

Packet Switch Architectures

*The following are (sometimes modified and rearranged slides) from an ACM Sigcomm 99 Tutorial by **Nick McKeown and Balaji Prabhakar**, Stanford University*

Slides used with permission from authors.
© 1999-2000. All rights reserved by authors.

Outline

- **Introduction:**
What is a Packet Switch?
- **Packet Lookup and Classification:**
Where does a packet go next?
- **Switching Fabrics:**
How does the packet get there?

Introduction

What is a Packet Switch?

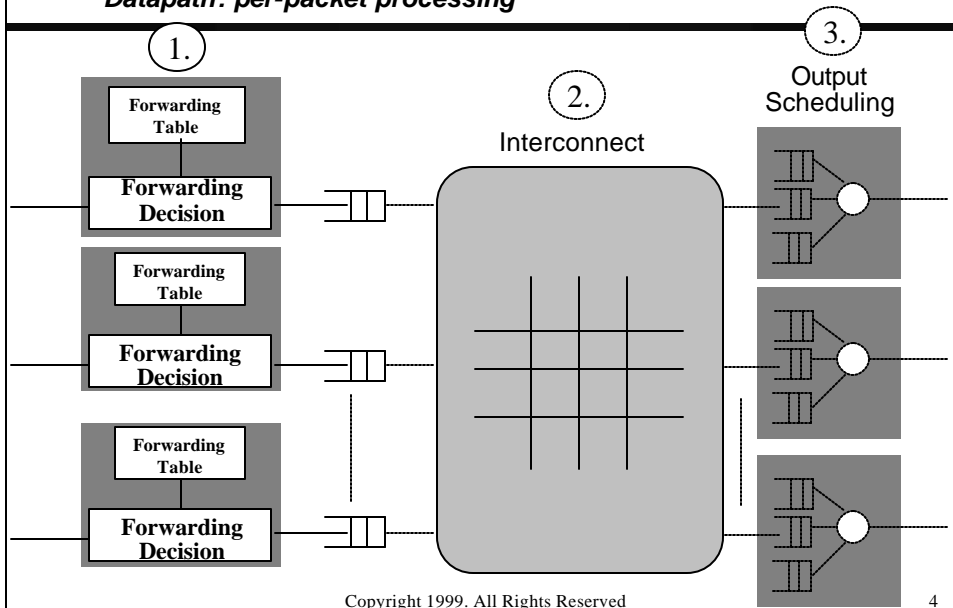
- Basic Architectural Components
- Some Example Packet Switches

Copyright 1999. All Rights Reserved

3

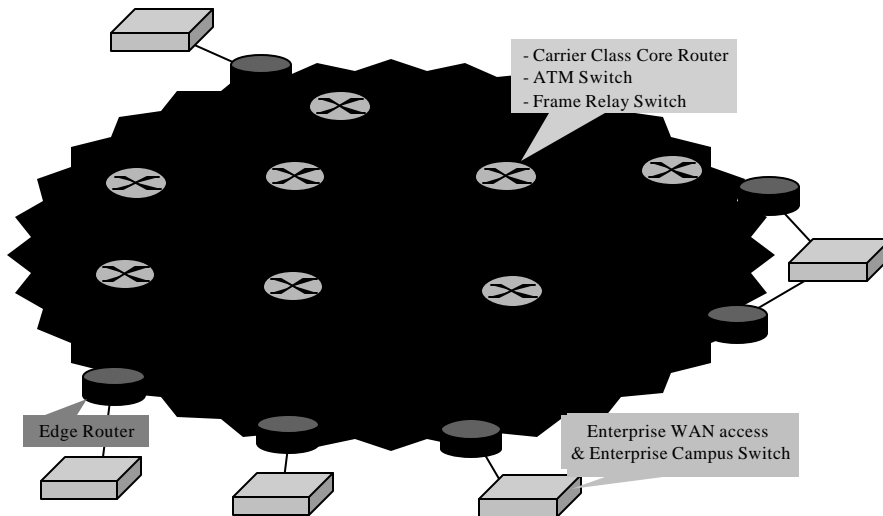
Basic Architectural Components

Datapath: per-packet processing



4

Where high performance packet switches are used



Copyright 1999. All Rights Reserved

5

Some Example Packet Switches

- Packet switches exist for different networking technologies
 - Internet: IP protocol suite
 - Ethernet: Ethernet switches
 - ATM (Asynchronous Transfer Mode): ATM switch
 - MPLS (Multiprotocol label switching): MPLS switch
- There are many similarities in the architecture of the switches

Copyright 1999. All Rights Reserved

6

Packet Lookup

Where does a packet go next?

- ATM and MPLS switches
 - Direct Lookup
- Bridges and Ethernet switches
 - Associative Lookup
 - Hashing
- IP Routers
 - Patricia trees/tries

Copyright 1999. All Rights Reserved

7

Lookup in an ATM Switch

- Lookup cell VCI/VPI in VC table.
- Replace old VCI/VPI with new.
- Forward cell to outgoing interface.
- Transmit cell onto link.

Copyright 1999. All Rights Reserved

8

Lookup in an Ethernet Switch

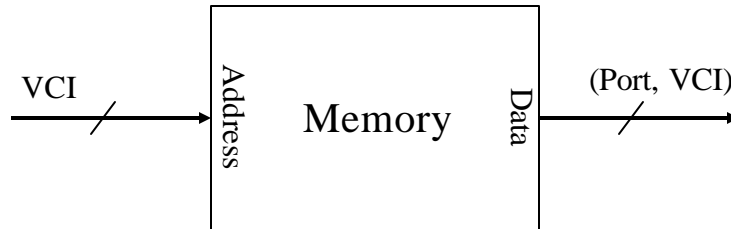
- Lookup frame DA in forwarding table.
 - If known, forward to correct port.
 - If unknown, broadcast to all ports.
- Learn SA of incoming frame.
- Forward frame to outgoing interface.
- Transmit frame onto link.

Lookup in an IP Router

- Lookup packet DA in forwarding table.
 - If known, forward to correct port.
 - If unknown, drop packet.
- Decrement TTL, update header Cksum.
- Forward packet to outgoing interface.
- Transmit packet onto link.

ATM and MPLS Switches

Direct Lookup

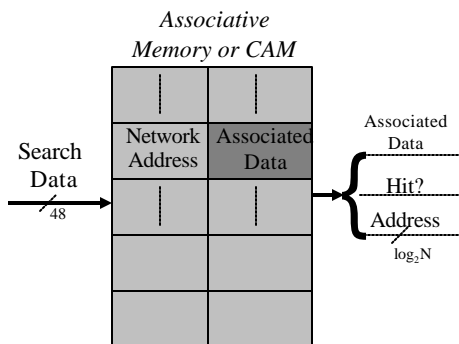


Copyright 1999. All Rights Reserved

11

Bridges and Ethernet Switches

Associative Lookups



Advantages:

- Simple

Disadvantages

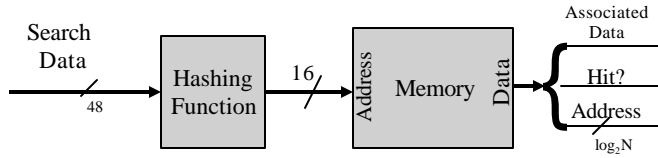
- Slow
- High Power
- Small
- Expensive

Copyright 1999. All Rights Reserved

12

Bridges and Ethernet Switches

Hashing

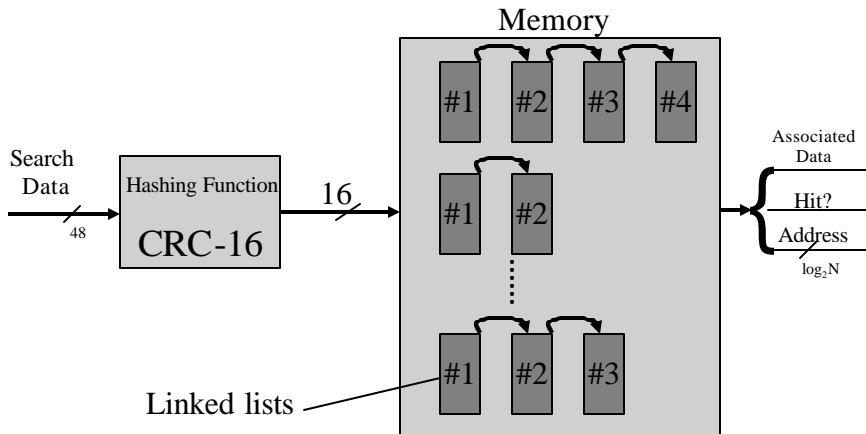


Copyright 1999. All Rights Reserved

13

Lookups Using Hashing

An example

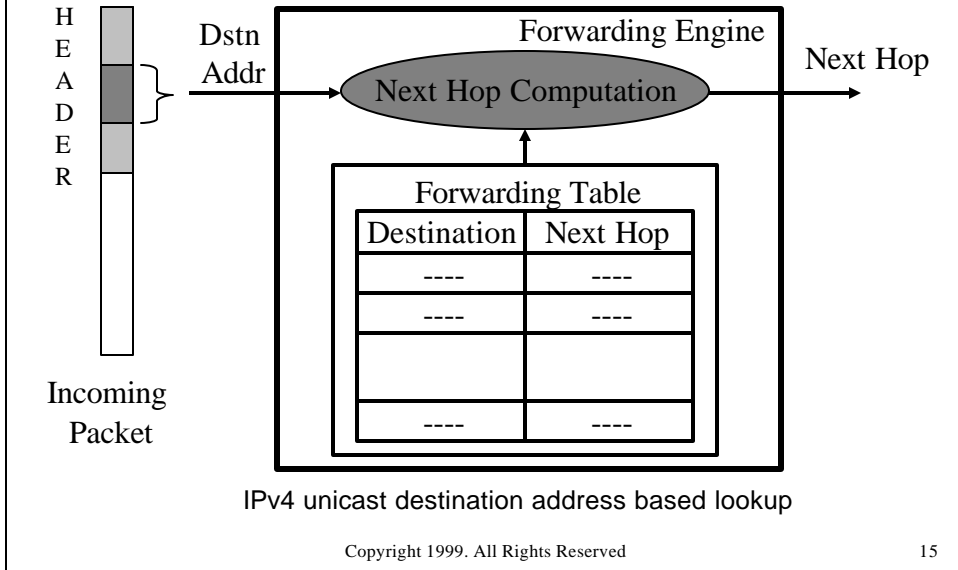


Copyright 1999. All Rights Reserved

14

IP Router

Lookup



IP Routers

Lookup

- Longest Prefix Matching

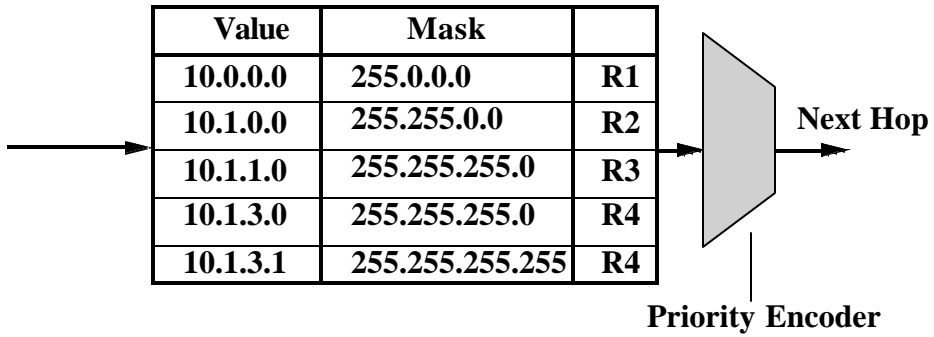
128.9.16.14

Prefix	Port
65/8	3
128.9/16	5
128.9.16/20	2
128.9.19/24	7
128.9.25/24	10
128.9.176/20	1
142.12/19	3

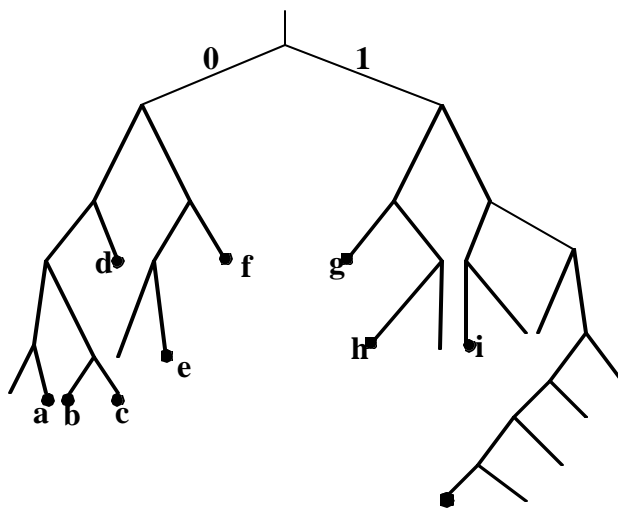
- Lookup time
- Storage space
- Update time
- Preprocessing time

Ternary CAMs

Associative Memory



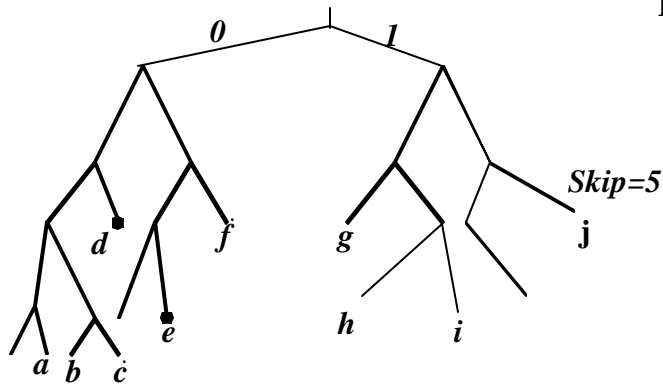
Binary Tries



Example Prefixes

- a) 00001
- b) 00010
- c) 00011
- d) 001
- e) 0101
- f) 011
- g) 100
- h) 1010
- i) 1100
- j) 11110000

Patricia Tree



Example Prefixes

- a) 00001
- b) 00010
- c) 00011
- d) 001
- e) 0101
- f) 011
- g) 100
- h) 1010
- i) 1100
- j) 11110000

Copyright 1999. All Rights Reserved

19

Switching Fabrics:

How does the packet get there?

- Output and Input Queueing
- Output Queueing
- Input Queueing

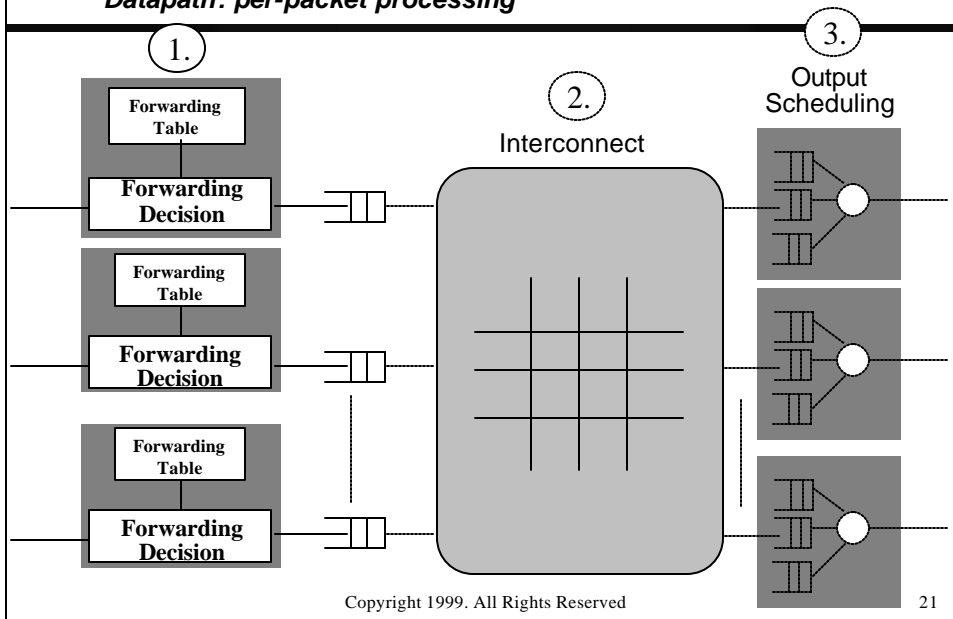
- Other non-blocking fabrics

Copyright 1999. All Rights Reserved

20

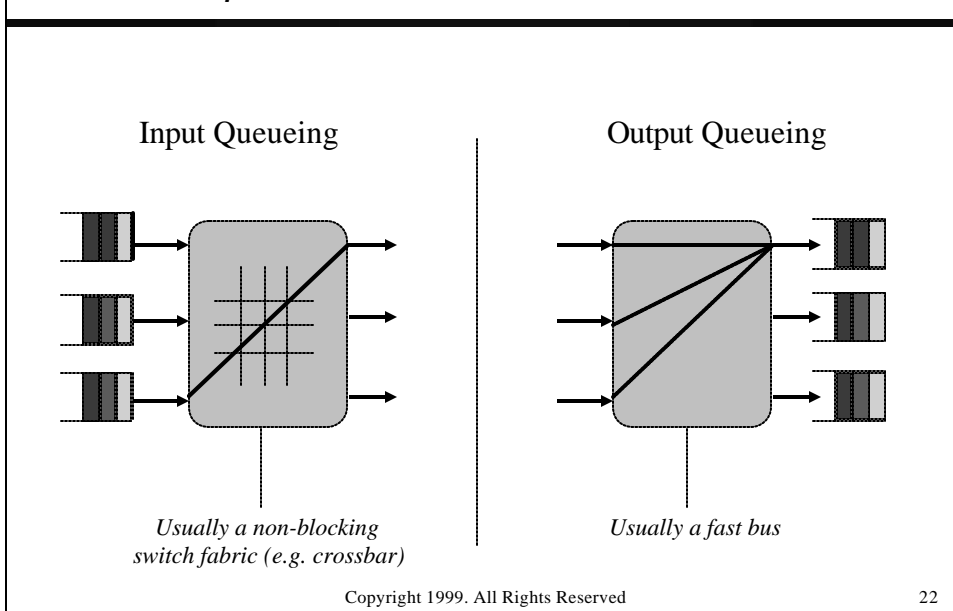
Basic Architectural Components

Datapath: per-packet processing



Interconnects

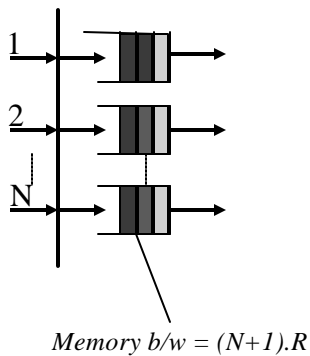
Two basic techniques



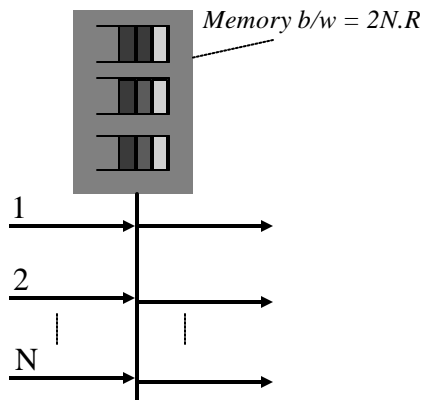
Interconnects

Output Queueing

Individual Output Queues



Centralized Shared Memory

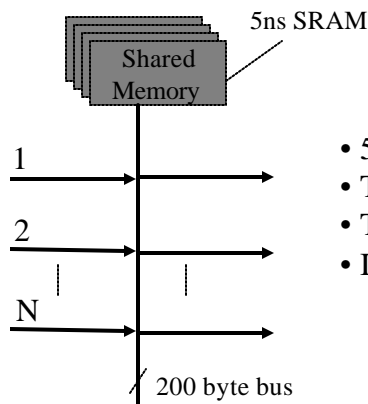


Copyright 1999. All Rights Reserved

23

Output Queueing

How fast can we make centralized shared memory?



- 5ns per memory operation
- Two memory operations per packet
- Therefore, up to 160Gb/s
- In practice, closer to 80Gb/s

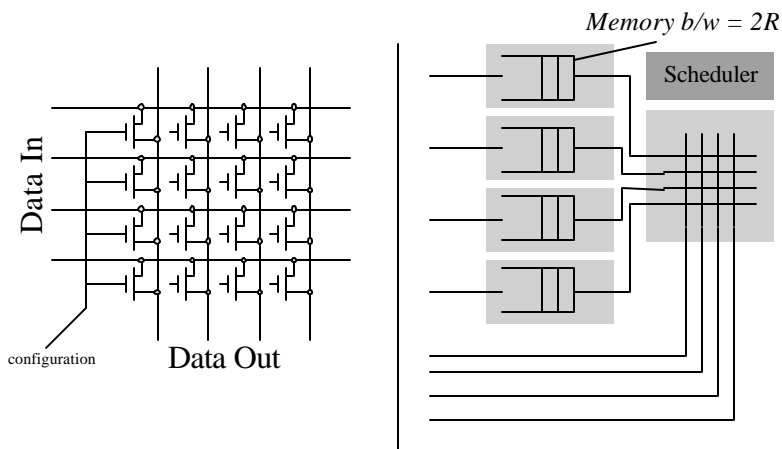
Copyright 1999. All Rights Reserved

24

Switching Fabrics

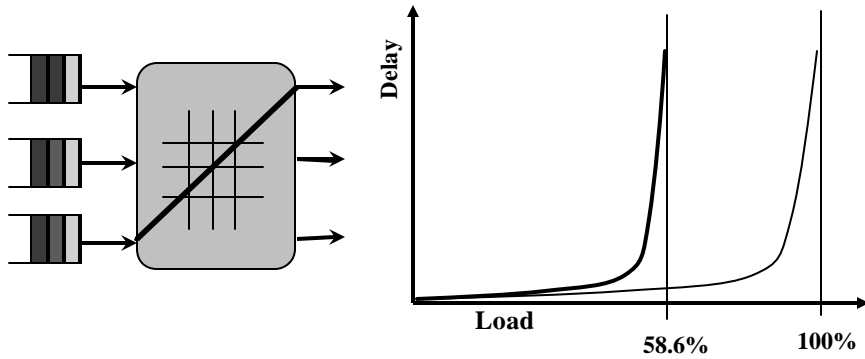
- Output and Input Queueing
- Output Queueing
- Input Queueing
 - Scheduling algorithms
 - Other non-blocking fabrics
 - Combining input and output queues
 - Multicast traffic

Input Queueing with Crossbar



Input Queueing

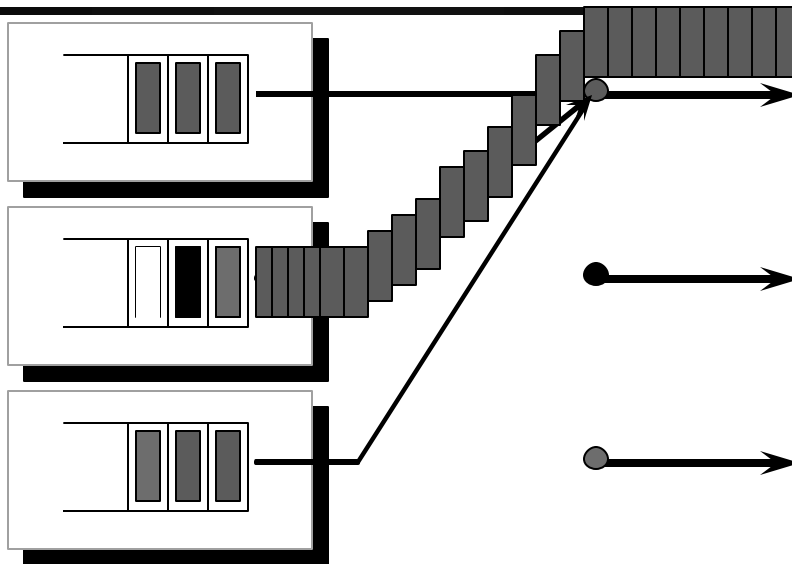
Head of Line Blocking



Copyright 1999. All Rights Reserved

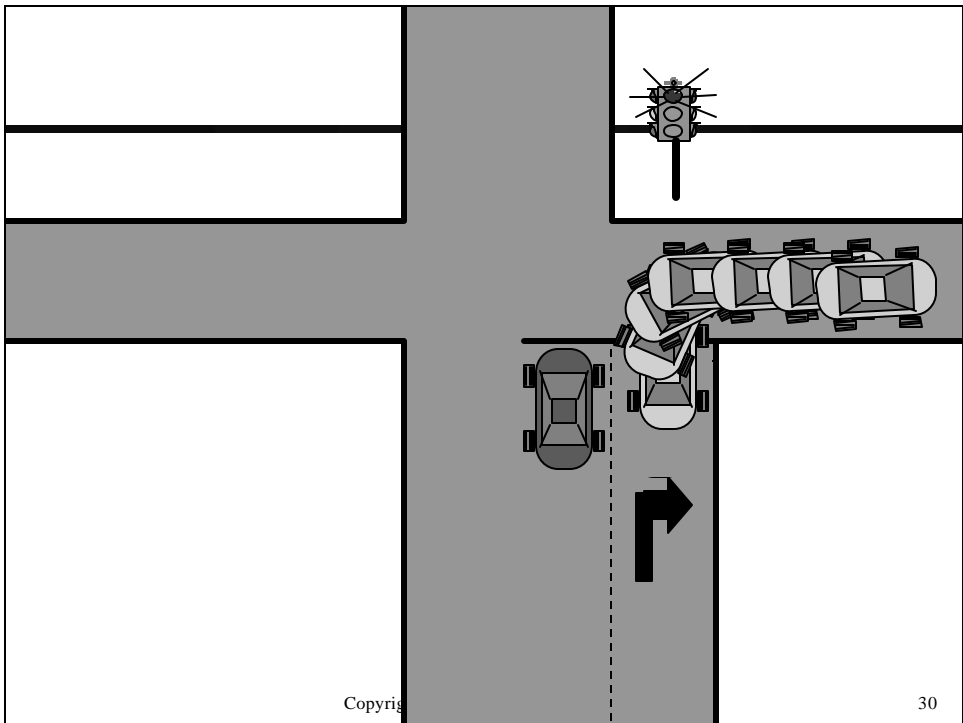
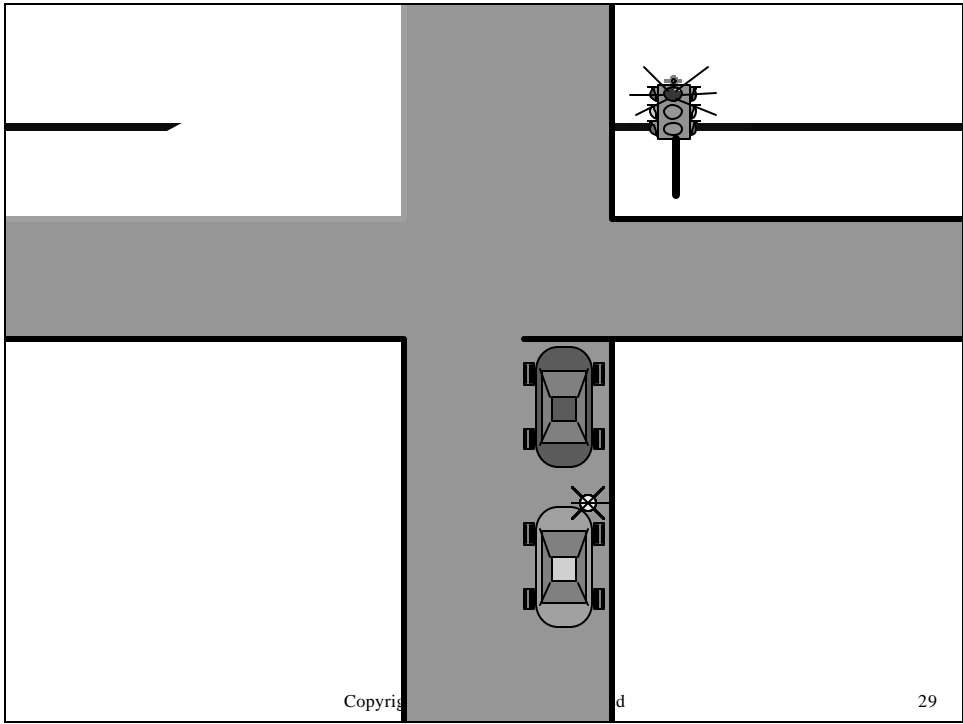
27

Head of Line Blocking



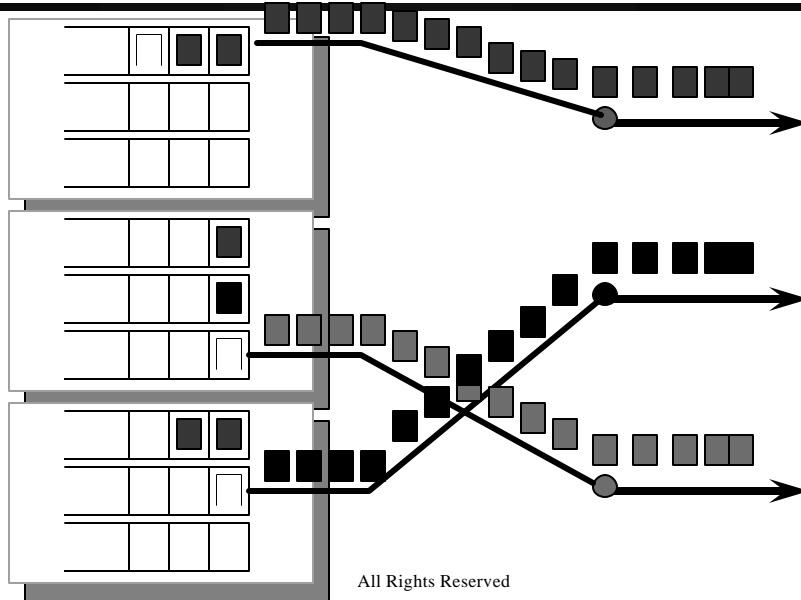
Copyright 1999. All Rights Reserved

28



Input Queuing

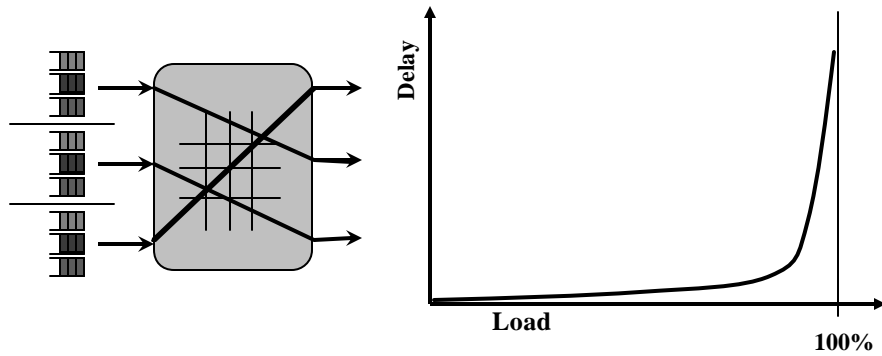
Virtual output queues



31

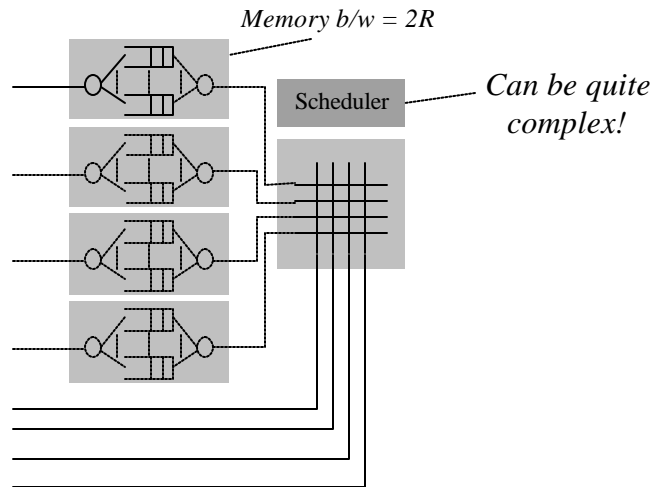
Input Queues

Virtual Output Queues



32

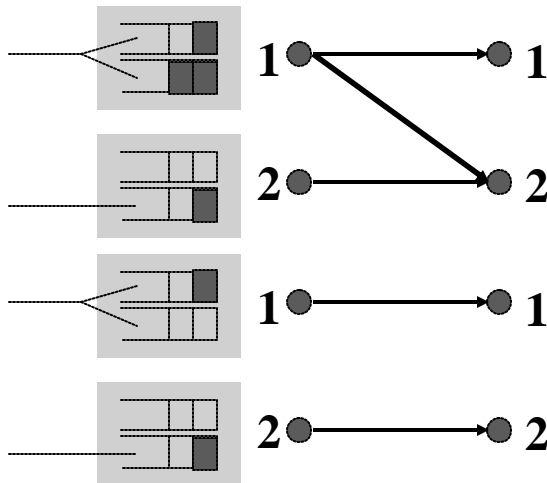
Input Queueing



Copyright 1999. All Rights Reserved

33

Input Queueing Scheduling

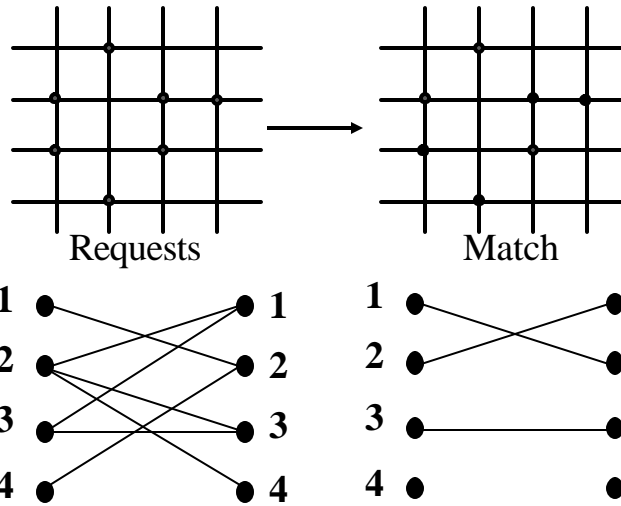


Copyright 1999. All Rights Reserved

34

Wave Front Arbiter

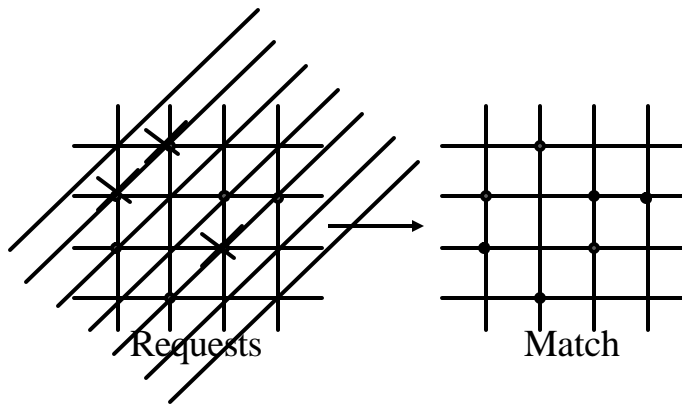
Scheduling Algorithm



Copyright 1999. All Rights Reserved

35

Wave Front Arbiter

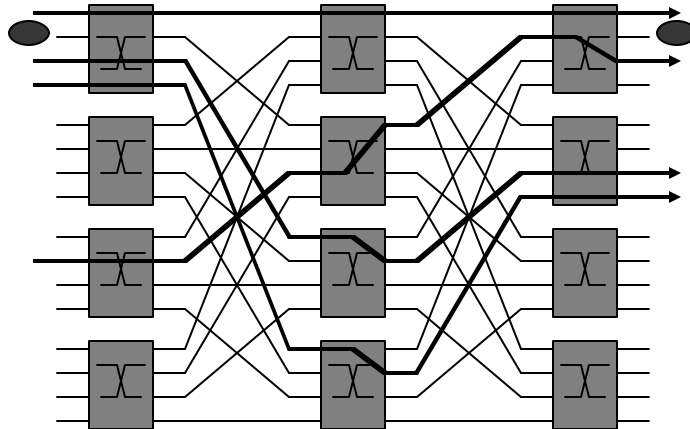


Copyright 1999. All Rights Reserved

36

Other Non-Blocking Fabrics

Clos Network



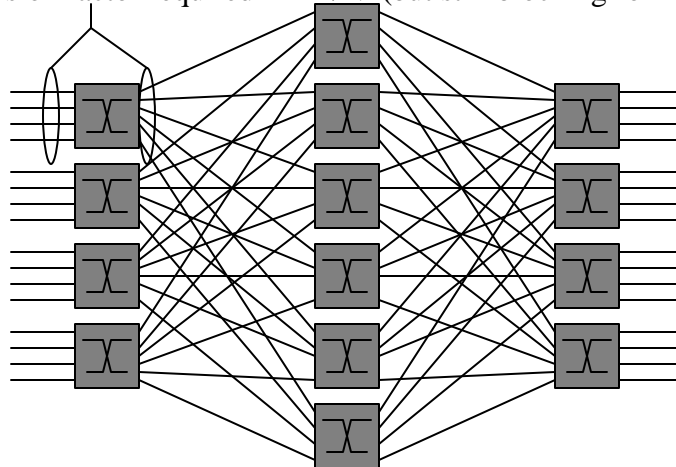
Copyright 1999. All Rights Reserved

37

Other Non-Blocking Fabrics

Clos Network

Expansion factor required = $2 - 1/N$ (but still blocking for multicast)

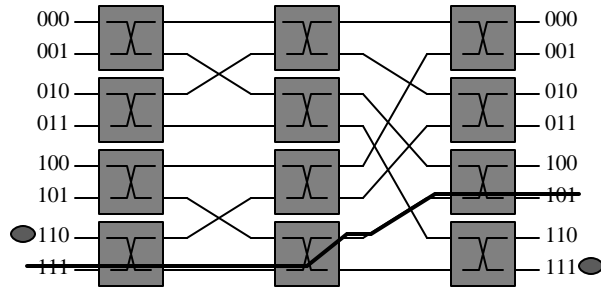


Copyright 1999. All Rights Reserved

38

Other Non-Blocking Fabrics

Self-Routing Networks



Copyright 1999. All Rights Reserved

39

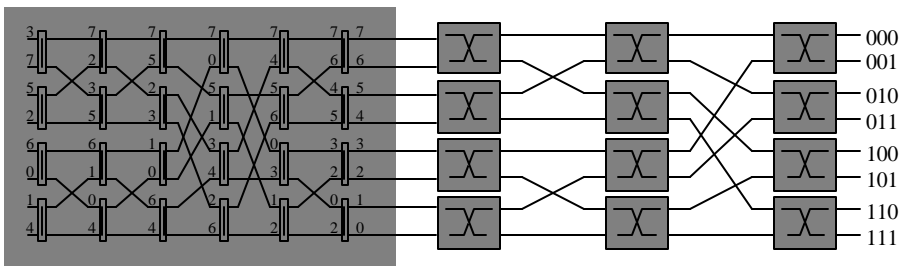
Other Non-Blocking Fabrics

Self-Routing Networks

The Non-blocking Batcher Banyan Network

Batcher Sorter

Self-Routing Network



- *Fabric can be used as scheduler.*
- *Batcher-Banyan network is blocking for multicast.*

Copyright 1999. All Rights Reserved

40