Telcordia Technologies
Performance from Experience

LTS

Next-Generation Data Networks: Architecture and Engineering

Seminar 1: Elements of IP Network Design

Stuart Wagner
ssw@research.telcordia.com

February 11, 2002

Seminar Schedule (Tentative)

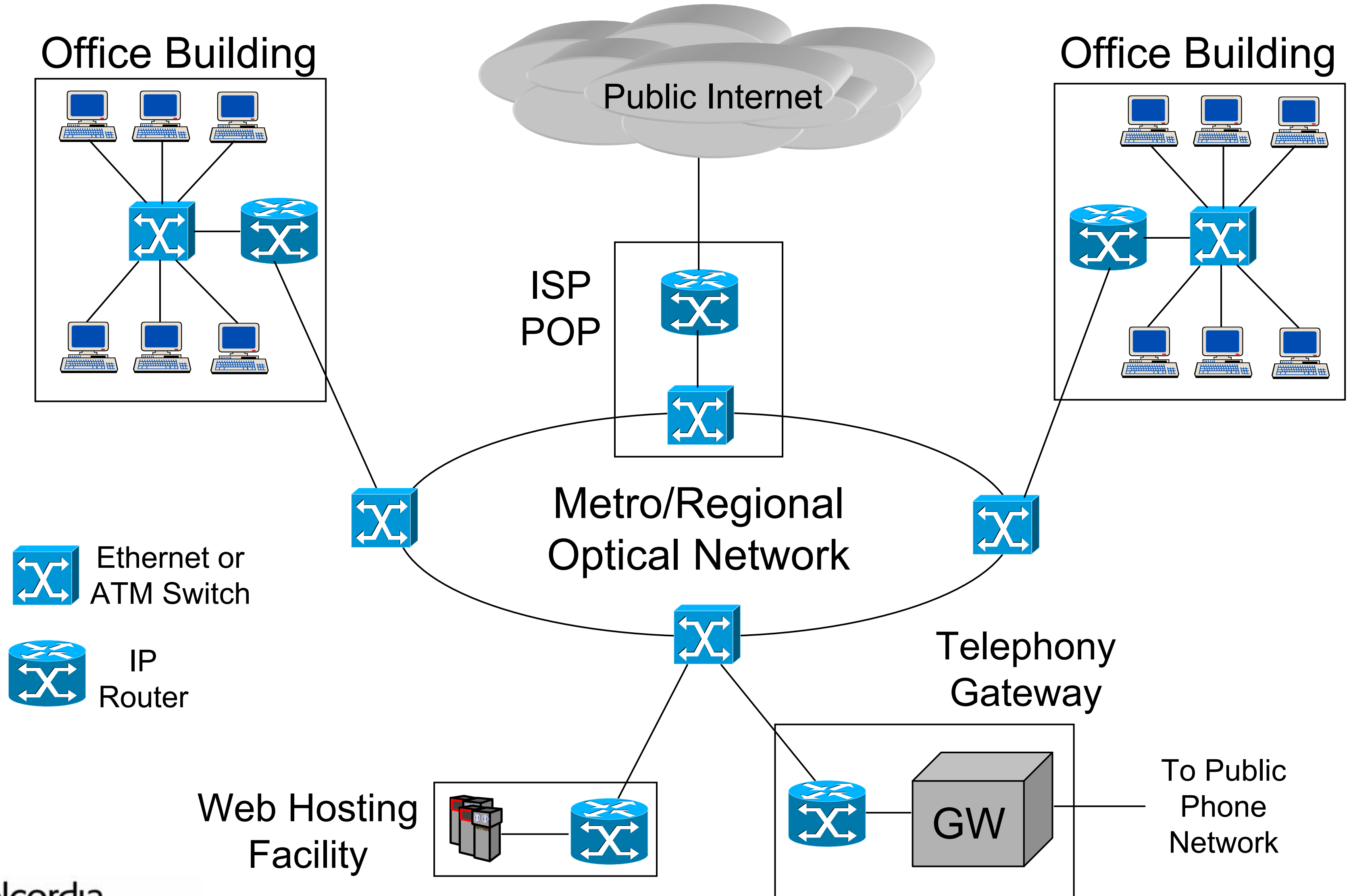
- IP Network Design (Feb. 11)
- Multiprotocol Label Switching (Feb. 25)
- Optical Networking (March 11)
- Gigabit Ethernet (April 1)
- Voice over IP (April 15)
- Wireless data networking (April 29)
- Data network security (May 13)

Seminar Series Objectives

- Highlight the fundamental principles and considerations governing data-network design
- Include perspectives on current trends within the commercial industry (carriers, equipment suppliers)
- Identify major research issues

Telcordia Technologies
Performance from Experience

Data Network Architecture Example



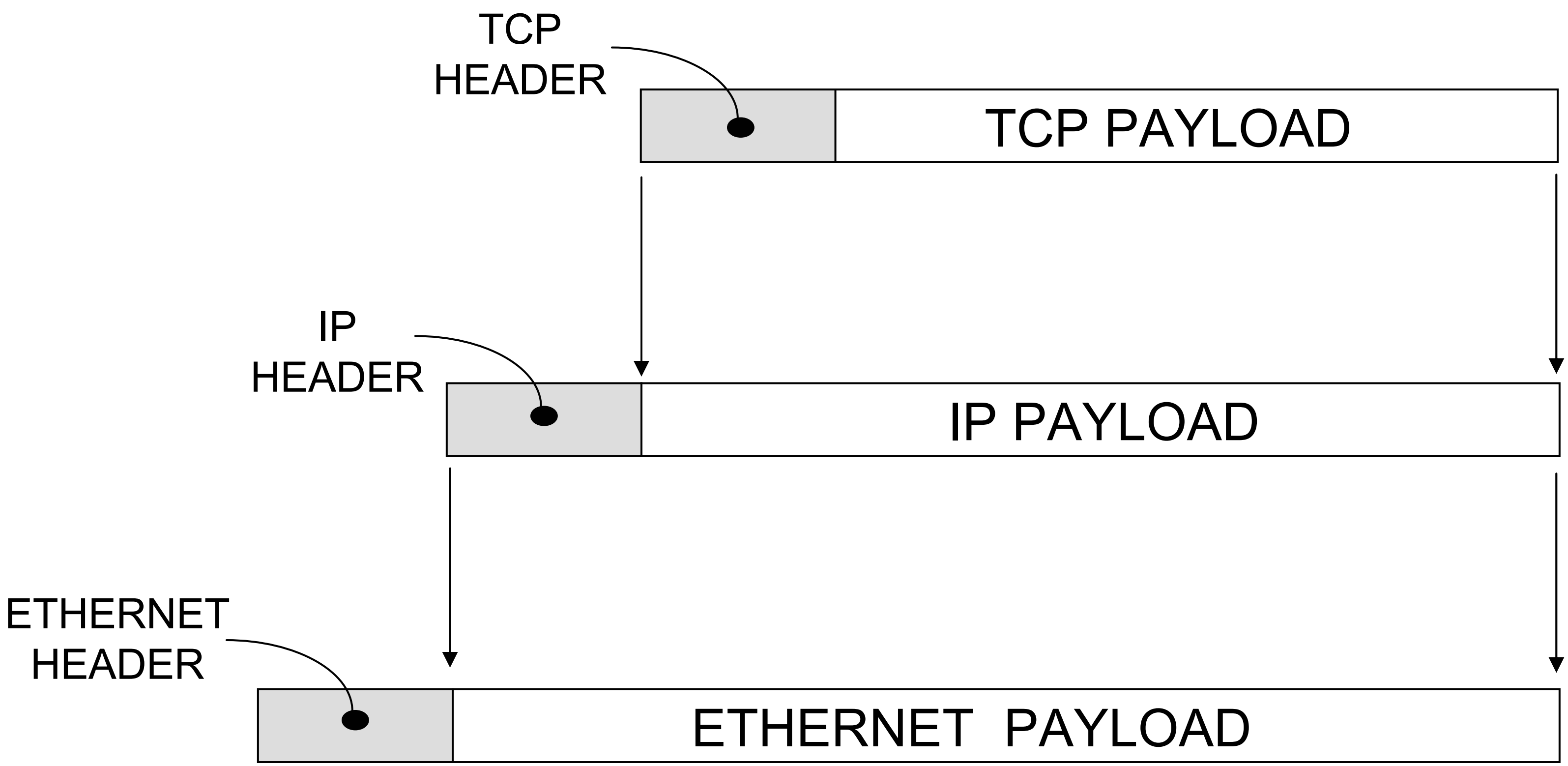
Telcordia Technologies
Performance from Experience

Breaking Up The Problem - Network Layering

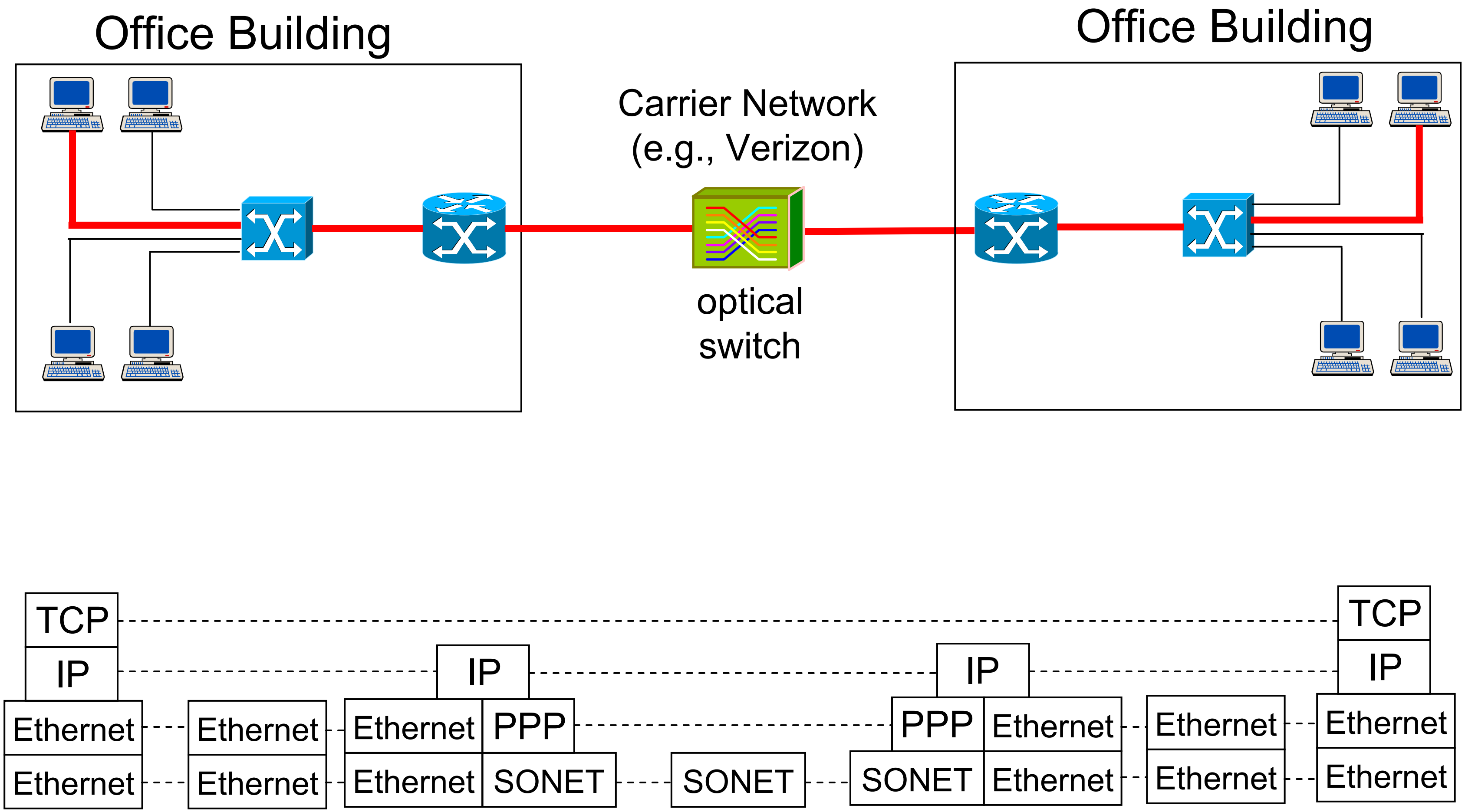
The “ISO 7-Layer Model”

7	Application	
6	Presentation	
5	Session	
4	Transport	*TCP, UDP, ICMP*
3	Network	*IP*
2	Data Link	*Ethernet, ATM, PPP, Frame Relay*
1	Physical	*Ethernet, SONET, Optical*

Layered Packet Format



Example of Network Layering

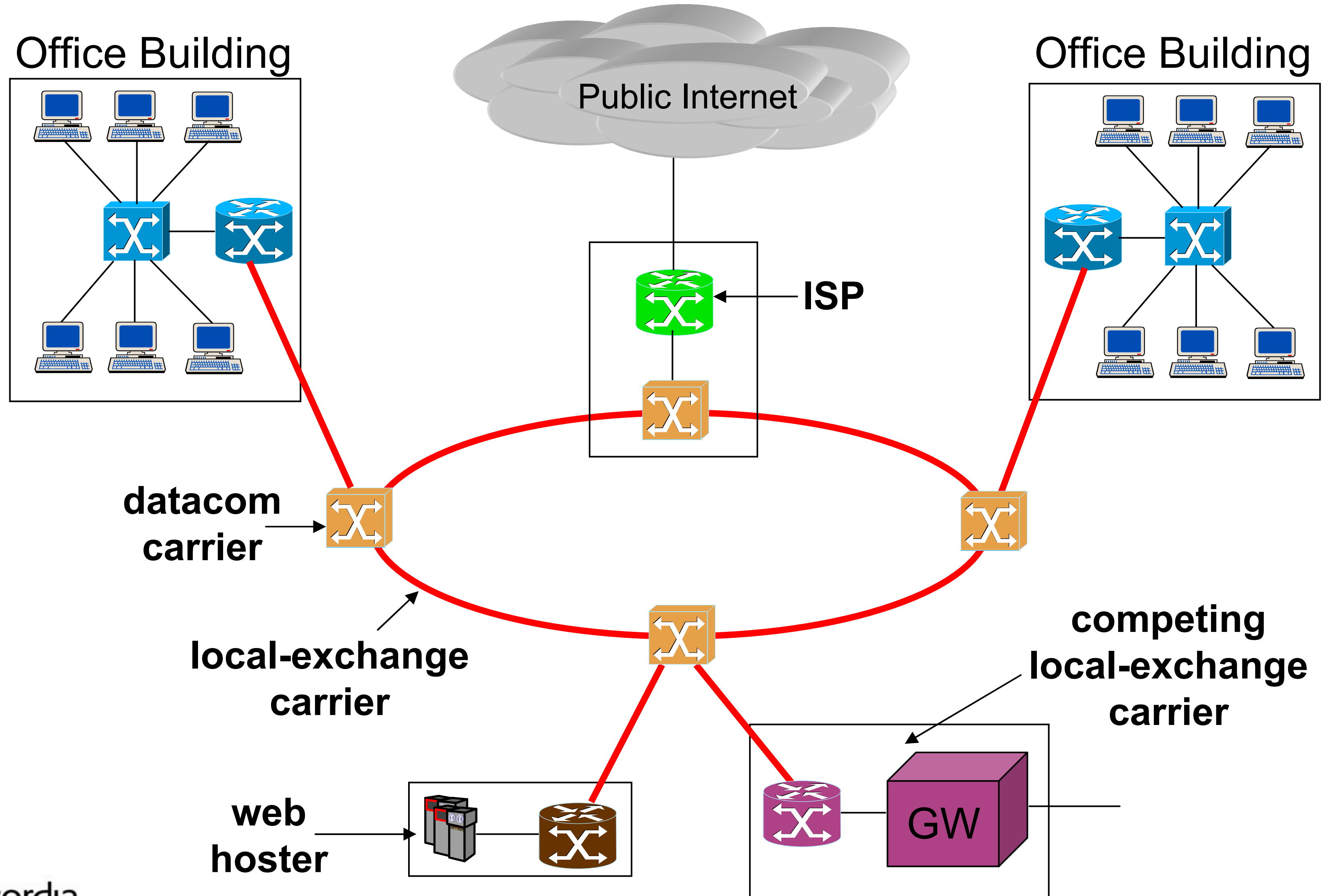


Observations on Network Layering

- Each layer has its own role and responsibilities
- Each layer depends on the ones below it, but can often detect and/or recover from errors in those lower layers
- Real networks do not always obey this strict layered model
  - Multiprotocol Label Switching (MPLS) is “layer 2.5”
  - Routers may perform processing based on layer-4 header content (firewall filtering, address translation, “layer-4 switching”, etc.)
- Different layers of the network may be owned and operated by separate businesses

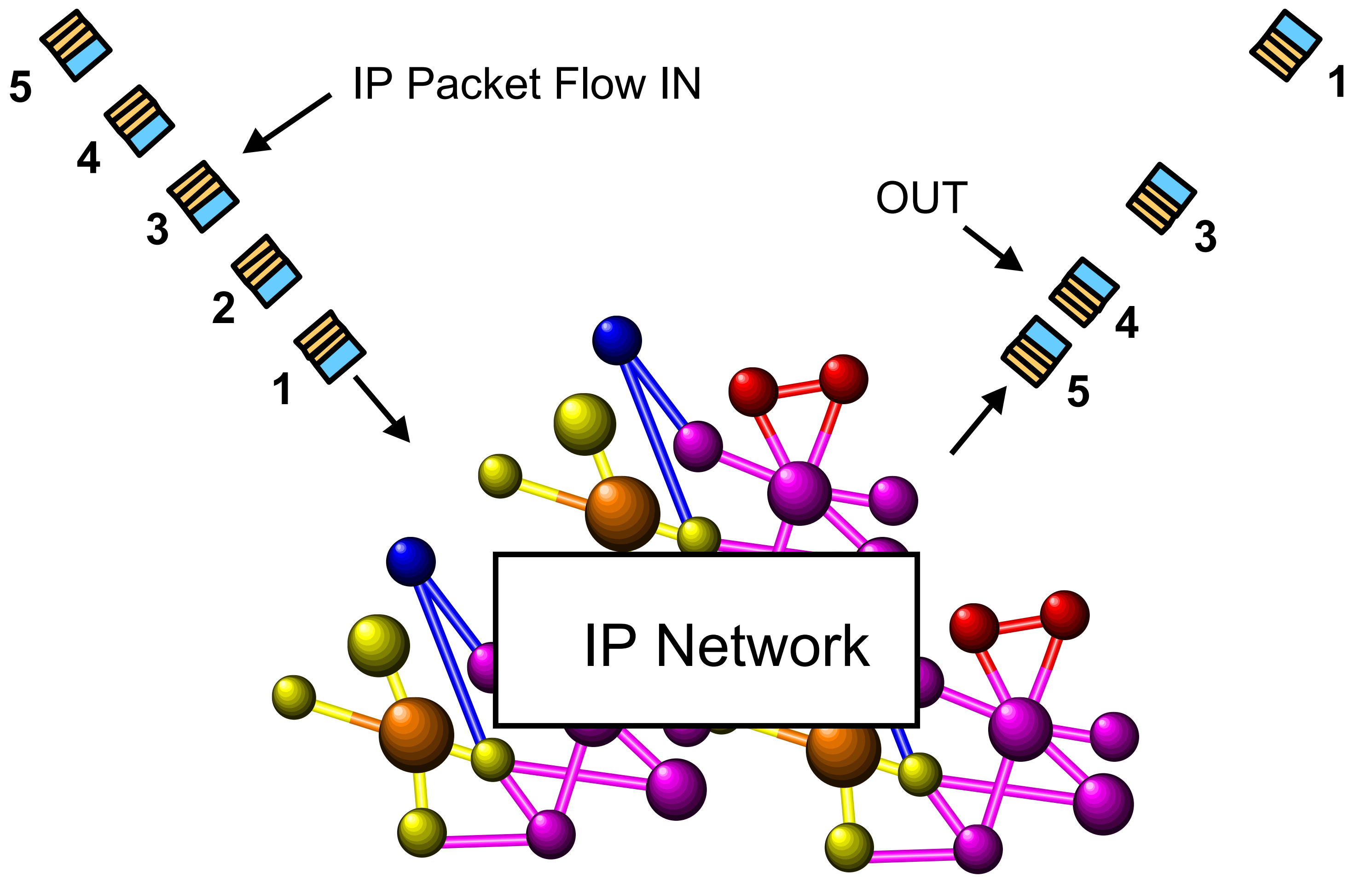
Telcordia Technologies
Performance from Experience

Datacom in an Deregulated World



Telcordia Technologies
Performance from Experience

Providing Service Quality in IP Networks



Providing Service Quality in IP Networks

How Do We Quantify It?

- Packet loss ratio
- End-to-end delay (average delay, delay jitter)
- Throughput and bandwidth measures
  - goodput (packets that are successfully delivered)
  - time-averaged offered load
  - burst tolerance
- Service reliability and availability
- Some applications place strict requirements on these parameters, particularly loss and delay

Telcordia Technologies
Performance from Experience

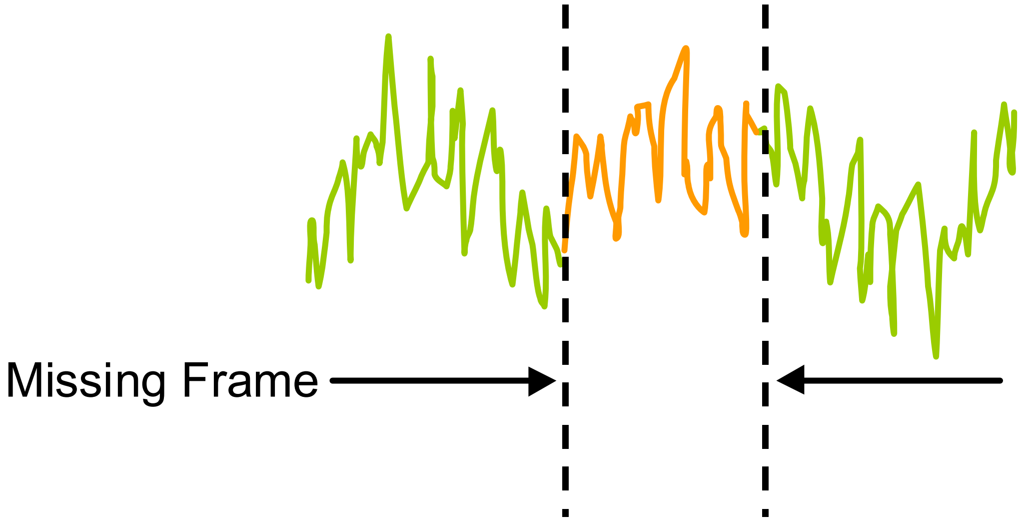
IP Service Example - Packet Audio

- Specific examples include IP telephony, Internet radio
- Uses UDP as the transport-layer (layer-4) protocol
  - no packet re-transmission; lost or mis-ordered packets are not recoverable
- Data streams have relatively low bandwidth (<10 kb/s average) but place tight constraints on performance
  - most codecs require packet loss <5%
  - packet delay (and delay jitter) are constrained as well

Internet Audio - Preserving Audio Quality

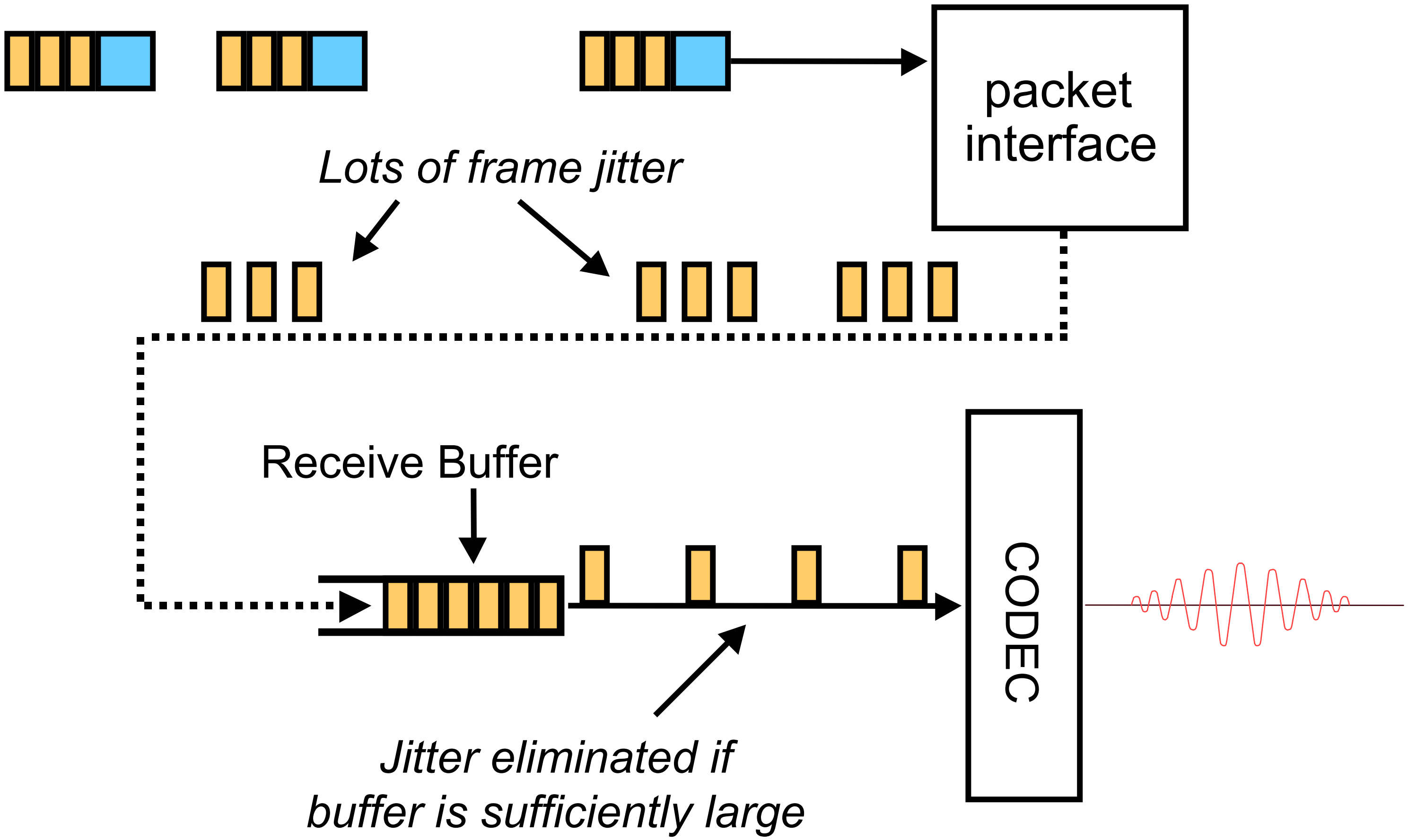
Recovering From Packet Loss

- CODEC Frame Loss Concealment Algorithms
  - Can attempt to conceal the lost frames of a lost packet
  - In essence, predicting and interpolating the missing sound in a "pleasing" way



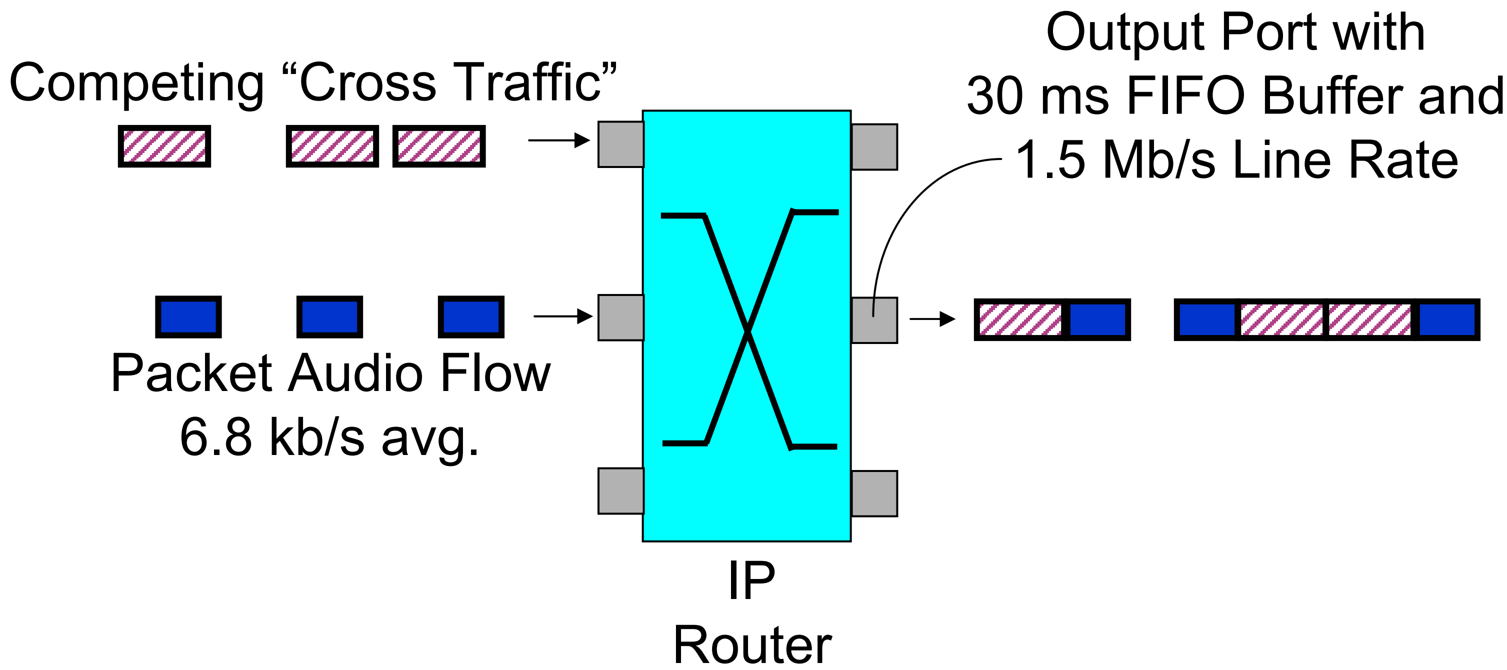
Internet Audio - Preserving Audio Quality

Overcoming Packet Delay Jitter



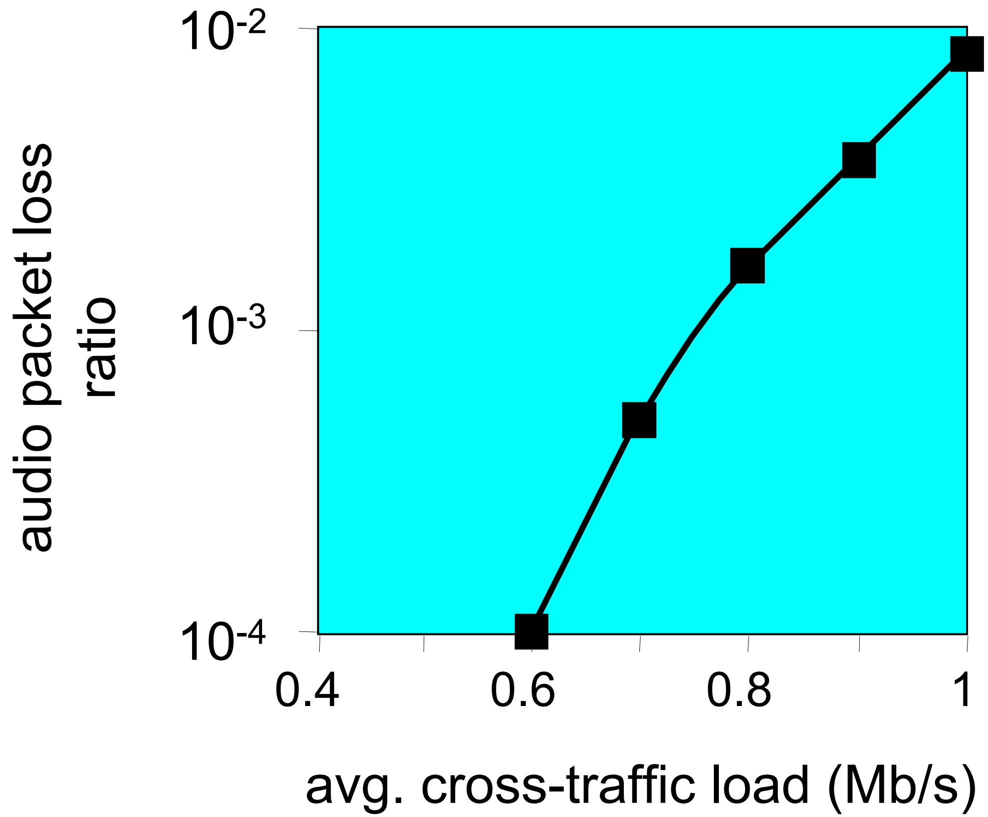
Telcordia Technologies
Performance from Experience

Packet Buffering and Congestion - Example



What happens to packet audio service quality as the volume of competing cross traffic increases?

Packet Loss Simulation Results



Is there a way to prevent the cross traffic from degrading the audio stream?

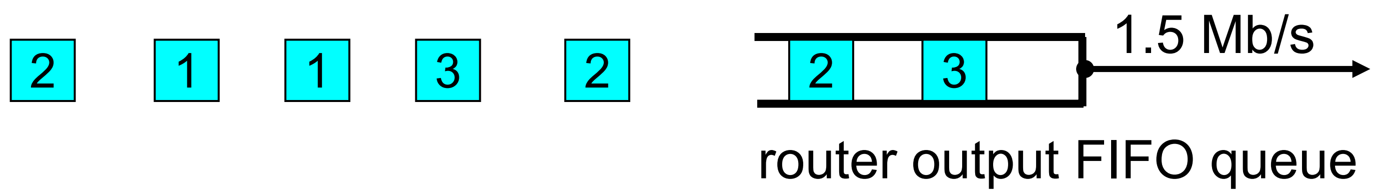
Telcordia Technologies

Performance from Experience

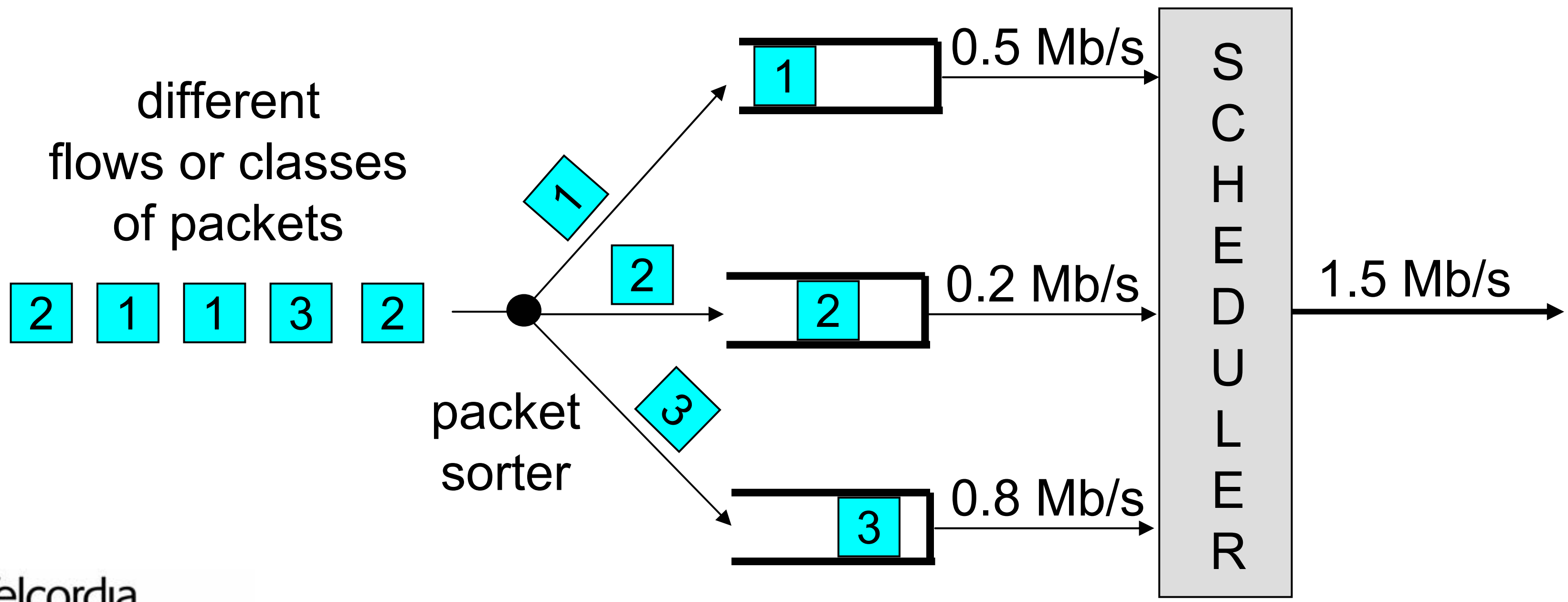
Courtesy Joel W. Gannett, Telcordia

Approaches to Improving Service Quality

Conventional FIFO Queueing



Weighted Fair Queueing (WFQ)



Telcordia Technologies
Performance from Experience

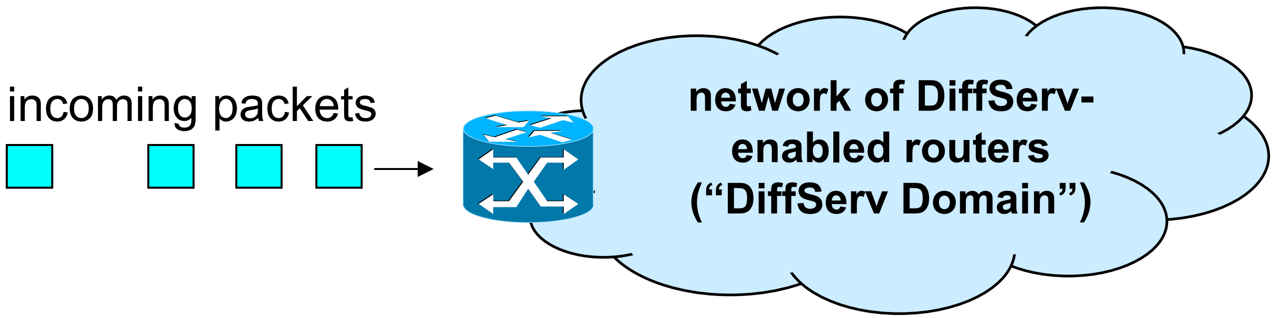
Approaches to Improving Service Quality

Bandwidth Partitioning and Differentiated Services

- As packets enter the network, classify them into a small number of service categories and mark them accordingly
- At each router interface, allocate bandwidth among the service categories using WFQ or similar techniques
- Bandwidth is allocated only to aggregations of flows; the network performs no per-flow processing
- This is the essence of the IETF's *Differentiated Services (DiffServ)* framework. DiffServ jargon:
  - A *"behavioral aggregate" (BA)* is a collection of flows that should receive the same service and that are marked in the same manner
  - A *"per-hop behavior" (PHB)* specifies the treatment that a BA should receive at a DiffServ router

Telcordia Technologies
Performance from Experience

DiffServ - Initial Packet Classification



- Edge router classifies each packet into a BA using
  - information in IP header (and/or higher-layer headers)
  - traffic metering information
  - other details specified by network operator
- The packets are marked with a *DiffServ Code Point (DSCP)* in the IP header, using the six most significant bits of the IPv4 "type of service" (TOS) octet
- The edge router may also perform traffic conditioning (e.g., selective dropping of packets) on incoming flows

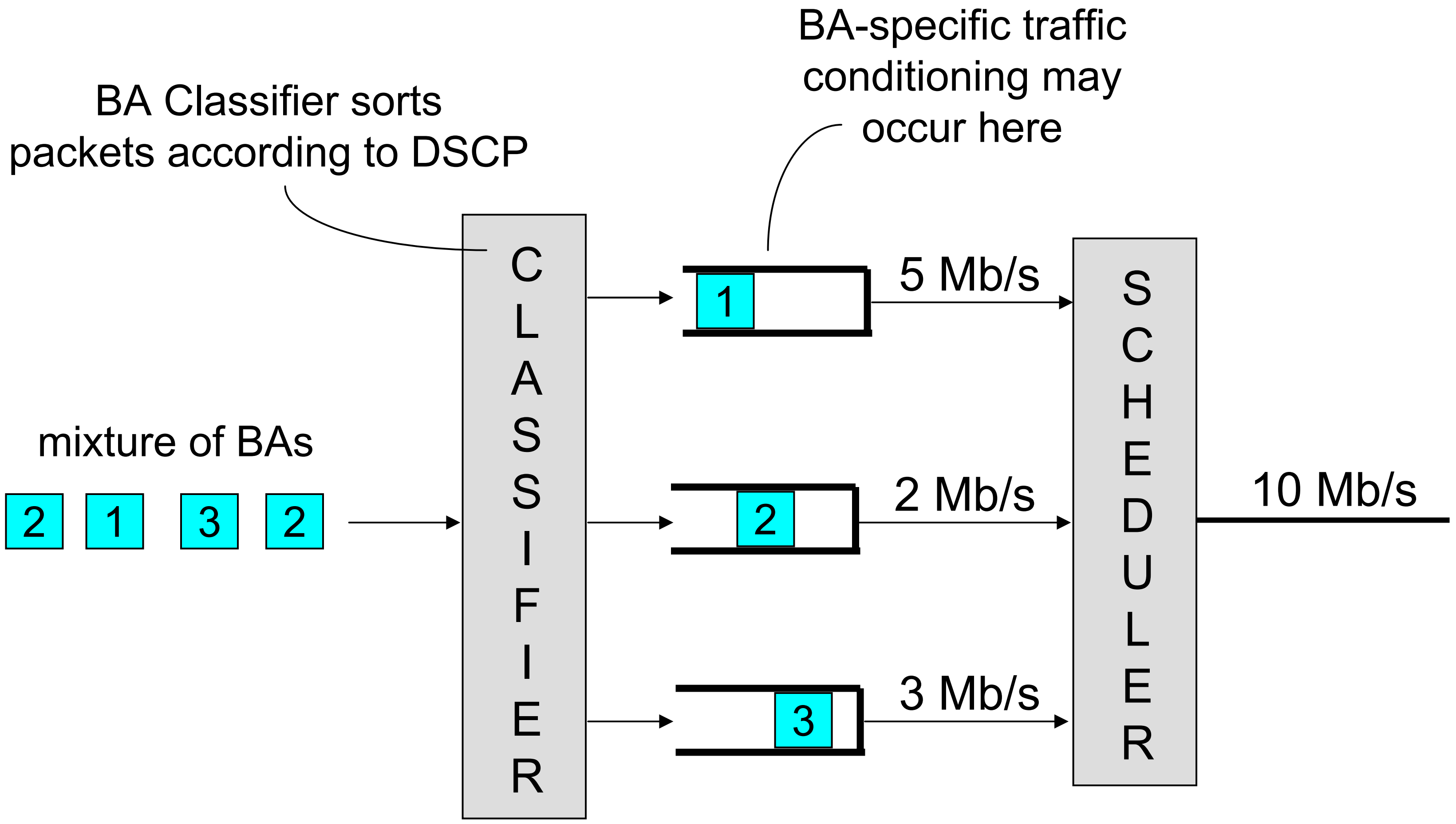
DiffServ - Per-Hop Behaviors (PHBs)

- PHBs can consist of bandwidth allocation and/or traffic conditioning actions at each DiffServ node, as dictated by the network operator
- Each BA is mapped to a PHB, which determines its treatment at each node
- PHBs typically utilize minimal processing in the interior of the network, to enhance scalability
- The IETF has defined certain PHBs, such as “Assured Forwarding” and “Expedited Forwarding”
- More information:

http://www.ietf.org/html.charters/diffserv-charter.html

Telcordia Technologies
Performance from Experience

DiffServ Implementation

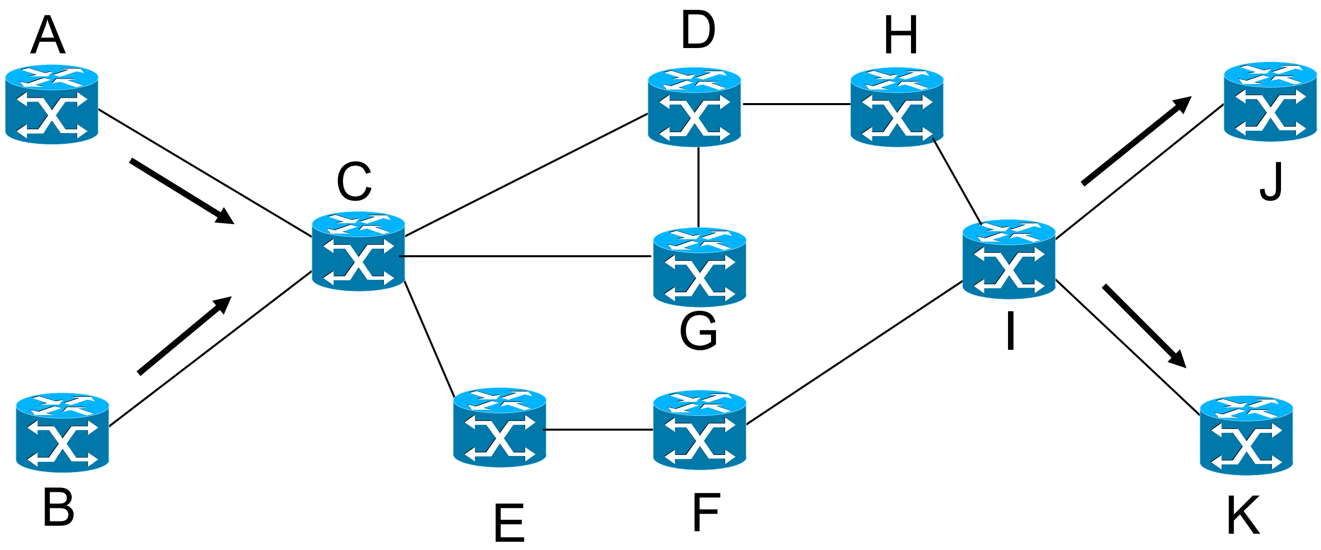


Does DiffServ Solve the IP QoS Problem?

- DiffServ divides resources among traffic types and helps to prevent BAs from affecting each others' service quality
- DiffServ is a useful building block but is not a complete solution for achieving adequate QoS, at least for some traffic types
- Significant problems remain:
  - We cannot be sure how traffic will be routed
  - If traffic in a particular BA exceeds its allocated bandwidth, that BA may suffer congestion and packet loss
  - Packets can get lost even before they reach the DiffServ domain

Telcordia Technologies
Performance from Experience

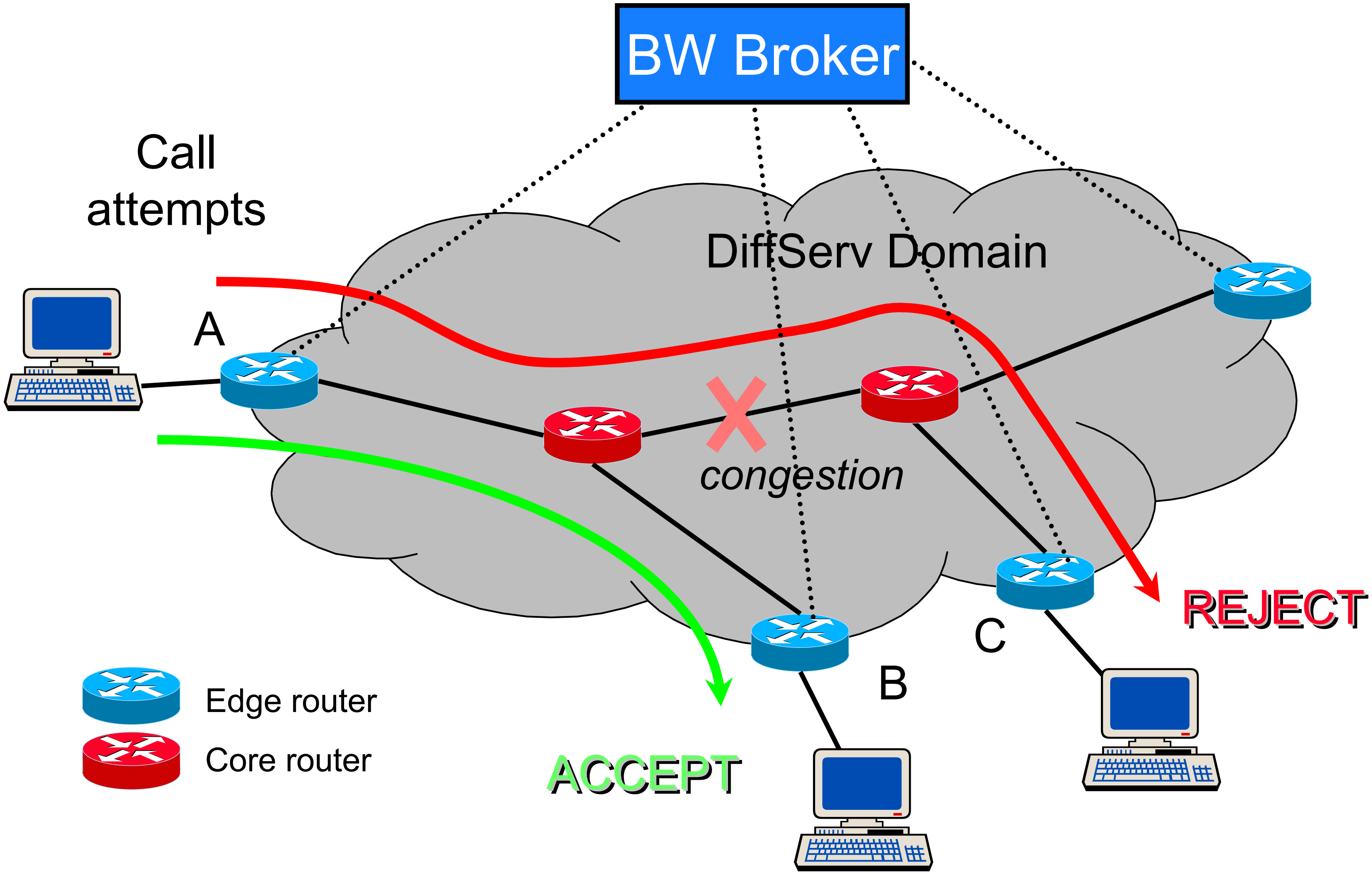
Routing and QoS in a Connectionless Network



- Of the possible paths from C to I, router C will identify one as the “shortest” and will use that path for all traffic from C to I
- Traffic from A and B will flow over the same path to I, congesting some links while leaving others under-utilized
- If the chosen path fails, the new path may be difficult to predict

Controlling Traffic and QoS Within a BA

Admission Control and "Bandwidth Brokers"



Courtesy R. R. Talpade, Telcordia

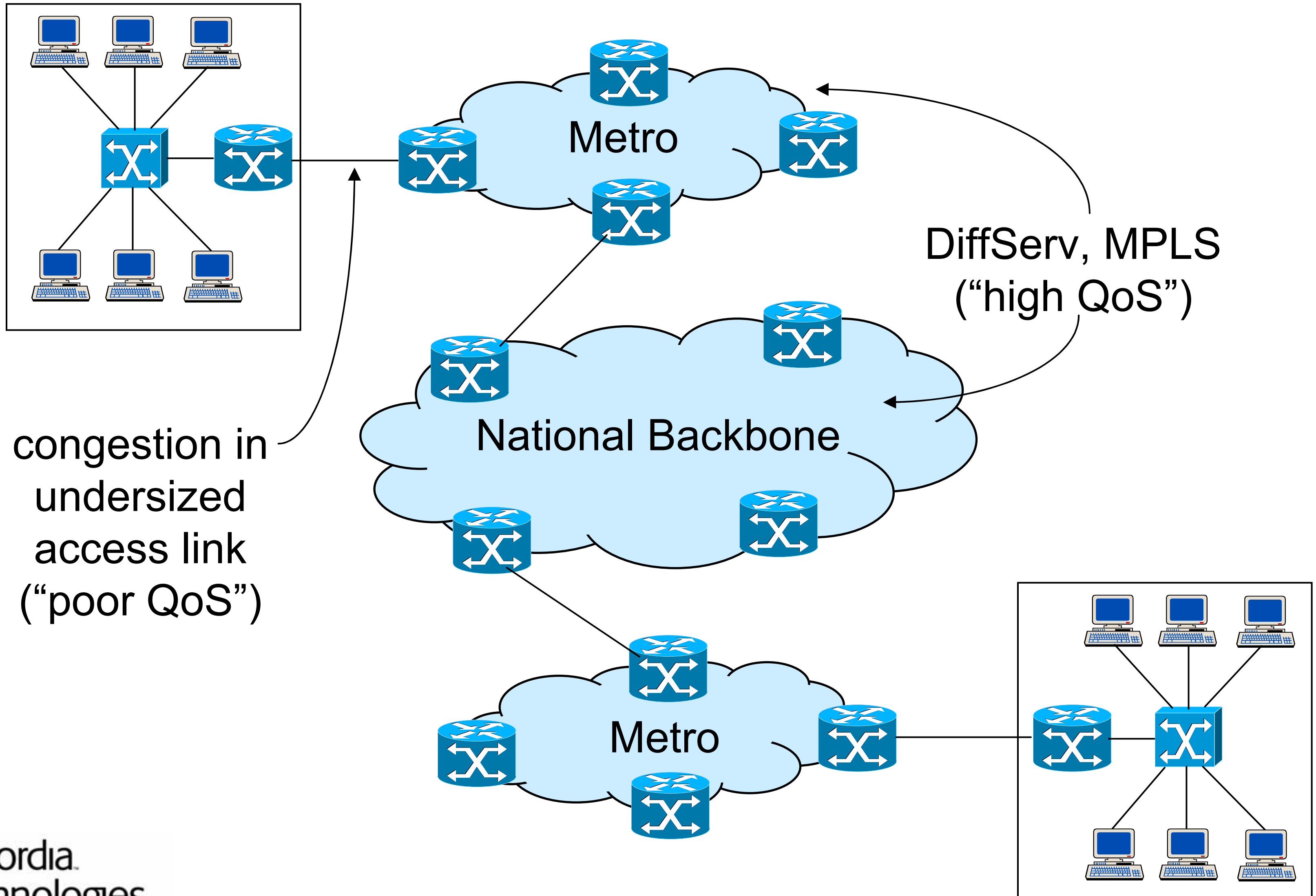
Telcordia Technologies
Performance from Experience

Bandwidth Broker

- It bases its admission decisions on
  - network and user policies (e.g., priorities, acceptable loading)
  - its knowledge of the state of the network (connectivity, current load)
- It has several ways to enforce admission decisions:
  - adjustment of traffic filters (classifiers) on edge routers
  - direct communication with hosts (if they are trusted)
  - communication with other management systems, such as voice-over-IP “softswitches,” to indirectly control traffic entry
- An active area of research
  - admission control algorithms for connectionless networks
  - admission control for multimedia, multiparty sessions
  - proactive or reactive network reconfiguration to overcome congestion

Tiered Structure of Data Transport

Where Do Packets Get Lost?



How Much Bandwidth Does Data Traffic Need?

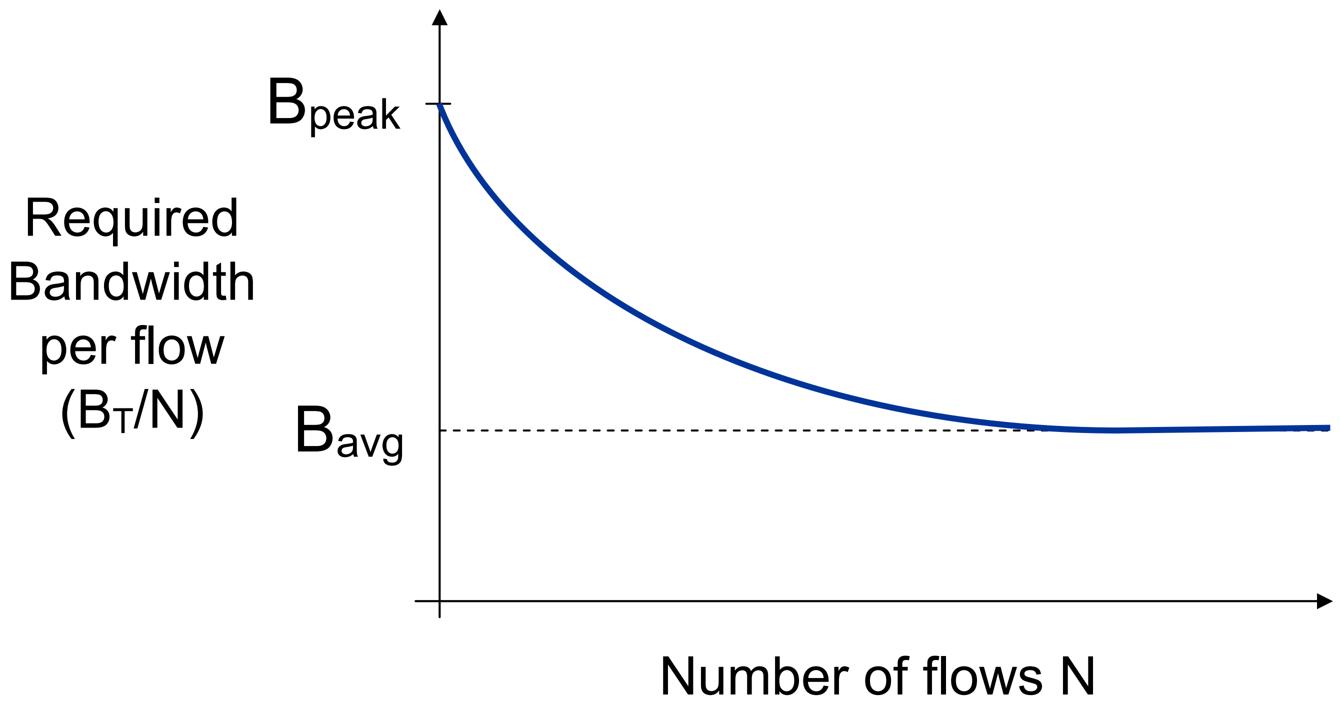
- A packet flow can be characterized by
  - peak bandwidth B_{peak}
  - average bandwidth B_{avg}
  - other temporal and statistical properties (duration, burstiness)
- A single isolated packet flow may require transmission bandwidth $B \sim B_{peak}$ for adequate QoS
- N multiplexed flows will require a total bandwidth B_T

$$NB_{peak} > B_T > NB_{avg}$$

- This is called *statistical multiplexing*, and relies on a “smoothing” of the traffic’s burstiness as N increases

Telcordia Technologies
Performance from Experience

Statistical Multiplexing Illustration



Engineering challenge: determine what B_T is required for a given traffic volume (i.e., how close is B_T to NB_{avg}?)

Bandwidth Estimation for Real Traffic

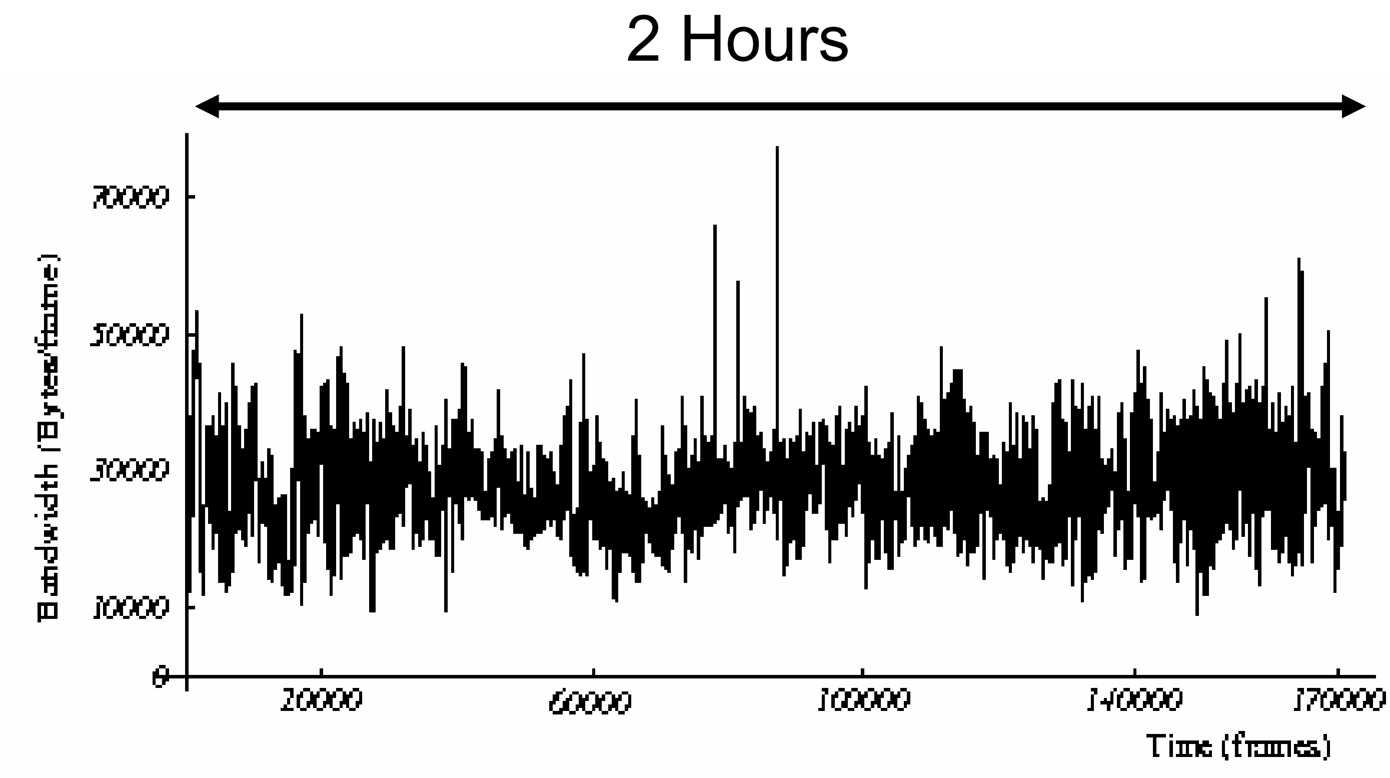
- "Classical" models of data traffic (e.g., Poisson) suggest that smoothing occurs very quickly
- These models are wrong for most types of data traffic
- Real traffic exhibits "self-similarity" and is much burstier
  - substantial, long-range correlations within the data streams
  - bursts lengths can vary by orders of magnitude
- Self-similar traffic smooths, but much more slowly than for conventional traffic models would suggest

Telcordia Technologies

Performance from Experience

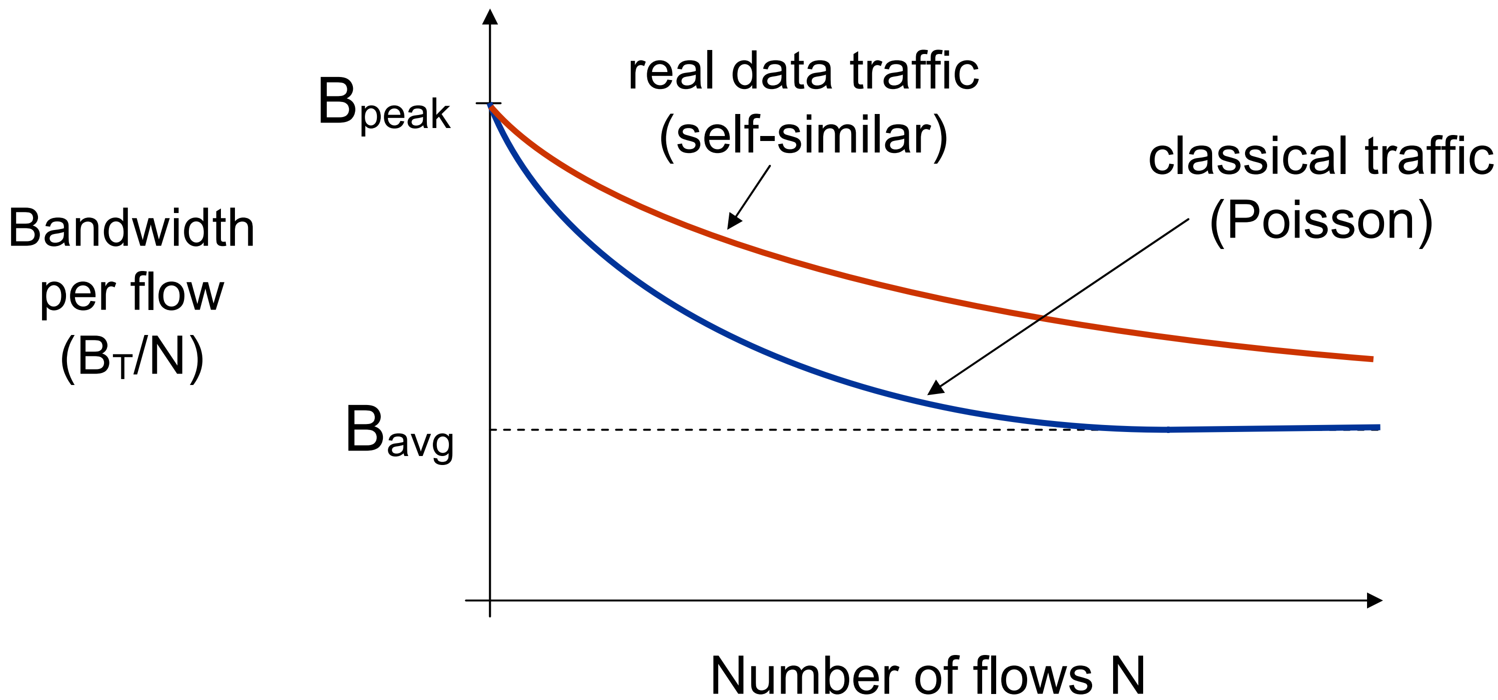
Traffic Trace Showing Self-Similarity

Variable-Rate Coded Video



courtesy Mark W. Garrett, Telcordia

Statistical Multiplexing Illustration



Real traffic required more bandwidth than conventional models would predict

Some Harsh Realities

- We rarely have good information about values for N, B_{avg} and other flow characteristics
- Data-network engineering is often based on close monitoring of aggregate traffic levels and heuristic rules about loading
  - "We try to keep our average loads at 50% during peak usage"
- Evidence of significant packet loss (e.g., from SLA monitoring tools) triggers installation of additional network bandwidth
- Luckily, some QoS-sensitive applications such as packet voice are not self-similar and have well-known statistical properties

Telcordia Technologies
Performance from Experience

Summary - Lessons Learned

- Data networks are heterogeneous
  - multiple layers and technologies
  - diverse mix of services and performance requirements
  - multiple administrative domains
- Providing service quality for data traffic remains challenging
  - connectionless nature of IP networks
  - traffic is highly bursty and difficult to characterize/predict
  - tools are available (e.g. DiffServ) but only for large traffic aggregates
- Newer technologies can help out, but introduce their own complexities
  - MPLS
  - dynamically configurable networks

Telcordia Technologies
Performance from Experience

References

- M. W. Garrett, W. Willinger, "Analysis, Modeling and Generation of Self-Similar VBR Video Traffic" ACM Comp Comm. Review, Vol 24, No 4, pp. 269-80, Oct 1994. (also Proc. ACM SigComm, London, September 1994.)
- W. E. Leland, M. S. Taqqu, W. Willinger and D. V. Wilson, “On the Self-Similar Nature of Ethernet Traffic", Proc. ACM SIGComm, San Francisco, Calif., pp. 183–193, Sept. 1993. (Extended Version: IEEE/ACM Trans. on Networking, Feb 1994.)
- http://www.ietf.org: DiffServ and related working groups in the Transport Area
- Kim, Mouchtaris, Samtani, Talpade, Wong, "A Bandwidth Broker Architecture for VoIP QoS," in Proceedings of SPIE's Intl Symp onConvergence of IT and Communications (ITCom), Colorado, Aug'01.
- http://www1.worldcom.com/global/about/network/ - information about the IP network of Worldcom (UUNet), a major ISP and data network operator