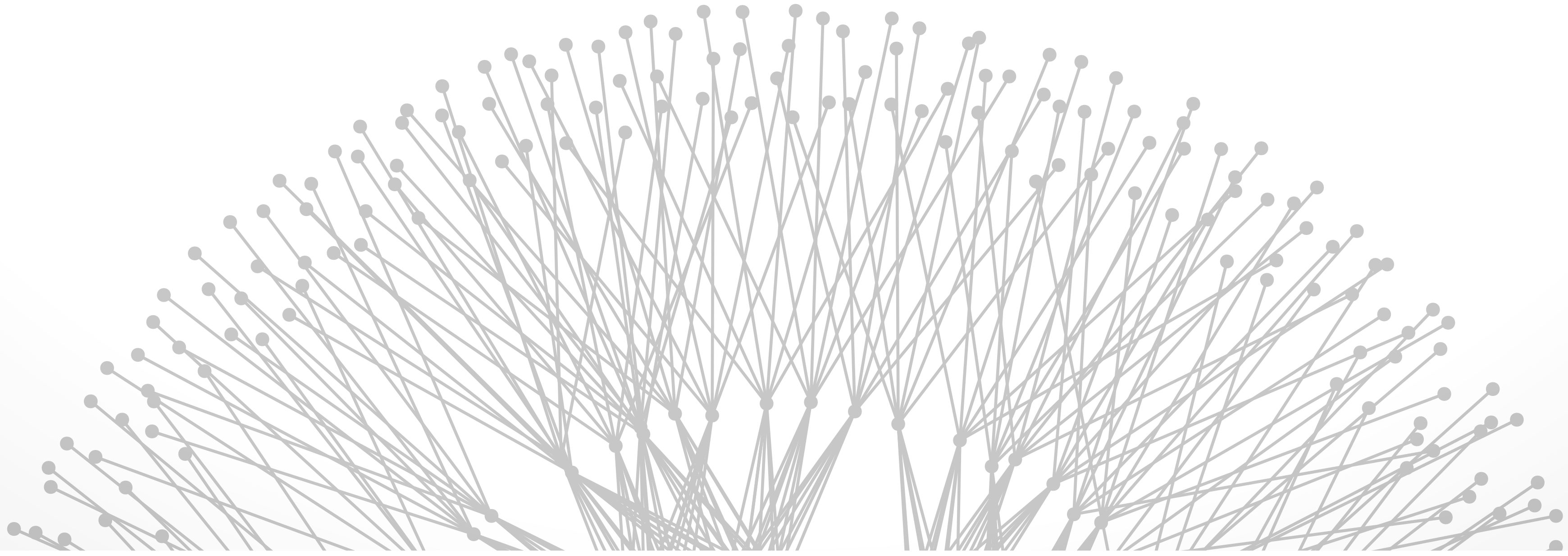
Forwarding Architecture

Brighten Godfrey
CS 538 February 14 2018



Building a fast router

Partridge: 50 Gb/sec router

A fast IP router

- well, fast at the time...

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

Routing vs. forwarding

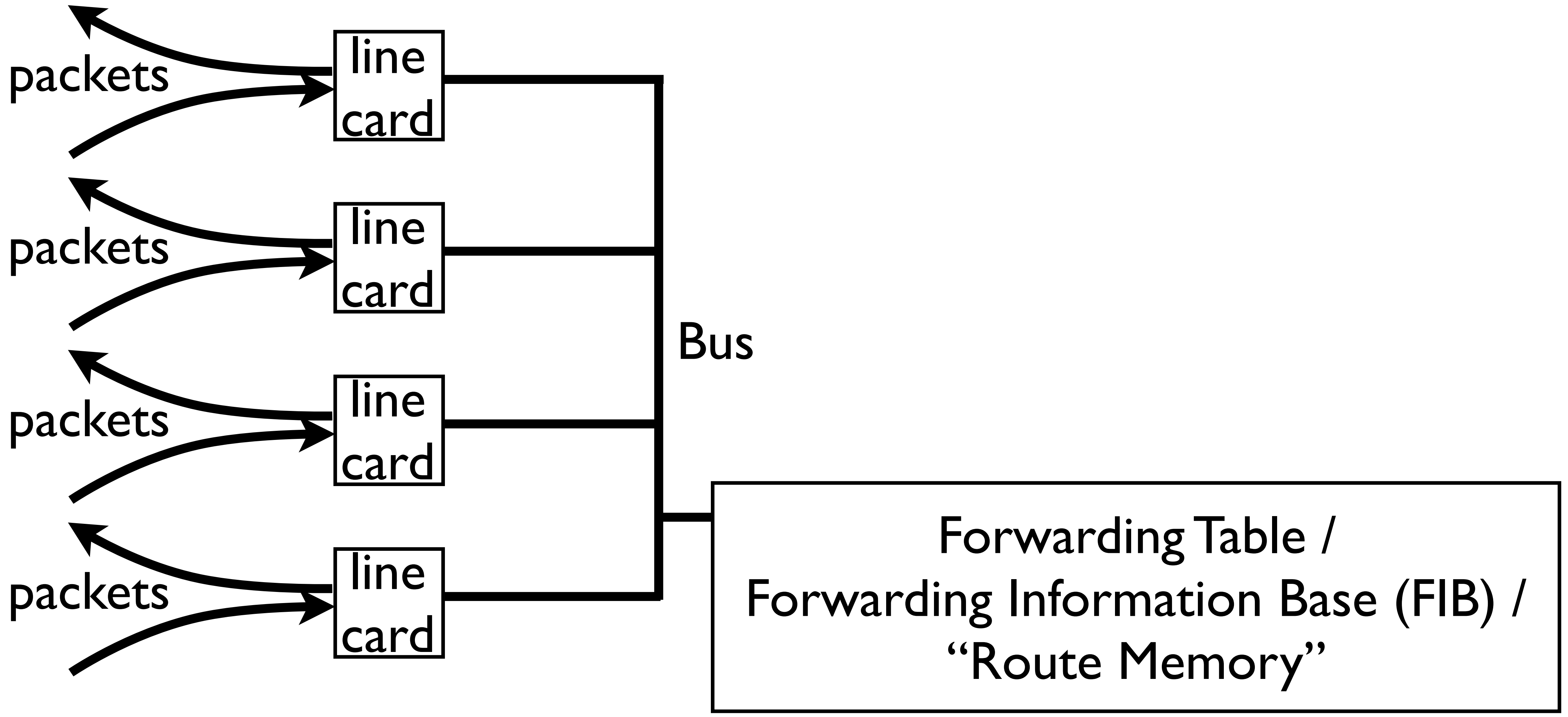
Control plane ⟵ Upcoming lectures

- Decides how data should flow through the network
- Uses operator configs & distributed routing protocols
- Can use commodity hardware (often Linux today)
- Output: Forwarding table
  - Contains instructions for what to do with each packet based on its header
  - Relatively simple data structure (lookup table, tree, ...)

Data plane ⟵ Today

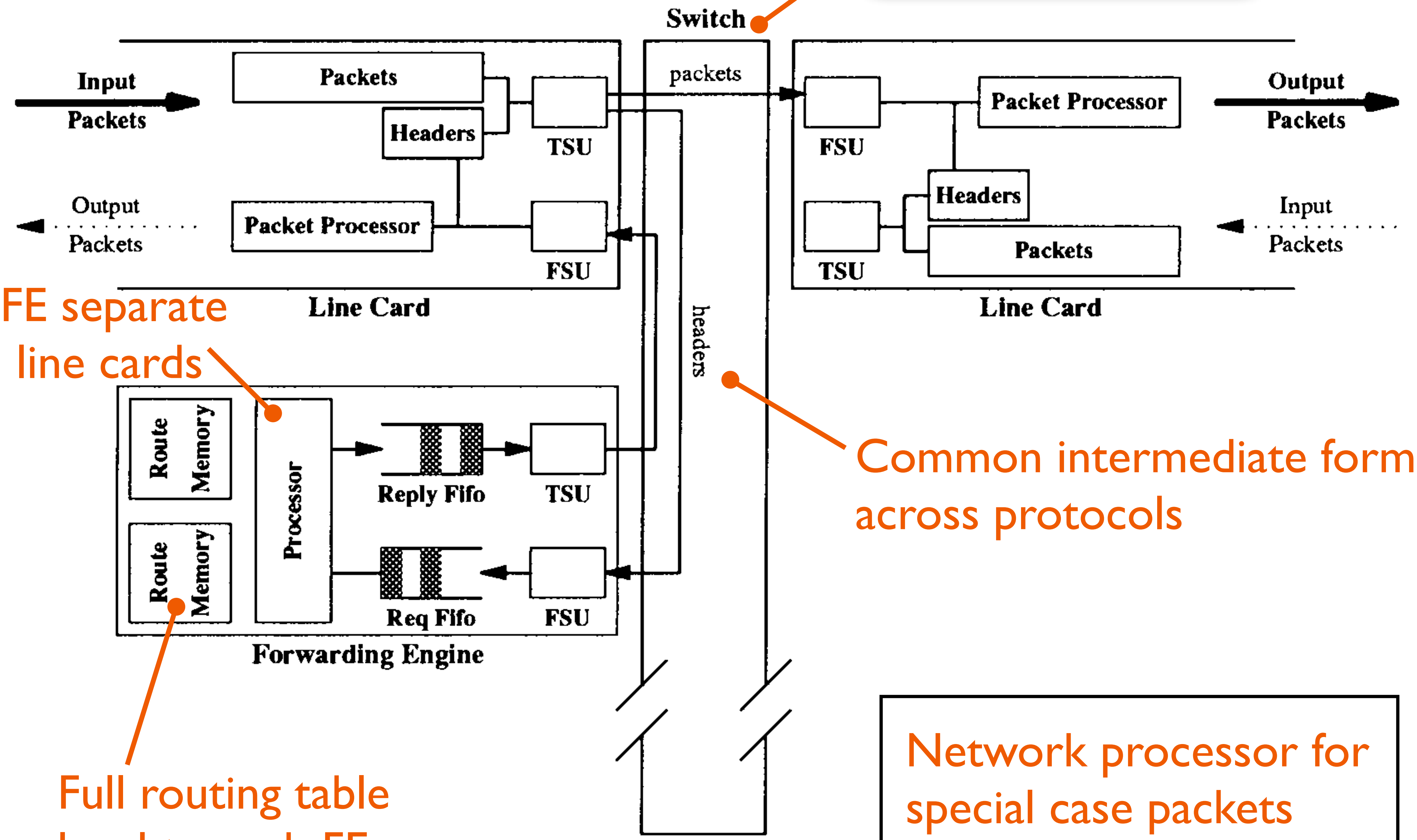
- Input: Forwarding table
- Forwards data through the network
- Typically specialized ASIC hardware for fast forwarding

Simplistic router diagram



What’s wrong with this picture?

Inside the router



Switched backplane

Fast FE separate from line cards

Common intermediate format across protocols

Full routing table local to each FE

Network processor for special case packets

[Partridge et al '98]

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

… 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

Beyond the 50 Gb/sec router

Many more problems to solve

Buffering (how big?)

Queueing (where? what order?)

FIB memory (how fast, large, and expensive?)

FIB algorithms (what data structure? how many memory accesses?)

Specialized functionality (how to do it efficiently?)

Specialized functionality

Modern routers are jacks-of-all-trades:

- Load balance across links
- Access control
- Filtering attacks
- Quality of service
- Accounting, traffic metering
- Virtual private networks
- Protocol support: IPv4, IPv6, MPLS, ethernet, ...
- ...

Can we make forwarding flexible, extensible?

Data plane flexibility over the ages

1990s: Software routers

- Zebra, later Quagga, Click, Xorp

1997: Label switching (MPLS)

- Set up explicit paths for classes of packets

1999: Active networks

- Packet header carries (pointer to) program code

2008: Software Defined Networks

- Open interface to data plane, programmed by software controller

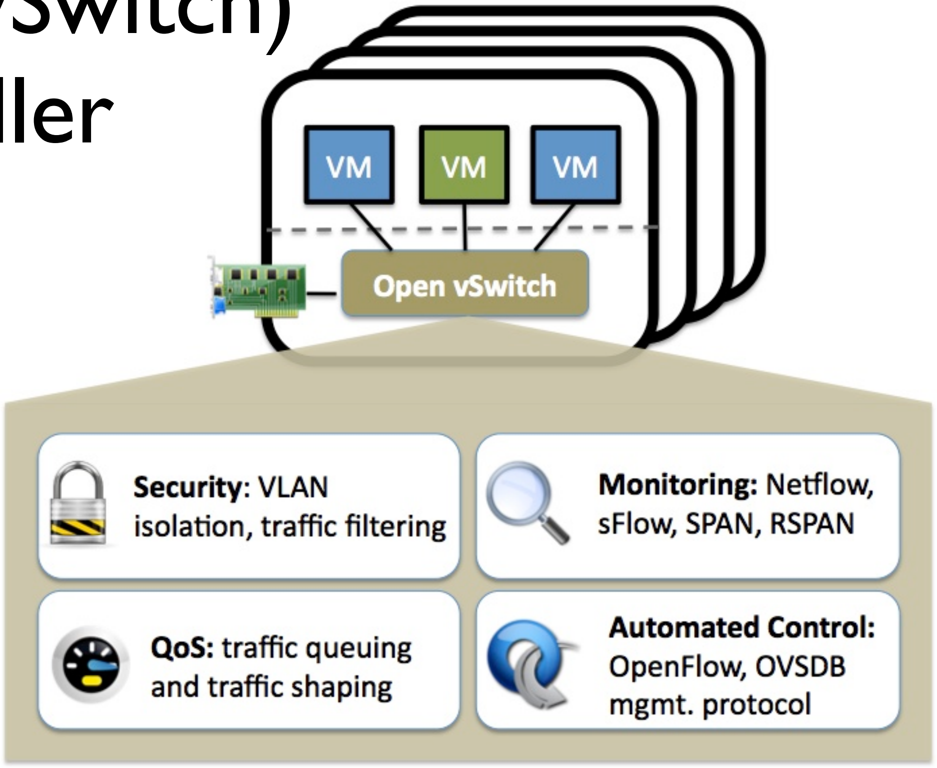
Data plane flexibility over the ages

2008: Software Defined Networks

- Open interface to data plane, programmed by software controller

2009: Network virtualization

- Data plane complexity moved to edge devices: software in servers (Hypervisor, Open vSwitch)
- Managed by SDN-style controller


VM
VM
VM
Open vSwitch
Security: VLAN isolation, traffic filtering
Monitoring: Netflow, sFlow, SPAN, RSPAN
QoS: traffic queuing and traffic shaping
Automated Control: OpenFlow, OVSDB mgmt. protocol

openvswitch.org

Data plane flexibility over the ages

2008: Software Defined Networks

- Open interface to data plane, programmed by software controller

2009: Network virtualization

- Data plane complexity moved to edge devices: software in servers (Hypervisor, Open vSwitch)
- Managed by SDN-style controller

2014: Next-gen programmable switches

Scaling software routers

Efficiency vs. extensibility

Hardware routers

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

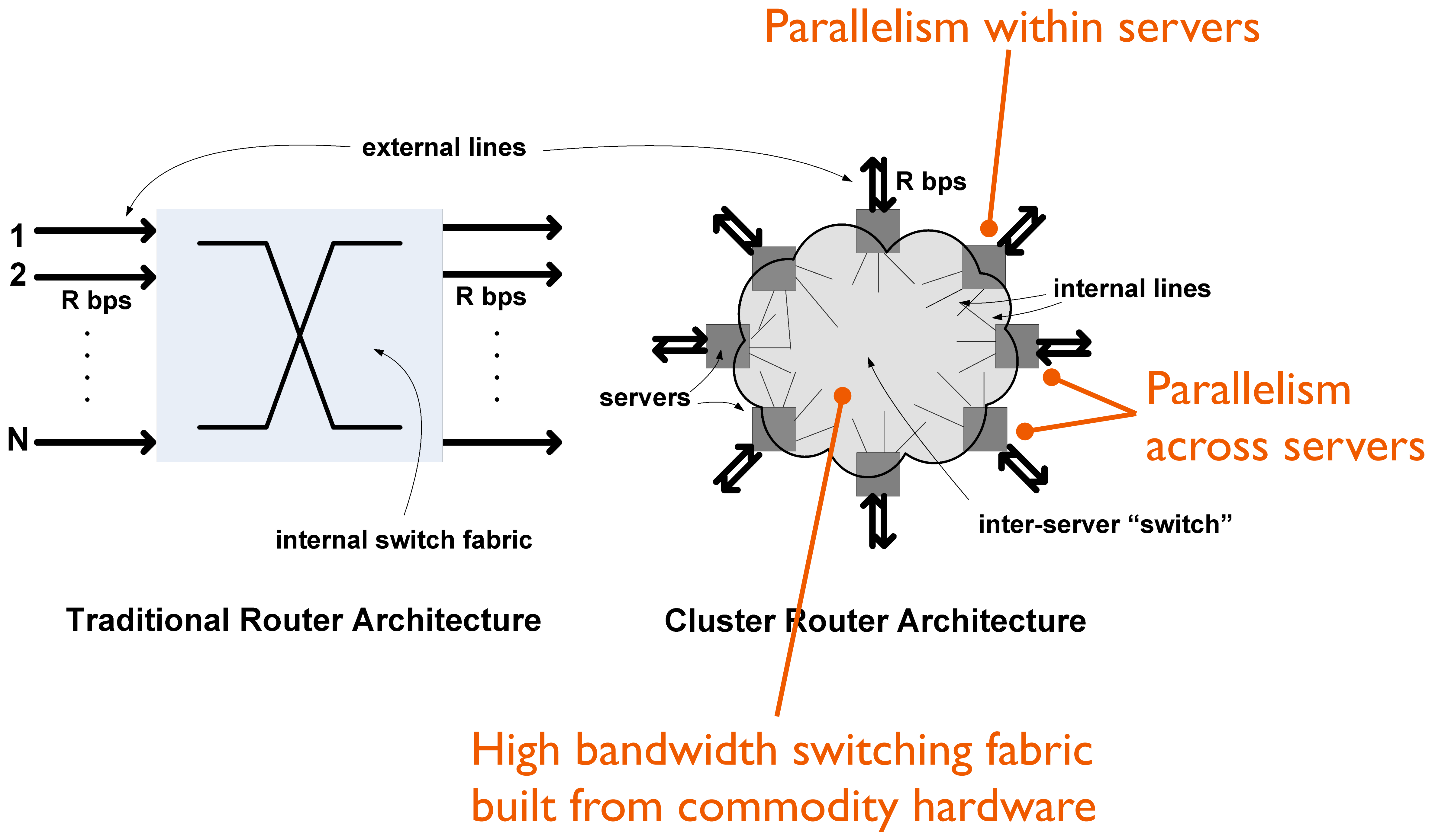
Software routers

- Slow
- Extensible

Can we get the best of both worlds?

RouteBricks approach

[Dobrescu, Egi, Argyraki, Chun, Fall, Iannaccone, Knies, Manesh, Ratnasamy, NSDI 2009]

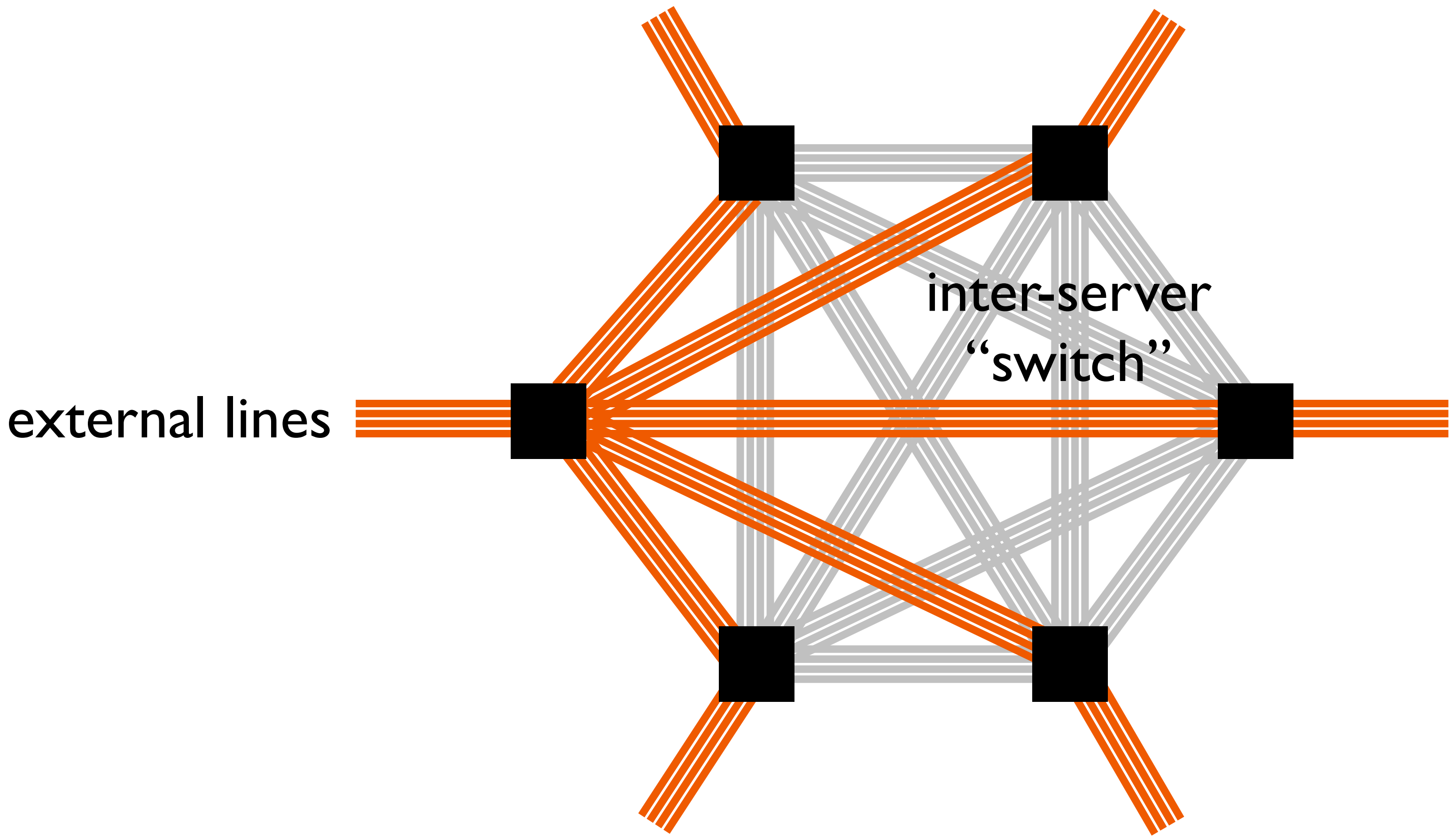

Parallelism within servers
external lines
R bps
1
2
R bps
N
R bps
internal lines
servers
Parallelism across servers
inter-server "switch"
internal switch fabric
Traditional Router Architecture
Cluster Router Architecture
High bandwidth switching fabric built from commodity hardware

Switching fabric challenges

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

Low degree (ports): we're using commodity hardware

Naïve approach:



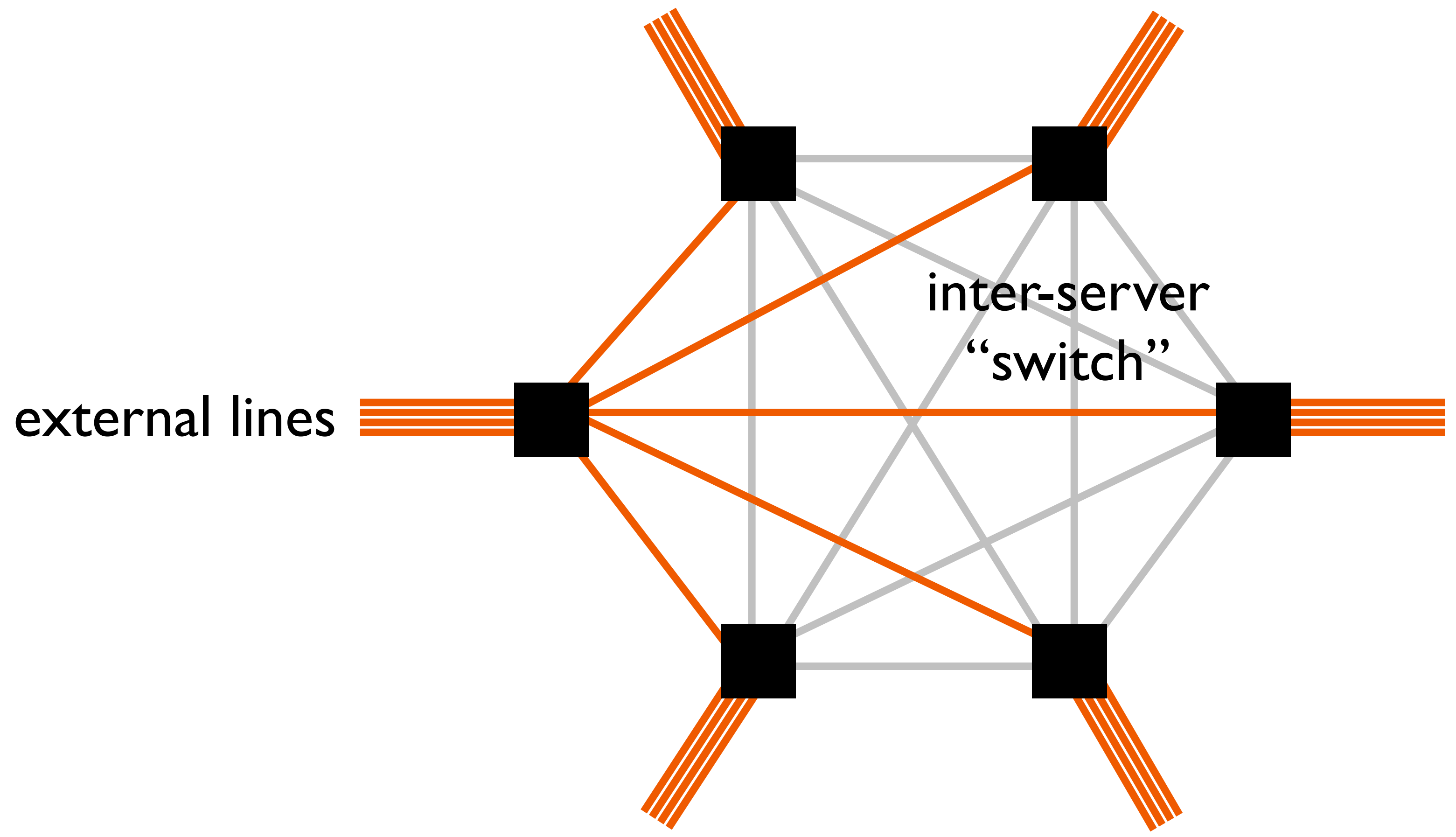
Useless: might as well just use one server!

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

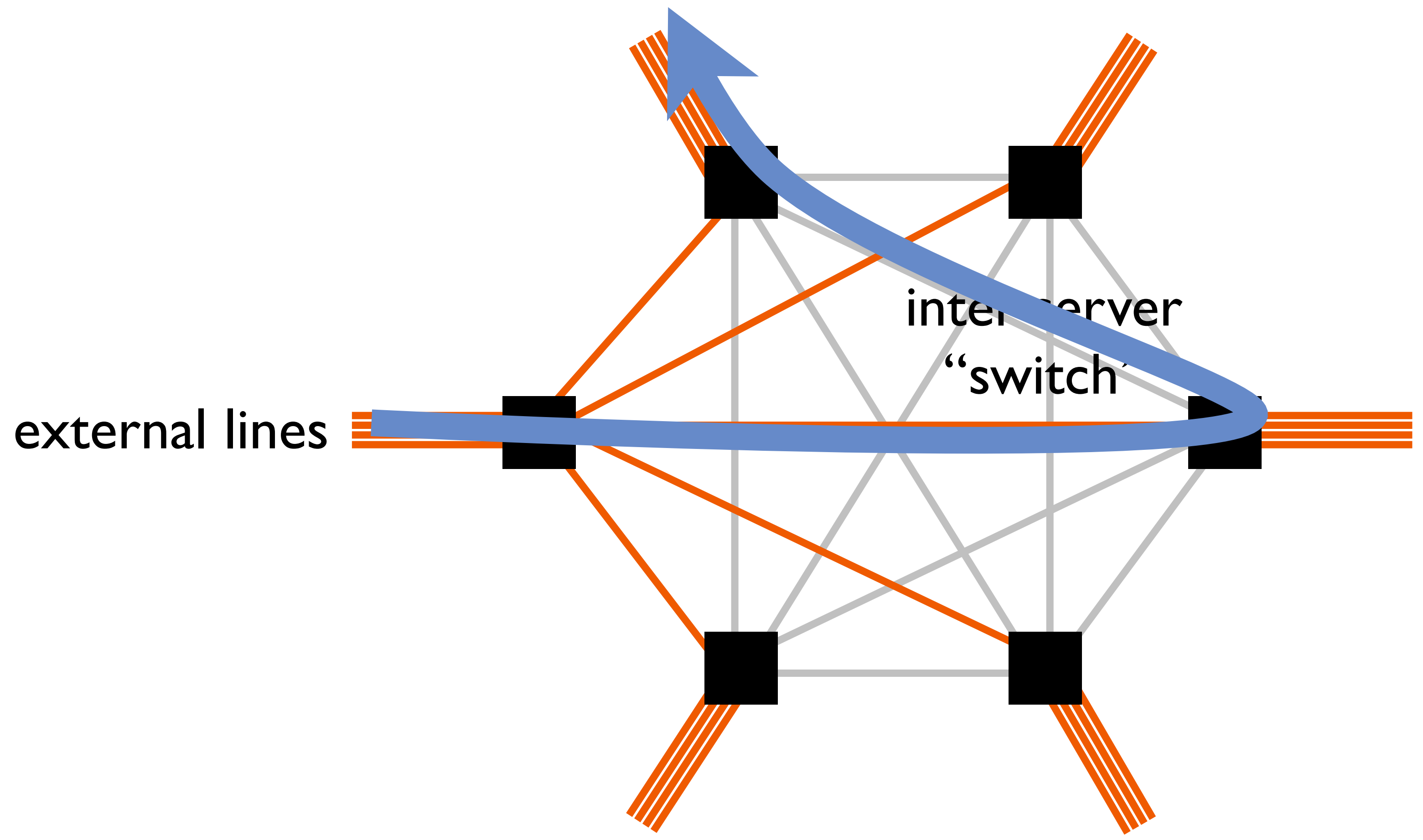


Low degree solution

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



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?

Retrospective

Software routers are not replacing big iron routers

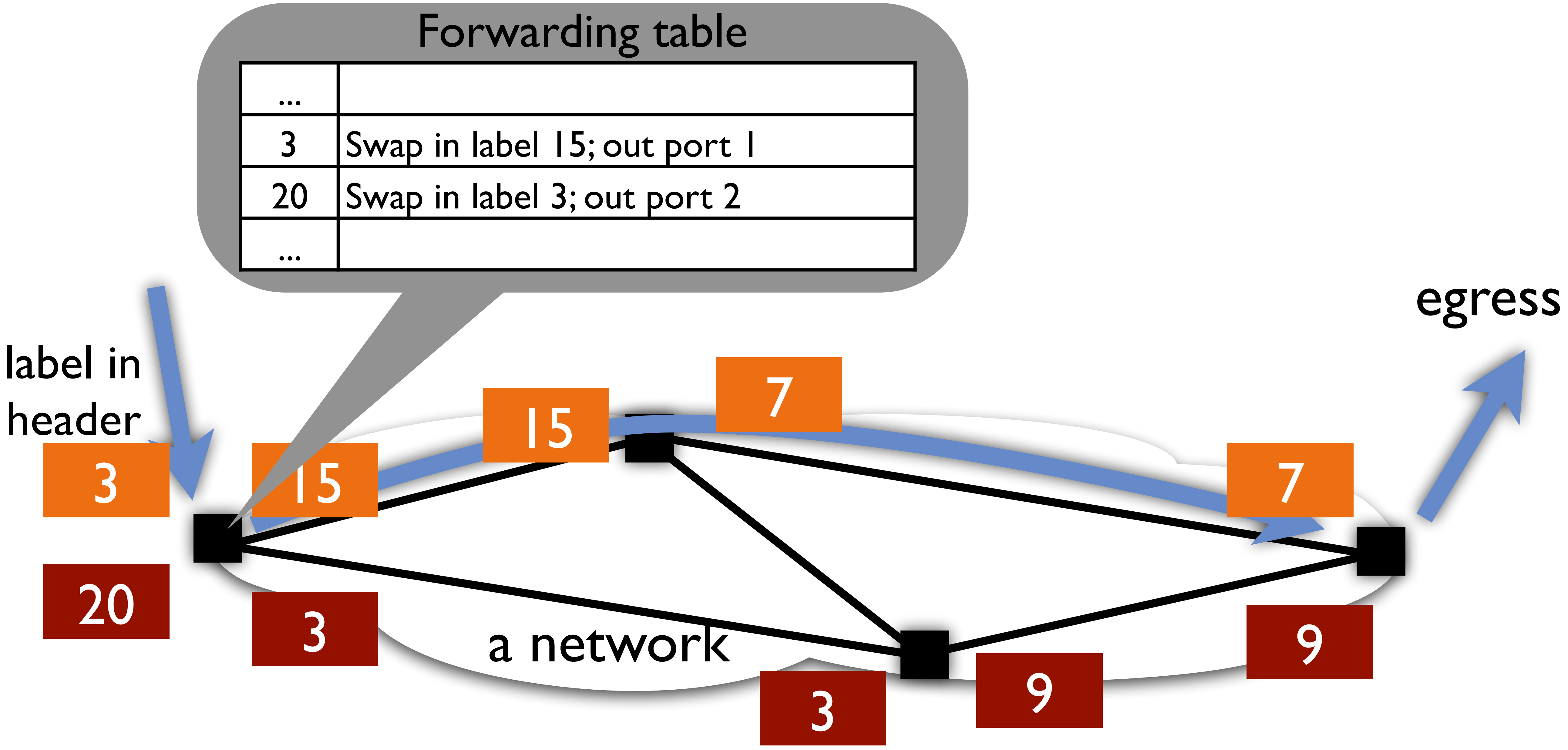
However, increasing use of software in the fast path

- 'Middleboxes' / Virtual Network Functions: load balancing, logging, firewalls, intrusion detection, ...
- Software switches in software-defined data centers

- Solutions from multiple vendors to speed software processing of data flows (e.g. 10 Gbps)

MPLS

MPLS design



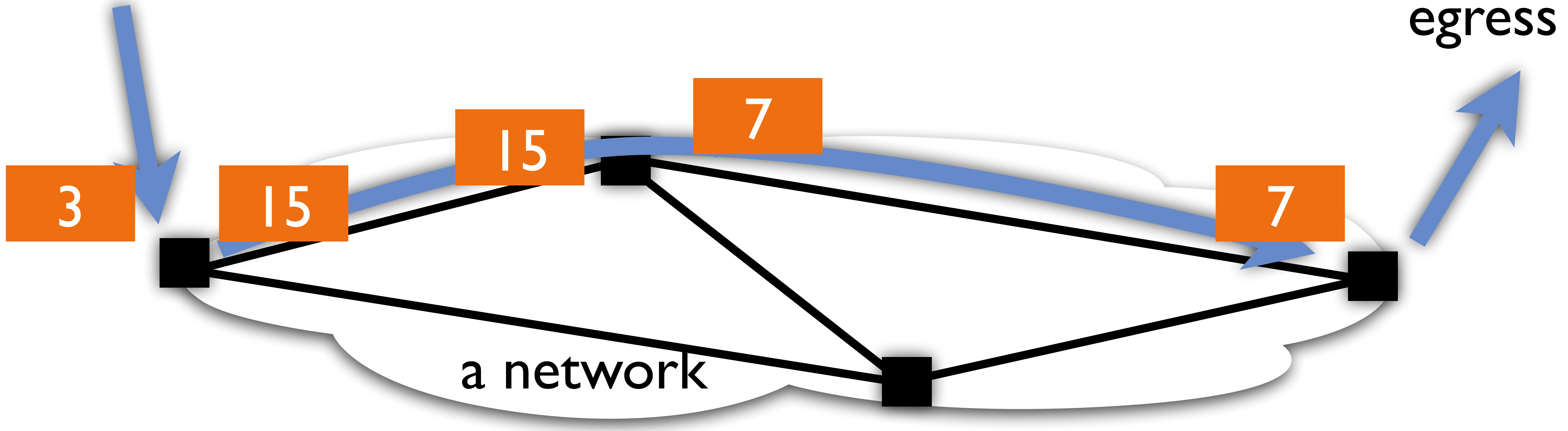
Forwarding table

...	
3	Swap in label 15; out port 1
20	Swap in label 3; out port 2
...	

Why is this more flexible than shortest path routing?

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

MPLS motivation

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 to achieve:

- Virtual Private Network (VPN) service along dedicated paths between enterprise sites
- Control backup paths with MPLS Fast ReRoute
- Traffic engineering (load balancing)

Announcements

Next Monday: Intradomain routing

- Papers will be posted by midnight tonight