

Packet Switch Architecture

The Hardware Architect's perspective on
High-Speed Networking Problems

Manolis Katevenis

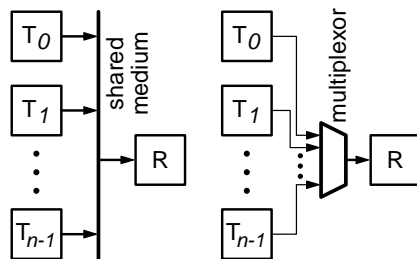
University of Crete & FORTH

(1984-now: course includes 20+ years of research...)

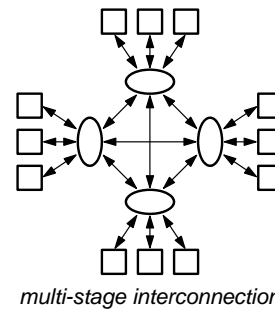
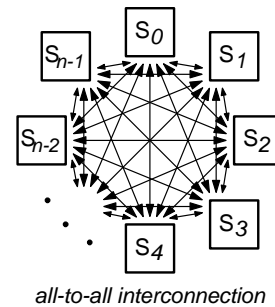
CS-534, Copyright Univ. of Crete

1

Communication Networks



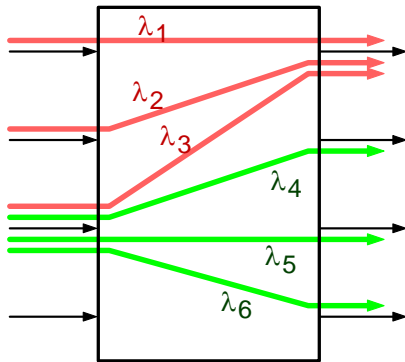
- Multi-party communication under resource constraints
- Receiver capacity is far below the aggregate capacity (rate) of all transmitters (and symmetrically for each transmitter relative to all rcvrs)



CS-534, Copyright Univ. of Crete

2

Interdependent Constraints



- $\lambda_1 + \lambda_2 + \lambda_3 \leq 100\%$
(output contention)
33%+33%+33% (fairness) ?
- $\lambda_3 + \lambda_4 + \lambda_5 + \lambda_6 \leq 100\%$
(input rate limitation)
25%+25%+25%+25% ?
- $\lambda_3 = 25\% \Rightarrow \lambda_3 + \lambda_4 = 37.5\% + 37.5\%$ (max-min fairness) ?
- or 50+50+0, 0+33+33+33 %
(maximum utilization) ?

CS-534, Copyright Univ. of Crete

3

Distributed Control Problem

- Who determines the solution? When? Where? How?
- Geographically distributed (traditional networking), or...
(long time scales may allow software solution)
- Microelectronic chips in a box or a room
(short time scales demand hardware speed)
- Inputs do not know of each other's intentions when they start transmitting
- Distance, speed, complexity preclude centralized solution

CS-534, Copyright Univ. of Crete

4

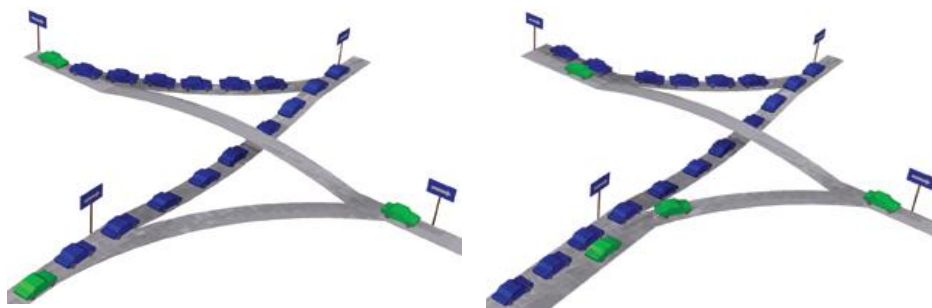
Reactions to Output Contention

- Short-Term (within a round-trip time)
 - buffer conflicting packets (in the network), or...
 - drop conflicting packets (and retransmit?)
 - Long-Term
 - flow control (congestion management) (after the fact)
 - admission control (beforehand)
- ⇒ High-speed Memories + Distributed Control

CS-534, Copyright Univ. of Crete

5

Head-of-Line Blocking

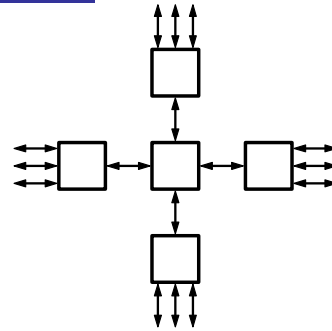
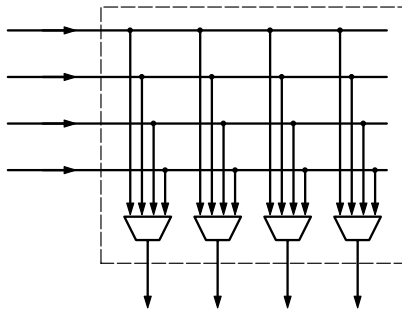


⇒ Multiple Queues within High-Speed Memories

CS-534, Copyright Univ. of Crete

6

Interconnection Fabrics



- Single-stage all-to-all (crossbar) performs well but its N^2 cost is too high for large N
- Multi-stage (hierarchical) fabrics trade cost for internal blocking (“all lines are busy –please try later”) (locality of traffic?)

CS-534, Copyright Univ. of Crete

7

Technology Outlook: Commodity Switches

- Ubiquitous, Switch-based Interconnection Networks
 - buses have inherent performance limitations
 - switch-based interconnects proliferate from WAN to LAN, then to SAN (storage-area or system-area), then to processor-memory-I/O interconnects and to Networks-on-Chip (NoC)
 - New Market: Next Generation IT Infrastructure
 - chip multiprocessors (CMP)
 - cluster/blade-based systems and servers
- ⇒ Commodity Switches: Mass Market, sharp price drop
- fabrics of inexpensive, mass-produced switches will replace the current very expensive, custom-made telco switches/routers (analogous to workstation clusters replacing supercomputers)
 - what should be their (“RISC-style”) architecture???

CS-534, Copyright Univ. of Crete

8