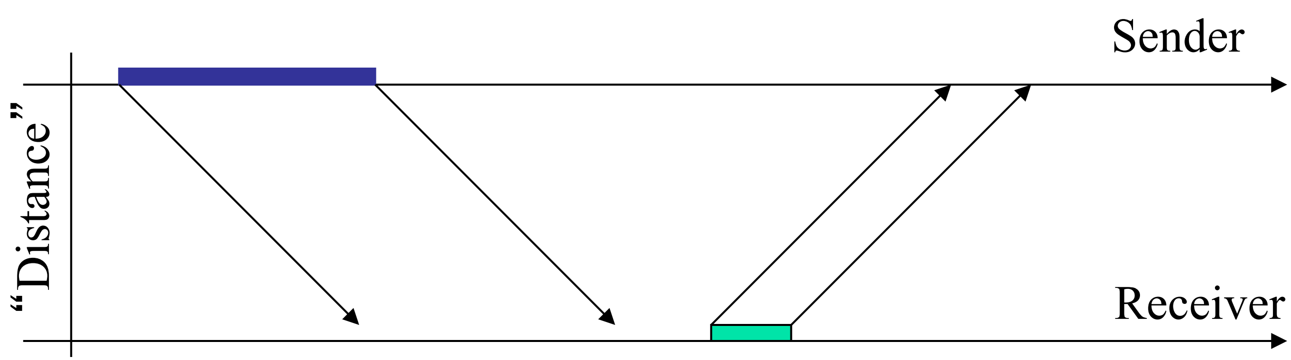
- Networks are engineered systems
  - Intended to serve users
- Performance appropriately measured in quantities the end-user cares about
  - First: connectivity
  - Delay, throughput
  - Other metrics derived from these
- More sophisticated metrics
  - Predictability of above metrics
  - Predictability of connectivity: Reliability / Survivability
  - Predictability of delay or throughput
    - Guarantees - Quality of Service contracts
  - Other emergent characteristics: e.g. Security

Delay

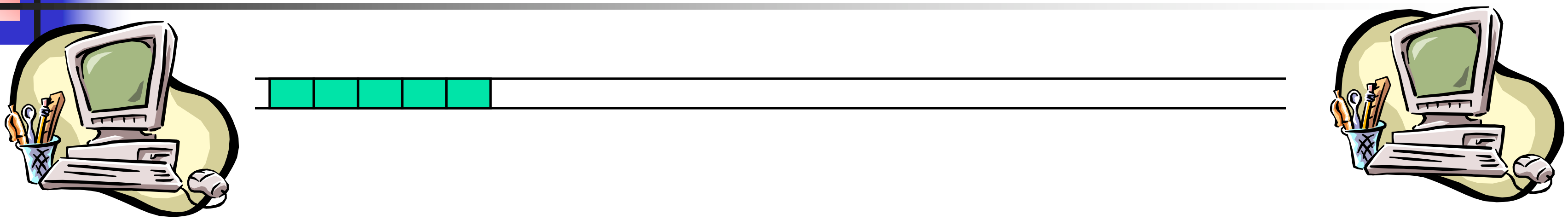




- Two meanings
  - How fast can successive bits be put into the pipe?
  - How long does a bit take to traverse the pipe?



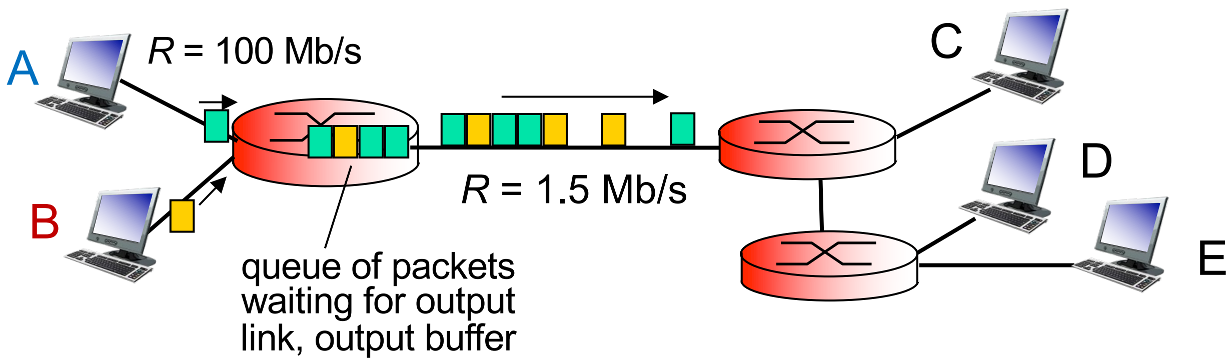
Throughput



- Total number of bits transferred, over given time
  - Related obviously to the transmission delay
  - Throughput sometimes referred to as “bandwidth”
  - “Bandwidth-delay product (BDP)” - filling the pipe

- Conceptually simply – need to be able to measure
  - *ping* can measure round-trip delay
  - Various tools such as *Iperf* for throughput

Queueing Delay, Loss



resource contention:

- aggregate resource demand (use of transmission link) can exceed amount available
- *congestion:*
  - packets will queue, wait for an available link use
  - packets can be dropped (lost) if no memory to store them

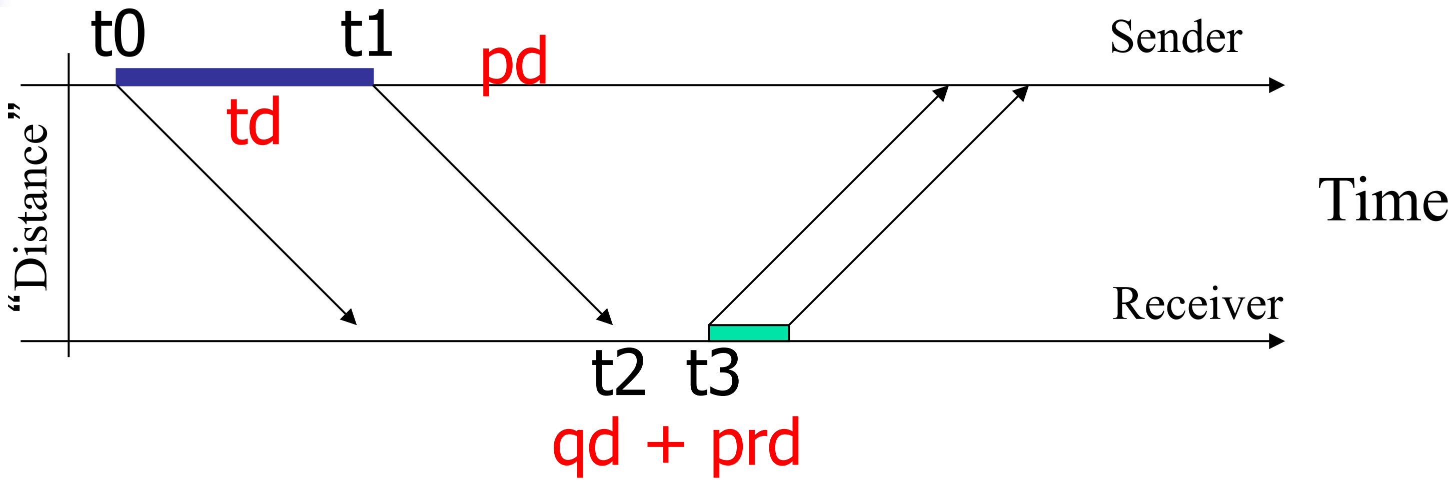
Bottleneck Link

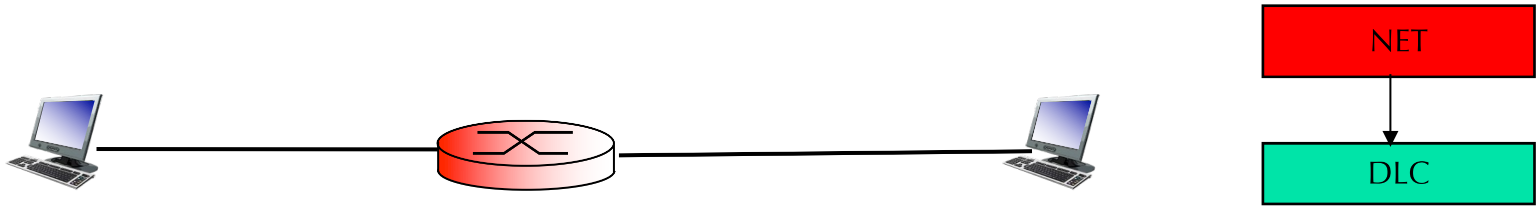
- The link with minimum "used" bandwidth along transmission path

More Delays

- Queuing (waiting) delay
  - Waiting to be served
  - Basic service is transmission on outgoing line
  - Input-buffer-process-buffer-output cycle
- Processing delay
  - Time taken for *other* processes data must go through before transmission (not transmission itself)
- These heavily depend on networking approach
  - Propagation delay is function of distance
  - Transmission delay is function of how much data (also modulation technology, medium, etc.)

Examples



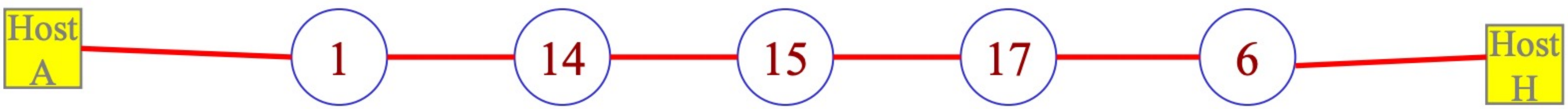


Processing delay: memory write/read 1) Delay from NET to shared memory;
2) Delay from shared memory to DLC.

Reducing Delays: To Reduce –

- Propagation delay
  - Get the ends closer together.
  - Usually not at luxury of network designer
- Transmission delay
  - Use faster transmission equipment
  - Has to be at both ends of link, high link capacity
  - Might need to upgrade cable (or medium)
- Queuing delay
  - Use faster memory for buffers, faster buses, higher clock-speed
  - Use fewer stages of buffering
  - Reduce buffer size (but probability of loss increases)
- Processing delay
  - Use faster computing equipment
  - Do more of the processing in dedicated hardware rather than in custom software on general purpose hardware

Question



Assume: All links: 2.5 km; C = 100Mbps; propagation speed = 200m/microsec.
queuing delay = processing delay =0; packet size = 1000 bytes

What is the round-trip latency?

Network Traffic

- Ultimately, networks exist to serve traffic (enable traffic to be carried)

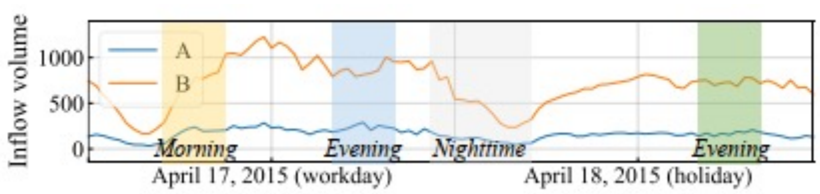
- What is traffic?
  - That which occupies / is carried by links (A stream of data)

- Traffic is offered to the network by/at network nodes
  - Network is made of end nodes, intermediate nodes, and links
  - All traffic ultimately originated/consumed by end-nodes

Traffic Characterization

- Traffic - "Demand" for networking services

- Magnitude (bandwidth)
  - Could vary with time, if "reasonably long" life
- Lifetime
  - How long it is resident in the network
- Arrival and departure patterns
  - Call (like telephony) arrival and departure
  - Increment and decrement
  - Periodic (scheduled)
  - Static (long-term)



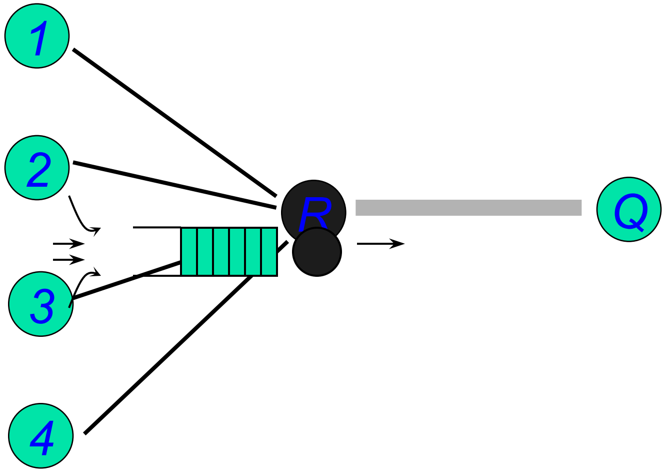
J., Ji etc

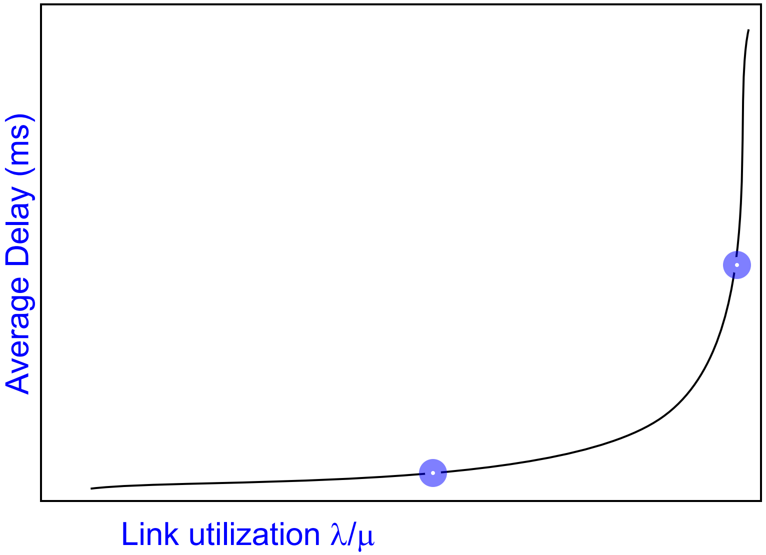
Traffic Demand Determines Packet Loss

- Loss may occur on the link
  - Usually very little in guided medium - ignore
  - Usually handled by L2 transmissions
- Loss may occur at intermediate nodes
  - Store-and-forward buffers are finite - may overflow
  - Other mechanism at intermediate node may discard
- Does retransmission occur?
  - May not be required / desired
  - If desired,
    - May be at L2, on link
    - May be E2E, at L4

Statistical TDM Performance

- Statistical TDM
- Delay is lower on average: “Statistical Multiplexing Gain”
  - But unpredictable for individual packet - prediction is statistical

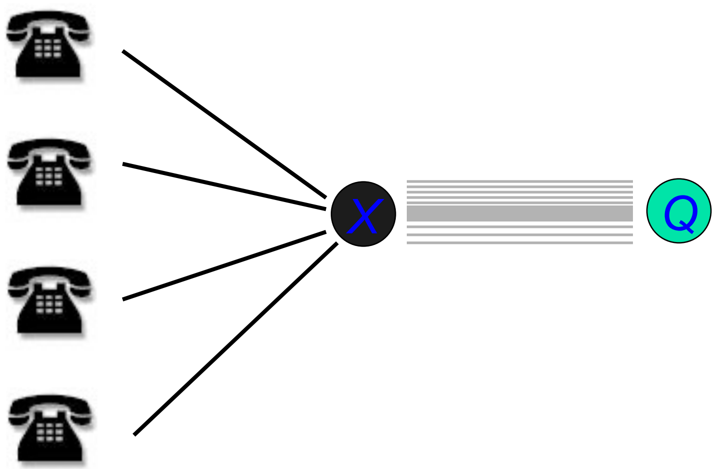




Blocking in Telephony

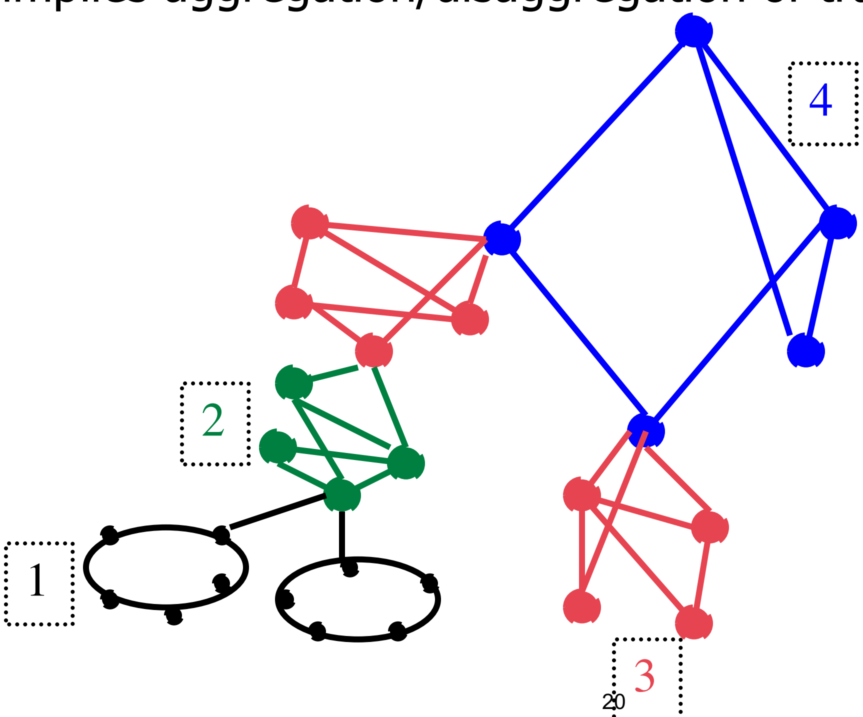
- Delay - very small and constant, operative quantity is blocking ratio
- Average call rate λ
- Average holding time τ
- Offered traffic load or intensity a = $\lambda\tau$
- $$B(a,c) = \frac{a^c / c!}{\sum_{k=0}^{c} a^k / k!}$$

Erlang B formula



Scaling, Hierarchy, Aggregation

- Because of scalability, hierarchy seems inevitable
  - Connectivity is always less than full (esp. in large networks)
  - Nature of end-nodes and intermediate nodes vary
  - All links are TDM (FDM modeled as separate links)
- Hierarchy implies aggregation/disaggregation of traffic

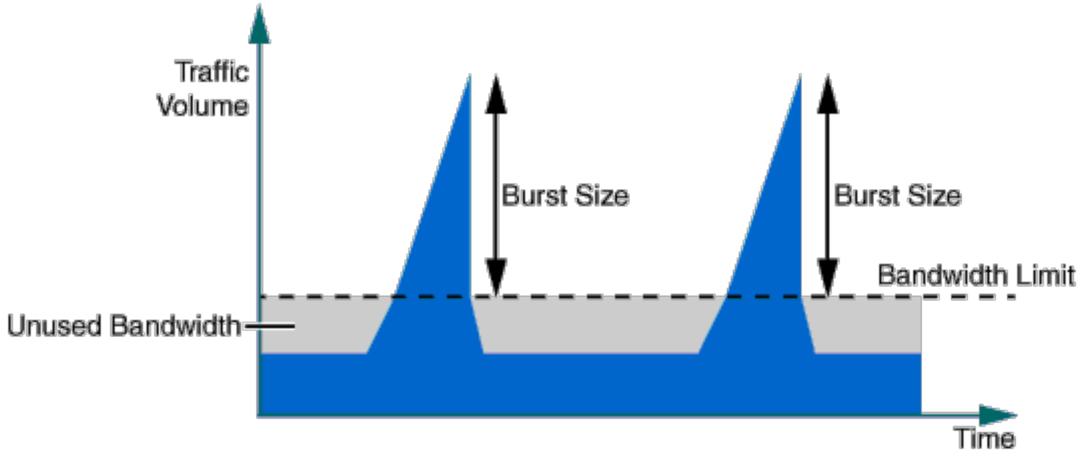


Traffic Aggregation - Static Traffic

- Assume each station injects traffic steadily
  - Number of bits injected per time unit is constant for each source
- Due to aggregation, magnitude increases as traffic climbs hierarchy
  - But constant nature of traffic remains
- Aggregation/dis-aggregation process is straightforward for intermediate nodes
  - Effectively same as slotted TDM
- Therefore static traffic is stable - remains static at higher levels of hierarchy

Bursty Traffic

- Traffic is generated intermittently at each end node
  - Assume (peak) rates are known



- Average bitrate vs. Peak bitrate
- Higher provision (waste of resources)
- Lower provision (may cause congestion and loss)

Contributing Factors

- "Elasticity" of traffic
  - Source-to-destination traffic flows in the Internet are *not* static as generated
  - Congestion in network will slow down bursts
  - In response, flow *duration* will *increase*

- "Filling up" of pipes
  - Provisioning levels are not perfect
    - At some levels, flows may be at capacity
  - At next level, contributes a static flow

Agenda

1. Layers in a Router/Switch
2. Delay and Throughput
3. Network Design

Network Design

- Various aspects of the network must be determined/chosen/configured
- Network resources - nodes and links
- Nodes
  - Physical connection interface
  - Buffers, scheduling, routing/forwarding, protocol
- Links
  - Circuit enablement, packet switching, bandwidth (bitrate capacity), protocol

Design Goals

- Goals are in terms of network performance (experienced by traffic)
- Basic goal: Connectivity
  - Basic design methodology: e.g., Routing (path finding)
  - Output of routing: Forwarding Information Bases
  - The one indispensable piece of information in making a forwarding decision: destination
  - Simplest routing approach: map from destination to next-hop
- Next goal: Performance metrics
  - Basic design: how much resource to place where
  - Dimensioning or provisioning

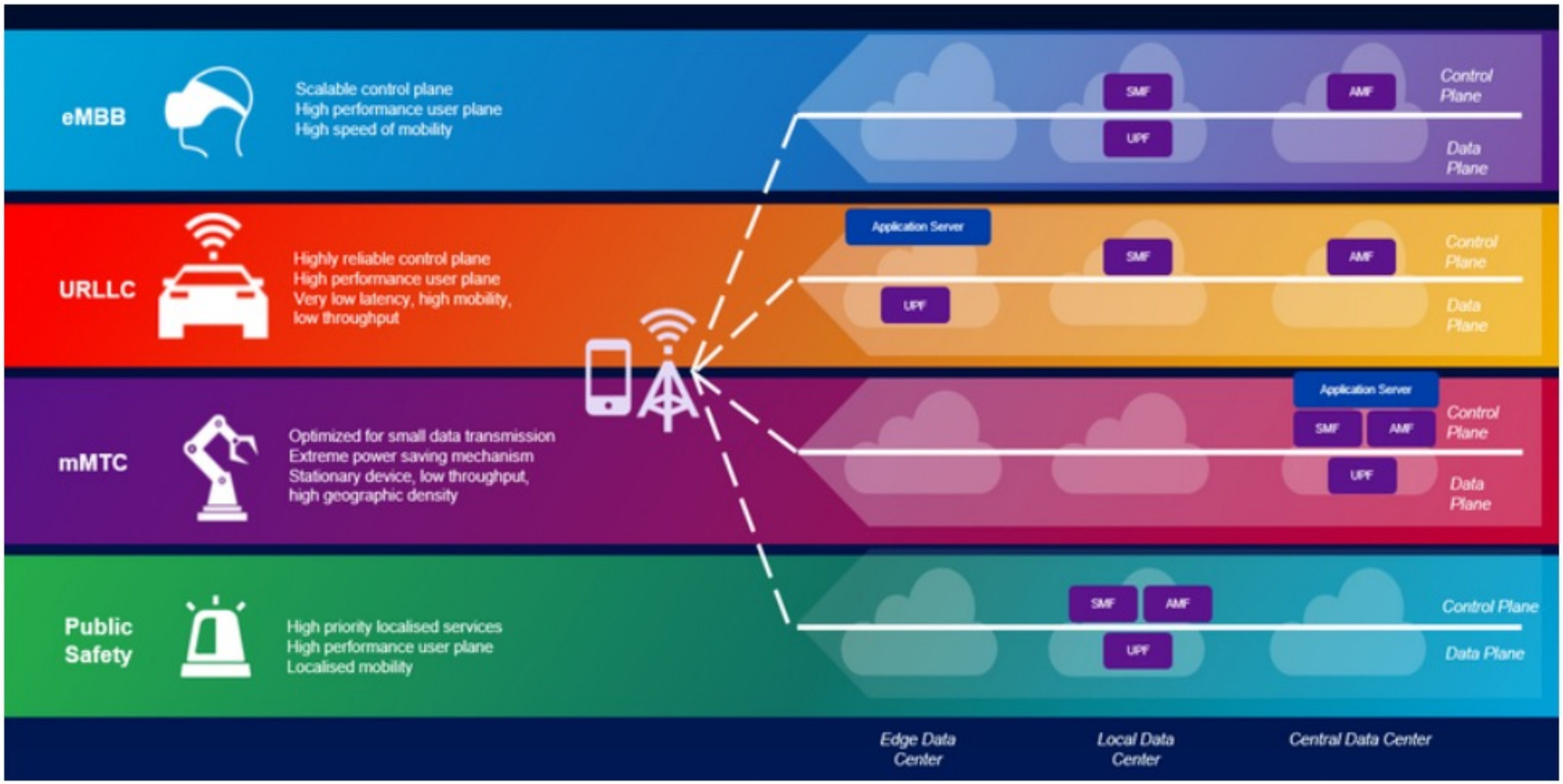
Orchestration

- We have introduced a large number of varieties of devices
- We have argued ourselves into attempting precise coordinated behavior of them

- How shall we configure all devices to obtain the coordinated behavior?
  - Need to be flexible and responsive to traffic → continuous reconfigurations

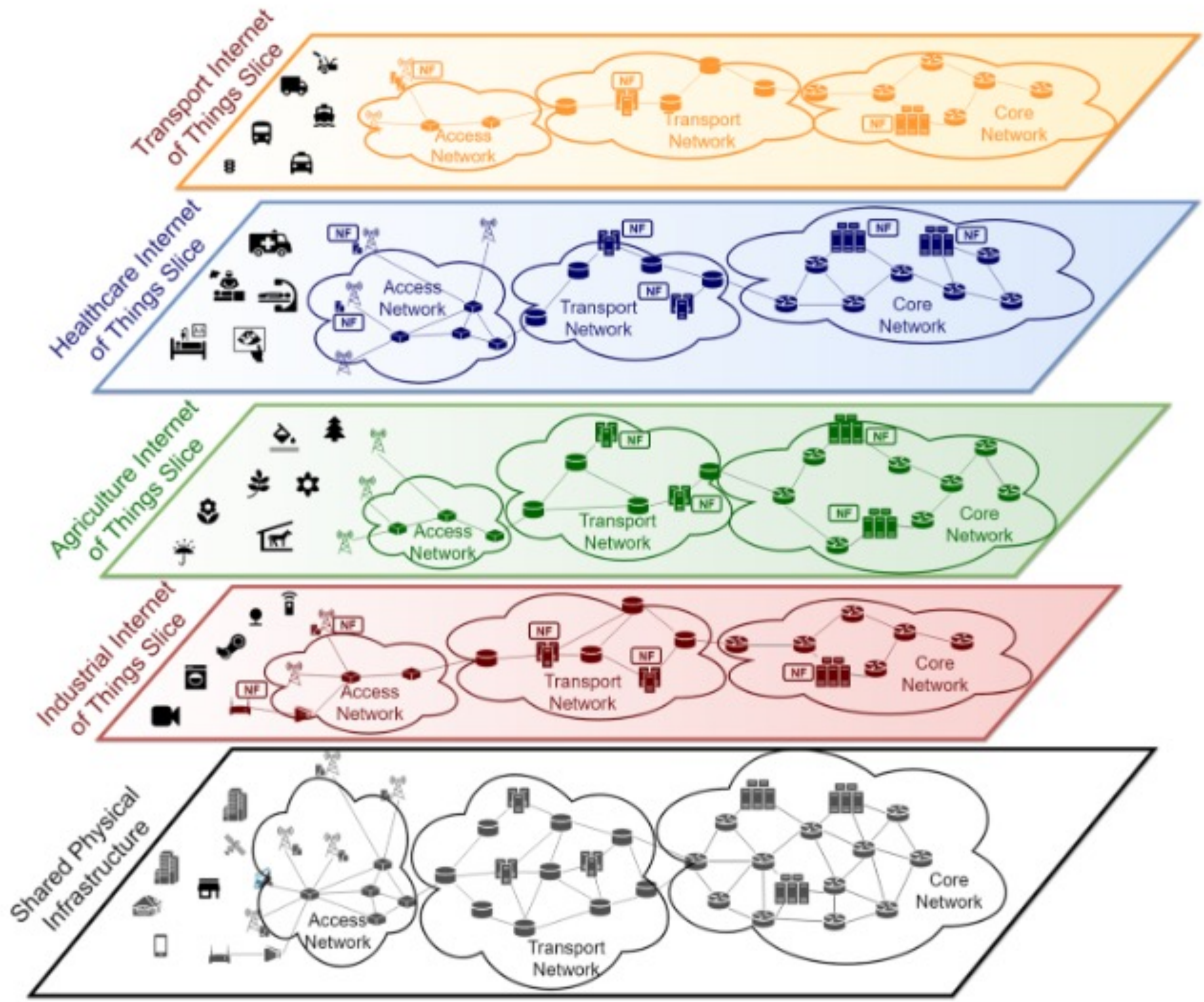
- Autoconfiguration → control signals
  - Control "plane"
  - Network administration/management/engineering/architecture
  - "Management plane"

Network Slicing



https://www.viavisolutions.com/en-us/5g-network-slicing

Orchestration



Summary

- Networking may be more complicated than obvious at first look
  - Additional complications introduced by need for predictable performance
- Networks must be carefully designed (resource provisioned)
- Design should proceed on the basis of
  - What is network use for
  - What behavior is expected or desired from network
- Different answers to above questions
  - Will result in different approach to design

Next Lecture

1. Network Simulation