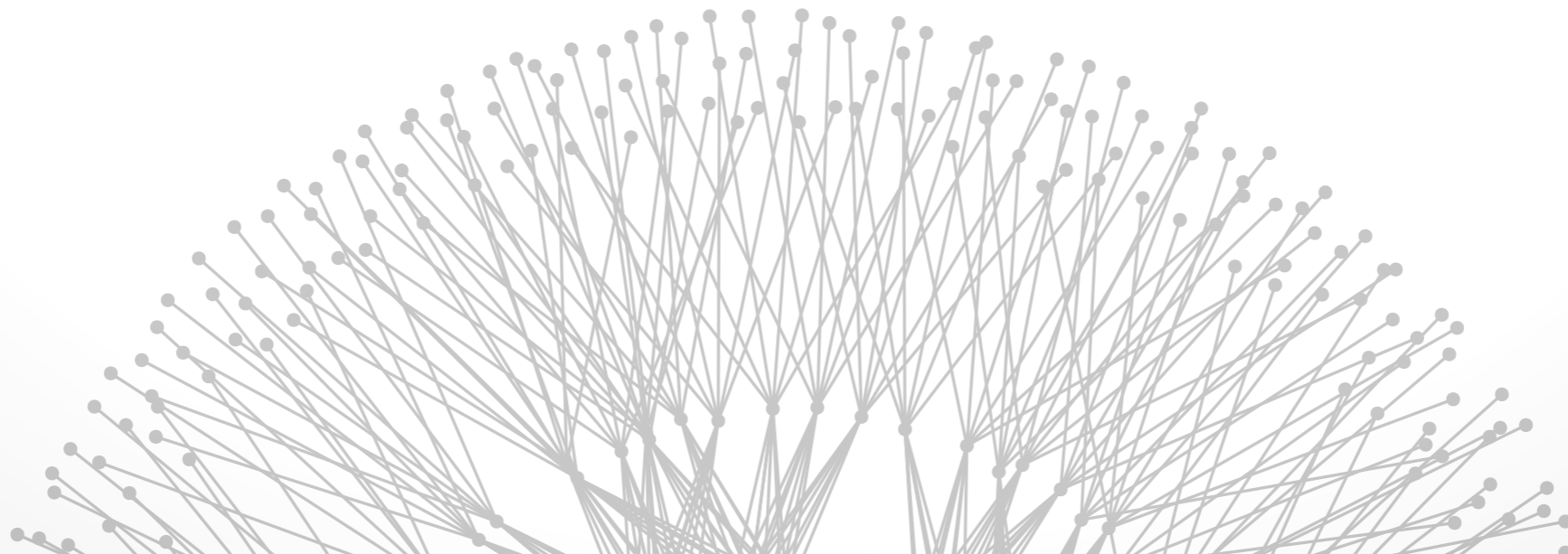


# Forwarding Architecture

Brighten Godfrey  
CS 538 September 13 2011



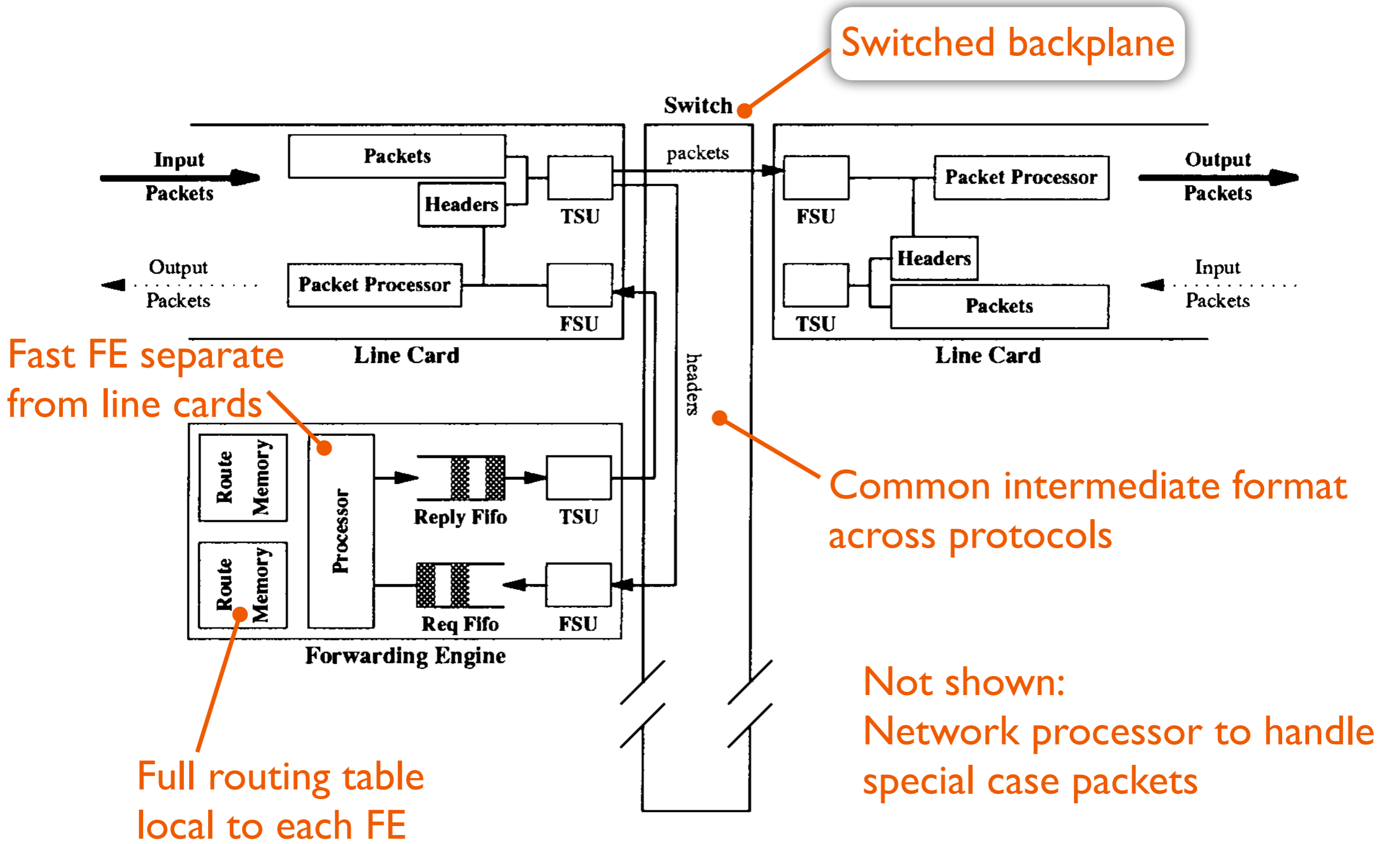
# Partridge: 50 Gb/sec router



A fast IP router

Good exhibition of the guts of a router and problems to be solved in router architecture

# Inside the router



Fast FE separate from line cards

Switched backplane

Common intermediate format across protocols

Full routing table local to each FE

Not shown:  
Network processor to handle special case packets

# Switching fabric



## Operates in **epochs**

- 128 bytes sent by each line card to next-hop line card
- Each line card can send to only one other card, and can receive from only one other card

... to outputs

Inputs ready to send...

	1	2	3	4	5	6	7
1	0	0	1	0	1	0	0
2	1	1	1	1	0	0	0
3	0	1	1	1	0	0	1
4	1	0	1	0	0	1	0
5	0	0	0	0	0	0	0
6	0	0	0	0	1	0	1
7	0	1	1	0	0	1	0

Allocator assigns inputs to outputs & tells line cards

What fundamental problem is being solved?

# Maximum bipartite matching



Maximize number of matched input-output pairs, such that each input & output only matched once

50 Gbit/s router uses approximate solution

Inputs ready to send...

... to outputs

	1	2	3	4	5	6	7
1	0	0	1	0	1	0	0
2	1	1	1	1	0	0	0
3	0	1	1	1	0	0	1
4	1	0	1	0	0	1	0
5	0	0	0	0	0	0	0
6	0	0	0	0	1	0	1
7	0	1	1	0	0	1	0

# Routers are jacks-of-all-trades



## Many functions in modern routers

- access control
- quality of service
- accounting, traffic metering
- IPv4, IPv6, MPLS, ethernet, ...
- Virtual private networks
- ...



## Hardware routers

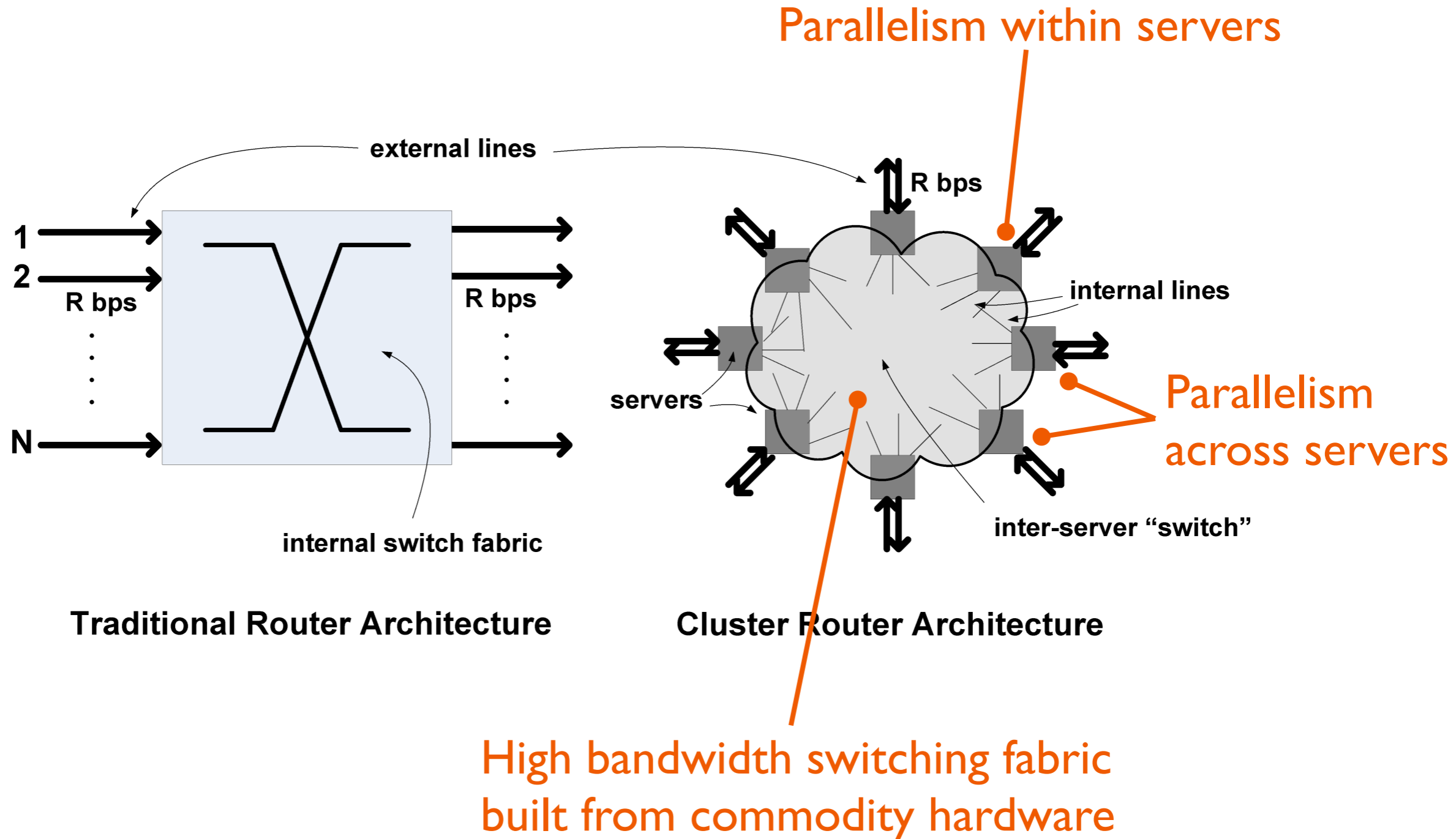
- Fast
- Specific functionality
- Result: many physical devices (routers, firewalls, intrusion detection, ...)

## Software routers

- Slow
- Extensible

Can we get the best of both worlds?

# RouteBricks approach





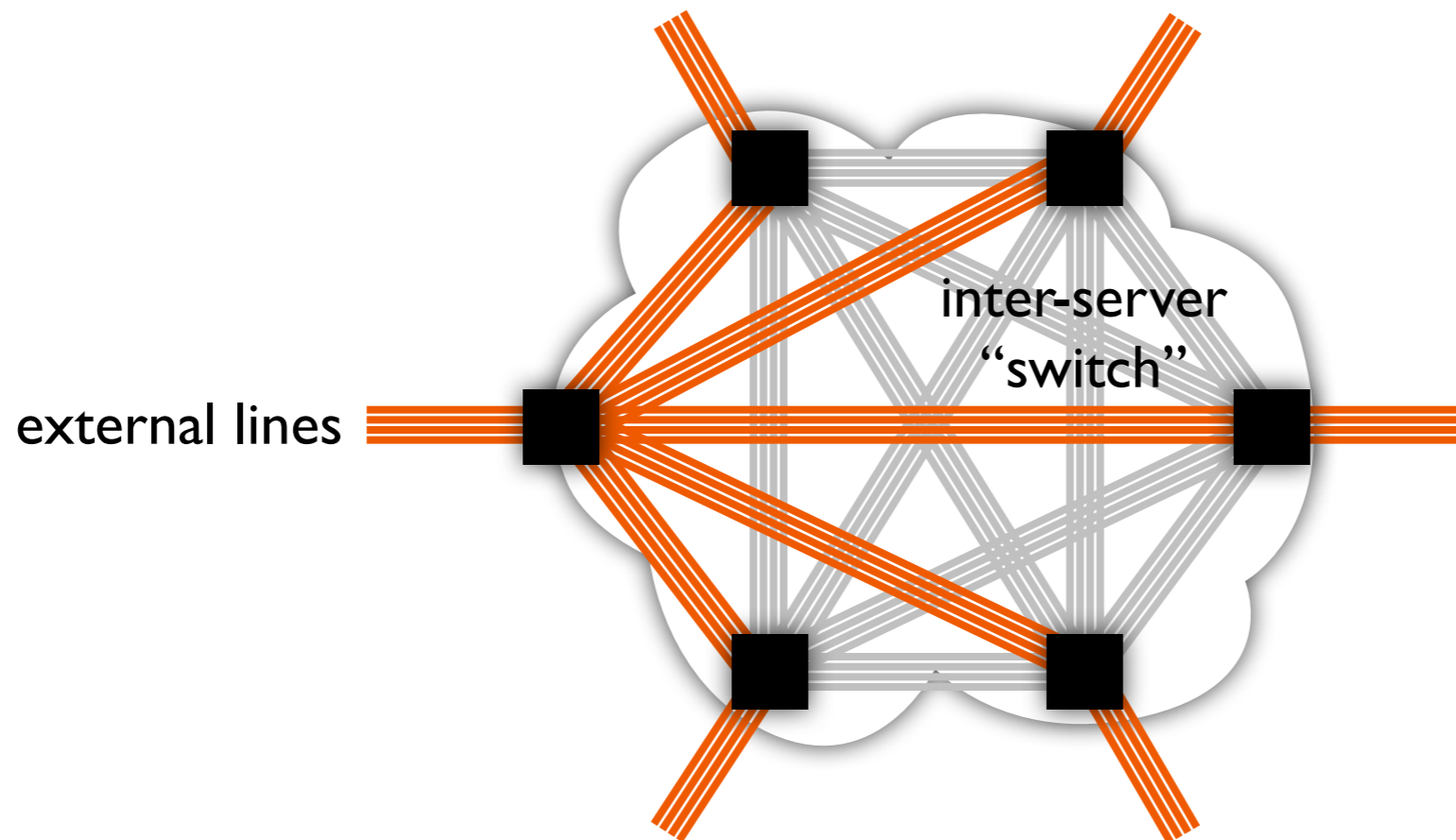
# Switching fabric challenges



**Handle any traffic pattern:** for example, all input traffic at a server might go to any one output server

**Low degree:** we're using commodity hardware

Naïve approach:



Useless: might as well just use one server!

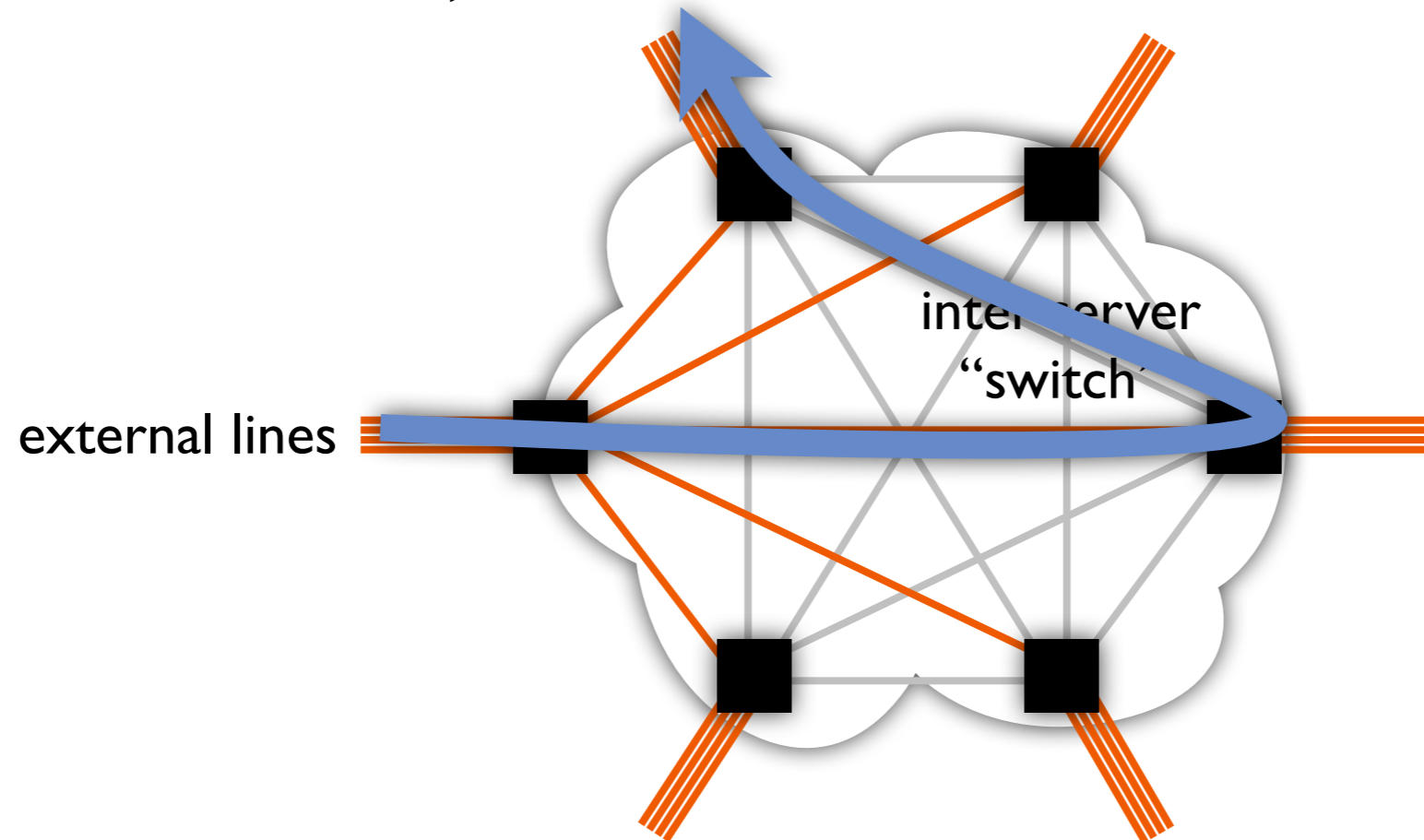
# Low degree solution



Just one link out for each link in

Total out b/w enough, but doesn't go where we need

Solution (**Valiant load balancing**): send packet to random intermediate node, then on to destination



# VLB guarantees & questions



Guaranteed to nearly full throughput for **any** traffic demands!

- “nearly” = 2x. Why?
- So, switch fabric needs to be 2x as fast as external links to provide guarantees

Why does sending to a random intermediate node work?

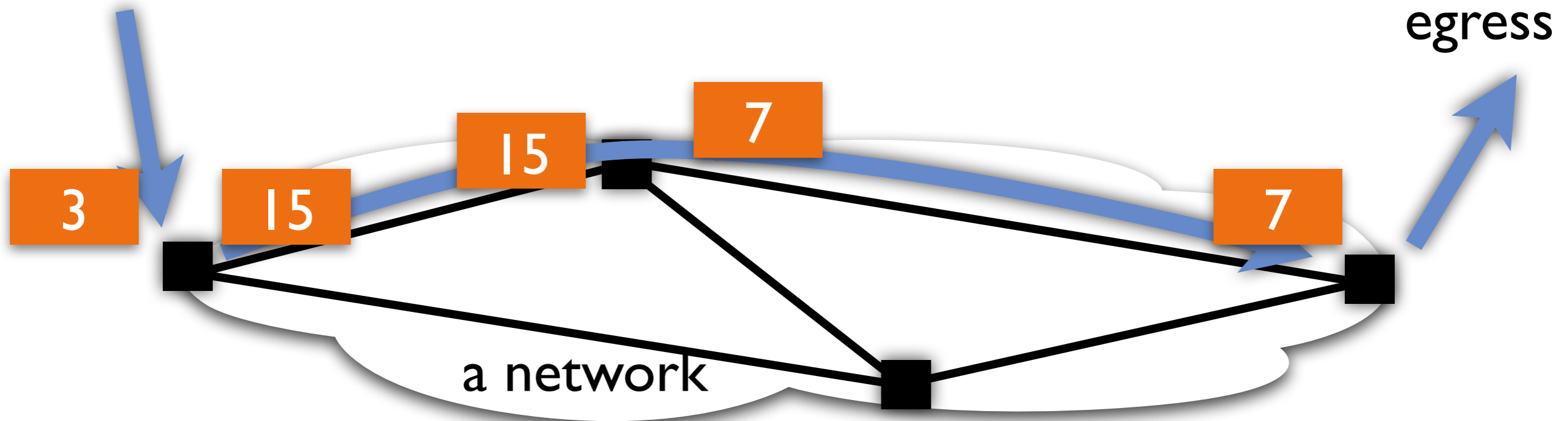
Still using one port per server. What if # servers > # ports available?

# MPLS design



Ingress:

Traffic classification, label packets (“forwarding equivalence class”)



Control plane constructs label-switched paths and coordinates labels

Can also stack labels = concatenate paths



## In the design doc

- High performance forwarding
- Minimal forwarding requirements, so can interface well with many types of media such as ATM
- Flexible control of traffic routing

What matters today? **Flexibility**. Widely used for:

- Virtual Private Network (VPN) service along dedicated paths between enterprise sites
- Traffic engineering on per-“flow” granularity
- Control backup paths with MPLS Fast ReRoute

# Announcements



## By tonight:

- Submit project proposal
- Fill out presentation topic survey

## Thursday:

- **BGP routing policies in ISP networks**  
(Caesar and Rexford, 2005)
- Class ends 4:30 p.m.