Network Layer: Router Design, Protocols

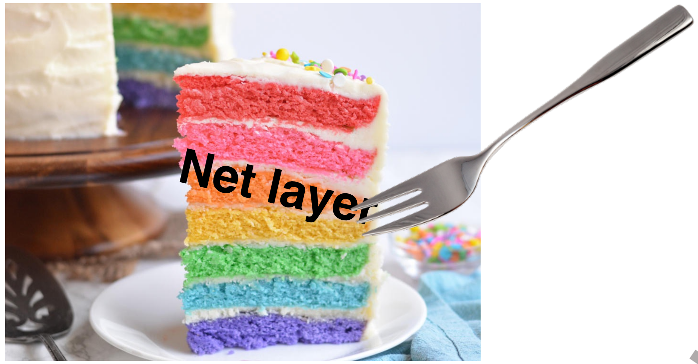
Lecture 20

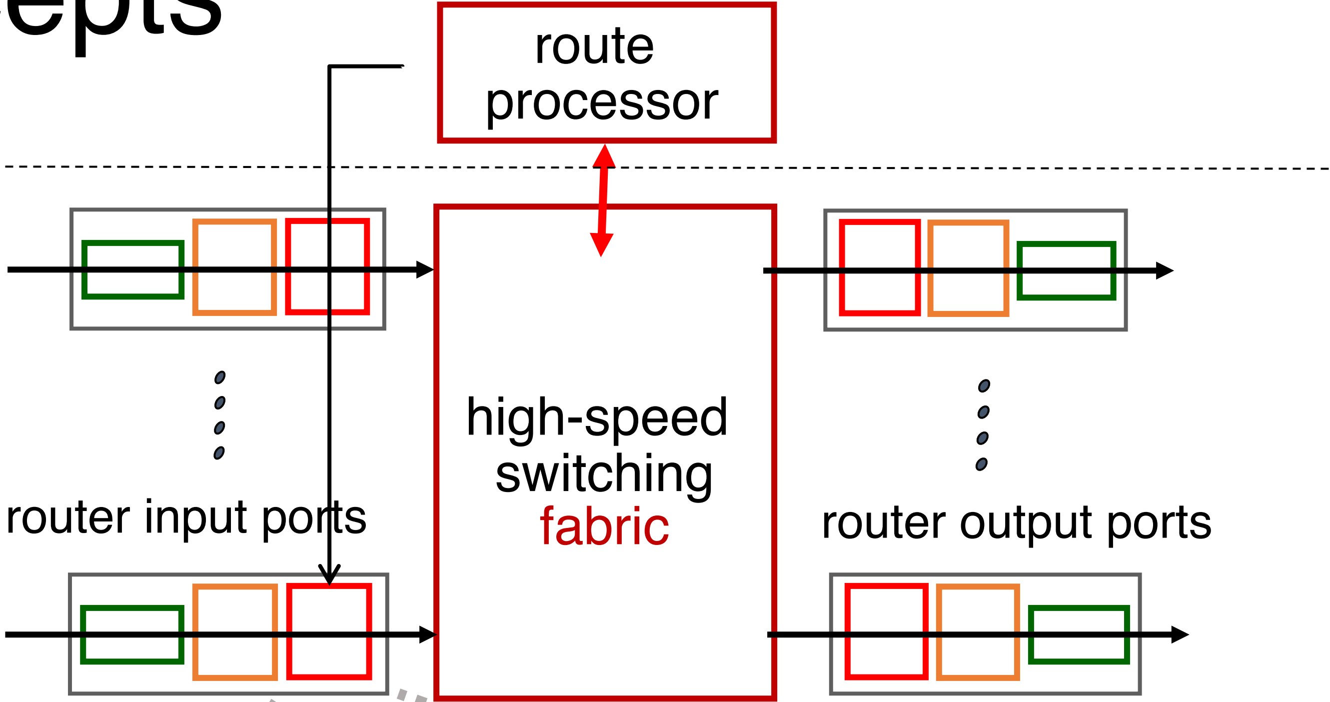
http://www.cs.rutgers.edu/~sn624/352-S22

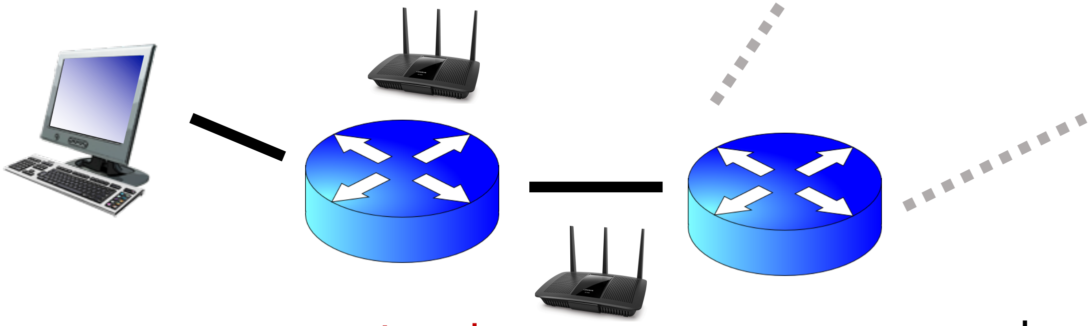
Srinivas Narayana

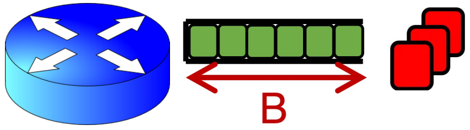
RUTGERS
UNIVERSITY | NEW BRUNSWICK

Quick recap of concepts









network part | host part

11001000 00010111 00010000 00000000

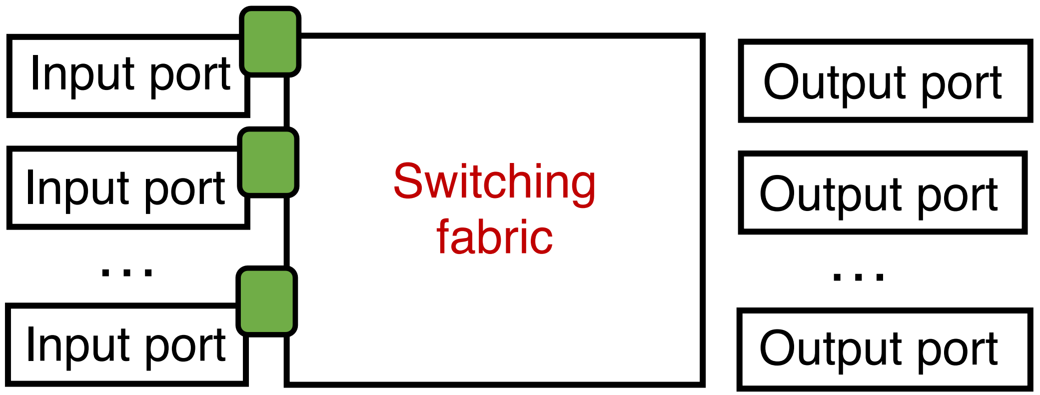
200.23.16.0/23

Forwarding Table

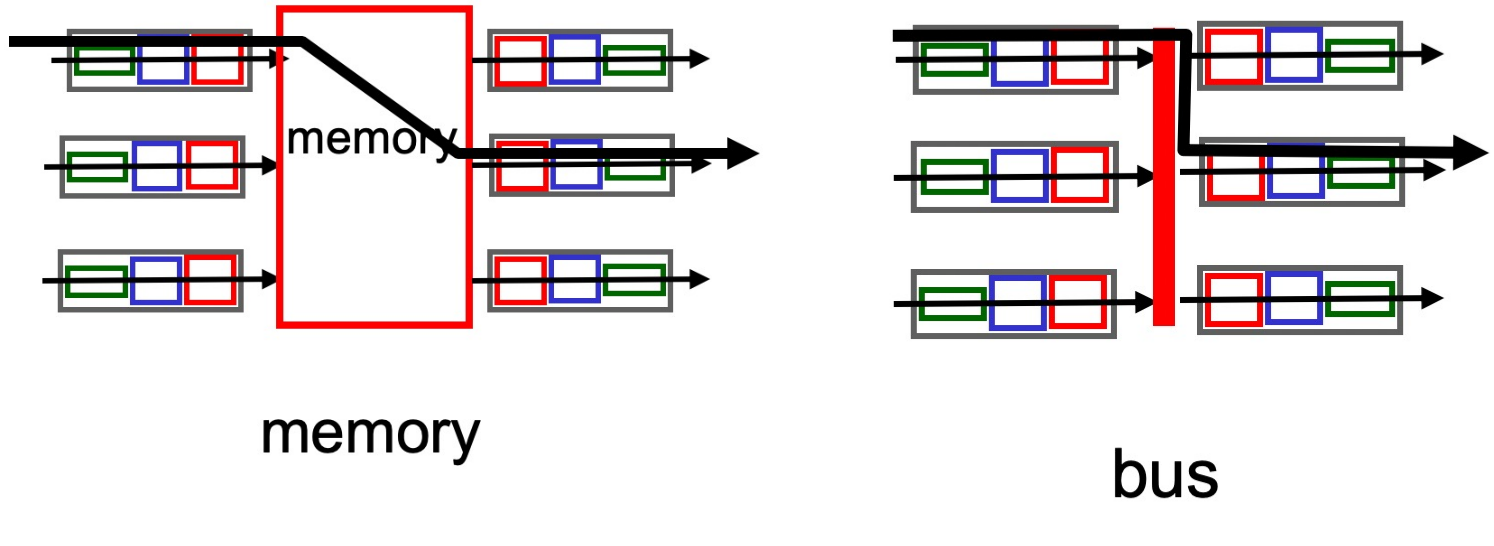
Dst-network	Port
65.0.0.0/8	3
128.9.0.0/16	1
⋮	⋮
149.12.0.0/19	7

The Internet uses destination IP based forwarding.

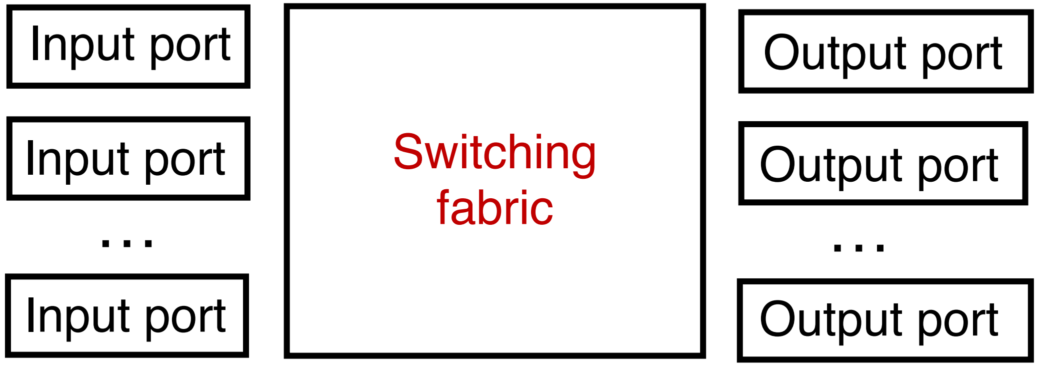
Review: Fabrics



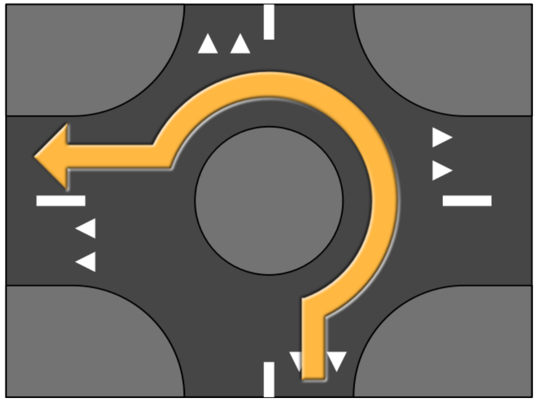
Fabric goal: Ferry as many packets as possible from input to output ports as quickly as possible.



Nonblocking fabrics



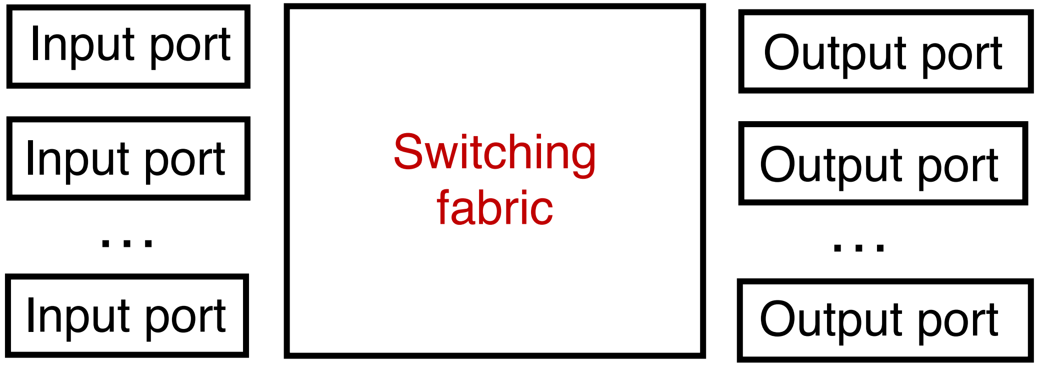
- High-speed switching fabrics designed to be nonblocking:
  - If an output port is “available”, an input port can always transmit to it without being blocked by the switching fabric itself



- Crossbars are nonblocking by design

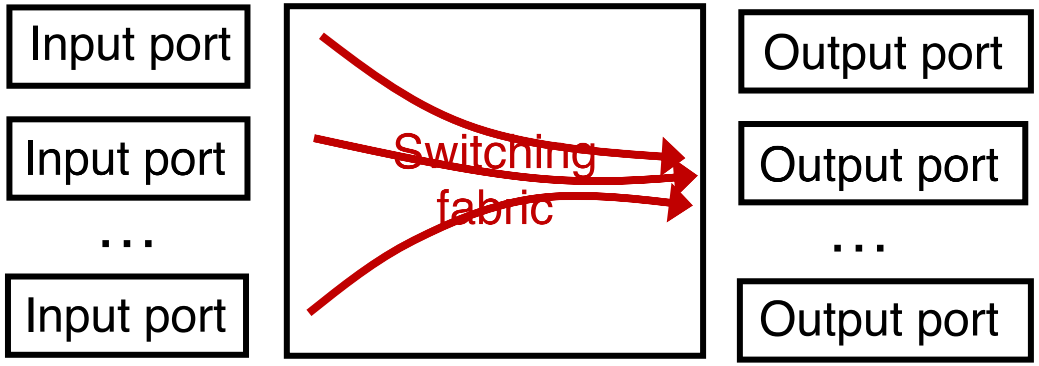
- Shared memory can be designed to be nonblocking if memory is optimized to be fast enough

Nonblocking fabrics



- With a nonblocking fabric, queues aren't formed due to the switching fabric.
- With a nonblocking fabric, there are no queues due to inefficiencies at the input port or the switching fabric
- Queues only form due to contention for the output port
  - Fundamental, unavoidable, given the route

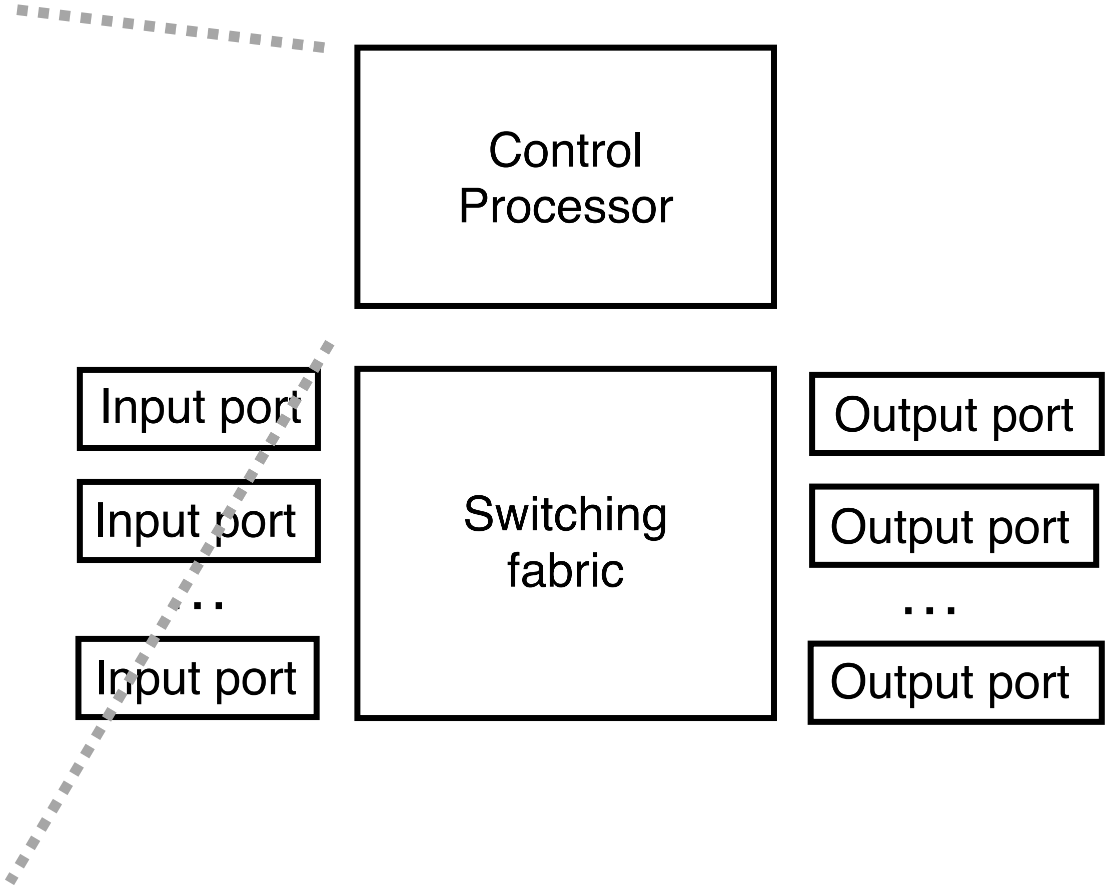
Nonblocking fabrics



- With a nonblocking fabric, queues aren't formed due to the switching fabric.
- With a nonblocking fabric, there are no queues due to inefficiencies at the input port or the switching fabric
- Queues only form due to contention for the output port
  - Fundamental, unavoidable, given the route
- Typically, these queues form on the output side
  - But can also "backpressure" to the input side if there is high contention for the output port
  - i.e.: can't move pkts to output Qs since buffers full, so buffer @ input

Control (plane) processor

- A general-purpose processor that "programs" the data plane:
  - Forwarding table
  - Scheduling and buffer management policy
- Implements the routing algorithm by processing routing protocol messages
  - Mechanism by which routers collectively solve the Internet routing problem
  - More on this soon.



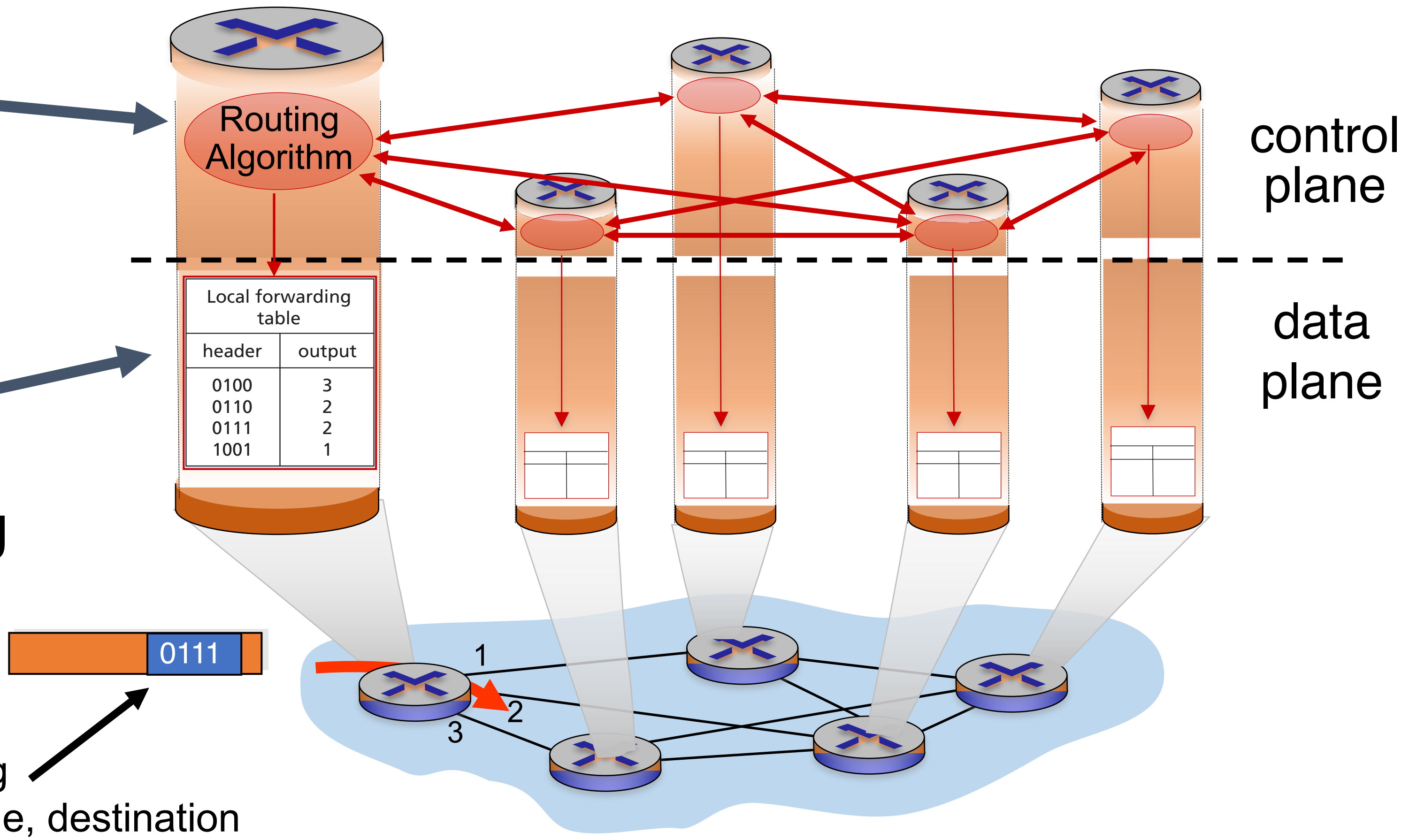
Router design: the bigger picture

Control plane

Traditional distributed routing: per route-change processing (~ a few tens of seconds)

Data plane

per-packet processing (~ tens of nanoseconds)

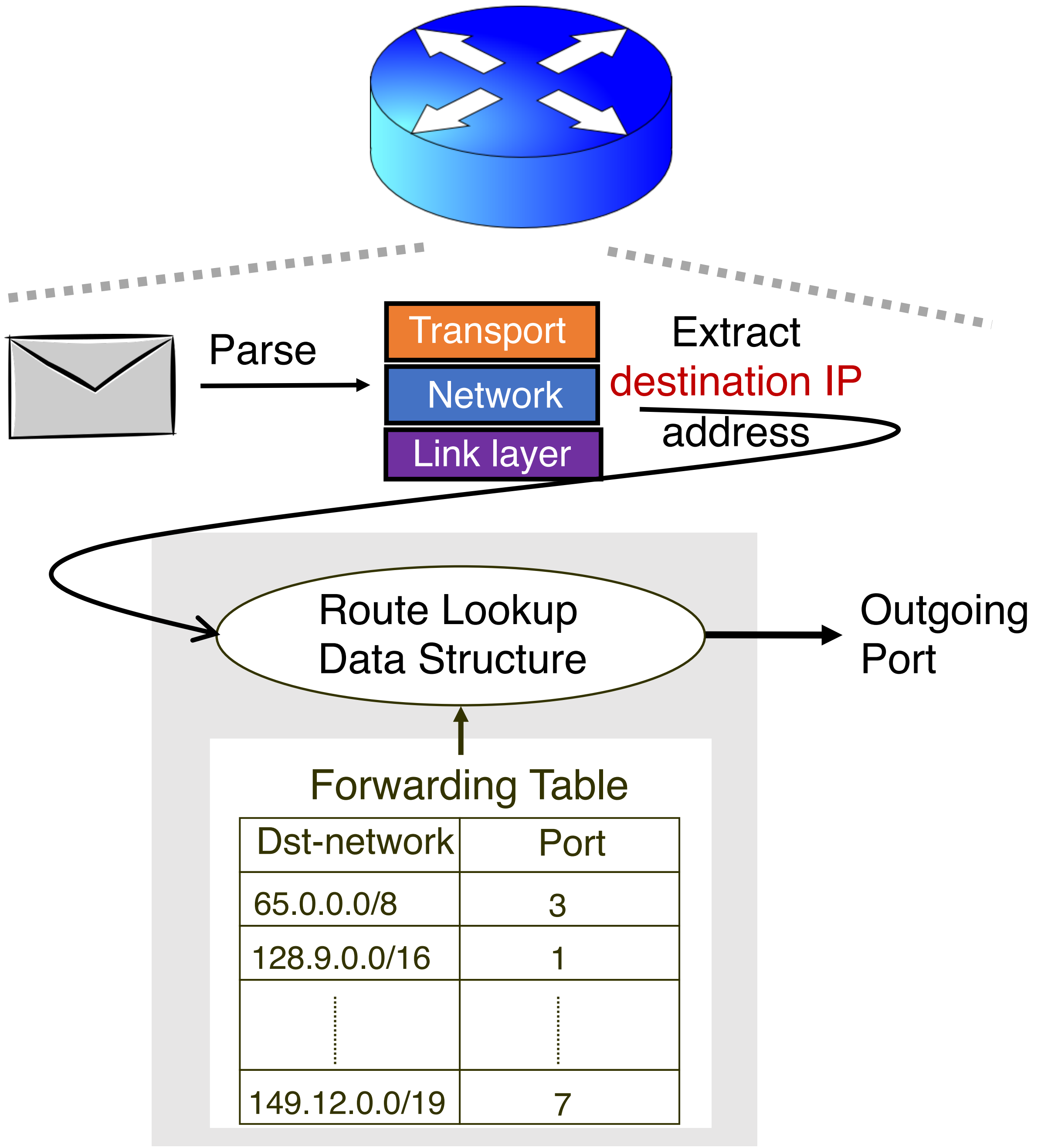


values in arriving packet header, i.e, destination IP address

Longest Prefix Matching

Review: Route lookup

- Table lookup matches a packet against an IP prefix
  - Ex: 65.12.45.2 matches 65.0.0.0/8
- Prefixes are allocated to organizations by Internet registries
- But organizations can reallocate a subset of their IP address allocation to other orgs

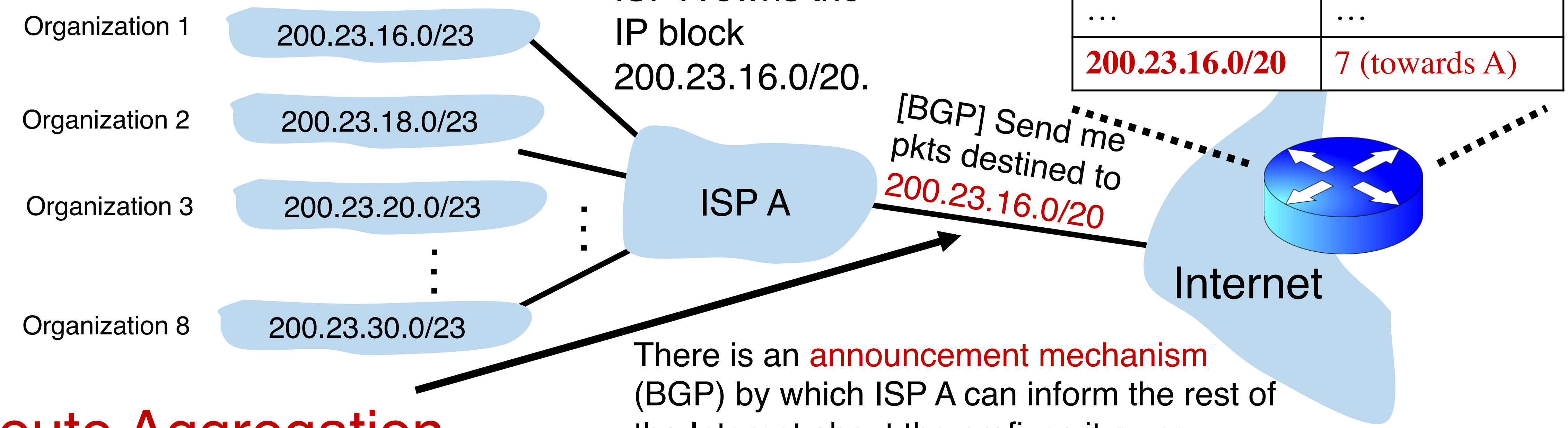


Example of IP block reallocation

Suppose ISP A reallocates a part of its IP block to orgs 1… 8

Dst IP Prefix	Output port
65.0.0.0/8	3
128.9.0.0/16	1
…	…
200.23.16.0/20	7 (towards A)

ISP A owns the IP block 200.23.16.0/20.



Route Aggregation

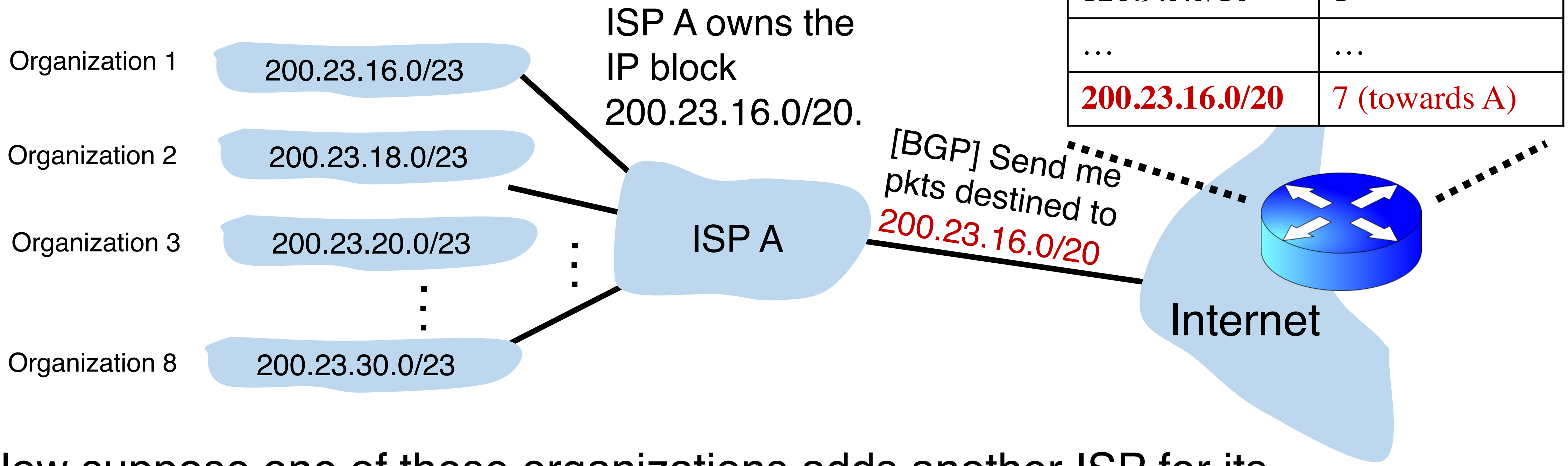
Save forwarding table memory
Fewer routing protocol msgs

There is an announcement mechanism (BGP) by which ISP A can inform the rest of the Internet about the prefixes it owns.
It is enough to announce a coarse-grained prefix 200.23.16.0/20 rather than 8 separate sub-prefixes.

Example of IP block reallocation

Suppose ISP A reallocates a part of its IP block to orgs 1… 8

Dst IP Prefix	Output port
65.0.0.0/8	3
128.9.0.0/16	1
…	…
200.23.16.0/20	7 (towards A)



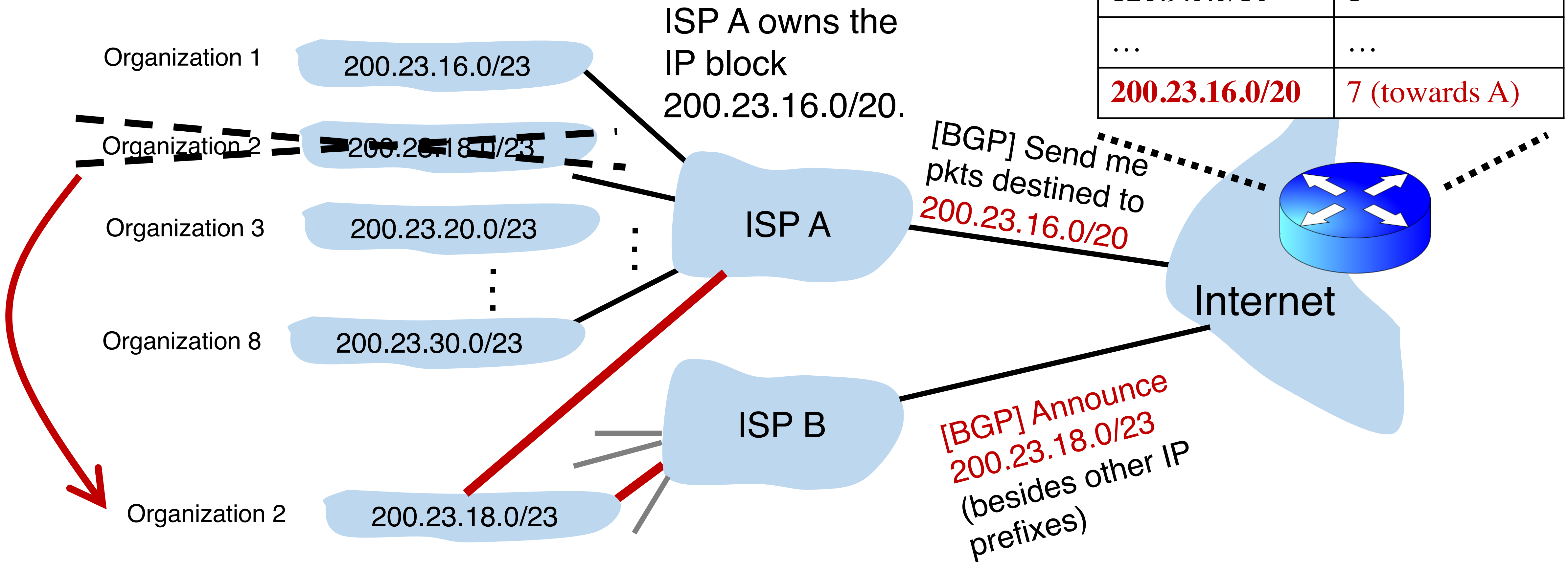
Now suppose one of these organizations adds another ISP for its Internet service and prefers using the new ISP.

Note: it's possible for the organization to retain its assigned IP block.

Example of IP block reallocation

Suppose ISP A reallocates a part of its IP block to orgs 1… 8

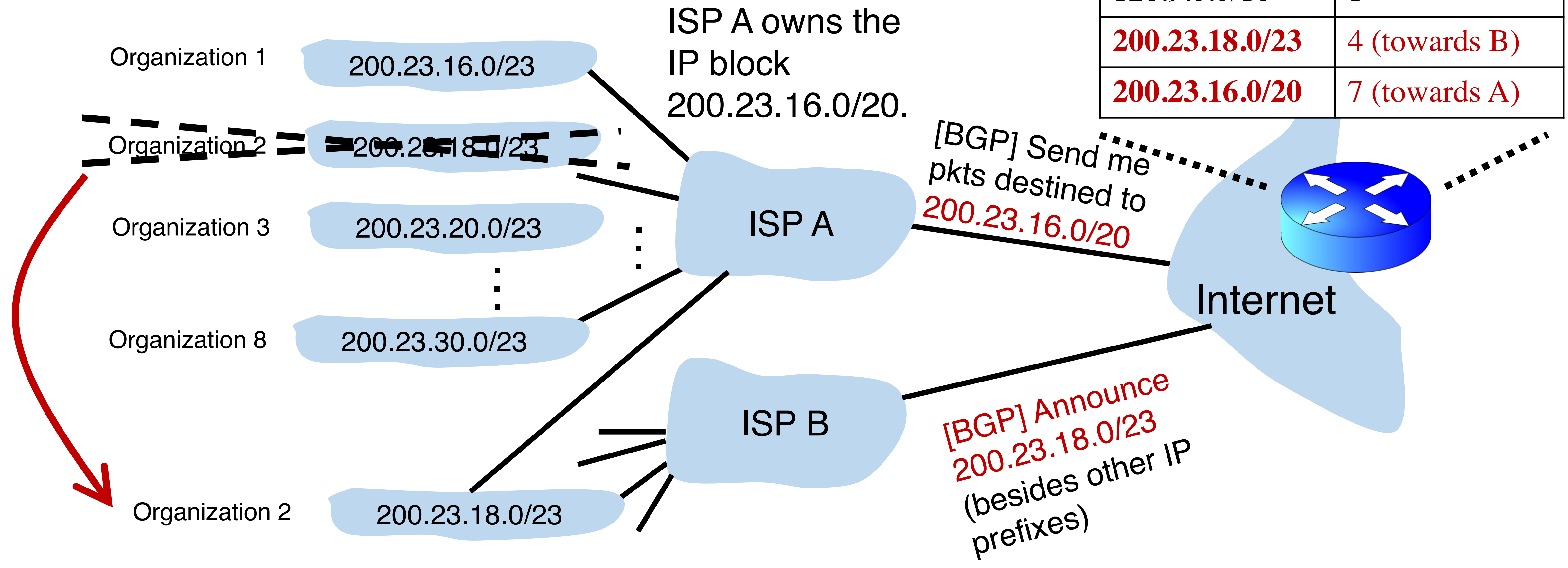
Dst IP Prefix	Output port
65.0.0.0/8	3
128.9.0.0/16	1
…	…
200.23.16.0/20	7 (towards A)



Example of IP block reallocation

Suppose ISP A reallocates a part of its IP block to orgs 1… 8

Dst IP Prefix	Output port
65.0.0.0/8	3
128.9.0.0/16	1
200.23.18.0/23	4 (towards B)
200.23.16.0/20	7 (towards A)



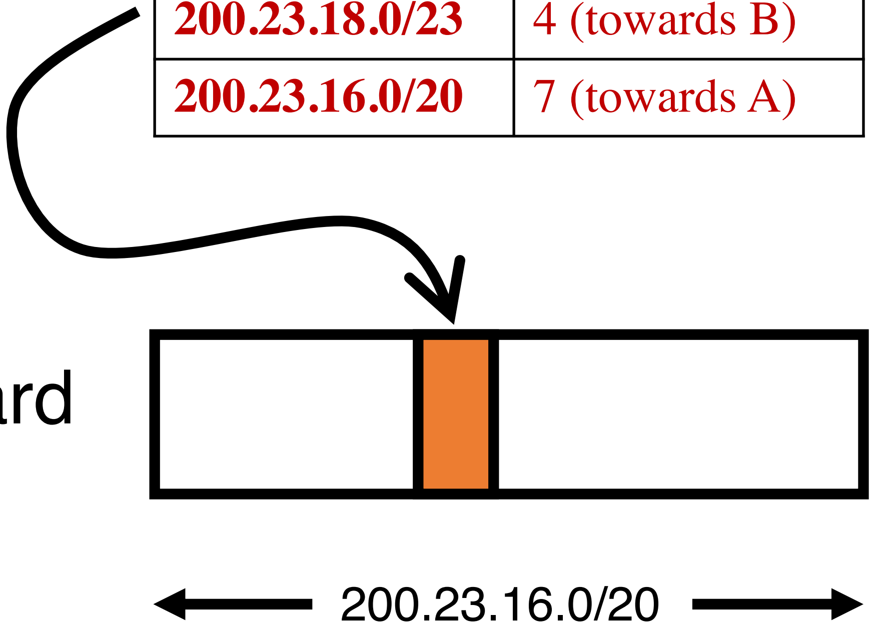
A closer look at the forwarding table

- 200.23.18.0/23 is inside 200.23.16.0/20

- A packet with destination IP address 200.23.18.xx is in both prefixes
  - i.e., both entries match

- Q: How should the router choose to forward the packet?
  - The org prefers B, so should choose B

Dst IP Prefix	Output port
65.0.0.0/8	3
128.9.0.0/16	1
200.23.18.0/23	4 (towards B)
200.23.16.0/20	7 (towards A)



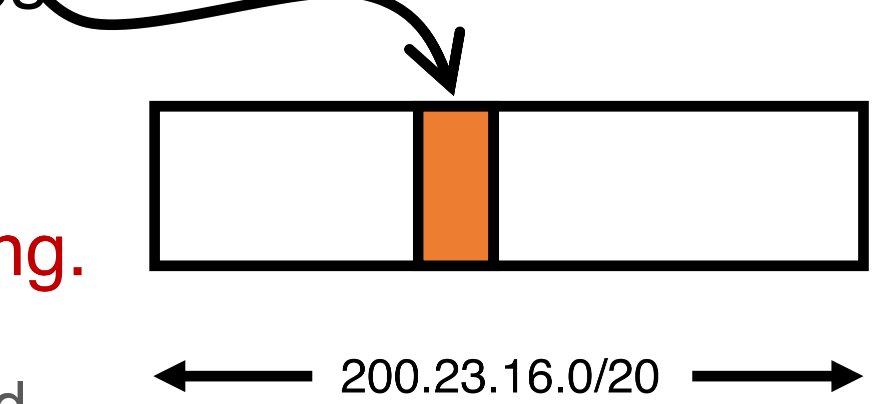
Longest Prefix Matching (LPM)

- Use the longest matching prefix, i.e., the most specific route, among all prefixes that match the packet.

- Policy borne out of the Internet's IP allocation model: prefixes and sub-prefixes are handed out

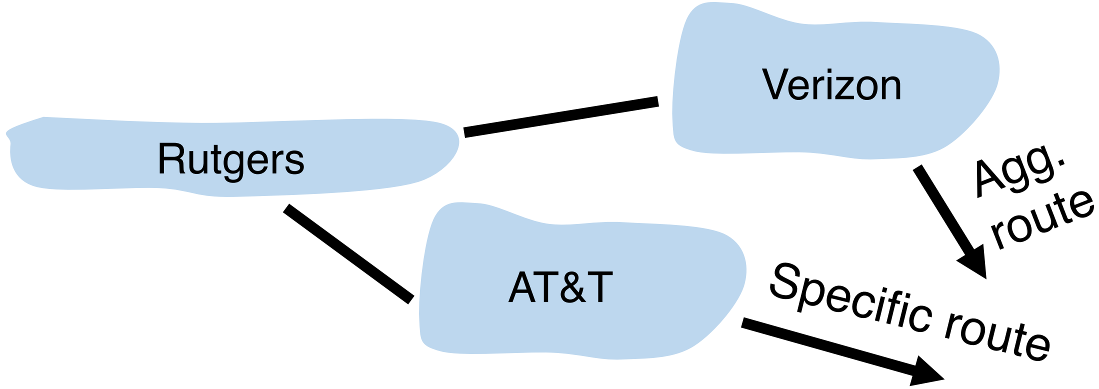
- Internet routers use longest prefix matching.
  - Very interesting algorithmic problems
  - Challenges in designing efficient software and hardware data structures

Dst IP Prefix	Output port
65.0.0.0/8	3
128.9.0.0/16	1
200.23.18.0/23	4 (towards B)
200.23.16.0/20	7 (towards A)



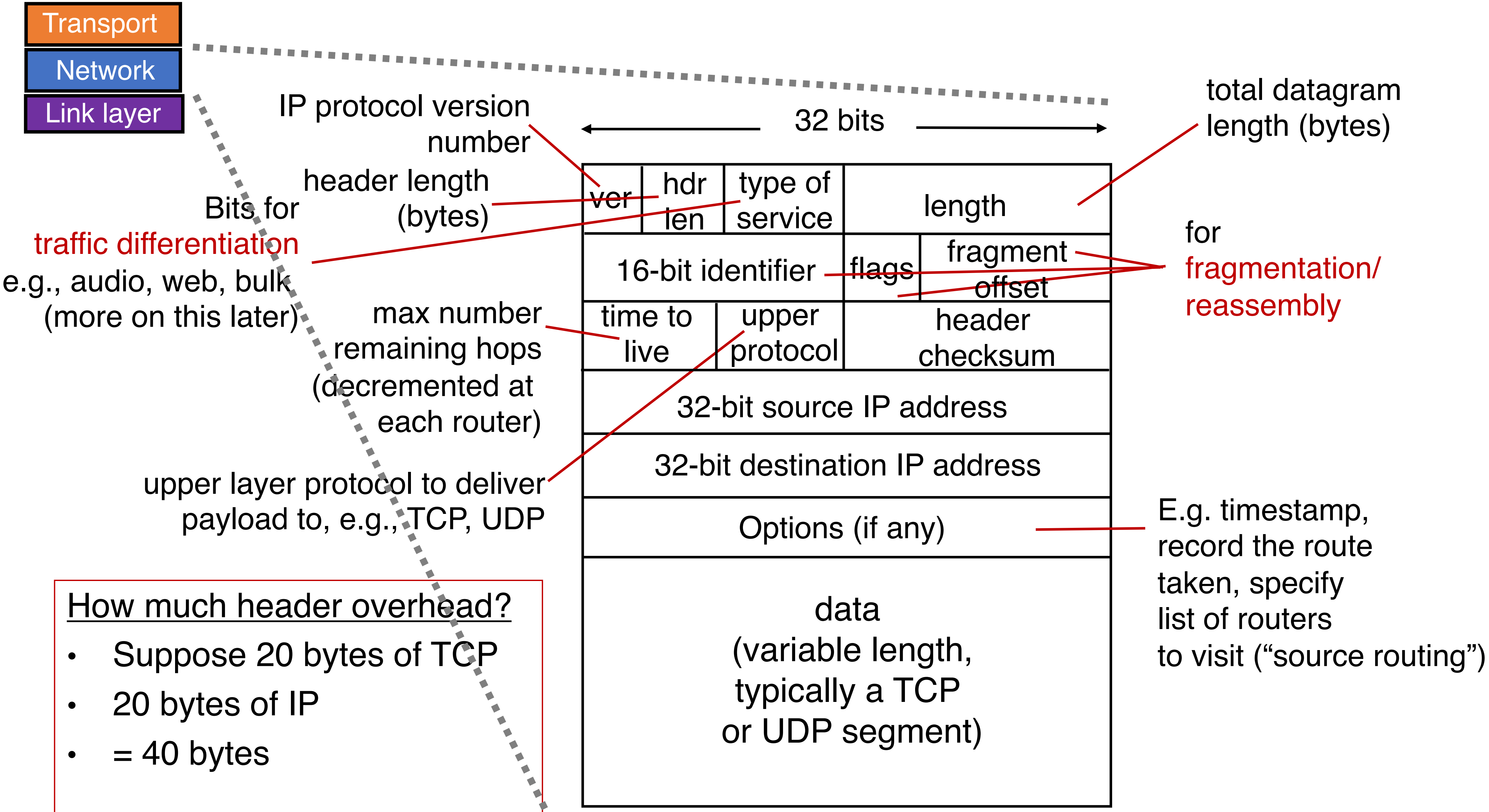
Internet routers perform longest-prefix matching on destination IP addresses of packets.

Why is LPM prevalent?



- An ISP (e.g., Verizon) has allocated a sub-prefix (or “subnet”) of a larger prefix that the ISP owns to an organization (e.g., Rutgers)
- Further, the ISP announces the aggregated prefix to the Internet to save on number of forwarding table memory and number of announcements
- The organization (e.g., Rutgers) is reachable over multiple paths (e.g., through another ISP like AT&T)
- The organization has a preference to use one path over another, and expresses this by announcing the longer (more specific) prefix
- Internet routers forward based on the longer prefix

IPv4 Datagram Format



<u>How much header overhead?</u>

- Suppose 20 bytes of TCP
- 20 bytes of IP
- = 40 bytes

The rest of this lecture and the next

- We'll talk about some support protocols and mechanisms for the network layer
  - Protocols: DHCP, ICMP, ARP
  - Mechanisms: NAT
  - We'll also talk about IP version 6 (IPv6)

- Some of these protocols use an IP header underneath their own header (ICMP) or replace the IP header with their own (ARP)
  - But these shouldn't be construed as transport/network protocols
  - They are fundamental to supporting IP/network layer functionality
  - More appropriately discussed as support protocols for the network layer

The network layer is all about reachability. Every protocol we'll see solves a sub-problem.

How does an endpoint get an address?

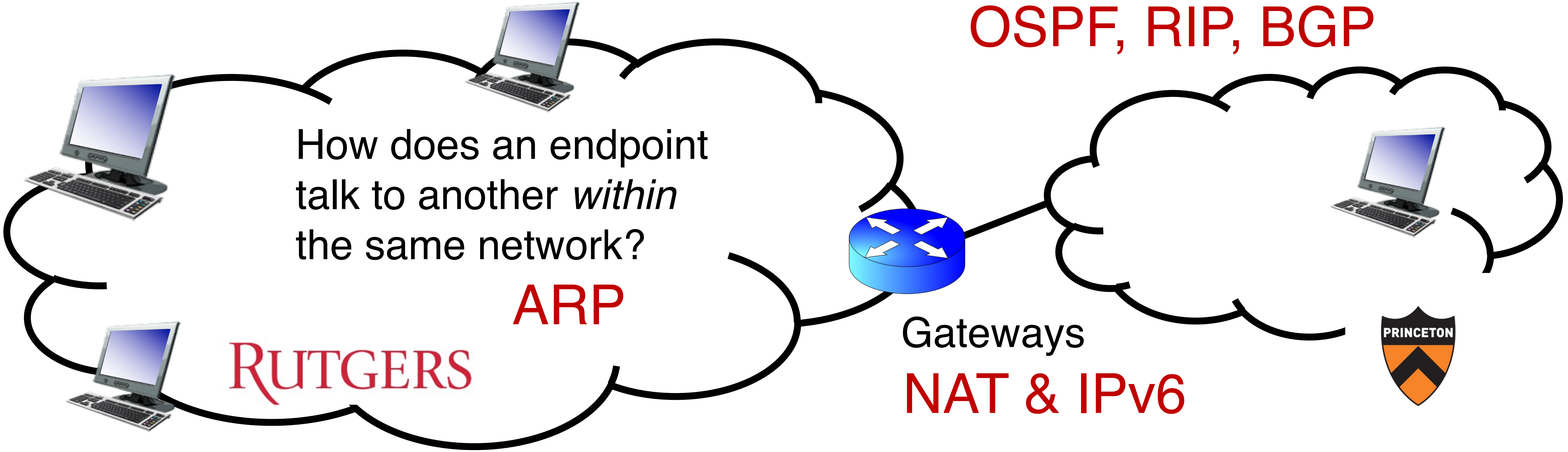
DHCP

Debugging?

ICMP

How does an endpoint talk to another *outside* its network?

Routing protocols
OSPF, RIP, BGP



Dynamic Host Configuration Protocol (DHCP)

How does an endpoint get its IP addr?

- One possibility: hard-code the IP address on the endpoint
    - e.g., a system admin writing addresses in a file
    - Linux: /etc/network/interfaces
    - Mac OS X (10.14.6): system preferences > Network > name of interface > advanced > TCP/IP > “Manually”

- Another possibility: dynamically receive an address “from the network”
    - DHCP: Dynamic Host Configuration Protocol
    - Provide plug-and-play functionality for endpoints (e.g., phones, laptops)

Many similar bootstrapping problems

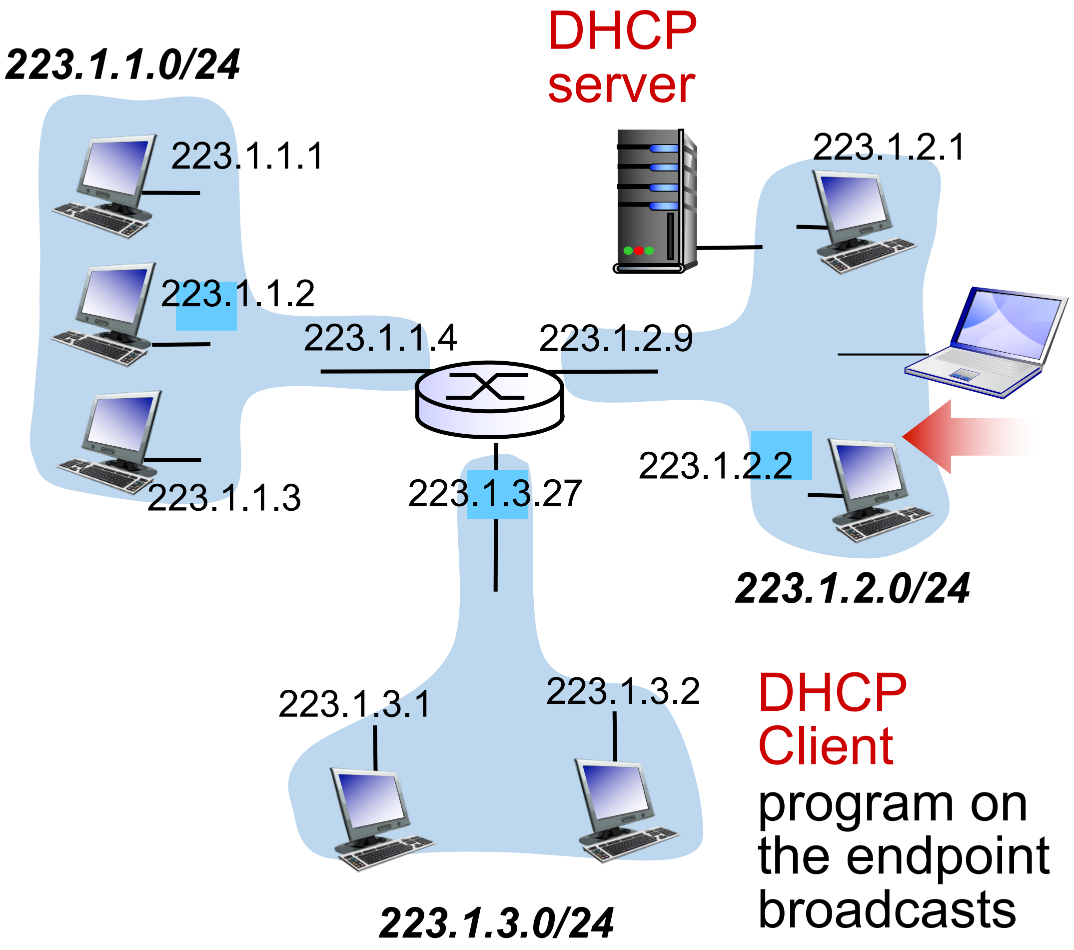
- How does a host get its IP address?
- How does a host know its local DNS server?
- How does a host know its netmask?
  - i.e., so that it can know which other hosts are in the same network
  - Note: the details how A and B talk to each other changes significantly when A and B are in the same network vs. different network
- How does a host know how to reach other networks?
  - i.e., which router is at the "border" of the current network?
  - This router is also called the gateway router: crucial for an endpoint to communicate with another endpoint external to the network

How DHCP works

- An endpoint that just joined a network knows nothing about it
  - Endpoint doesn’t even have an IP address for its point of attachment
- We solved a similar bootstrapping problem before:
  - Domain Name Service (DNS) to retrieve addresses
- Often, it makes little sense to have the endpoint contact a “known” server to receive an IP address
  - E.g., connecting to a brand-new network you’ve never been in
- The only idea that really works is to ask everyone
  - Broadcast a query

How DHCP works

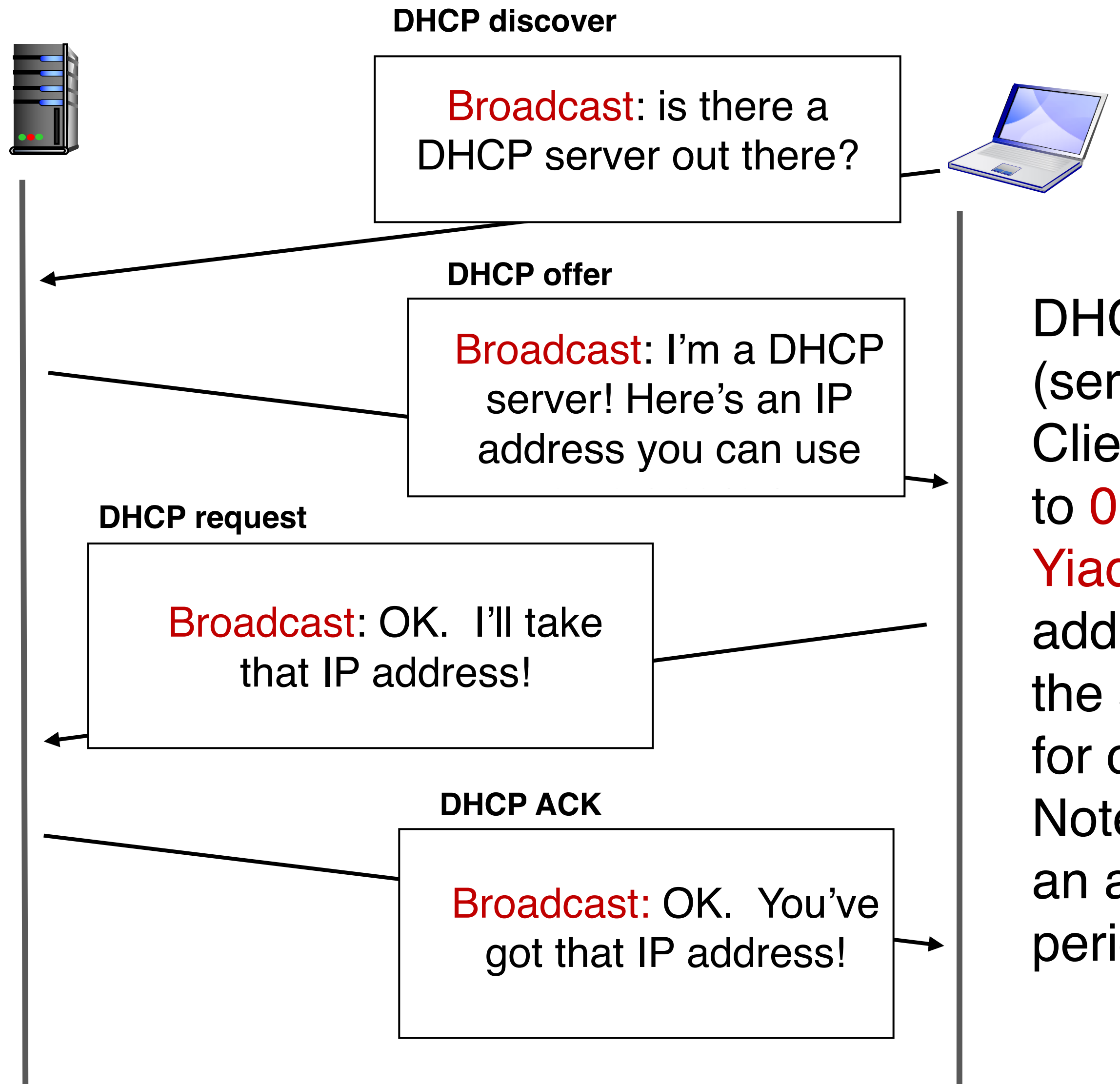
- DHCP allows a host to dynamically obtain its IP address from a server on a network when it joins the network
- DHCP can allow a host to be mobile across different networks, obtaining IP addresses as needed
- DHCP uses leases on addresses
  - Host must renew lease periodically
  - Allows network to reuse an IP with an expired lease, reclaiming addresses from inactive hosts



DHCP Client program on the endpoint broadcasts message requesting an IP address.

DHCP client-server scenario

DHCP server: 223.1.2.5



223.1.2.4

Arriving client

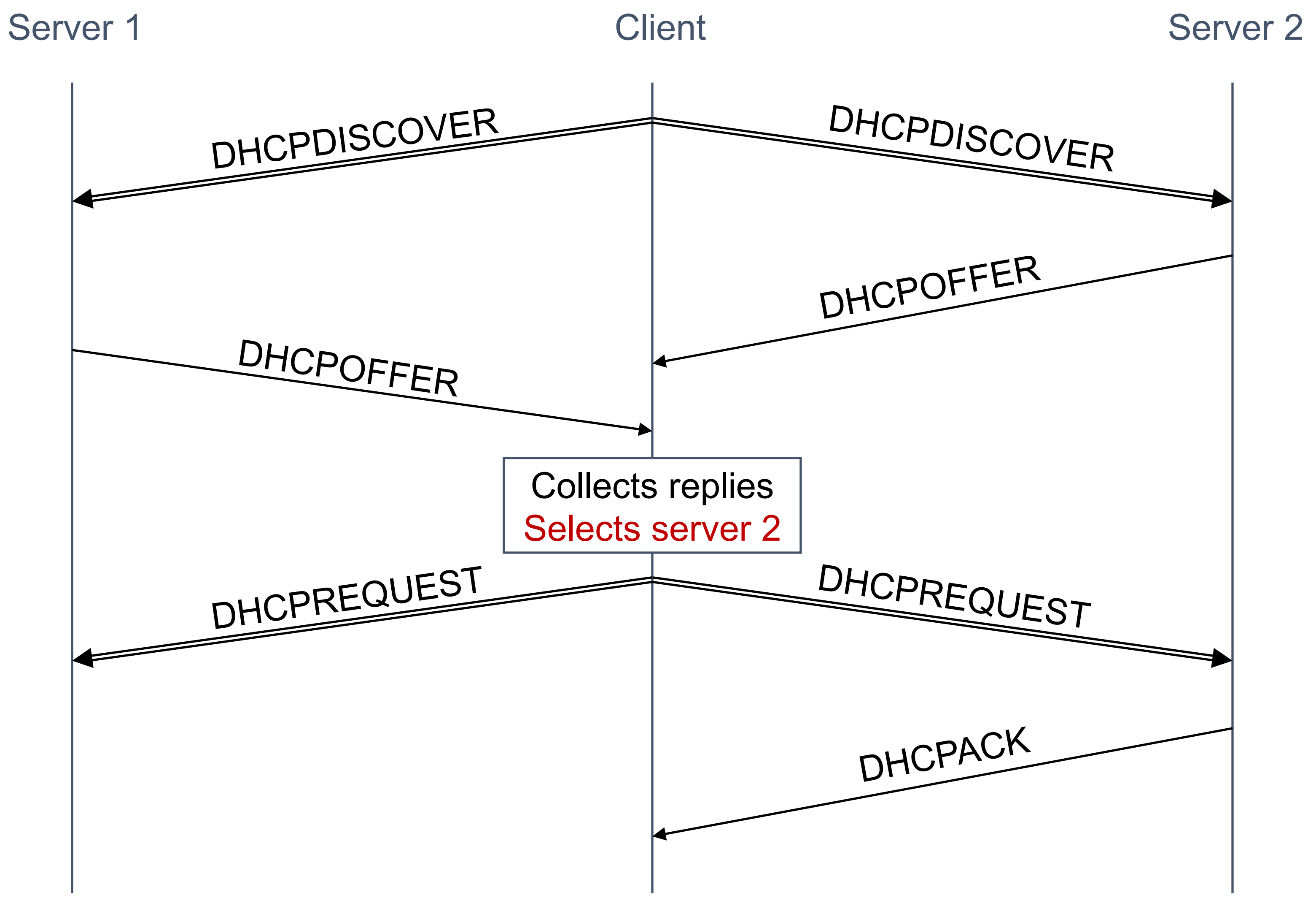
DHCP runs on UDP ports 67 (server) and 68 (client)

Client's initial IP address is set to 0.0.0.0

Yiaddr stands for "your IP address" – an address value the server sends to the client for consideration

Note that the IP allocation has an associated lifetime (lease period)

Multiple DHCP servers can coexist



DHCP returns more than an IP address

- Name and IP address of the local DNS server
- Netmask of the IP network the host is on
  - Useful to know whether another endpoint is inside or outside the current IP network
- Address of the gateway router to enable the endpoint to reach other IP networks

Your home router runs DHCP

- Likely, your home devices (laptops, tablets, phones) are all using DHCP-assigned IP addresses

- The DHCP server is running on the control processor of your home's access router (e.g., WiFi router)

- You can access the DHCP client program on Linux using the command `dhclient` and on Linux using `sudo ipconfig <interface> DHCP`

Summary of DHCP

- Want endpoints to have plug and play functionality
  - Avoid tedious manual configuration of IP addresses and other information
- DHCP: a general bootstrapping mechanism for critical information required for network layer functionality
- Hosts can be simple: receive information from DHCP servers by broadcasting over the network