Lecture 27: Networks & Interconnect—Introduction, Implementation, Performance

Professor Randy H. Katz

Computer Science 252

Spring 1996

Review: Storage System Issues

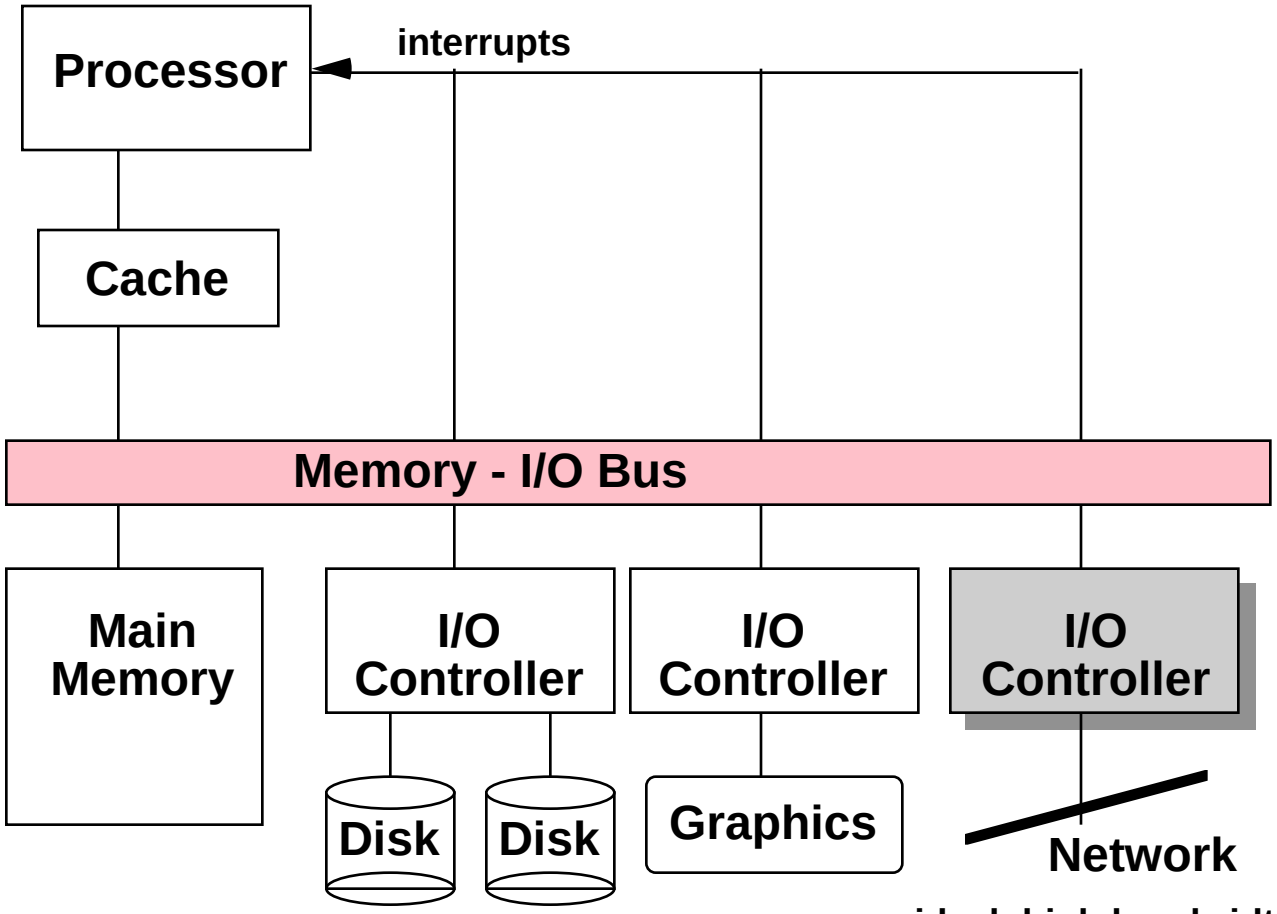
- Historical Context of Storage I/O
- Storage I/O Performance Measures
- Secondary and Tertiary Storage Devices
- A Little Queuing Theory
- Processor Interface Issues
- I/O & Memory Buses
- Show and Tell
- ABCs of UNIX File Systems
- I/O Benchmarks
- Comparing UNIX File System Performance
- RAID
- Tertiary Storage Possibilities

Review: RAID

RAID sales, "The Independent RAID Report" May/June 1994, p. 15 (dwilmot@crl.com, 510-938-7425)

- **1993: $3.4 billion on 214,667 arrays (≈ $15,000 / RAID)**
- **1996 forecast: $11 billion**
- **1997 forecast: $13 billion on 837,155 units**
  - **Source: DISK/TREND, 5/94 (415-961-6209)**

I/O to External Devices and Other Computers

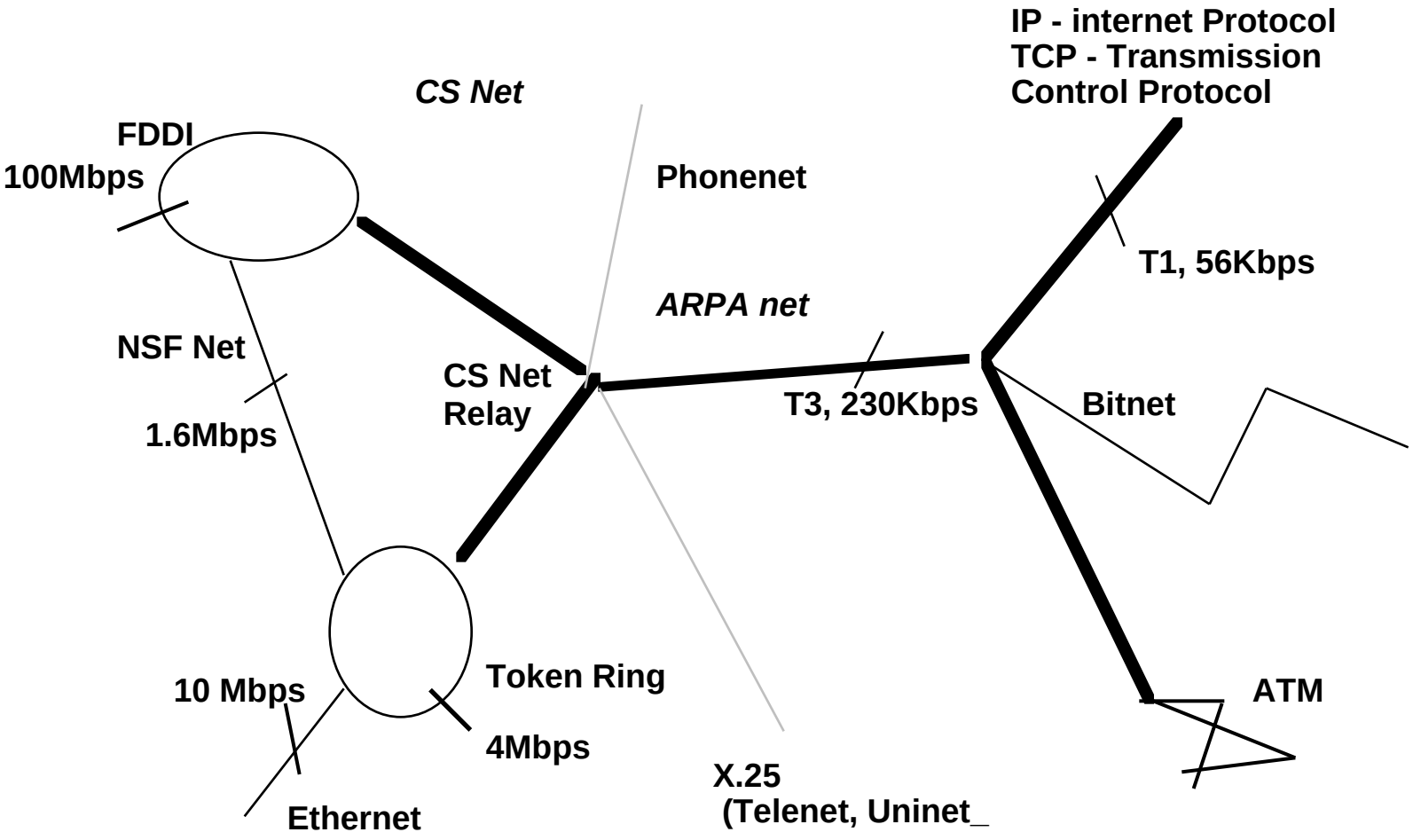


ideal: high bandwidth, low latency

Networks

- **Goal: Communication between computers**
- **Eventual Goal: treat collection of computers as if one big computer, distributed resource sharing**
- **Theme: Different computers must agree on many things**
  - **Overriding importance of standards and protocols**
- **Warning: Terminology-rich environment**

Current Major Networks

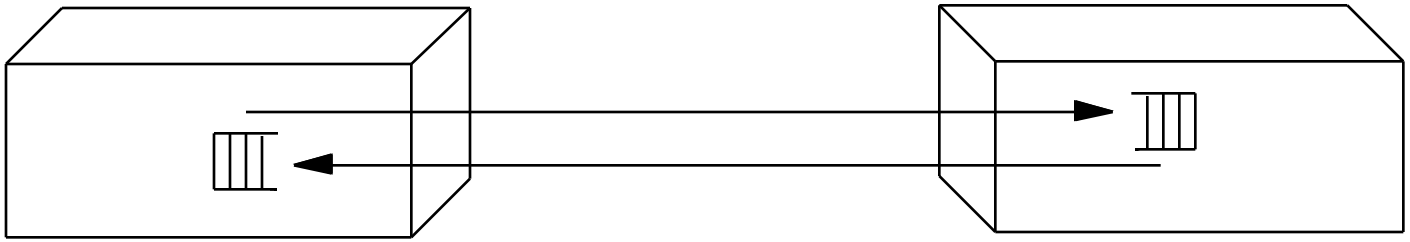


Networks

- **Facets people talk a lot about:**
  - **direct (point-to-point) vs. indirect (multi-hop)**
  - **topology (e.g., bus, ring, DAG)**
  - **routing algorithms**
  - **switching (aka multiplexing)**
  - **wiring (e.g., choice of media, copper, coax, fiber)**
- **What really matters:**
  - **latency**
  - **bandwidth**
  - **cost**
  - **reliability**

ABCs of Networks

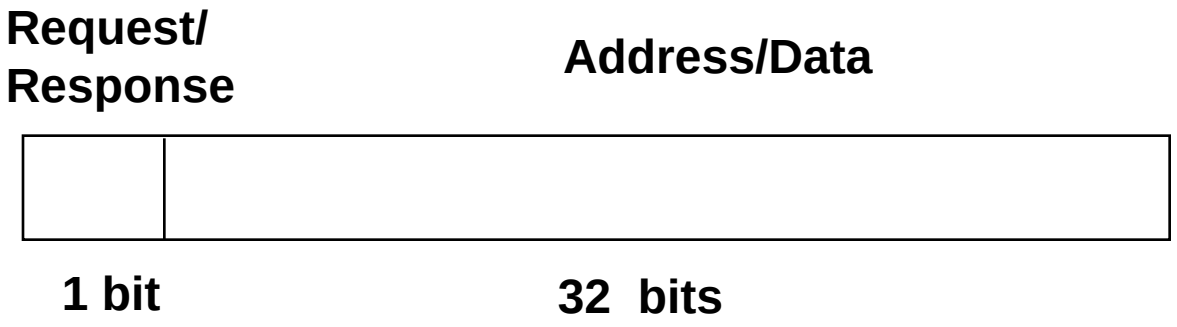
- **Starting Point: Send bits between 2 computers**



- **Queue on each end**
- **Can send both ways (“Full Duplex”)**
- **Rules for communication? “protocol”**
  - **Inside a computer:**
    - » **Loads/Stores: Request (Address) & Response (Data)**
    - » **Need Request & Response signaling**
  - **Name for standard group of bits sent: Packet**

A Simple Example

- **What is the format of packet?**
  - **Fixed? Number bytes?**



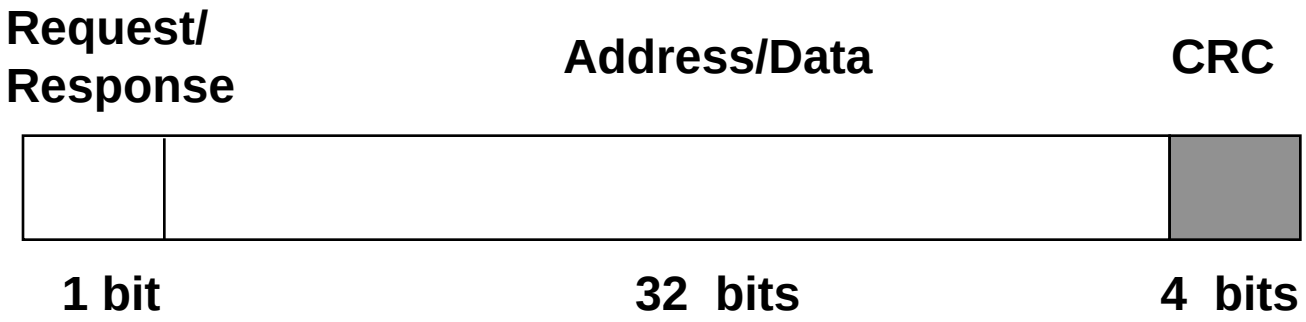
0: Please send data from Address
1: Packet contains data corresponding to request

Questions About Simple Example

- **What if more than 2 computers want to communicate?**
  - **Need computer address field (destination) in packet**
- **What if packet is garbled in transit?**
  - **Add error detection field in packet (e.g., CRC)**
- **What if packet is lost?**
  - **More elaborate protocols to detect loss (e.g., NAK, ARQ, time outs)**
- **What if multiple processes/machine?**
  - **Queue per process**
- **Questions such as these lead to more complex protocols and packet formats**

A Simple Example Revisted

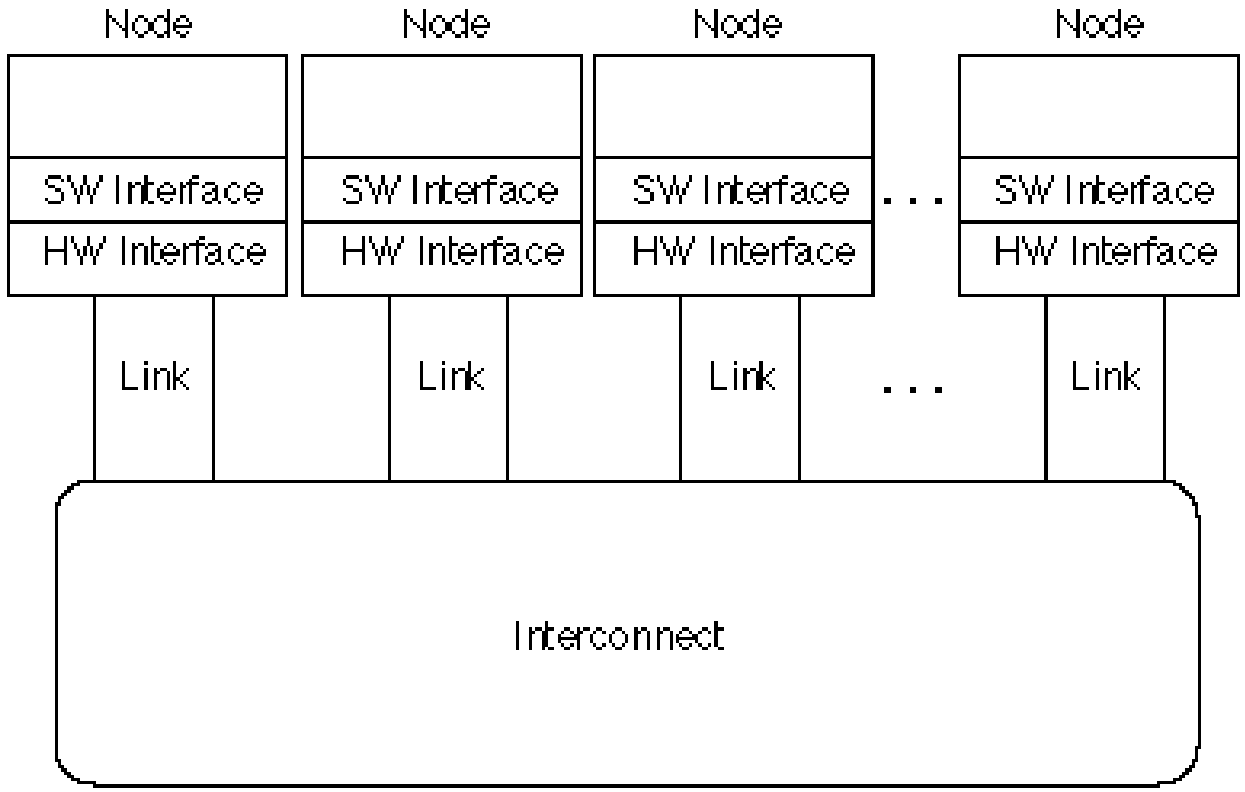
- **What is the format of packet?**
  - **Fixed? Number bytes?**



00: Request—Please send data from Address
01: Reply—Packet contains data corresponding to request
10: Acknowledge request
11: Acknowledge reply

Interconnections (Networks)

- **Examples:**
  - **MPP networks (CM-5): 1000s nodes; ≤ 25 meters per link**
  - **Local Area Networks (Ethernet): 100s nodes; ≤ 1000 meters**
  - **Wide Area Network (ATM): 1000s nodes; ≤ 5,000,000 meters**



Interconnect Issues

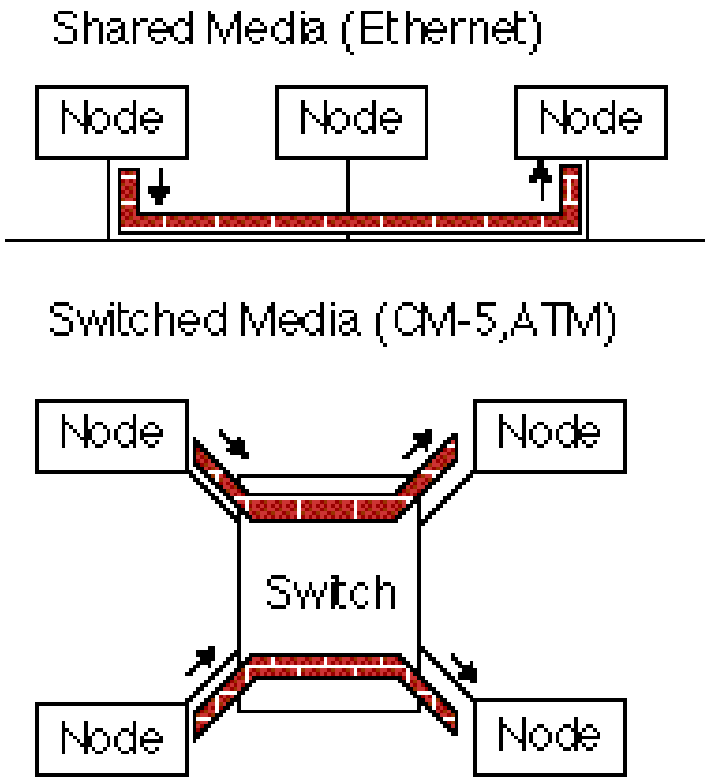
- Implementation Issues
- Performance Measures
- Architectual Issues
- Practical Issues

Implementation Issues

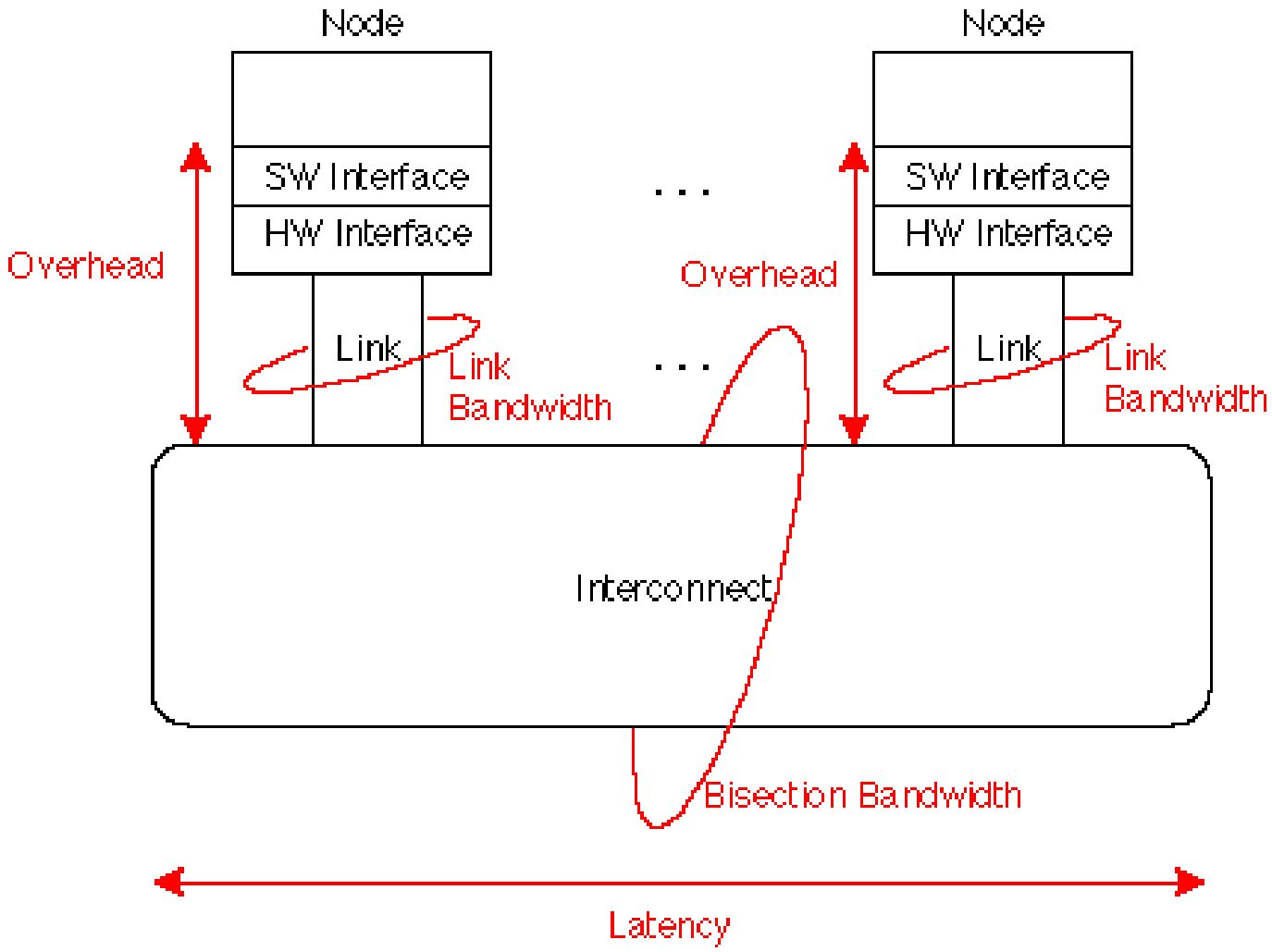
Interconnect	*MPP*	*LAN*	*WAN*
Example	CM-5	Ethernet	ATM
Maximum length between nodes	25 m	500 m; ≤5 repeaters	copper: 100 m optical: 2 km—25 km
Number data lines	4	1	1
Clock Rate	40 MHz	10 MHz	≥ 155.5 MHz
Shared vs. Switch	Switch	Shared	Switch
Maximum number of nodes	2048	254	> 10,000
Media Material	Copper	Twisted pair copper wire or Coaxial cable	Twisted pair copper wire or optical fiber

Implementation Issues

- **Advantages of Serial vs. Parallel lines:**
  - **No synchronizing signals**
  - **Higher clock rate and longer distance than parallel lines (e.g., 60 MHz x 256 bits x 0.5 m vs. 155 MHz x 1 bit x 100 m)**
    - » **Imperfections in the copper wires or integrated circuit pad drivers can cause skew in the arrival of signals, limiting the clock rate, and the length and number of the parallel lines.**
- **Switched vs. Shared Media: pairs communicate at same time: "point-to-point" connections**



Network Performance Measures

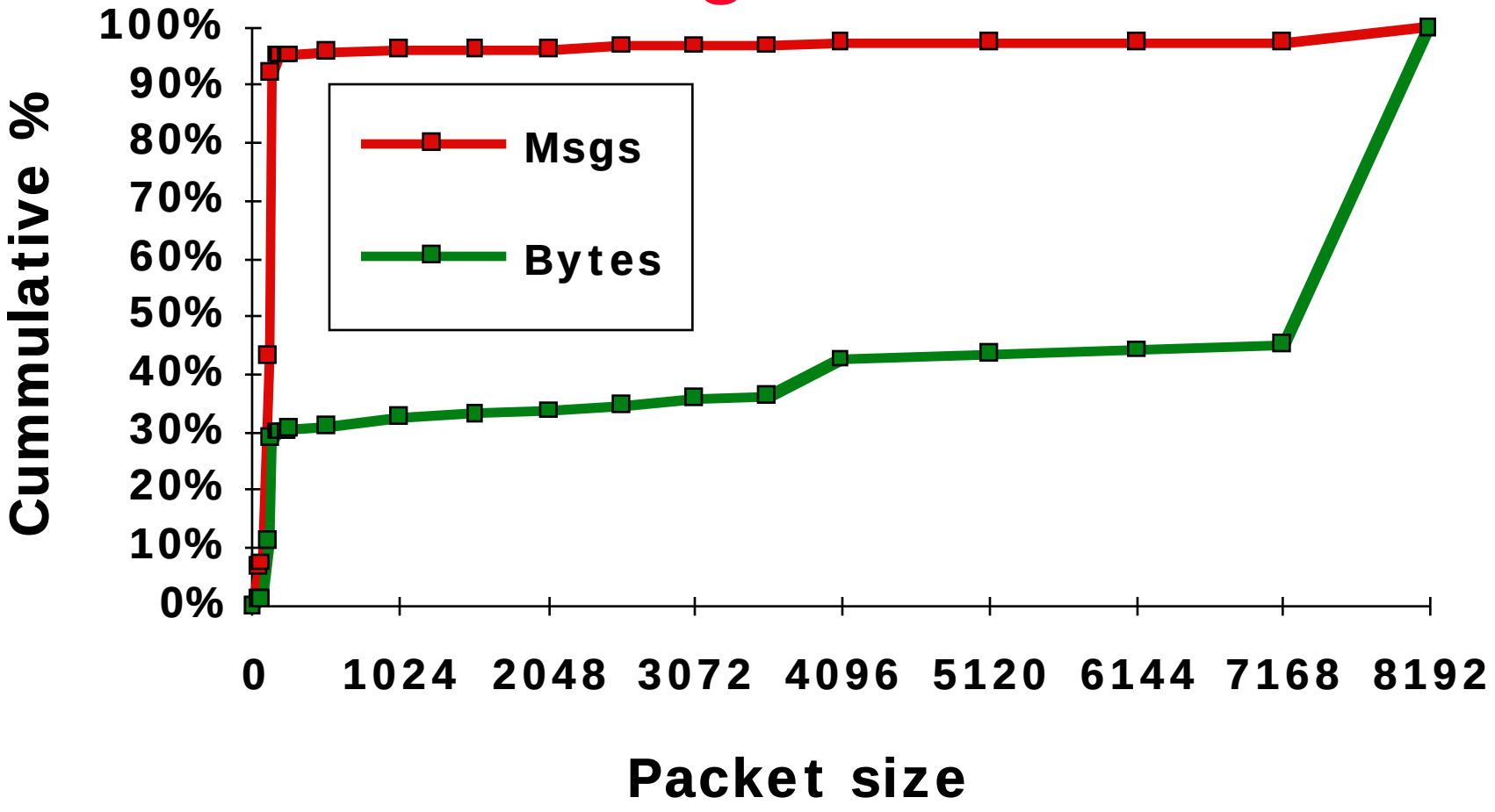


- **Overhead: latency of interface vs. Latency: network**

Example Performance Measures

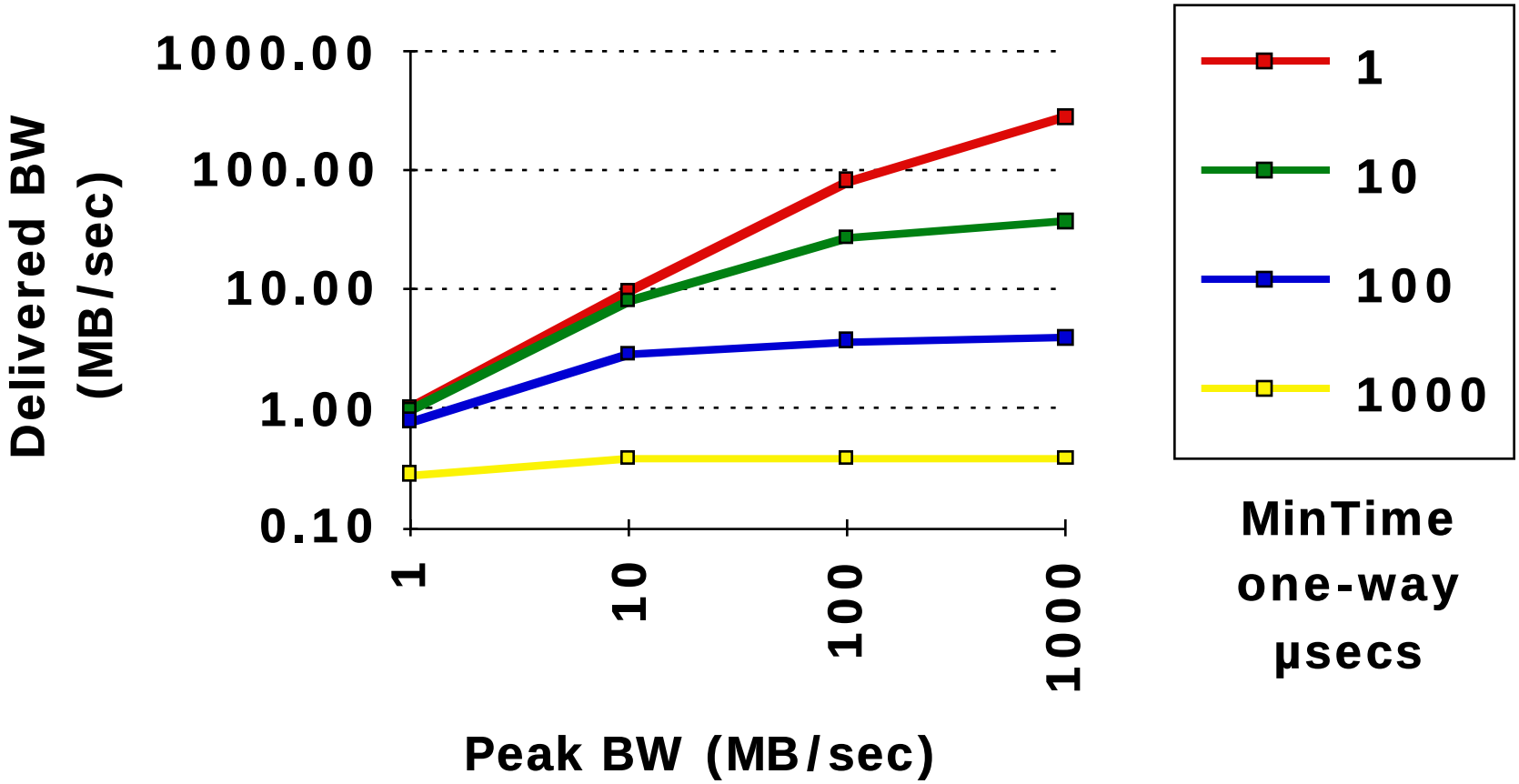
Interconnect	*MPP*	*LAN*	*WAN*
Example	CM-5	Ethernet	ATM
Bisection BW	N x 5 MB/s	1.125 MB/s	N x 10 MB/s
Int./Link BW	20 MB/s	1.125 MB/s	10 MB/s
Latency	5 µsec	15 µsec	50 to 10,000 µs
HW Overhead to/from	0.5/0.5 µs	6/6 µs	6/6 µs
SW Overhead to/from	1.6/12.4 µs	200/241 µs	207/360 µs
		(TCP/IP on LAN/WAN)	

Measurement: Sizes of Message for NFS



- 95% Msgs, 30% bytes for packets ≤ 200 bytes
- > 50% data transfered in packets = 8KB

Impact of Overhead on Delivered BW



- BW model: Time = overhead + msg size/peak BW
- > 50% data transfered in packets = 8KB

Summary: Interconnections

- **Communication between computers**
- **Packets for standards, protocols to cover normal and abnormal events**
- **Implementation issues: length, width, media**
- **Performance issues: overhead, latency, bisection BW**
- **Topologies: many to chose from, but (SW) overheads make them look the alike; cost issues in topologies**