Chapter 8: Interconnect Technology

Interconnected computers are probably the true "computer revolution"

Subject of course sequence: CS 640 & 740

Performance & Network Types

Connections & Packets

Flow/Congestion Control & Practical Issues

(Examples)

Performance

Bandwidth

- Link vs. (minimum) bisection
- Peak vs. sustained
- MB/s vs. Mb/s

Latency

- SW overhead for send
  HW overhead for send
- Interconnect latency
- HW overhead for receive
  SW overhead for receive

Network Types

Massively Parallel Processor (MPP)

- Proprietary with little fault tolerance
- E.g., CM-5

Local Area Network (LAN)

- Standard with medium fault tolerance
- E.g., Ethernet

Wide Area Network (WAN)

- Standard with good fault tolerance
- Requires "internetworking"
- E.g., Asynchronous Transfer Mode (ATM)

Network Types, cont.

Network	**CM-5**	**Ethernet**	**ATM**
Length (m)	25	500 (w/o repeaters)	100 (copper) 1000 (fiber)
Data lines	4	1	1
Clock (MHz)	40	10	155.5+
Media	switched	shared	switched
Link BW (MB/s)	20	1.125	10
Bisection BW (MB/s for n nodes)	5n	1.125	10n
Interconnect Latency (μsec)	5	15	50-10,000 (distance)
HW Overhead	0.5	6	6
SW Overhead (send) (send+receive)	1.6+12.4	200+241 (TCP/IP)	207+360 (TCP/IP)

Topology

Bus

E.g., Ethernet

\+ simple, passive

\- bisection = link BW

Ring E.g., FDDI (token ring)

\- not much better than bus

Fully-connected

E.g., small, proprietary connections

\- too expensive: approx $n^2/2$ connections

Topology, cont.

Use interdirection: intermediate node

Crossbar: almost fully-connected (draw picture)

E.g., inside of a switch
\+ fast, only output conflicts
\- expensive

k-D Grid/Torus (torus has wrap-around)

E.g., Intel Paragon (k=2), Cray T3D (k=3)
\+ fixed number of ports per switch
\- latency is order $n^{1/k}$

Hypercube: connect each node to $\log_2 n$ others

E.g., Intel iPSC/2
\+ latency is order $\log_2 n$
\- variable number of ports per switch

Topology, cont.

t-ary Omega Network

E.g., BBN Butterfly
\+ latency is $\log_t n$
\- must layout

t-ary Fat tree: each child has multiple parents

E.g., CM-5
\+ latency is $\log_t n$ with short cuts
\- must layout

Summary

More interesting in theory than practice

Must layout into chips, boards, backplanes

SW & HW overheads can dominate

Fixed bisection BW requires that links grow faster than n

Connections

Connection-Oriented: establish connection & pass data on it

E.g., telephone, ATM

Better for long messages

Phones use time division multiplexing (synchronous transfer mode)

Connectionless: route each datum separately

E.g., postal mail, IP datagrams

Each "packet" must contain destination address

Packets

Show packet format picture

- CM-5
- Ethernet
- ATM (cell)

Flow & Congestion Control

Flow: from source to destination

- Need back pressure (negative feedback)
- E.g., acks w/ N-packet window

Congestion: from anywhere to switch

- Harder than flow control
- Limit sum of all flows through switch
- Discard packets
- Choke packets for back pressure

Practical Issues

LANs and WANs must be

- Standardized
- Fault tolerant
- Hot insert

Key problem with today's machines

- Poor HW interfaces
  (not rewarded by SPECmarks)
- Too high SW overhead
  (must call OS in common case)

Practical Issues

Must confront protocol complexity with layers

- Illusion of peer-to-peer communication
- But all but lowest layer goes down
- E.g., communication between generals
- C.f., OSI 7-layer model

WANs must support internetworking

- Use IP packets (datagrams)
- Transmission protocol on top to provide end-to-end guarantees (TCP most common)
- Application protocols layered above
  (e.g., ftp, rlogin, telnet)