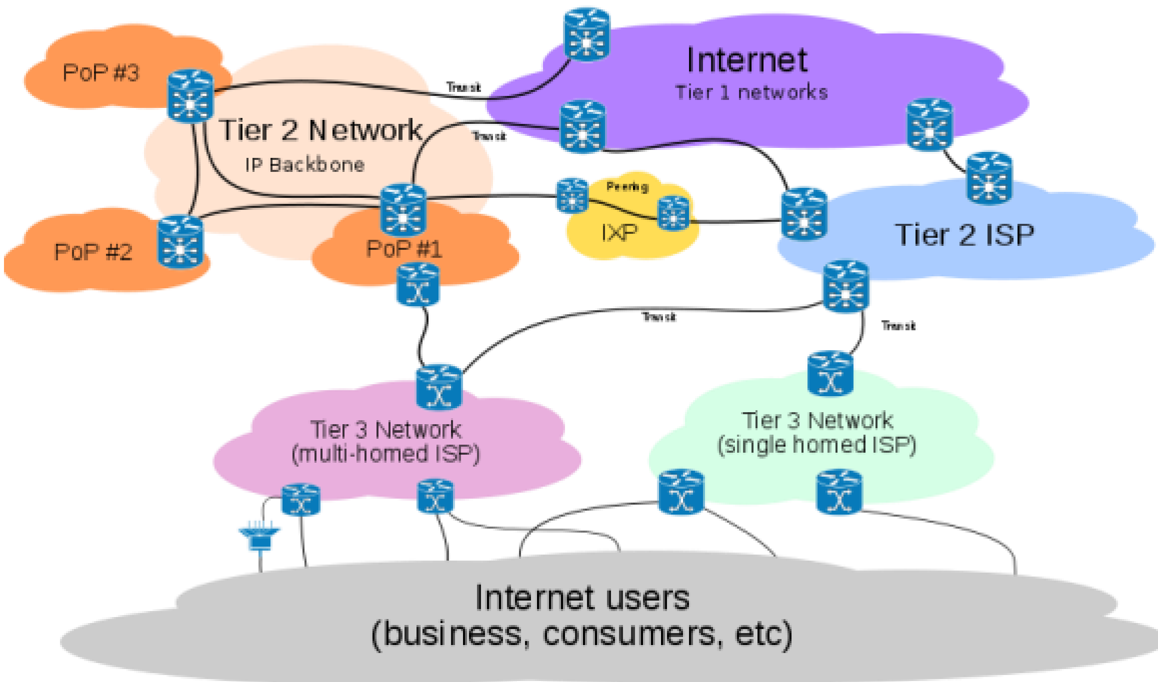
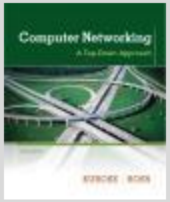
Network core and metrics



latency

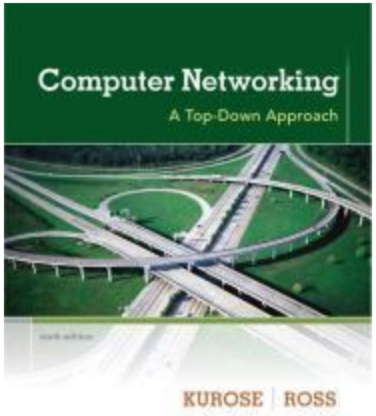
propagation	**transmit**	**queue**

Computer Networking: A Top Down Approach
6th edition
Jim Kurose, Keith Ross
Addison-Wesley

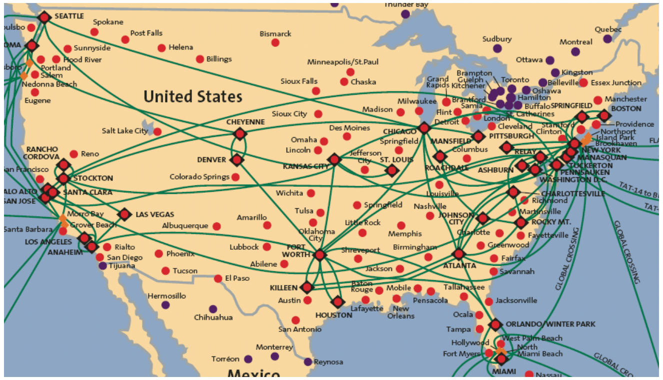


Overview

- Chapter 1: Introduction
  - Quick overview of field
  - Learn some terminology

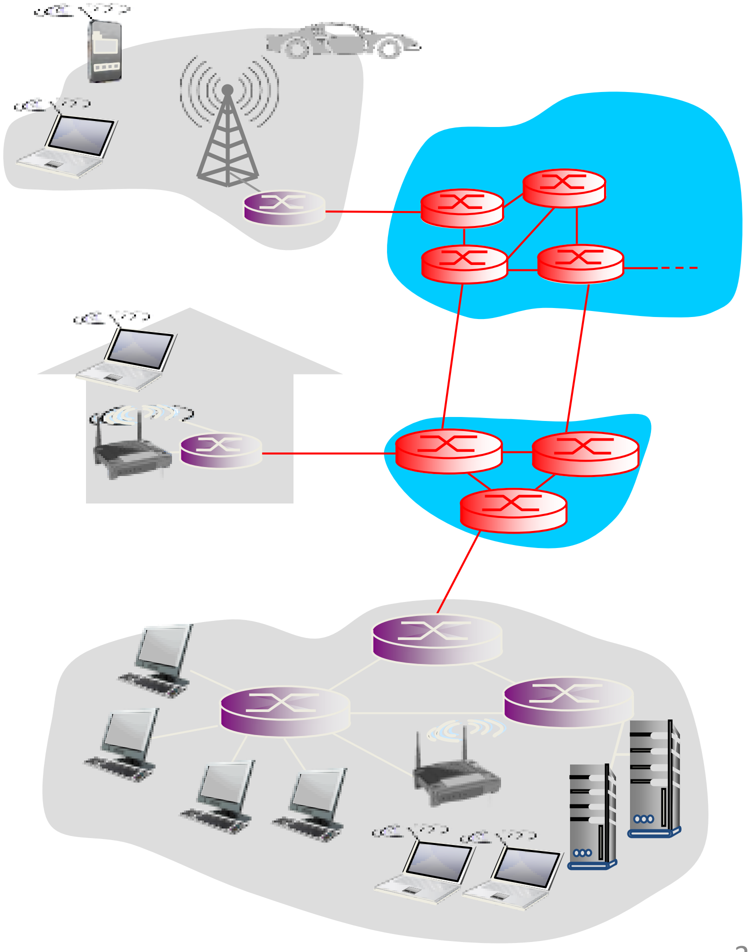


- Network core
  - Mesh of routers and links connecting end systems
- Metrics
  - Measuring performance of the network

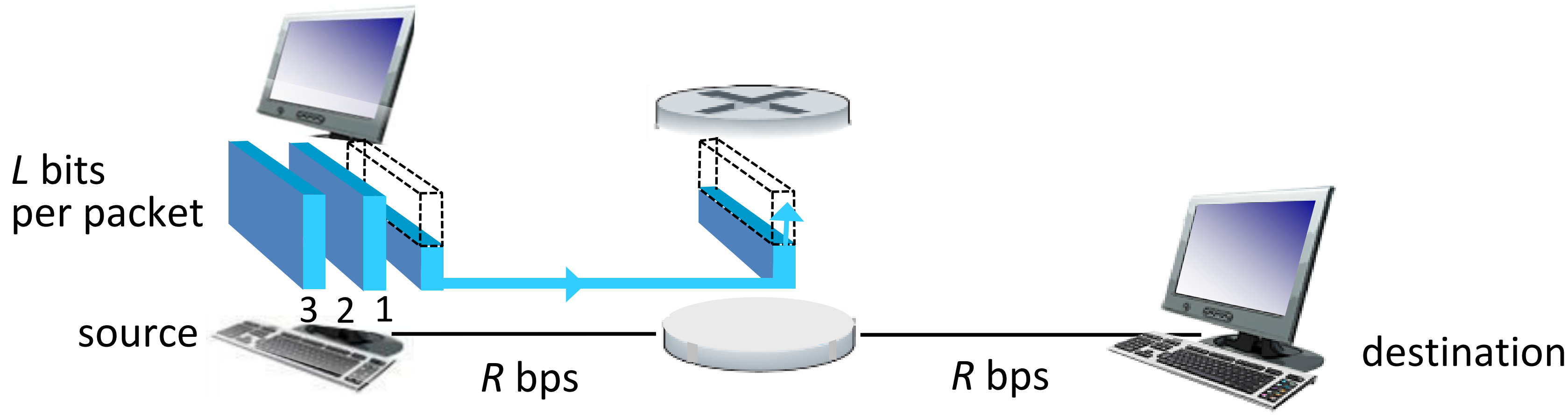


The network core

- Mesh of interconnected routers
- Packet-switching
  - Break application-layer messages into *packets*
  - Forward packets from one router to the next, across links on path from source to destination
  - Packets transmitted at full link capacity



Packet-switching: store-and-forward



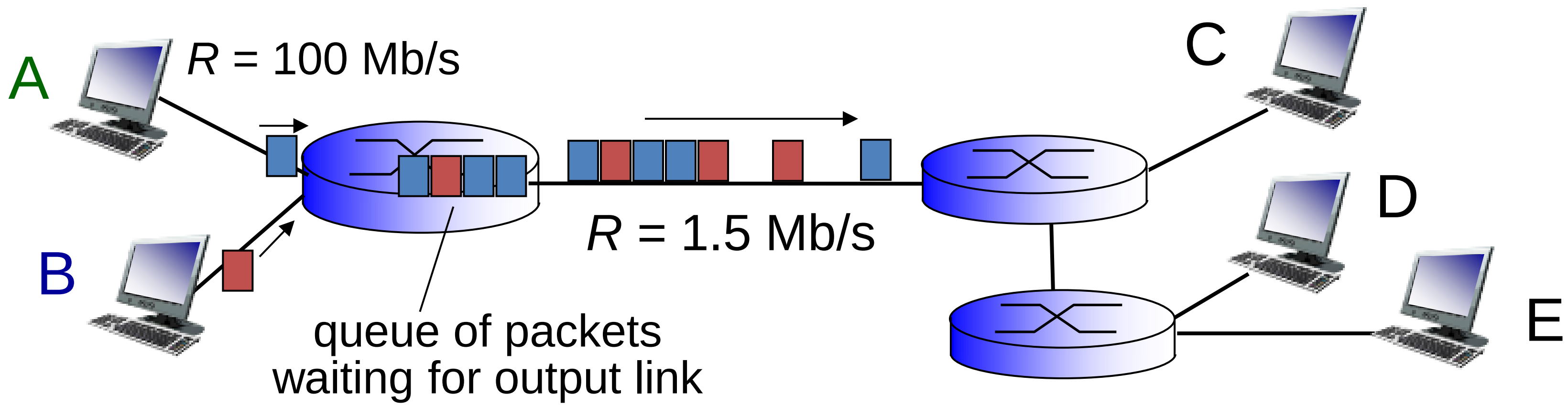
- *L*/*R* seconds to transmit (push out) *L*-bit packet into link at *R* bps
- *Store and forward:*
  - Entire packet must arrive at router before it can be transmitted on next link

one-hop numerical example:
- *L* = 7.5 Mbits
- *R* = 1.5 Mbps
- one-hop transmission delay = 5 sec

- end-end delay = 2*L*/*R* (assuming zero propagation delay)

} more on delay shortly …

Packet-switching: queuing delay, loss



Queuing and loss:

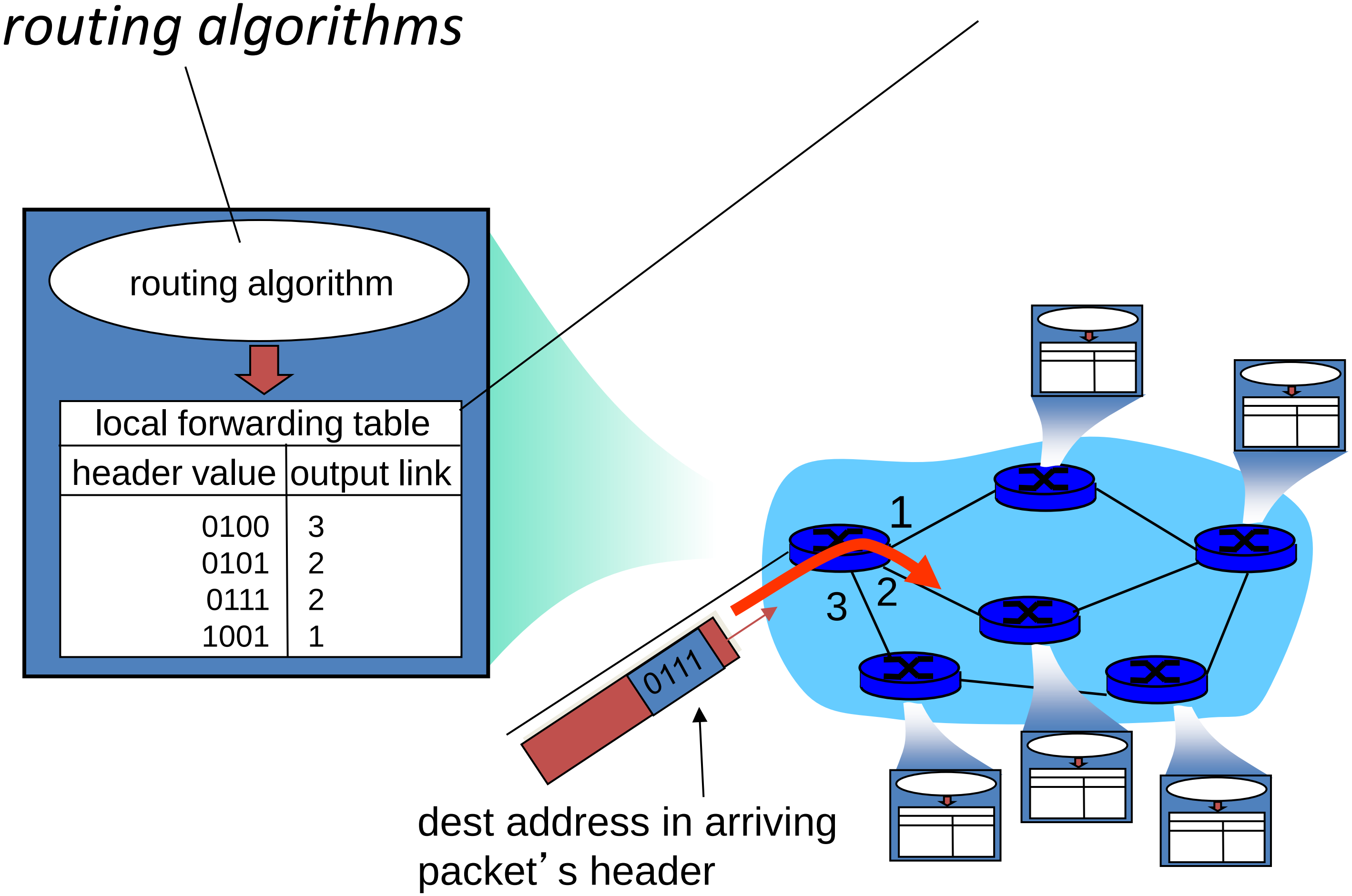
- If arrival rate (in bits) exceeds transmission rate of link:
  - packets will queue, wait to be transmitted
  - packets can be dropped (lost) if memory (buffer) fills up

Two key network-core functions

Routing: Determines route from source to destination taken by packets

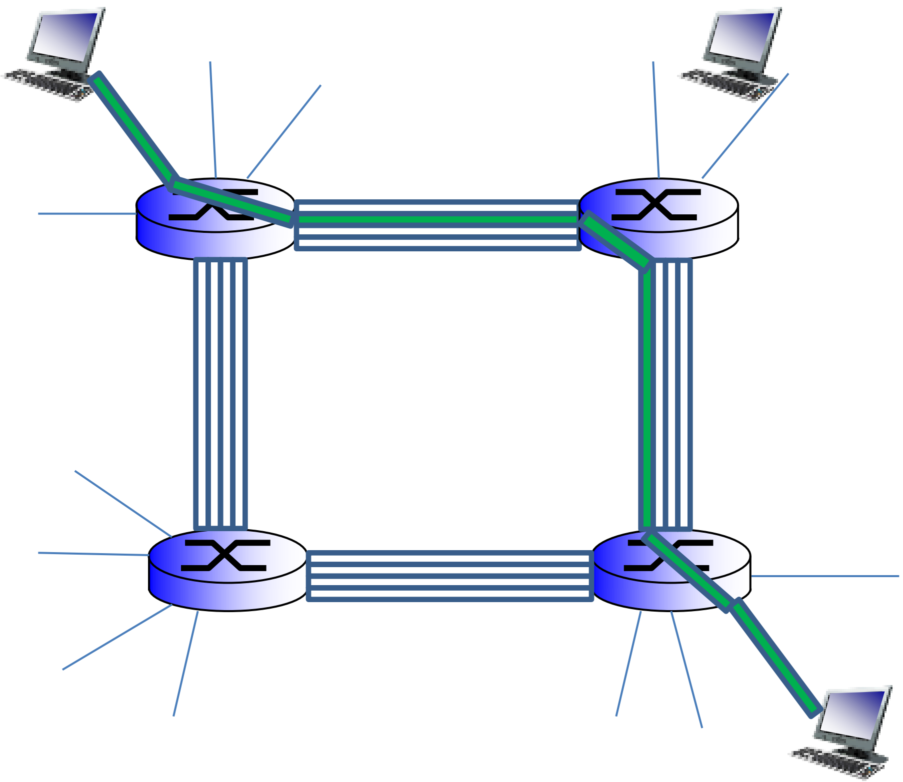
- *routing algorithms*

Forwarding: Move packets from router's input to appropriate router output



Alternative core: circuit switching

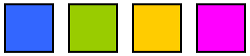
- Circuit switching
  - Resources reserved for "call" between source & dest
  - Dedicated resources: no sharing, idle if not in use
  - Circuit-like (guaranteed) performance
  - Commonly used in traditional telephone networks



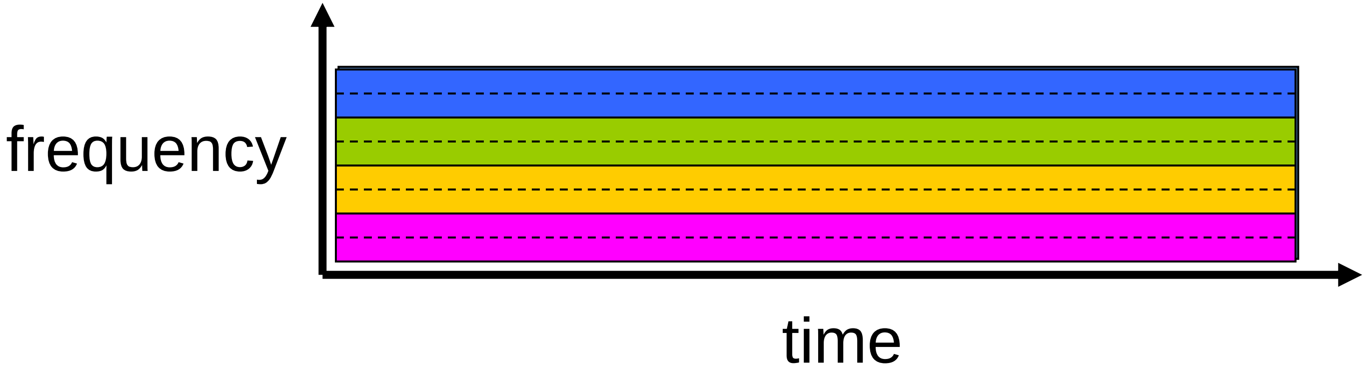
Circuit switching: FDM vs. TDM

Example:

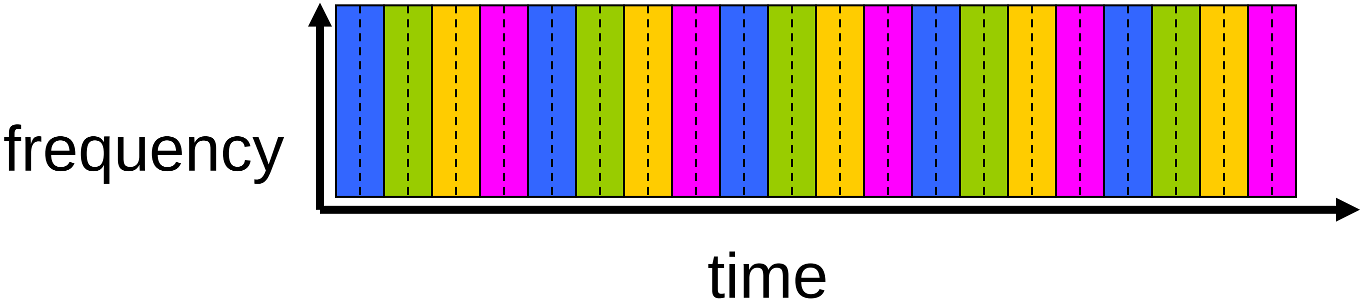
4 users



Frequency Division Multiplexing (FDM)



Time Division Multiplexing (TDM)



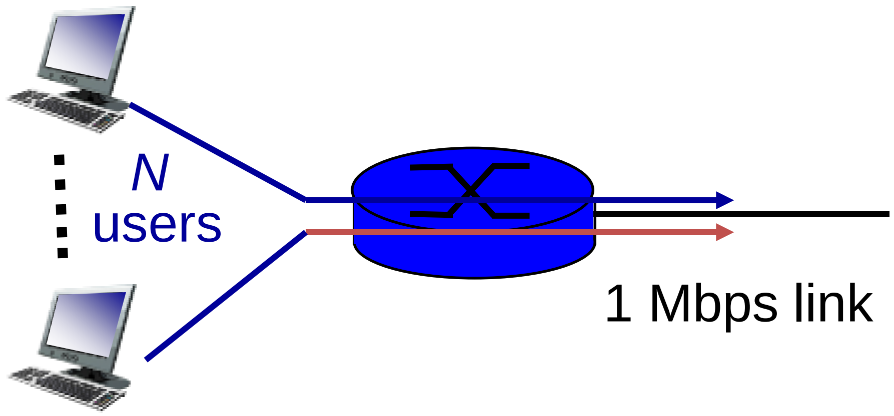
Packet switching vs. circuit switching

Packet switching allows more users to use network!

Example:

- 1 Mb/s link
- Each user:
  - 100 Kb/s when "active"
  - Active 10% of time

- *Circuit-switching:*
  - 10 users
- *Packet switching:*
  - 35 users, probability > 10 active at same time is less than .0004



Q: How did we get value 0.0004?

Q: What happens if > 35 users?

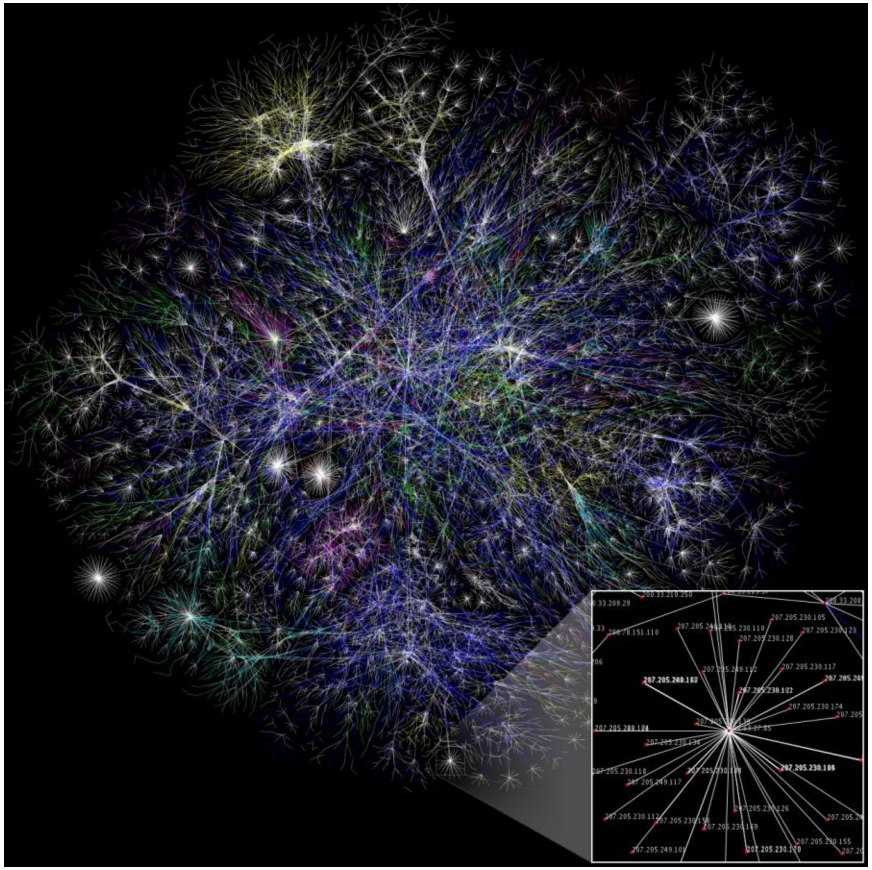
Packet switching vs. circuit switching

Is packet switching a "slam dunk" winner?

- Great for bursty data
  - Resource sharing
  - Simpler, no call setup
- Excessive congestion possible:
  - Packet delay and loss
  - Protocols needed for reliable transfer, congestion control
- *Q:* How to provide circuit-like behavior?
  - Bandwidth guarantees needed for audio/video apps
  - Still an unsolved problem (chapter 7)

Internet structure: network of networks

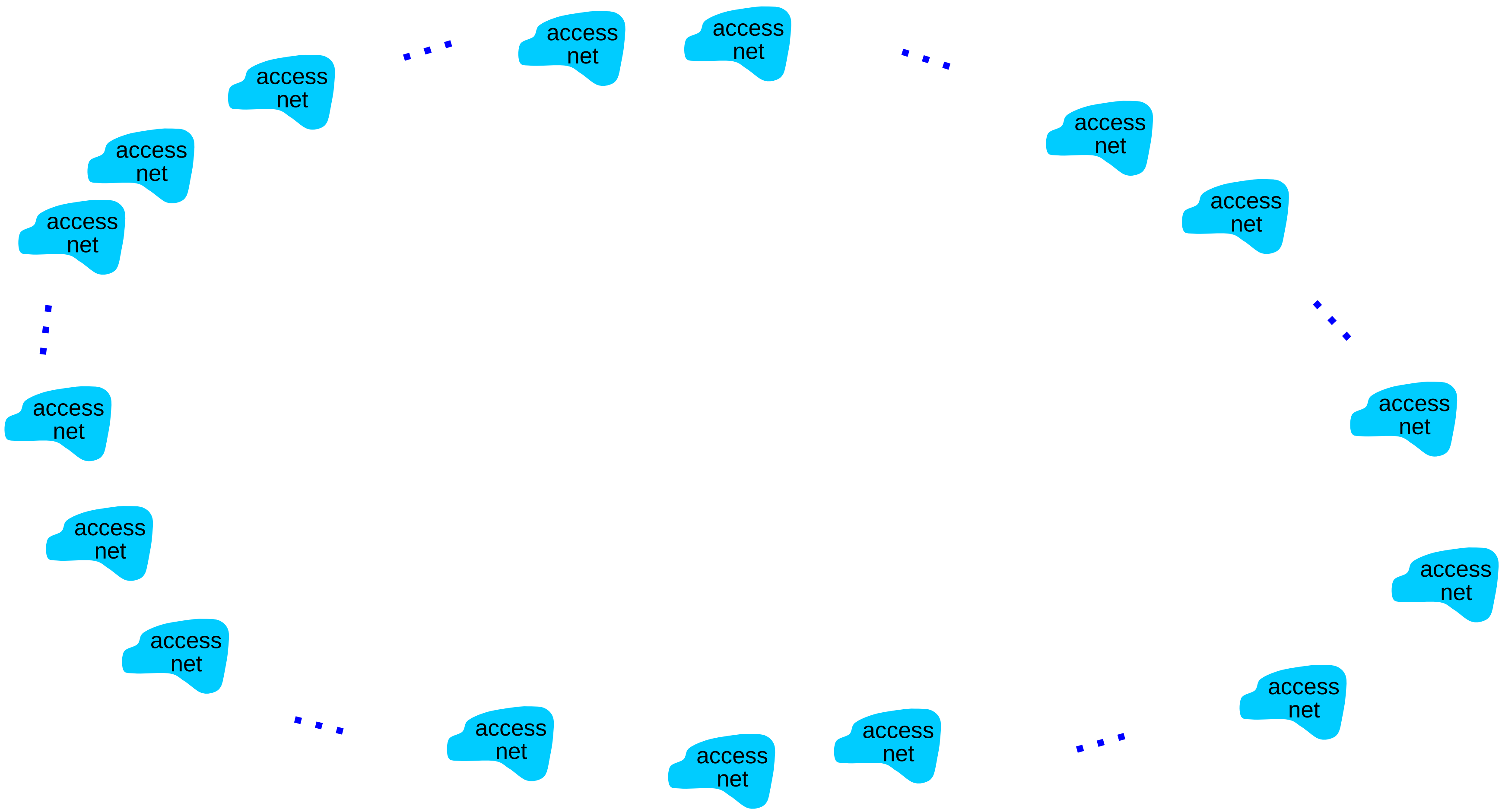
- End systems connect via access ISPs
  - Residential, company and university ISPs
- Access ISPs must be interconnected
  - So any two hosts can send packets to each other
- Resulting network of networks is very complex
  - Evolution driven by economics and national policies



Internet structure: network of networks

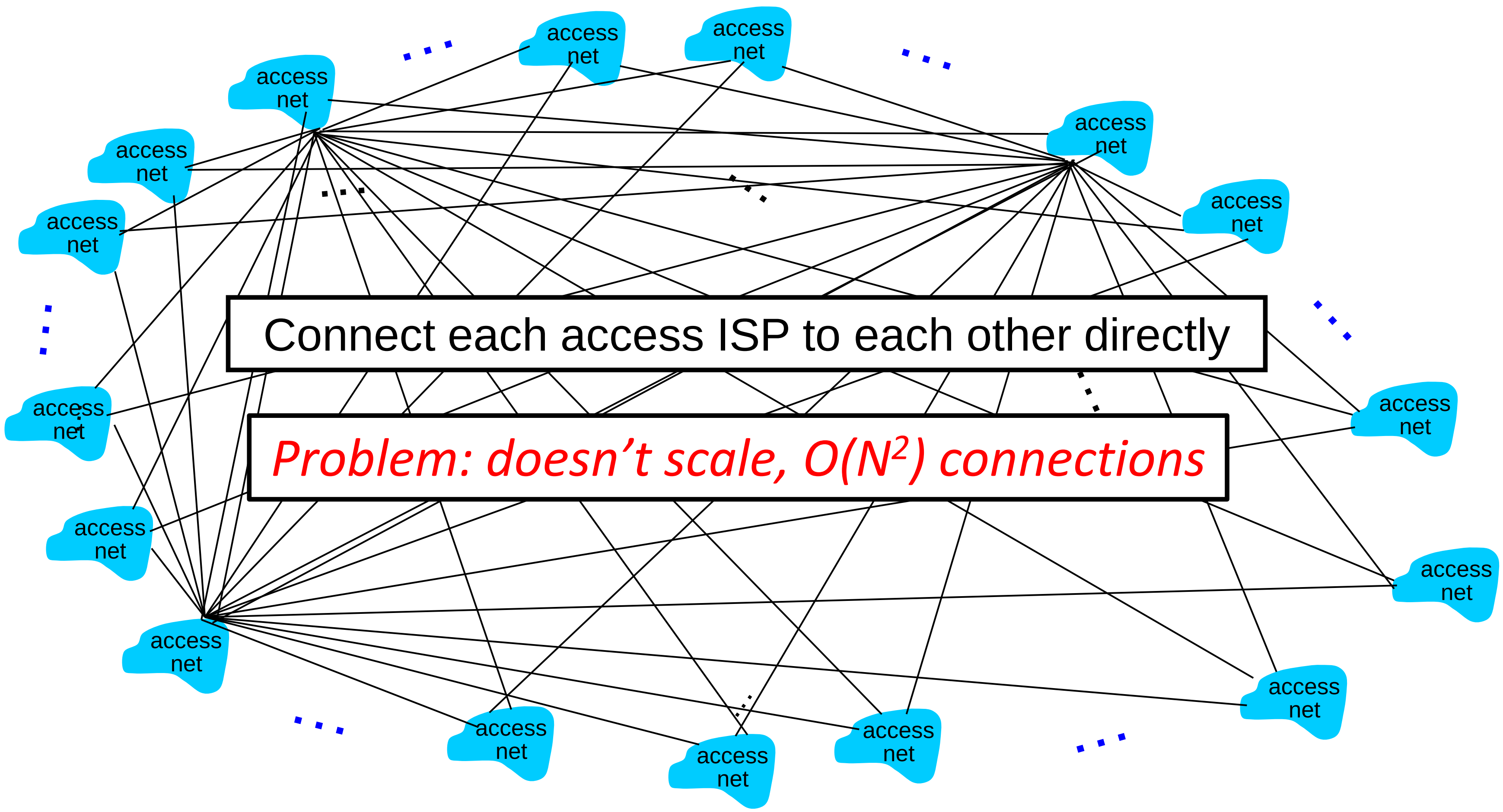
Question:

Given *millions* of access ISPs, how to connect them?



Internet structure: network of networks

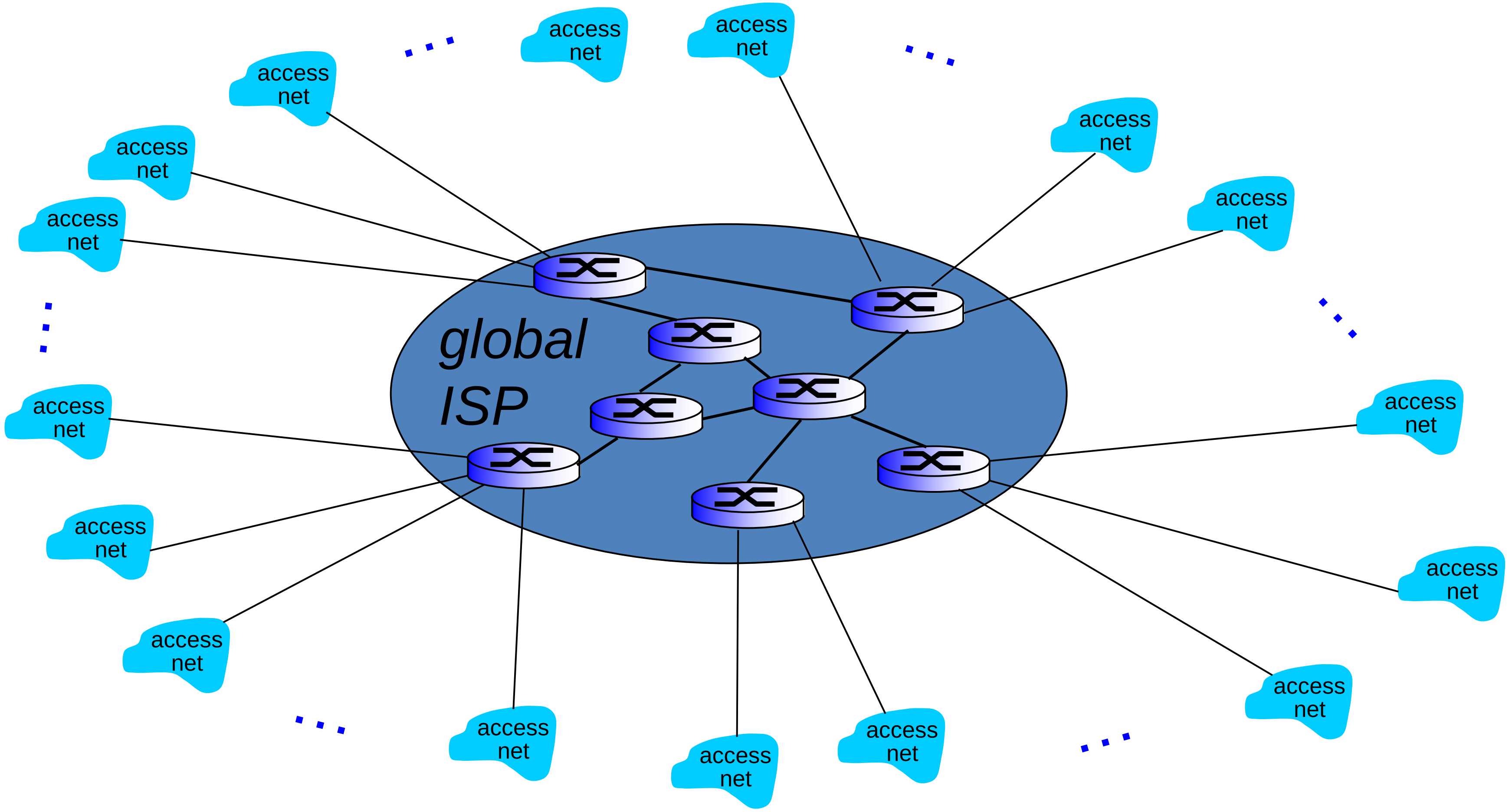
Option: Connect each access ISP to every other access ISP



Internet structure: network of networks

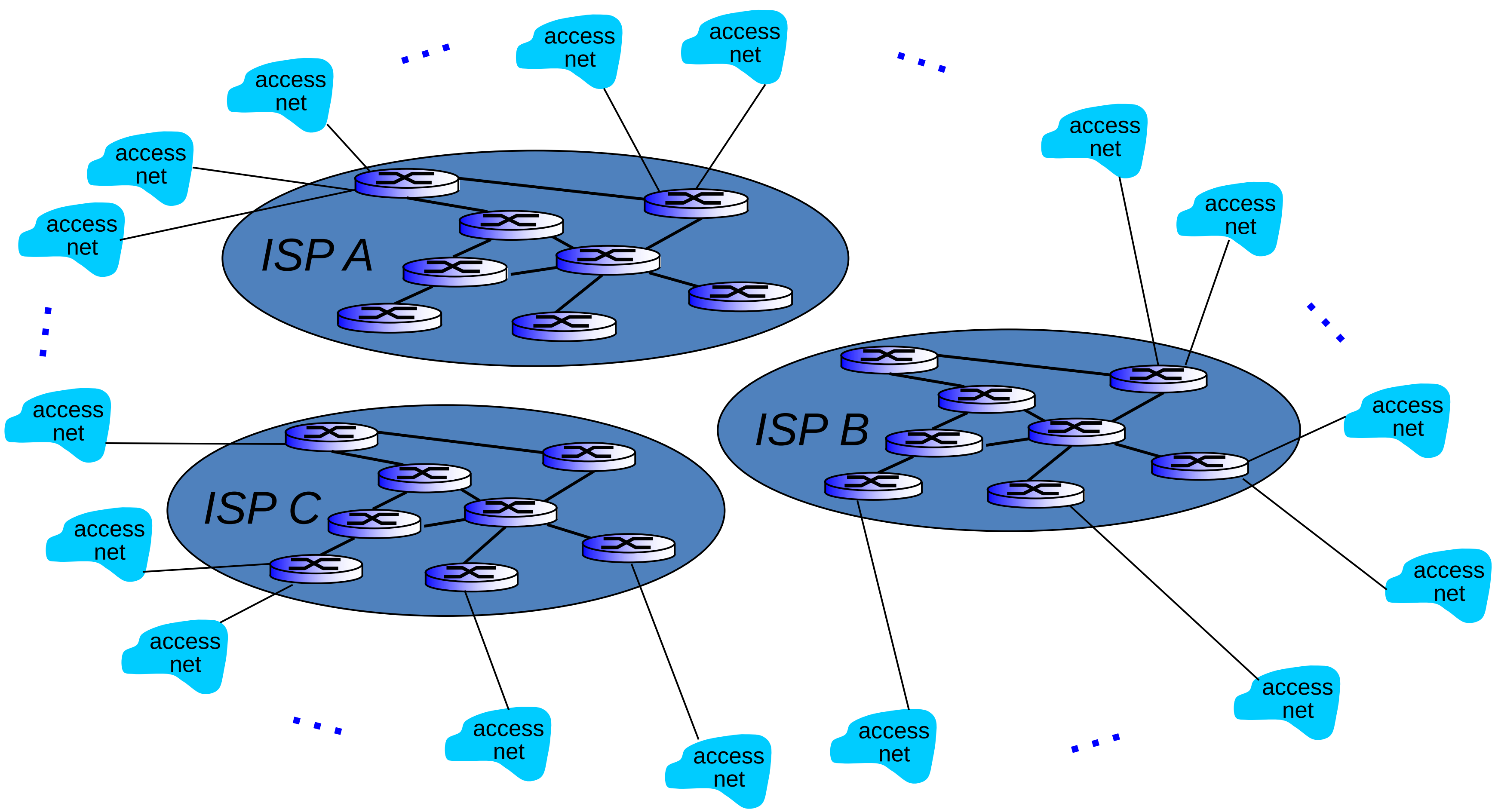
Option: Connect each access ISP to a global transit ISP

Customer and provider ISPs have economic agreement



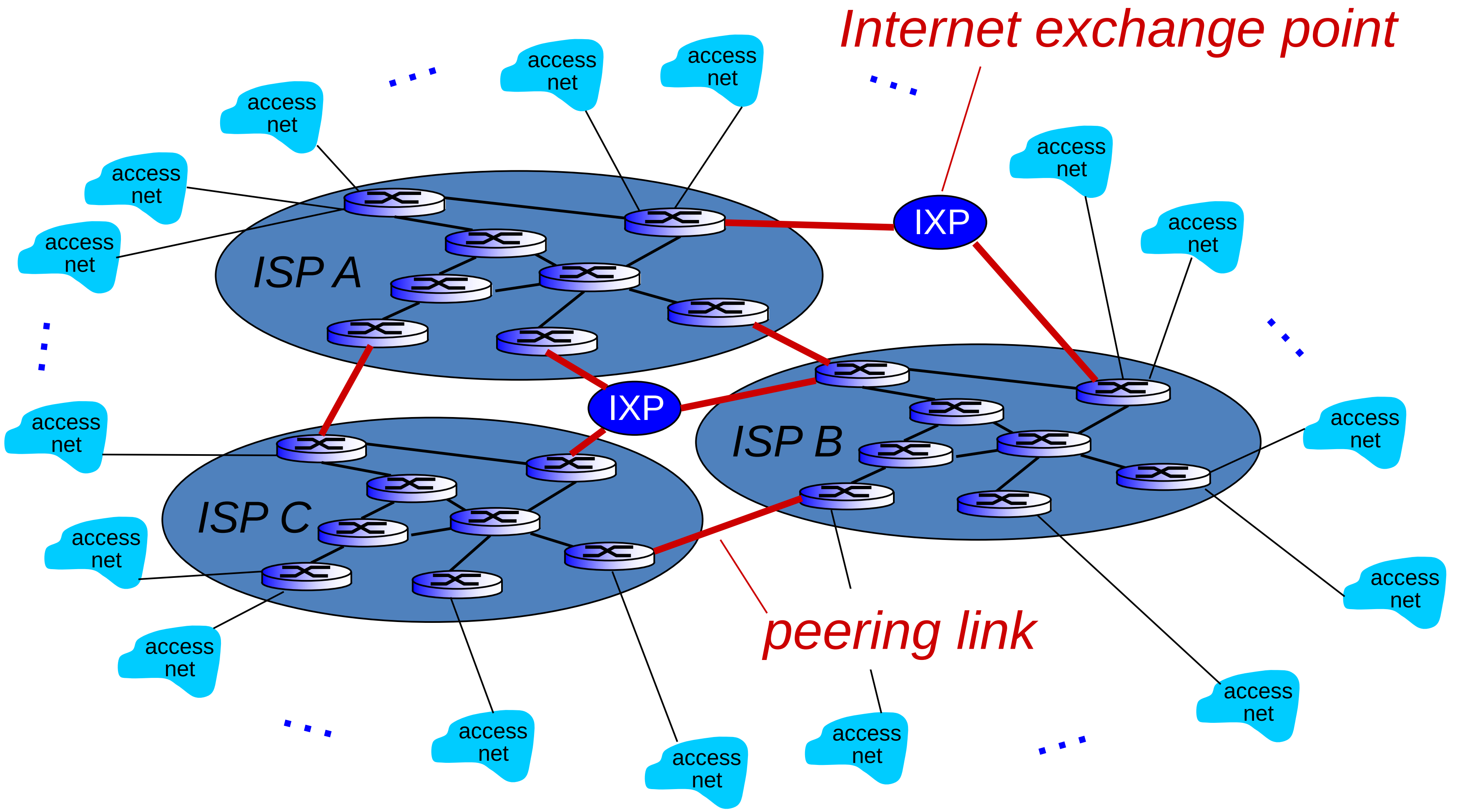
Internet structure: network of networks

But if one global ISP is viable business, there will be competitors ….



Internet structure: network of networks

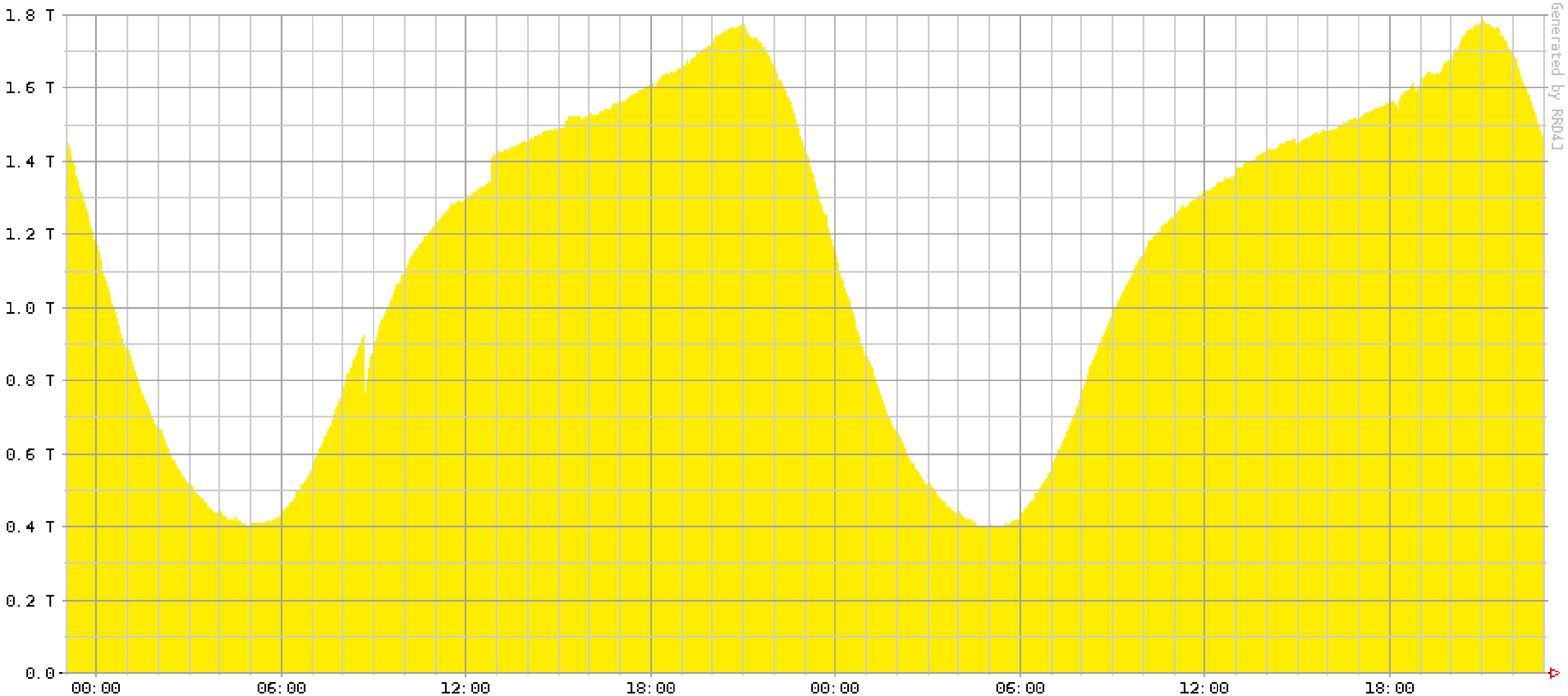
But if one global ISP is viable business, there will be competitors
…. which must be interconnected



Internet exchange point

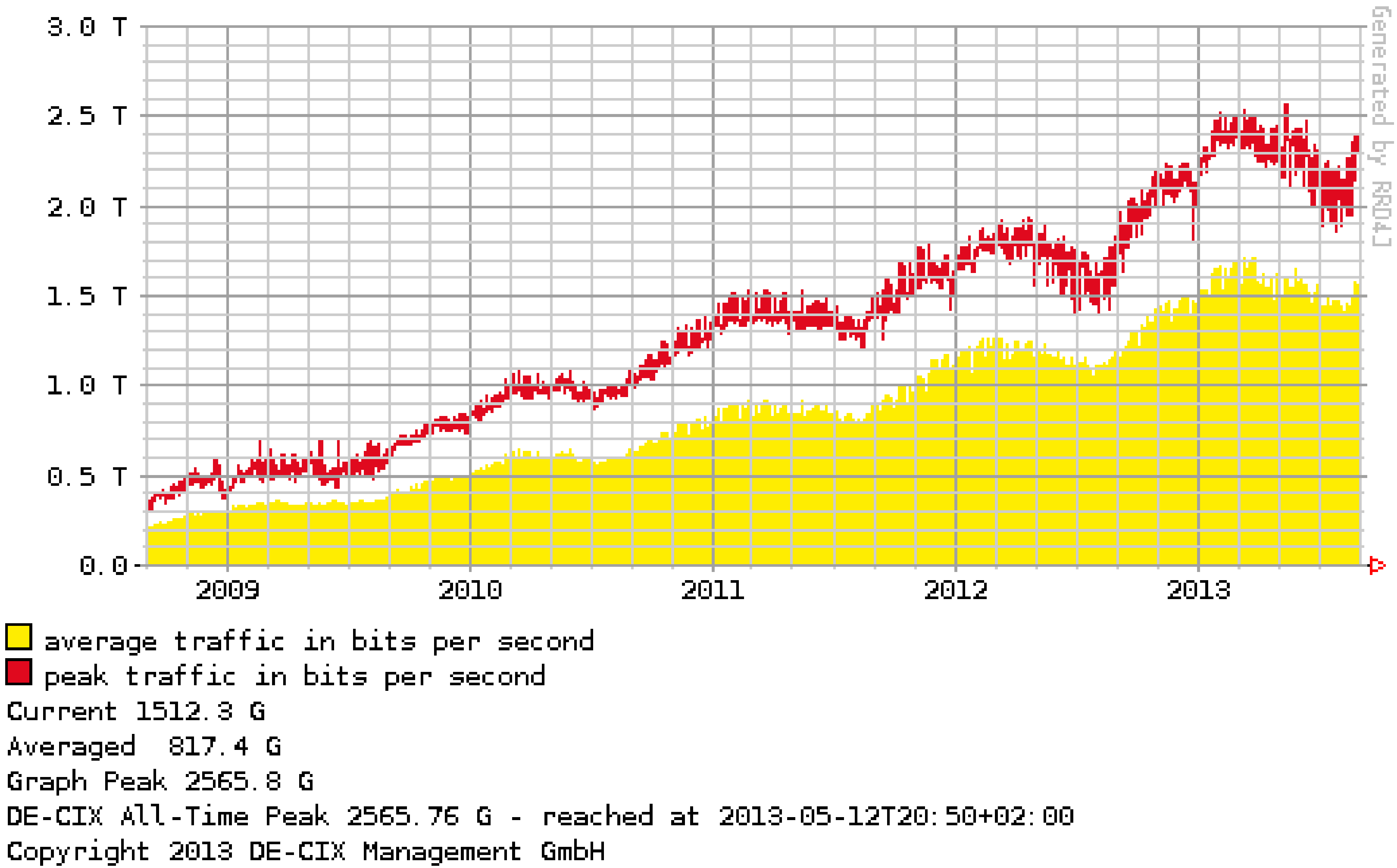
- Internet exchange point
  - Many networks come together in one location
  - Exchange traffic
    - reduce cost
    - improve performance
    - improve reliability
  - e.g. DE-CIX
    - One of the world's largest peering points
    - 465+ ISPs
    - 7 Tbps of capacity
    - 100% uptime since 2007

DE-CIX 2-day graph



average traffic in bits per second
Averaged 1144.1 G
Peak 1811.7 G
Current 1446.5 G

DE-CIX 5-year graph



DE-CIX Topology

DE-CIX Frankfurt relies on the most advanced platform in the industry. In 2013, DE-CIX implemented its new flagship, the **DE-CIX Apollon platform**.

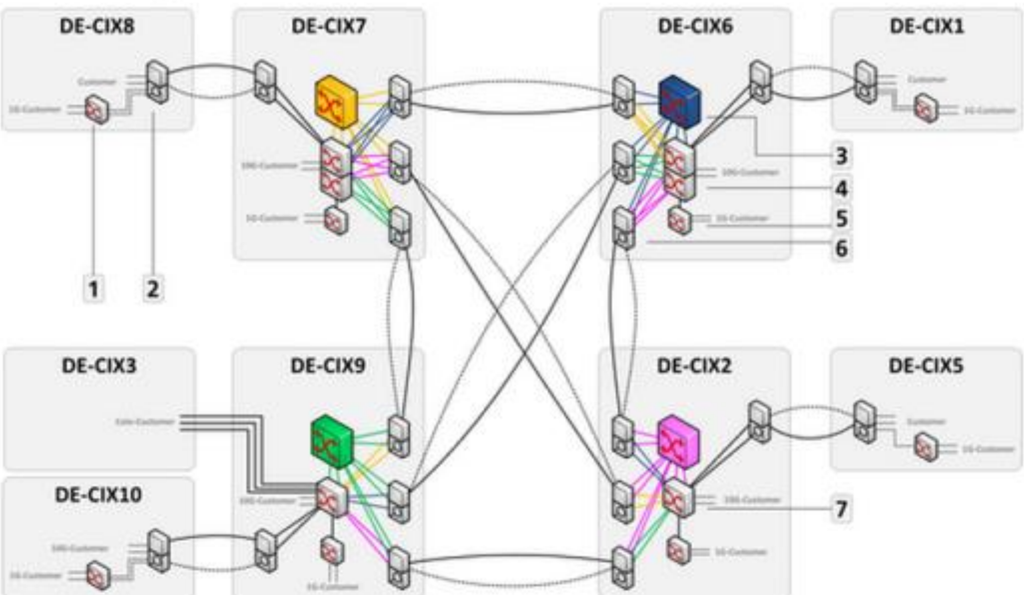
The platform utilizes the ADVA Optical Networking's FSP 3000 for the optical backbone, and Alcatel-Lucent's 7950 XRS.

The optical backbone has a total capacity of 12 terabits per second across a mesh-network topology and provides transport speeds of up to 2 terabits per second per fibre.

The Alcatel-Lucent Core Router 7950 XRS supports a world-leading port density of up to 80 100 Gigabit Ethernet ports. Compared to the old platform, port density has doubled: 320x 100 GE altogether – and is expandable.

DE-CIX Apollon is built of four supernodes, each of them being a combination of an ADVA optical node, an Alcatel-Lucent edge switch and an Alcatel-Lucent core switch. DE-CIX Apollon delivers a 3 to 1 redundancy: all four cores are live, one only for redundancy.

DE-CIX Apollon Platform

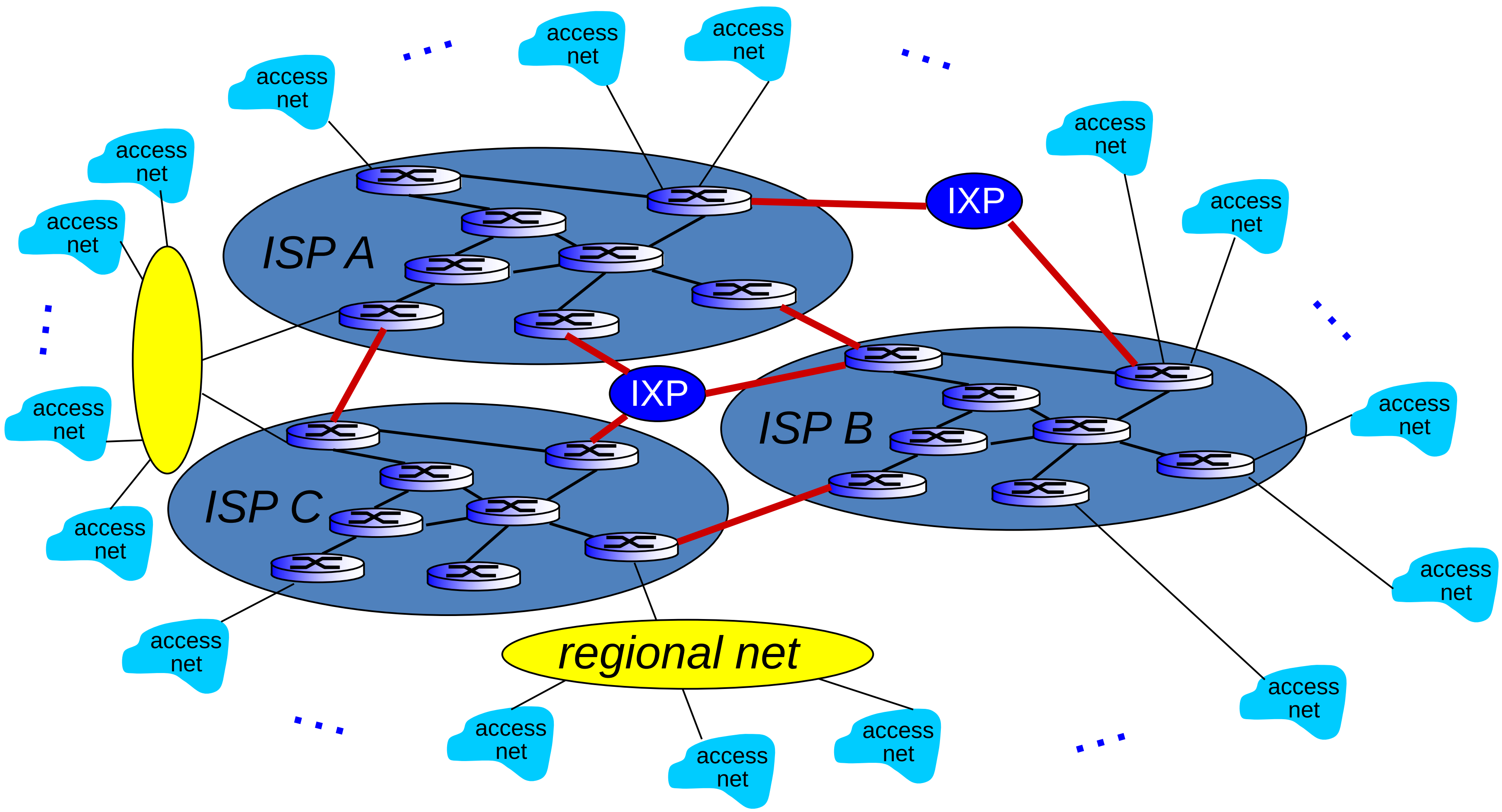


1 Alcatel-Lucent 7210 SAS-M
2 ADVA FSP3000R7 for Remote-Locations
3 Alcatel-Lucent 7950XRS20 Core-Node
4 Alcatel-Lucent 7950XRS40 Edge-Node
5 Alcatel-Lucent 7210 SAS-M
6 ADVA FSP3000R7 for Interconnect-Connections
7 Alcatel-Lucent 7950XRS20 Edge-Node

http://www.de-cix.net/about/topology/

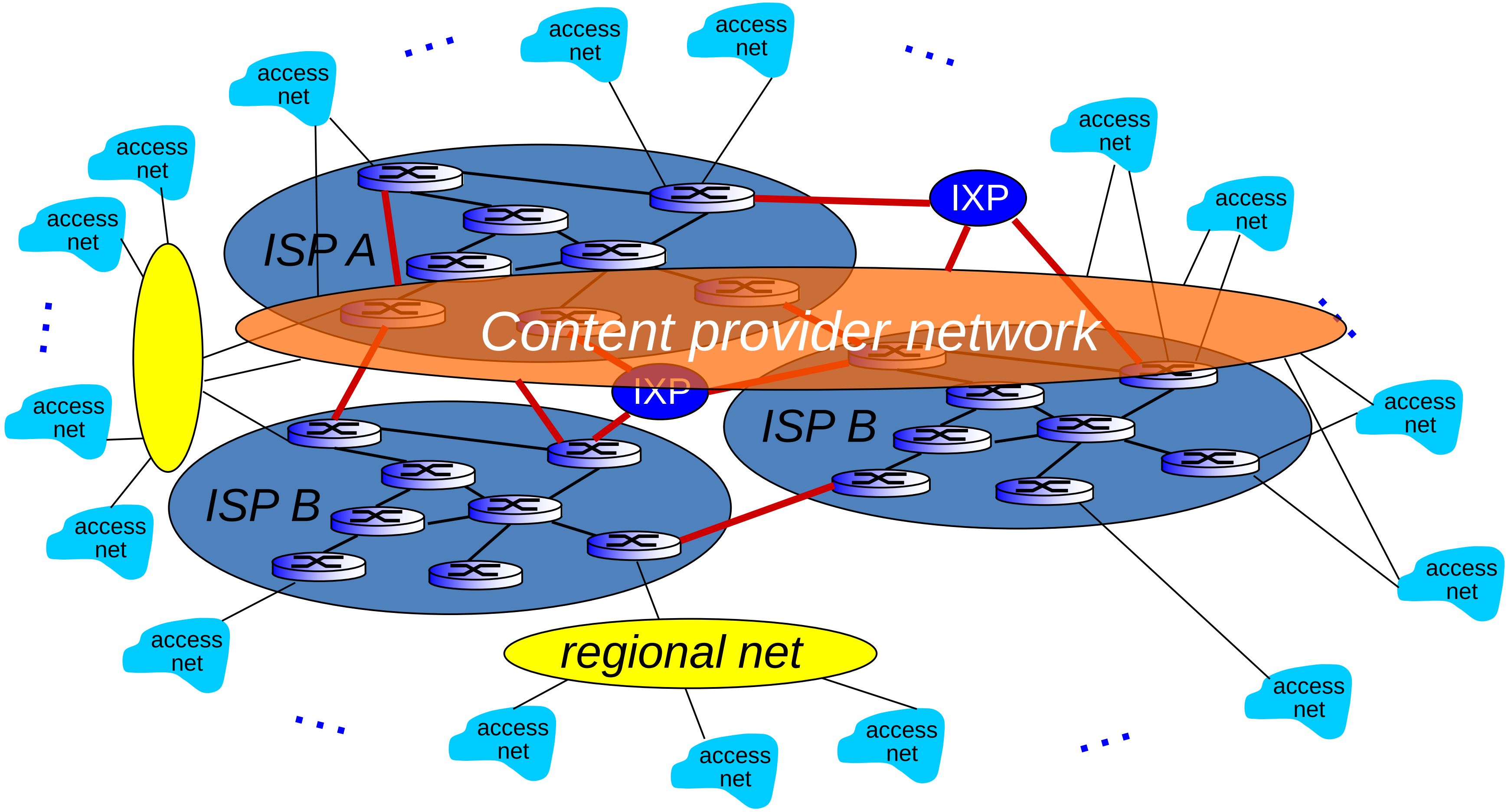
Internet structure: network of networks

... and regional networks may arise to connect access nets to ISPS

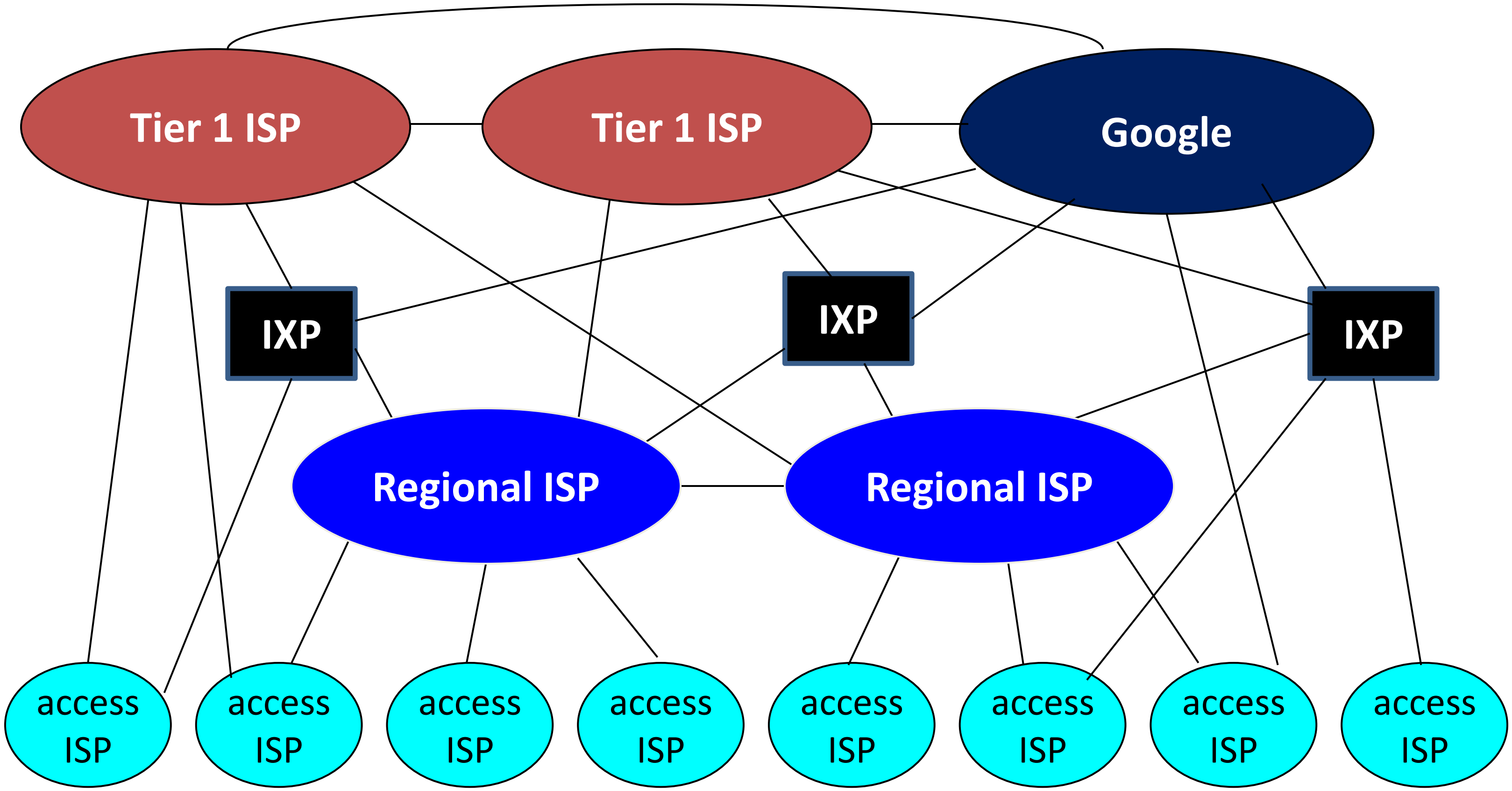


Internet structure: network of networks

… and content provider networks (e.g. Google, Microsoft, Akamai) may run their own network, to bring services, content close to end users

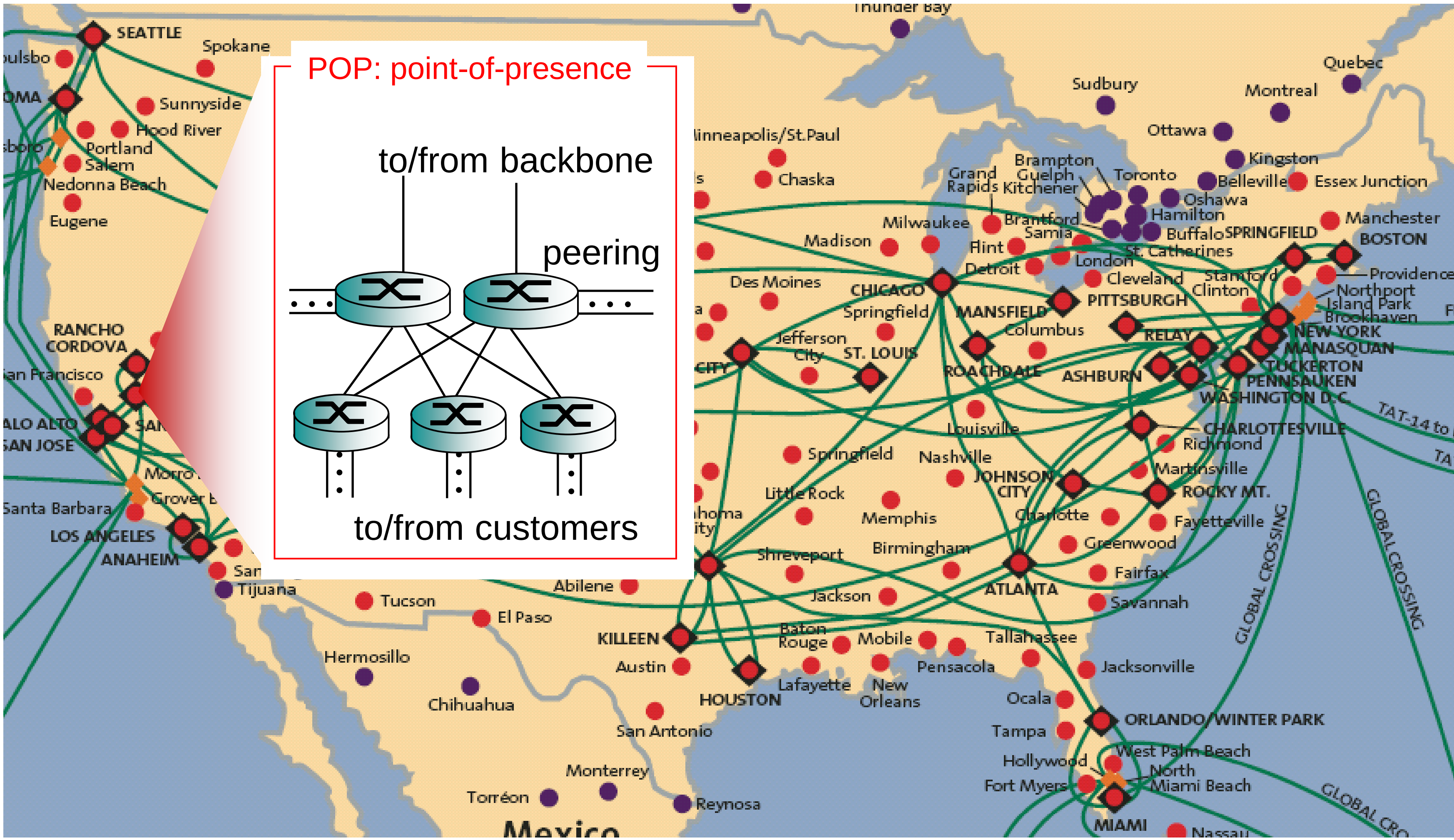


Internet structure: network of networks



- At center: small # of well-connected large networks
  - Tier-1 commercial ISPs (e.g., Level 3, Sprint, AT&T, NTT), national & international coverage
  - Content provider network (e.g, Google): private network that connects it data centers to Internet, often bypassing tier-1, regional ISPs

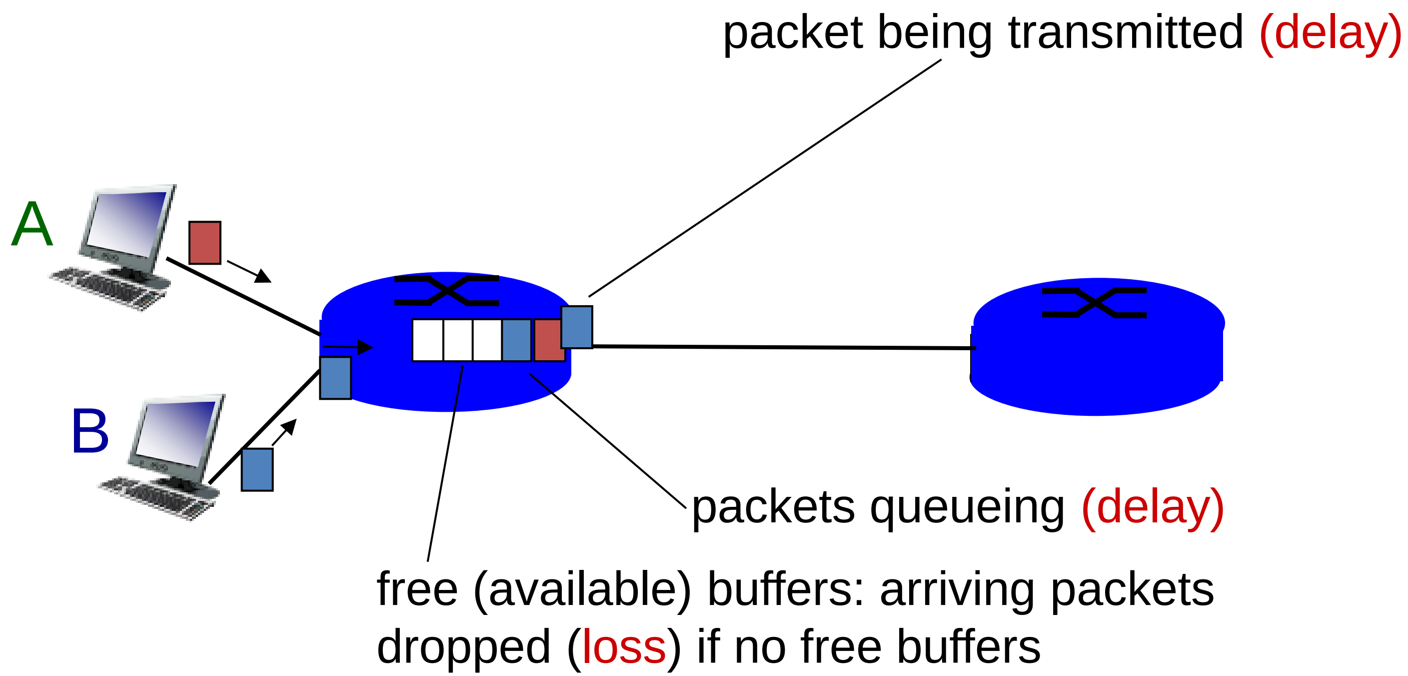
Tier-1 ISP: e.g. Sprint



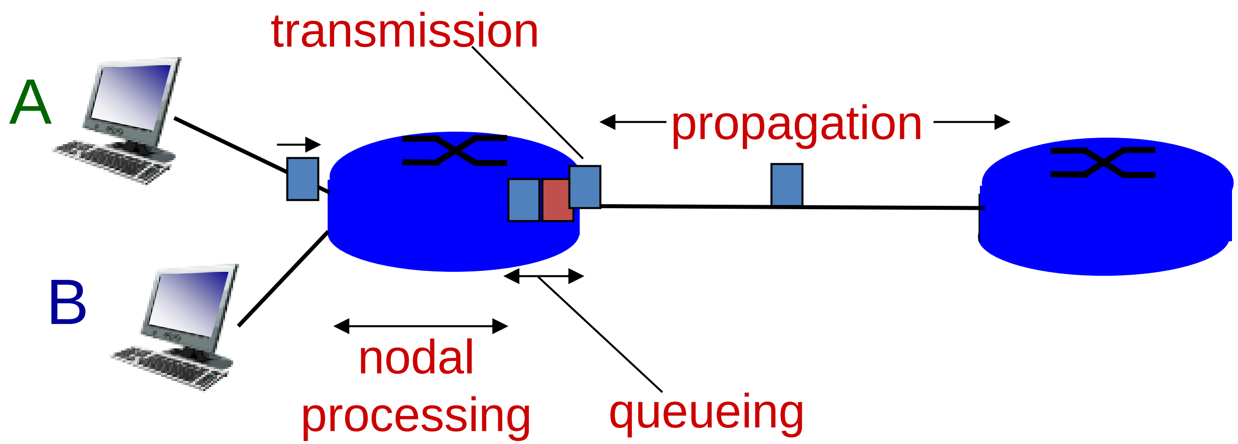
How do loss and delay occur?

Packets *queue* in router buffers

- Packet arrival rate (temporarily) exceeds output capacity
- Packets queue, wait their turn in router's buffer



Four sources of packet delay



$$d_{nodal} = d_{proc} + d_{queue} + d_{trans} + d_{prop}$$

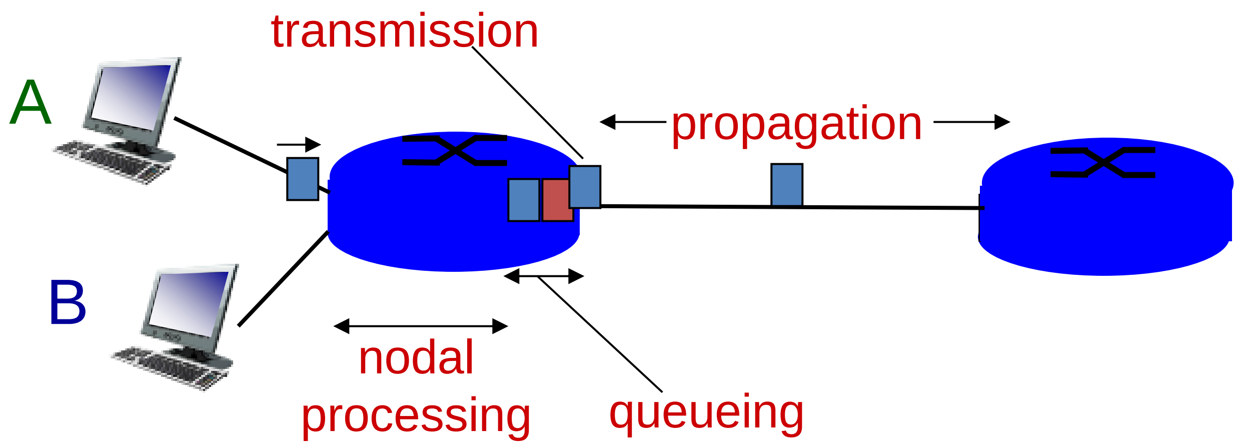
d_{proc}: nodal processing

- Check bit errors
- Determine output link
- Typically < msec

d_{queue}: queueing delay

- Time waiting at output link for transmission
- Depends on congestion level of router

Four sources of packet delay



$$d_{nodal} = d_{proc} + d_{queue} + d_{trans} + d_{prop}$$

d_{trans}: transmission delay

- L: packet length (bits)
- R: link bandwidth (bps)
- $d_{trans} = L / R$

d_{prop}: propagation delay

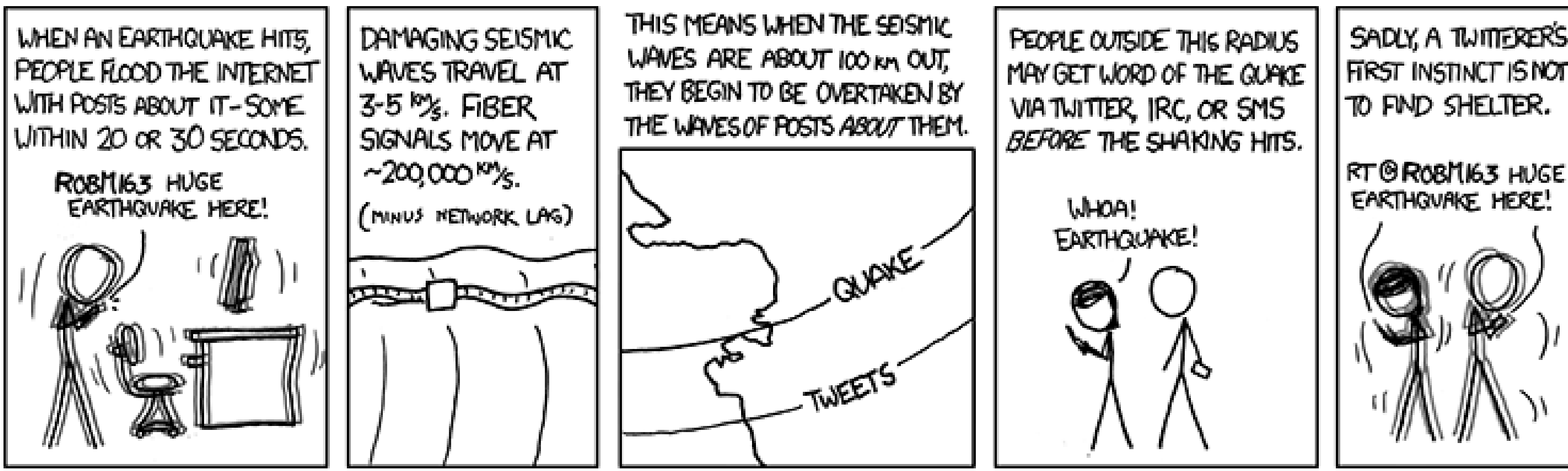
- d: length of physical link
- s: propagation speed in medium (~2×10^8 m/sec)
- $d_{prop} = d / s$

d_{trans} and d_{prop} *very* different

http://media.pearsoncmg.com/aw/aw_kurose_network_2/applets/transmission/delay.html

Speed of light

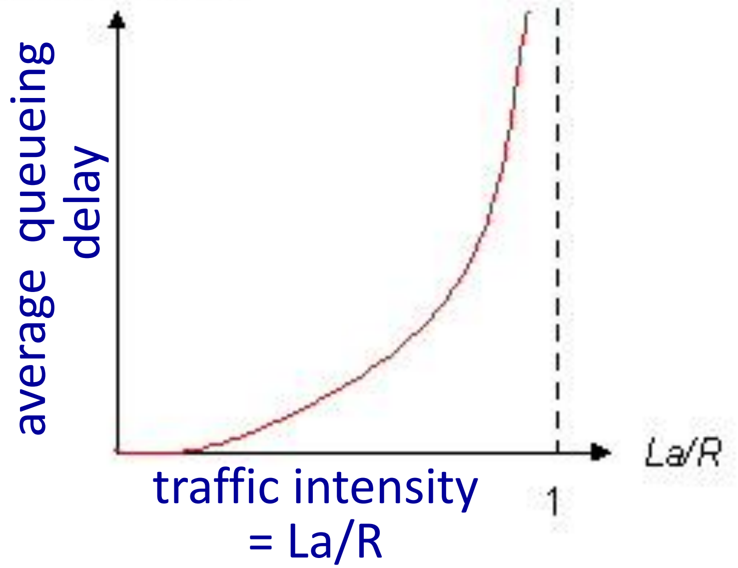
Medium	Speed of light
Vacuum	3.0×10^8 m/s
Copper cable	2.3×10^8 m/s
Optical fiber	2.0×10^8 m/s
Seismic waves	4.0×10^3 m/s



http://xkcd.com/723/

Queueing delay (revisited)

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



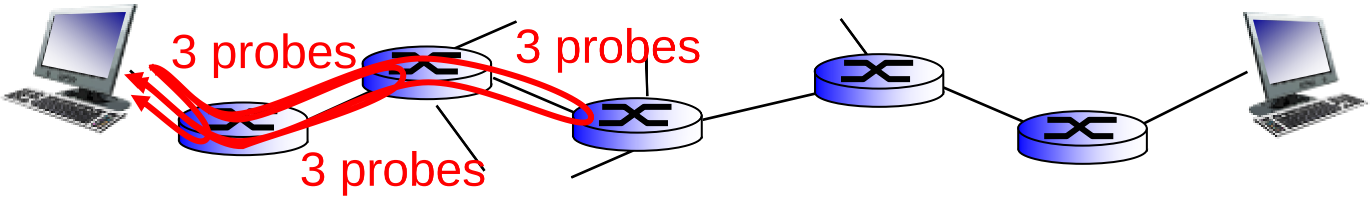
- *La/R* ~ 0: avg. queueing delay small
- *La/R* -> 1: avg. queueing delay large
- *La/R* > 1: more work arriving than can be serviced average delay infinite!

La/R ~ 0

La/R -> 1

"Real" Internet delays and routes

- What do "real" Internet delay & loss look like?
- traceroute program
  - Provides delay measurement from source to router along end-end Internet path towards destination.
  - For all *i:*
    - Sends three packets with time-to-live (TTL) of i
    - Reached router *i* on path towards destination
    - Router *i* will return packets to sender
  - Sender times between transmission and reply



"Real" Internet delays and routes

traceroute: gaia.cs.umass.edu to www.eurecom.fr

3 delay measurements from gaia.cs.umass.edu to cs-gw.cs.umass.edu

```
1  cs-gw (128.119.240.254)  1 ms  1 ms  2 ms
2  border1-rt-fa5-1-0.gw.umass.edu (128.119.3.145)  1 ms  1 ms  2 ms
3  cht-vbns.gw.umass.edu (128.119.3.130)  6 ms 5 ms 5 ms
4  jn1-at1-0-0-19.wor.vbns.net (204.147.132.129)  16 ms 11 ms 13 ms
5  jn1-so7-0-0-0.wae.vbns.net (204.147.136.136)  21 ms 18 ms 18 ms
6  abilene-vbns.abilene.ucaid.edu (198.32.11.9)  22 ms  18 ms  22 ms
7  nycm-wash.abilene.ucaid.edu (198.32.8.46)  22 ms  22 ms  22 ms
8  62.40.103.253 (62.40.103.253)  104 ms 109 ms 106 ms
9  de2-1.de1.de.geant.net (62.40.96.129)  109 ms 102 ms 104 ms
10  de.fr1.fr.geant.net (62.40.96.50)  113 ms 121 ms 114 ms
11  renater-gw.fr1.fr.geant.net (62.40.103.54)  112 ms  114 ms  112 ms
12  nio-n2.cssi.renater.fr (193.51.206.13)  111 ms  114 ms  116 ms
13  nice.cssi.renater.fr (195.220.98.102)  123 ms  125 ms  124 ms
14  r3t2-nice.cssi.renater.fr (195.220.98.110)  126 ms  126 ms  124 ms
15  eurecom-valbonne.r3t2.ft.net (193.48.50.54)  135 ms  128 ms  133 ms
16  194.214.211.25 (194.214.211.25)  126 ms  128 ms  126 ms
17  * * *
18  * * *
19  fantasia.eurecom.fr (193.55.113.142)  132 ms  128 ms  136 ms
```

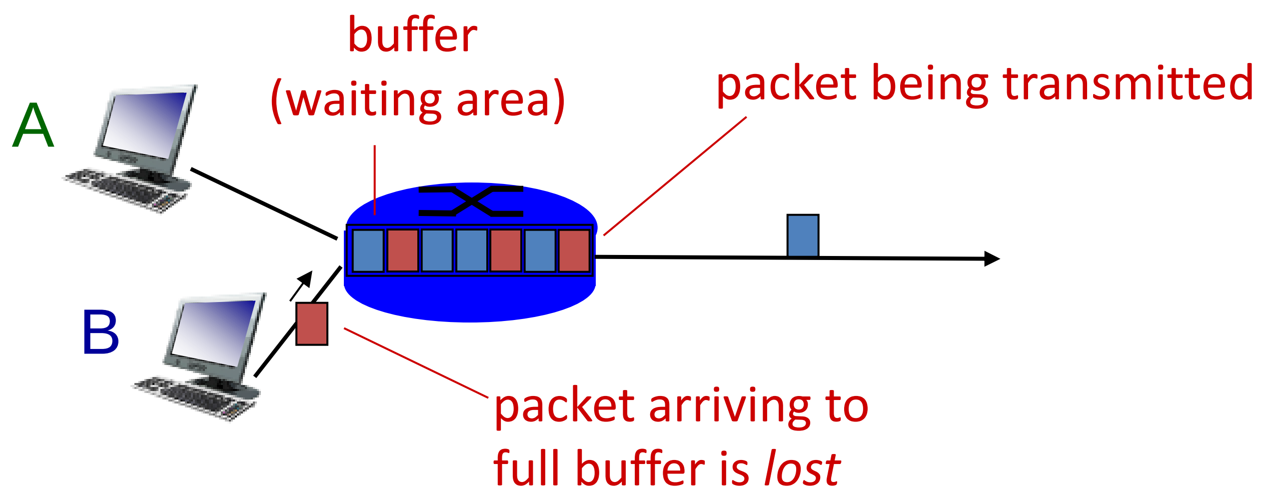
trans-oceanic link

* means no response (probe lost, router not replying)

http://traceroute.monitis.com

Packet loss

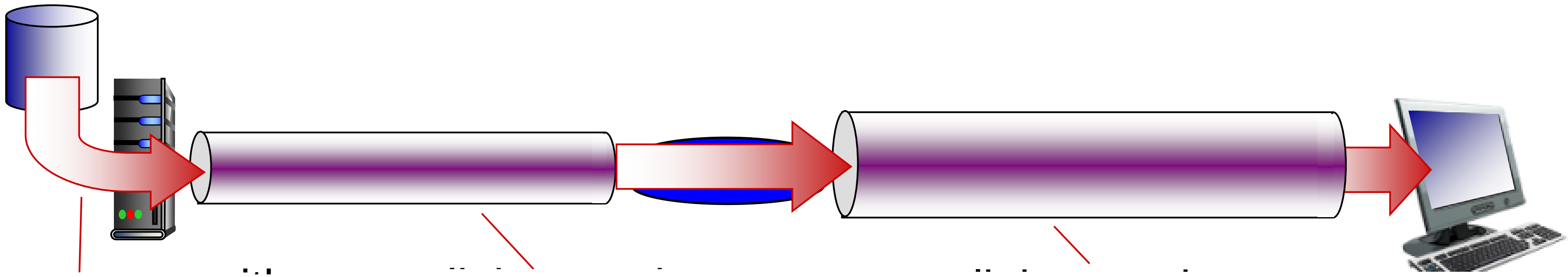
- Queue (aka buffer) preceding link has finite capacity
  - Packet arriving to full queue dropped (aka lost)
  - Lost packet may be retransmitted by previous node, by source end system, or not at all



http://media.pearsoncmg.com/aw/aw_kurose_network_2/applets/queuing/queuing.html

Throughput

- *Throughput:* rate (bits/time unit) at which bits transferred between sender/receiver
  - *instantaneous:* rate at given point in time
  - *average:* rate over longer period of time



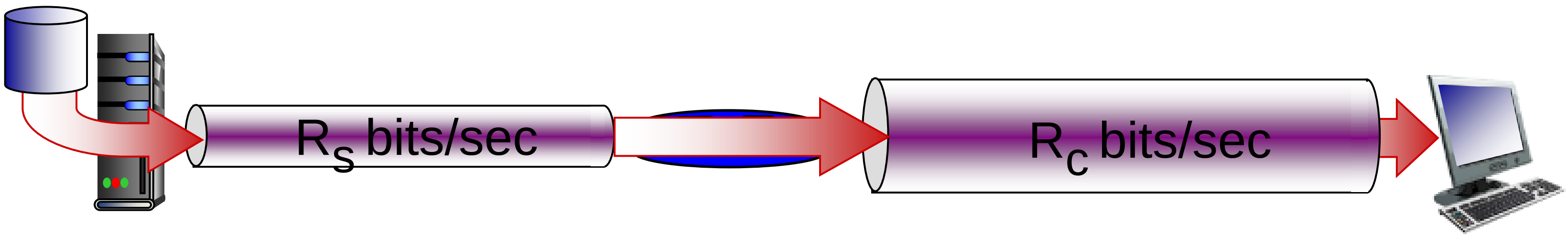
server sends bits (fluid) into pipe

pipe that can carry fluid at rate R_s bits/sec)

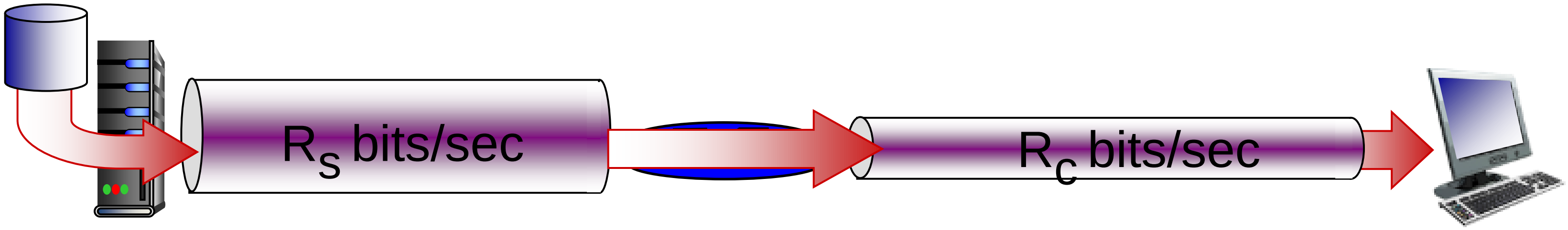
pipe that can carry fluid at rate R_c bits/sec)

Throughput

- $R_s < R_c$ What is average end-end throughput?



- $R_s > R_c$ What is average end-end throughput?

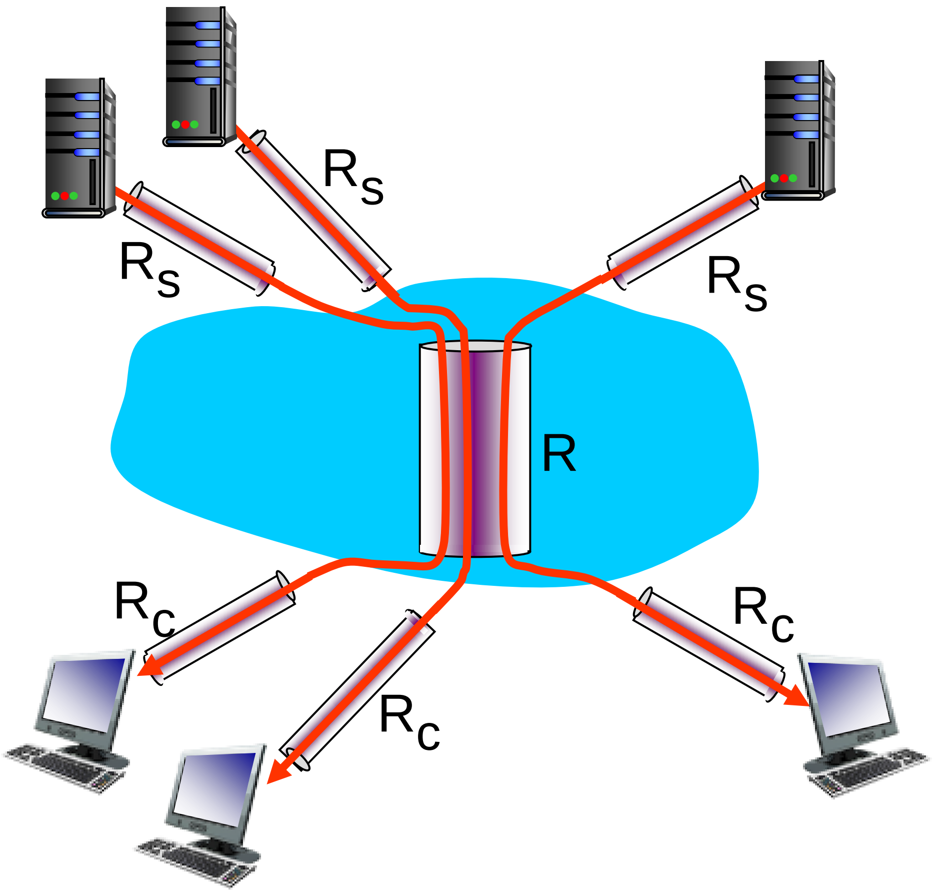


bottleneck link

link on end-end path that constrains end-end throughput

Throughput: Internet scenario

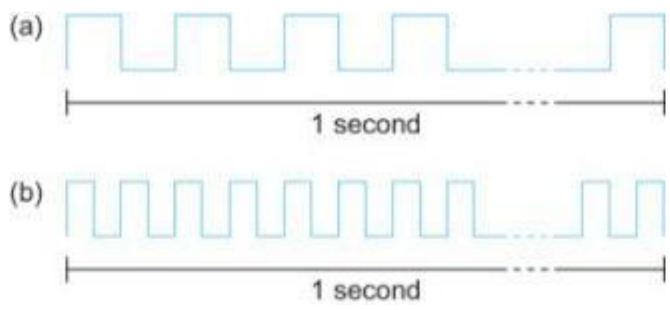
- Per-connection end-end throughput:
  - $\min(R_c, R_s, R/10)$
- In practice:
  - R_c or R_s is often bottleneck



10 connections (fairly) share backbone bottleneck link R bits/sec

Bandwidth

- ***Bandwidth*** - measure of the frequency band
  - e.g. voice telephone line, frequencies from 300-3300 Hz, bandwidth = 3000 Hz
- ***Bandwidth*** - bits transmitted per unit time
  - 1 Mbps = 1×10^6 bits/second
  - e.g. 802.11g wireless has a bandwidth of 54 Mbps
    - Bandwidth, mega = 1×10^6 = 1000000
    - File size, mega = 2^{20} = 1048576



- ***Throughput*** - actual obtainable performance
  - e.g. 802.11g wireless has a throughput of ~22 Mbps

Watch your units!

- ***Bandwidth***
  - gigabits (Gbps) $= 10^9$ bits/second
  - megabits (Mbps) $= 10^6$ bits/second
  - kilobits (Kbps) $= 10^3$ bits/second
- ***File sizes***
  - 8 bits / byte
  - gigabyte (GB) $= 2^{30}$ bytes
  - megabyte (MB) $= 2^{20}$ bytes
  - kilobyte (KB) $= 2^{10}$ bytes

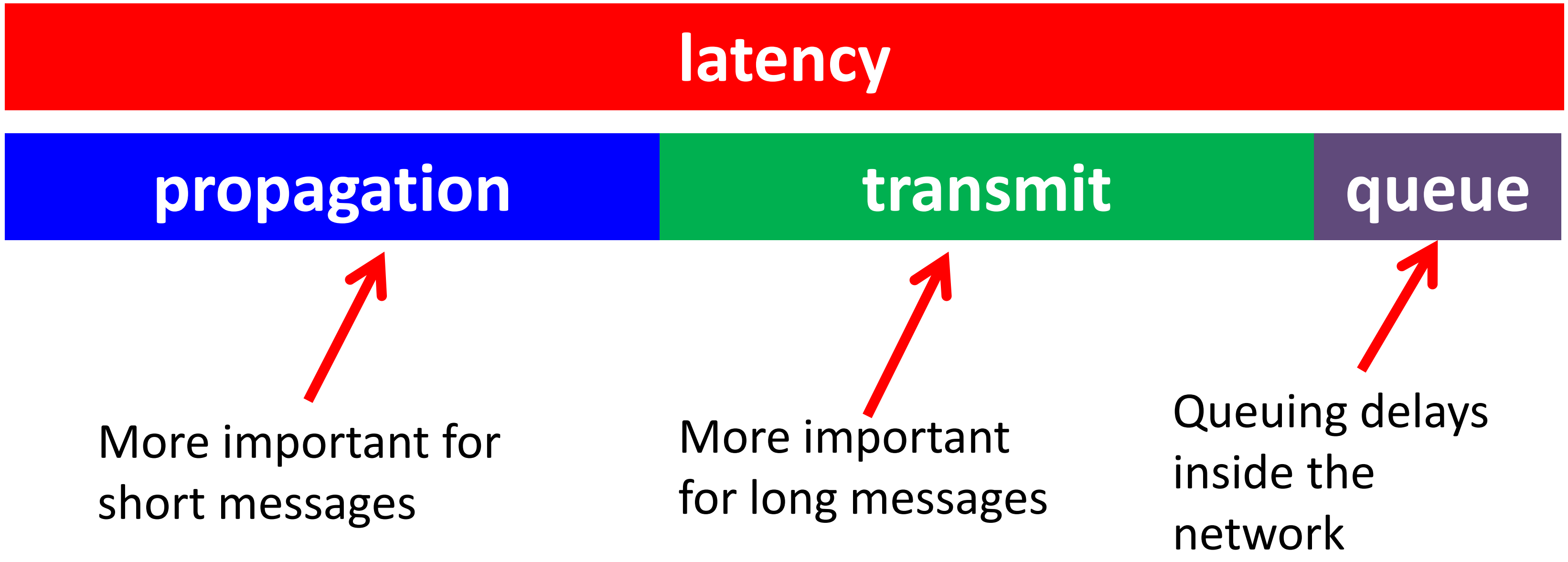
Multiples of bits				v · t · e
SI decimal prefixes		Binary usage	IEC binary prefixes	
Name (Symbol)	Value		Name (Symbol)	Value
kilobit (kbit)	10^3	2^{10}	kibibit (Kibit)	2^{10}
megabit (Mbit)	10^6	2^{20}	mebibit (Mibit)	2^{20}
gigabit (Gbit)	10^9	2^{30}	gibibit (Gibit)	2^{30}
terabit (Tbit)	10^{12}	2^{40}	tebibit (Tibit)	2^{40}
petabit (Pbit)	10^{15}	2^{50}	pebibit (Pibit)	2^{50}
exabit (Ebit)	10^{18}	2^{60}	exbibit (Eibit)	2^{60}
zettabit (Zbit)	10^{21}	2^{70}	zebibit (Zibit)	2^{70}
yottabit (Ybit)	10^{24}	2^{80}	yobibit (Yibit)	2^{80}
See also: Nibble · Byte · Multiples of bytes Orders of magnitude of data				

Latency

- ***Latency*** or ***delay*** - how long it takes a message to go from one end of network to other
  - Measured in units of time (often ms)
- ***Round-trip time (RTT)*** - how long from source to destination and back to source
- ***Jitter*** - variance in latency (affects time sensitive applications)

Latency

- latency = propagation + transmit + queue
- propagation = distance / speed of light
- transmit = size / bandwidth



Effect of file size

- Throughput = Transfer size / Transfer time
- Transfer time = RTT + 1/Bandwidth x Transfer size

File size (MB)	RTT	Bandwidth (Gbps)	Transmit time (ms)	Transfer time (ms)	Throughput (Mbps)
0.25	100	1	2.1	102.1	**19.6**
0.50	100	1	4.2	104.2	**38.4**
1	100	1	8.4	108.4	**73.8**
2	100	1	16.8	116.8	**137.0**
4	100	1	33.6	133.6	**239.6**
8	100	1	67.1	167.1	**383.0**
16	100	1	134.2	234.2	**546.5**

Summary

- Network core
  - Mesh of routers and links connecting end systems
  - Packet switching versus circuit switching
  - Network structure
    - Tier 1 ISPs, content providers, regional ISPs, access ISPs, Internet exchange points
- Metrics
  - Measuring performance of the network
    - Processing delay, transmission delay, queueing delay, propagation delay, throughput, latency, RTT, jitter
    - traceroute utility