Networking

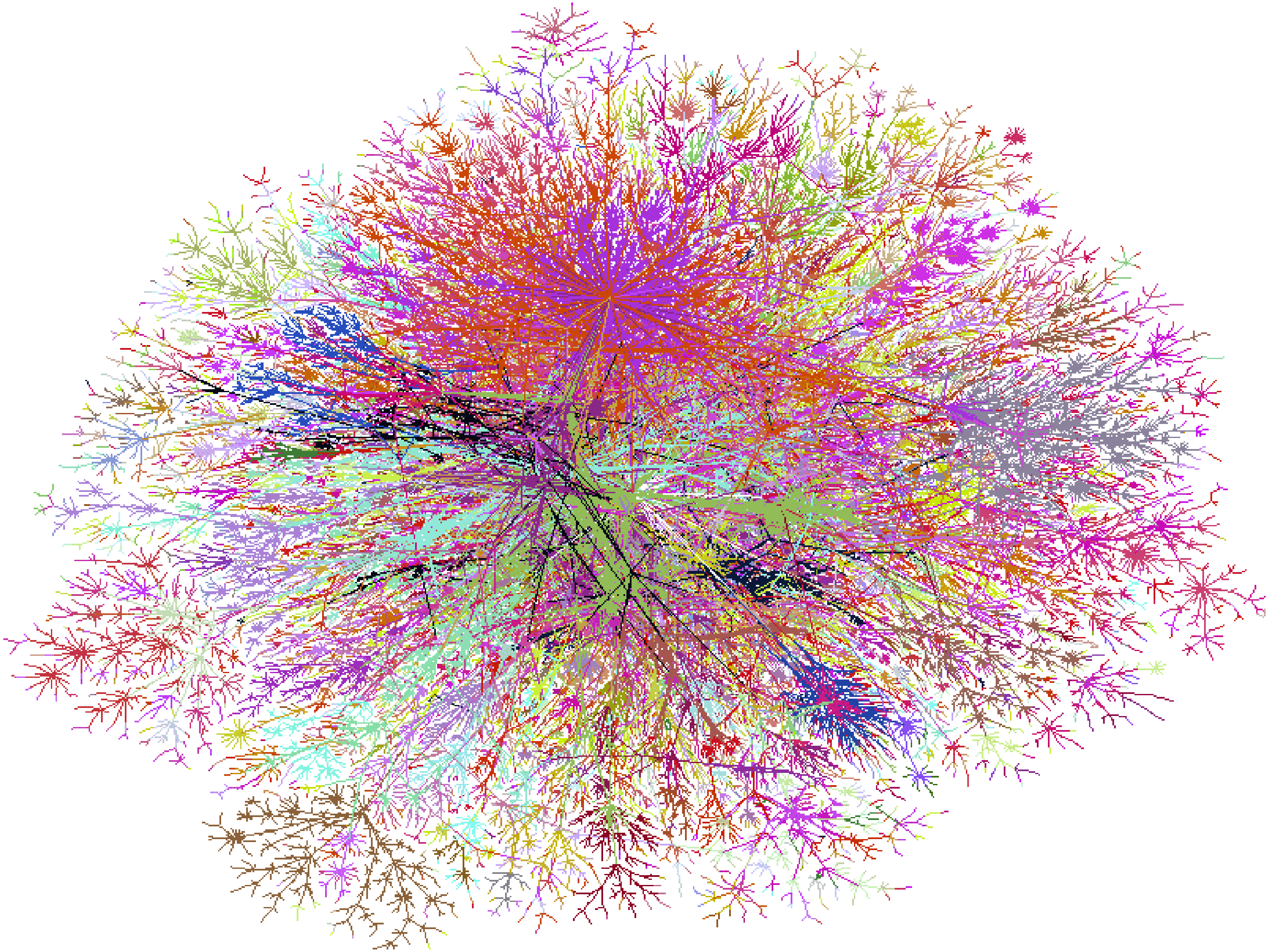
CS 4410
Operating Systems

Cornell CIS
COMPUTING AND INFORMATION SCIENCE

[R. Agarwal, L. Alvisi, A. Bracy, M. George, Kurose, Ross, E. Sirer, R. Van Renesse]

Routing

The Internet is Big…



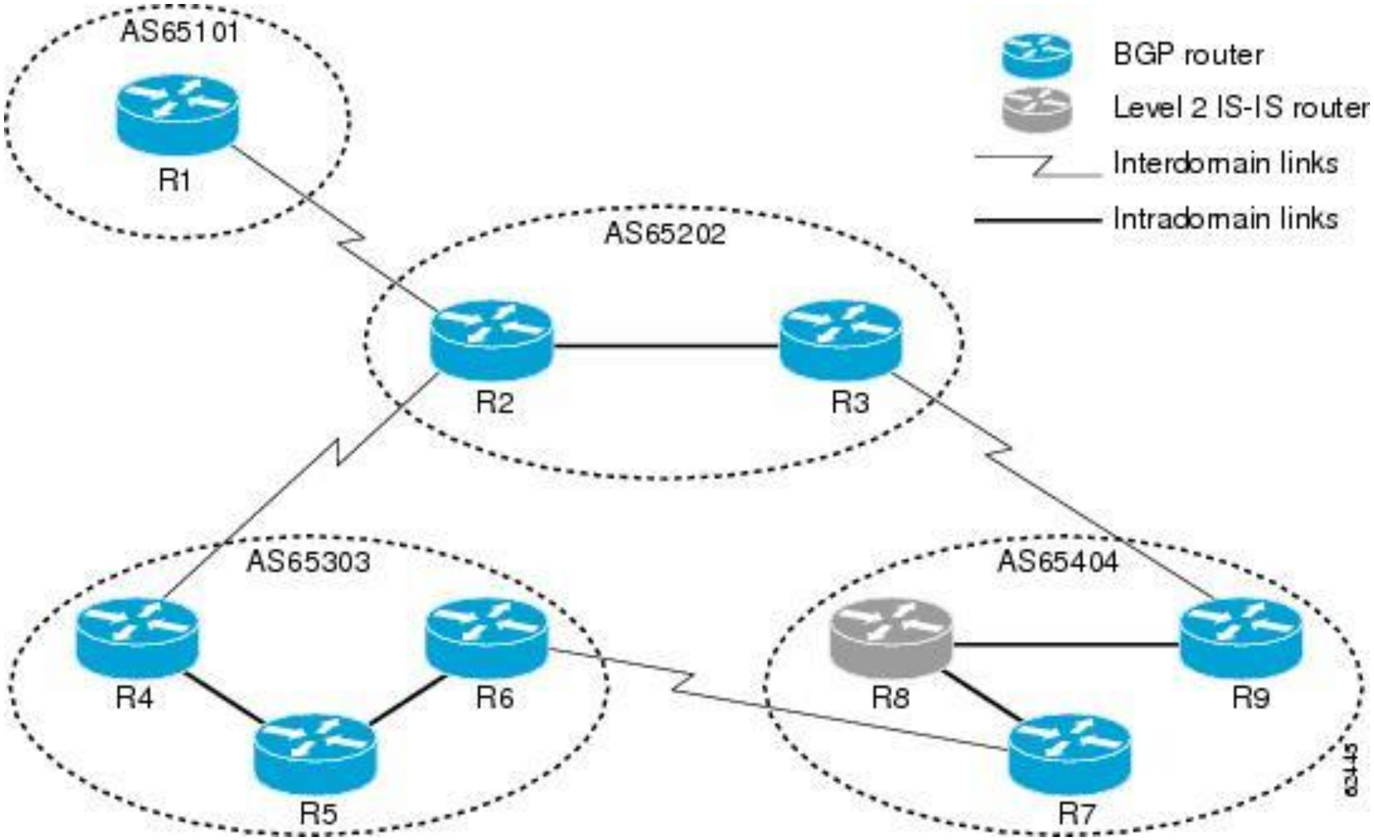
Routing

- How do we route messages from one machine to another?
- Subject to
  - churn
  - efficiency
  - reliability
  - economical considerations
  - political considerations

Internet Protocol (IP)

- The Internet is subdivided into disjoint Autonomous Systems (AS)

Graph of subgraphs



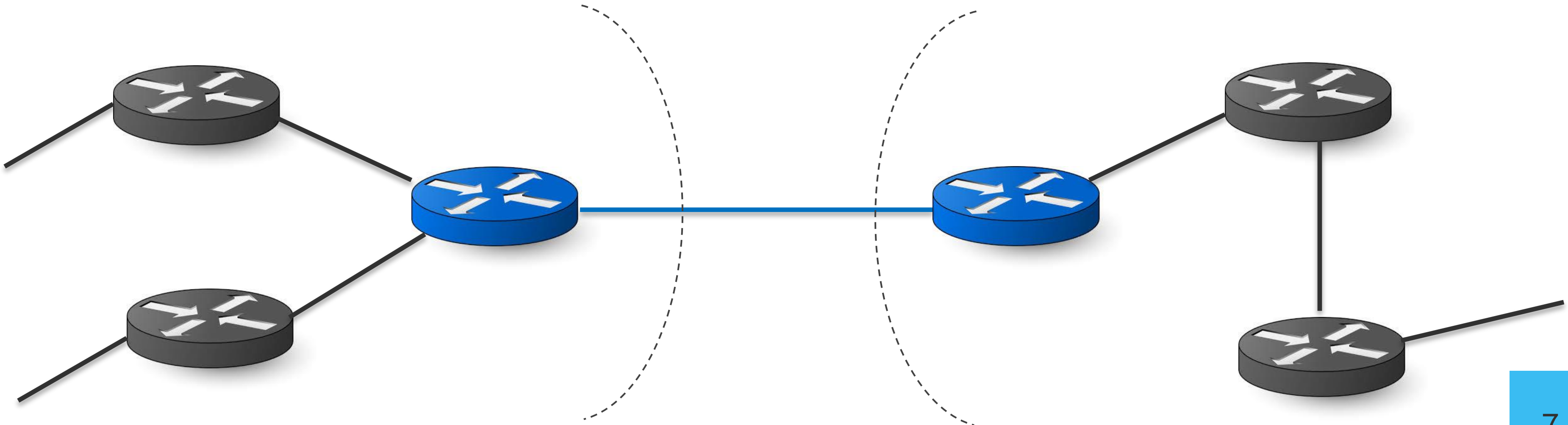
Autonomous Systems

- Each AS is a routing domain in its own right
  - has a private IP network
  - runs its own routing protocols
  - may have multiple IP subnets
    - each with their own IP prefix
  - has a unique “AS number”
- ASs are organized in a graph
  - routing between ASs using BGP (Border Gateway Protocol)

Thus routing is hierarchical!

Three steps:

1. A packet is first routed to an "edge router" at the source AS—using the internal routing protocol used by the source AS
2. Next the packet is routed to an edge router at the destination AS—determined by the destination address prefix—using BGP
3. The destination AS's edge router then forwards the packet to its ultimate destination—determined by the address suffix—using the internal routing protocol used by the destination AS



Internet Routing, observations

- There are no longer special “government” routers that route between ASs. Instead, each AS has one or more “edge routers” that are connected by interdomain links.
- Two types:
  - **Transit AS**: forwards packets coming from one AS to another AS
  - **Stub AS**: has only “upstream” links and does not do any forwarding

What’s an ISP?

- An ISP (Internet Service Provider) is simply an AS (or collection of ASs) that provides, to its customers (which may be people or other ASs), access to the “The Internet” (i.e. all the other ASs)
- Provides one or more PoPs (Points of Presence) for its customers.

Routers (Layer-3 Switches)

- Connects multiple LANs (subnets)
- Two classes:
  - Edge or Border router: Resides at the edge of an AS, and has two faces
    - one faces outside to connect to one or more per edge router in other ASs
    - one faces inside, connecting to zero or more other routers within the same AS
  - Interior router:
    - has no connections to routers in other ASs

Routing Table

- Maps IP address to interface or port and to MAC address
- Longest Prefix Matching
- Your laptop/phone has a routing table too!

Address/Mask	IF or Port	MAC
128.84.216/23	en0	c4:2c:03:28:a1:39
127/8	lo0	127.0.0.1
128.84.216.36/32	en0	74:ea:3a:ef:60:03
128.84.216.80/32	en0	20:aa:4b:38:03:24
128.84.217.255/32	en0	ff:ff:ff:ff:ff:ff
130.18/16	en1	c8:d4:58:1a:32:de

Prefix of address to match

Number of bits in prefix

Netmask: a "1" for each bit that matters
For /16, netmask is 255.255.0.0

Router Function

often implemented in hardware

```
for ever:
    receive IP packet p
    if isLocal(p.dest): return localDelivery(p)
    if --p.TTL == 0: return dropPacket(p)
    matches = { }
    for each entry e in routing table:
        if p.dest & e.netmask == e.address & e.netmask:
            matches.add(e)
    bestmatch = matches.maxarg(e.netmask)
    forward p to bestmatch.port/bestmatch.MAC
```

Destination: 128.84.216.33
Entry: 128.84.216.0/23
Netmask: 255.255.254.0

Dest & Netmask = 128.84.216.0
Entry & Netmask = 128.84.216.0

Routing Loops?

- Loop: Chain of routers forward to one another, packet never reaches destination
- In steady state, there should be no routing loops
- But steady state is rare. If routing tables are not in sync, routing loops can occur.
- To avoid problems, IP packets maintain a maximum hop count (TTL) that is decreased on every hop until 0 is reached, at which point a packet is dropped.

Constructing Routing Tables

- For end-hosts, mostly DHCP and ARP as discussed before
- For routers, use a “routing protocol”

Example: Distance Vector Routing

- State at each router: Distance to each destination, next hop to get there
- Routers periodically share tables with neighbors
- Upon receiving a table from router *r*, update local table:
  - If *r*'s table contains an entry for destination *d* with a shorter distance *n*
  - Then *d*'s next hop is now *r*, with distance *n+1*
- Build routing table from DV table

Distance Vector Routing

Router 1's table

Destination	Hops	Next Hop
128.84.216/23	4	Router 4
139.122/16	3	Router 3
138.43.121/24	1	Router 1

Destination	Hops	Next Hop
128.84.216/23	3	Router 2
139.122/16	3	Router 3
138.43.121/24	1	Router 1

Router 2's table

Destination	Hops	Next Hop
128.84.216/23	2	Router 6
139.122/16	4	Router 4
138.43.121/24	99	???

Destination	Hops	Next Hop
128.84.216/23	2	Router 6
139.122/16	4	Router 4
138.43.121/24	2	Router 1

Distance Vector Routing

Router 1's table

Destination	Hops	Next Hop
128.84.216/23	3	Router 2
139.122/16	3	Router 3
138.43.121/24	1	Router 1

Distance-Vector Table

Address	Port	MAC
128.84.216/23	1	c4:2c:03:28:a1:39
139.122/16	3	c4:2c:38:ef:10:bd
138.43.121.40	6	20:aa:4b:38:03:24
138.43.121.42	6	74:ea:3a:ef:60:03

Routing Table

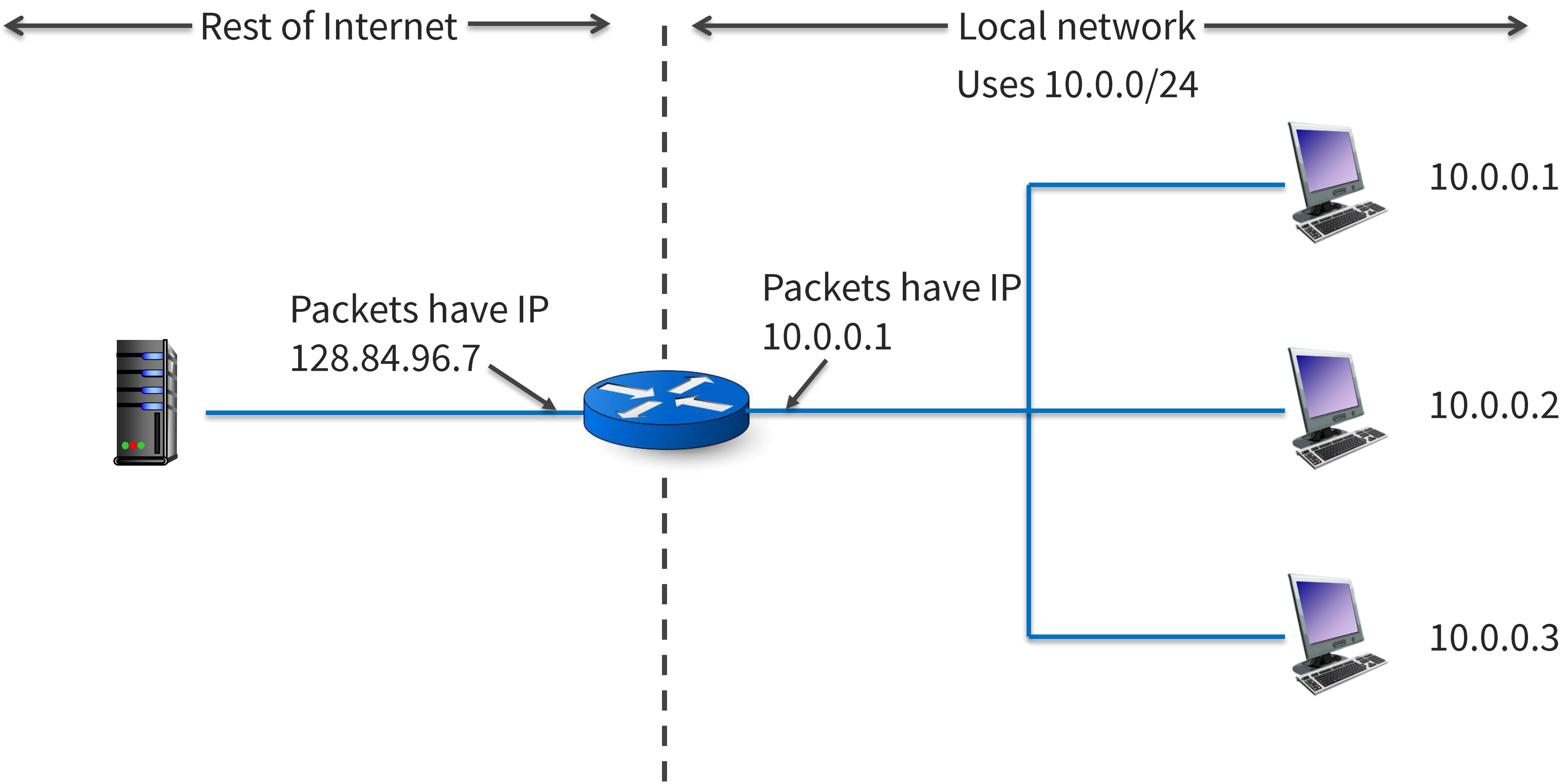
Network Address Translation

- IPv6 adoption is very slow, and IPv4 addresses have run out
- NAT allows entire sites to use a single globally routable IPv4 address for a collection of machines
  - exploits the sparsely used 16-bit TCP/UDP port number space
- A “NAT box” keeps a table that maps global TCP/IP addresses into local ones
- Overwrites the local source address with the globally addressable address

“Private” IP addresses

- The IPv4 addresses 10.x.x.x and 192.168.x.x are freely available for anybody to use
- Many machines have the IP address 192.168.0.100, for example

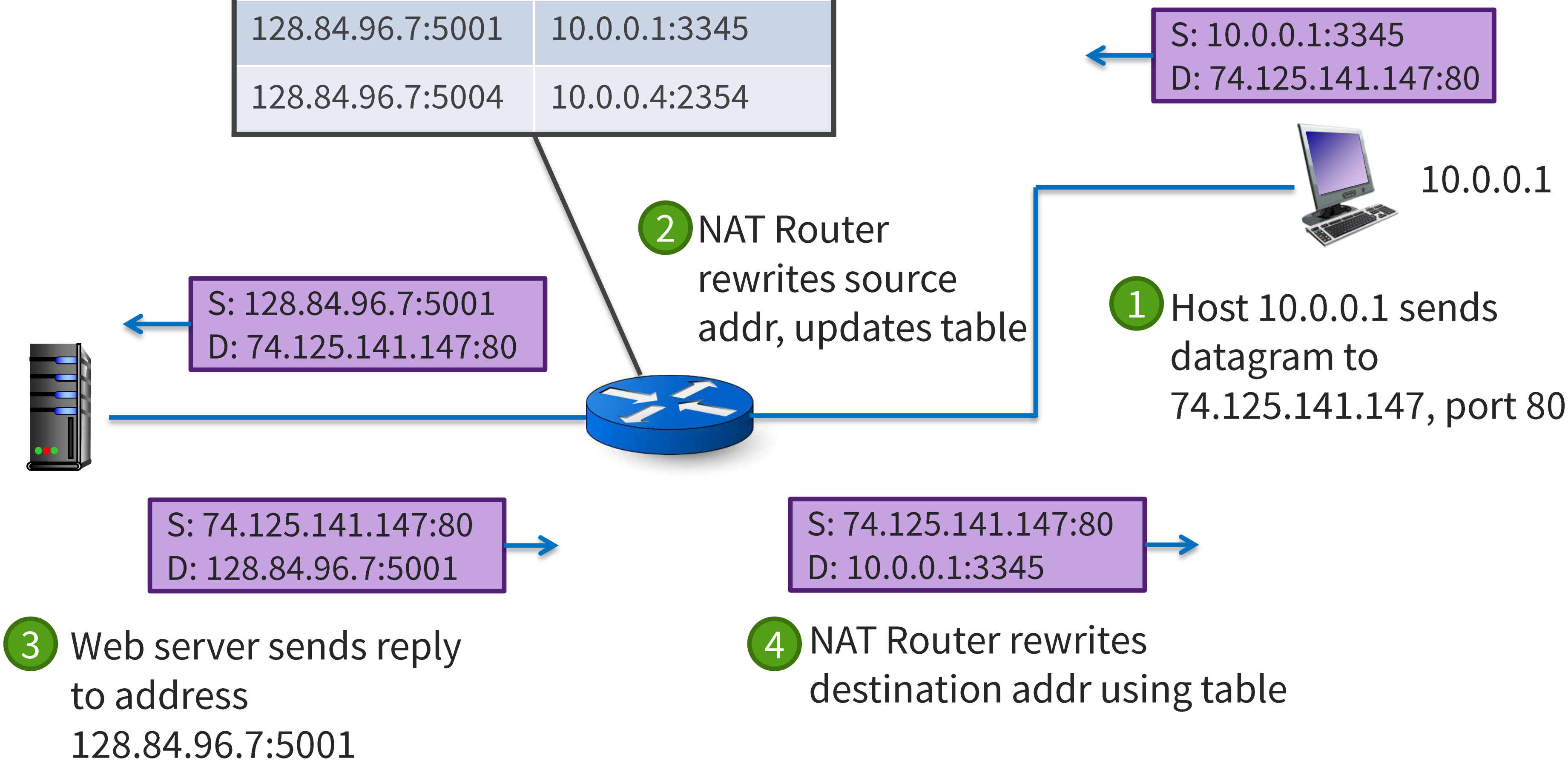
NAT Overview



How NAT Works

NAT Translation Table

WAN Address	LAN Address
128.84.96.7:5001	10.0.0.1:3345
128.84.96.7:5004	10.0.0.4:2354



Vice versa: punching holes or “game ports”

- When an external host tries to send a message to one of your machines in your house, it first arrives at the NAT box
  - Because you advertise your global IP address
- What if your machine wants to listen for incoming messages on a particular port?
  - e.g. port 80 for web, port 27015 for game traffic
- Your machine can request a specific port in the NAT table to be mapped to its IP address
- “Port forwarding”: Forward incoming traffic on port 80 to host 10.0.0.4 on port 80