Network Layer

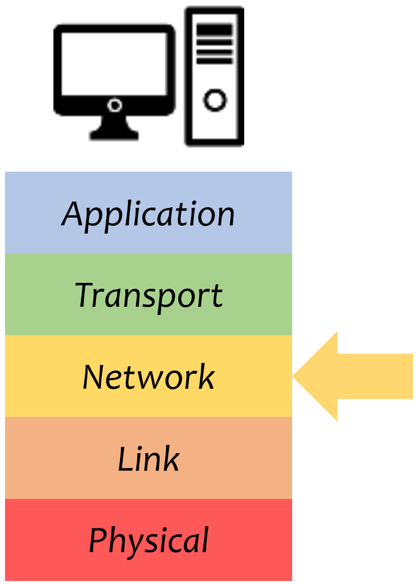
Prof. Anja Feldmann, Ph.D.

Dr. Oliver Gasser

INET @MPII

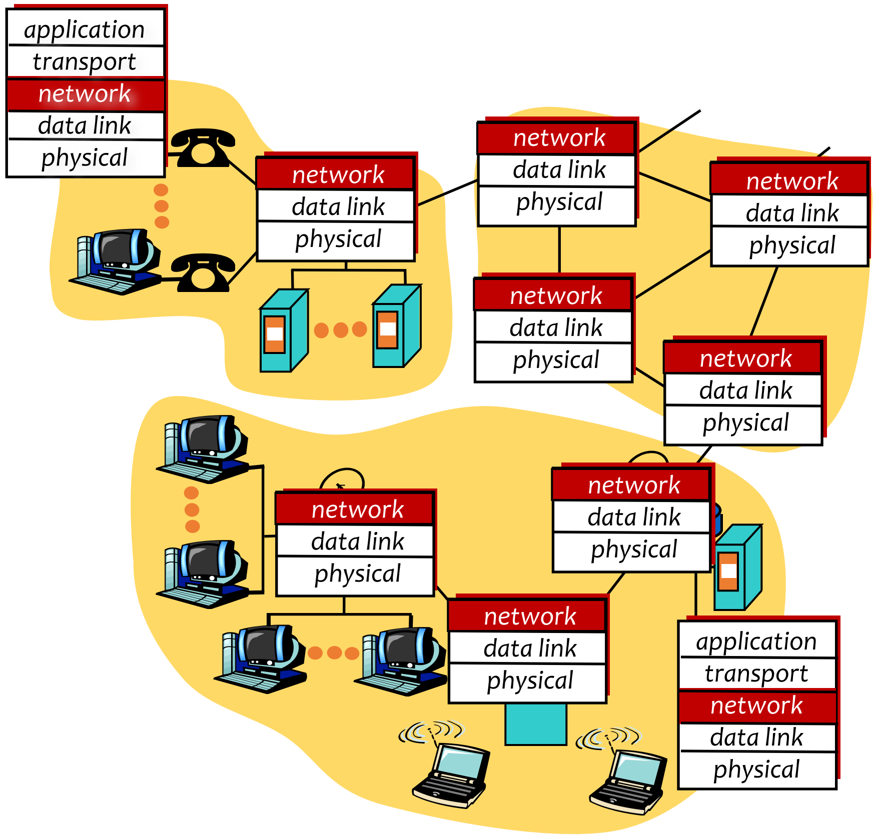
Hello, Network Layer!

Shifting our focus on to the next layer in the protocol hierarchy



Network Layer

- Transport packets from sending to receiving hosts
- Network layer protocols in every host, router



Three important functions:

- Addressing
- Path determination: Route taken by packets from source to destination → routing algorithms
- Switching/forwarding: Move packets from router's input to appropriate router output

IP Addressing

Prof. Anja Feldmann, Ph.D.

Dr. Oliver Gasser

Outline

- IPv4 and IPv6 addressing
- Routing and forwarding
- Network address translation

IP Interfaces

IP address:

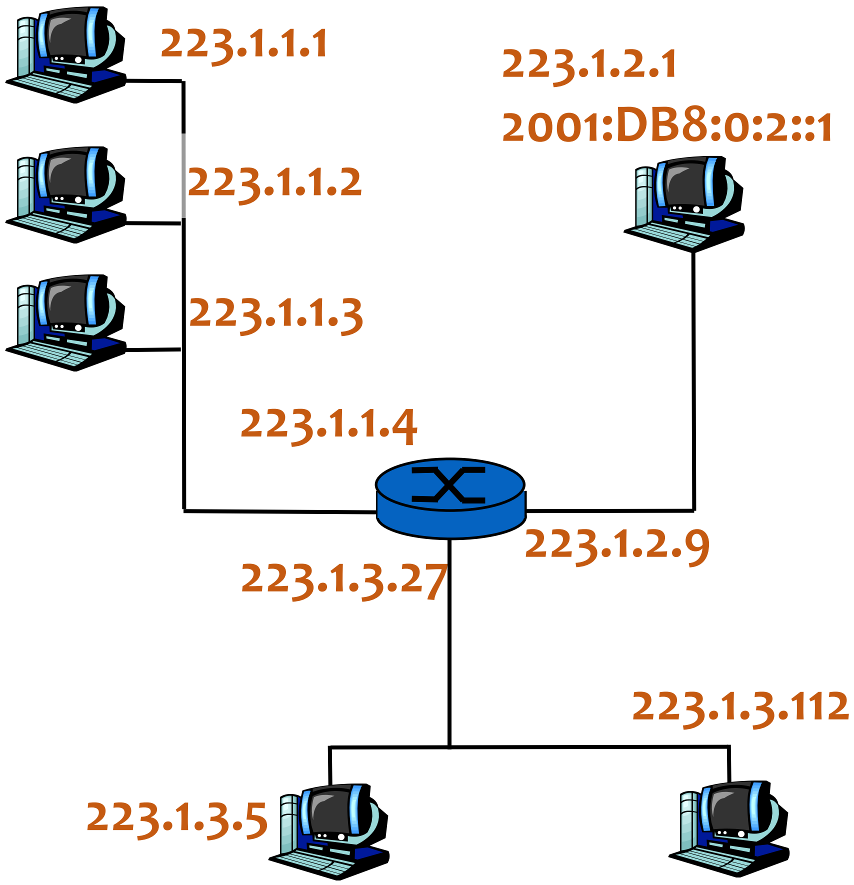
Identifier for host or router interface

- **IPv4:** 32 bits
- **IPv6:** 128 bits

Interface:

Connects a host or router to a physical link

- Routers typically have multiple interfaces
- Hosts may have multiple interfaces
- IP addresses are associated with interfaces, not hosts or routers



IPv4: Addressing

IP address:

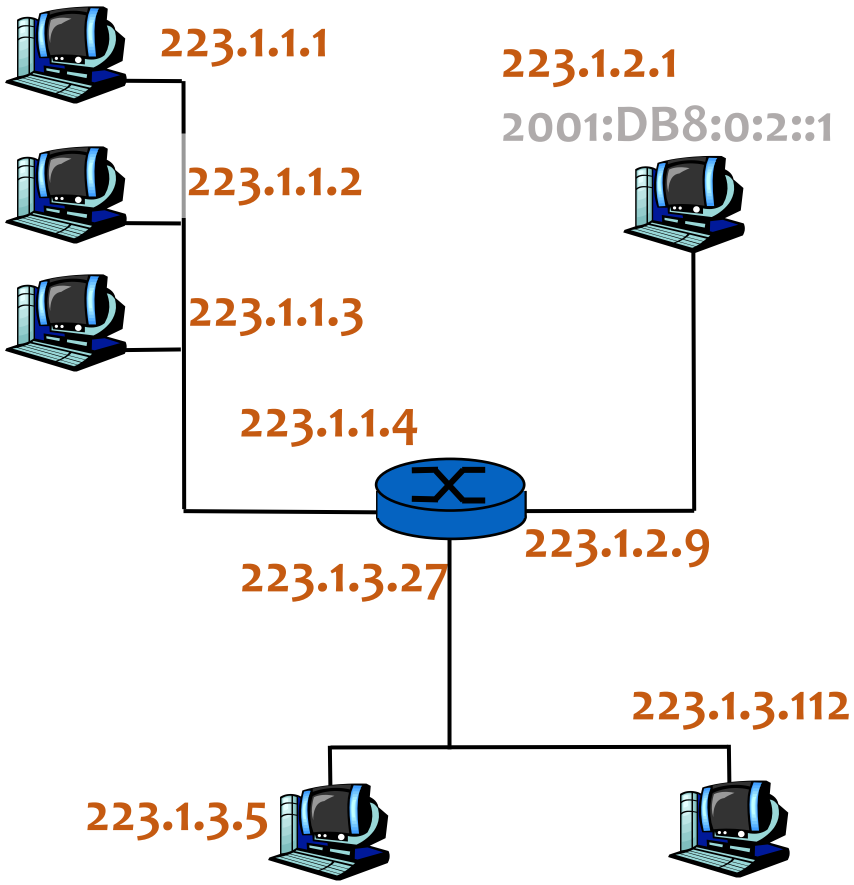
Identifier for host or router interface

IPv4 address:

- 32 bits
- Written as four 8-bit (octets) in decimal

223.1.1.1 = 11011111 00000001 00000001 00000001

11011111	00000001	00000001	00000001
223	1	1	1



IPv6: Motivation and History

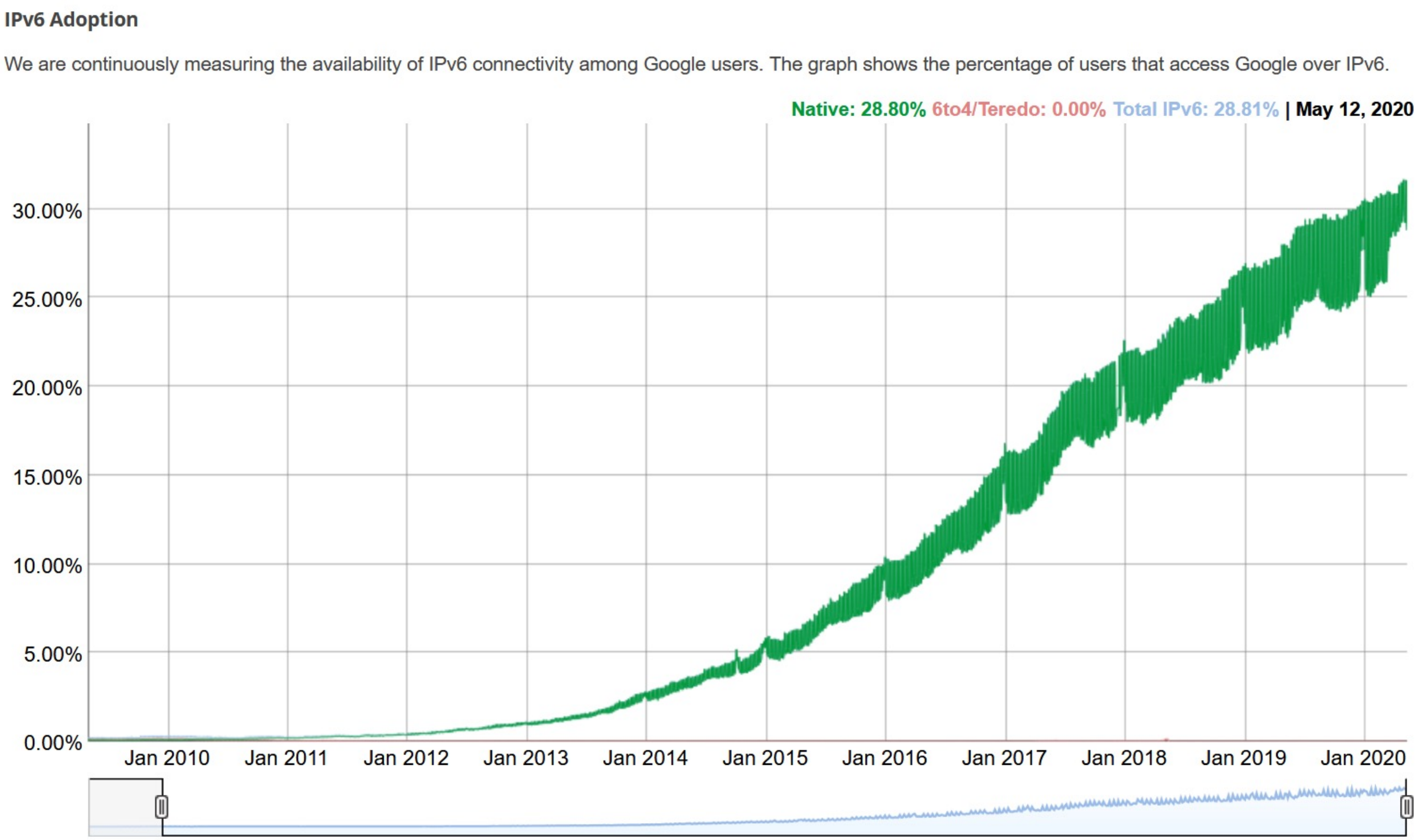
IPv4 address space is 32 bits

- Quite limited (32-bits? ~4 billion addresses)
- IPv4 was designed in the 1970s
- Some requirements changed!

Timeline:

- **1992** IETF begins discussion about IPv4 successor
- **1995** First IPv6 RFCs published
- **2000** 50% of IPv4 address space assigned
- **2007** All major OSes have IPv6 enabled by default
- **2011** IANA assigns last IPv4 block
  World IPv6 Day: Major sites test IPv6 for a day
- **2012** World IPv6 Launch Day: Major sites enable IPv6
- **2018** About 10% of all traffic is IPv6 at some Internet locations

Example: IPv6 Adoption seen by Google



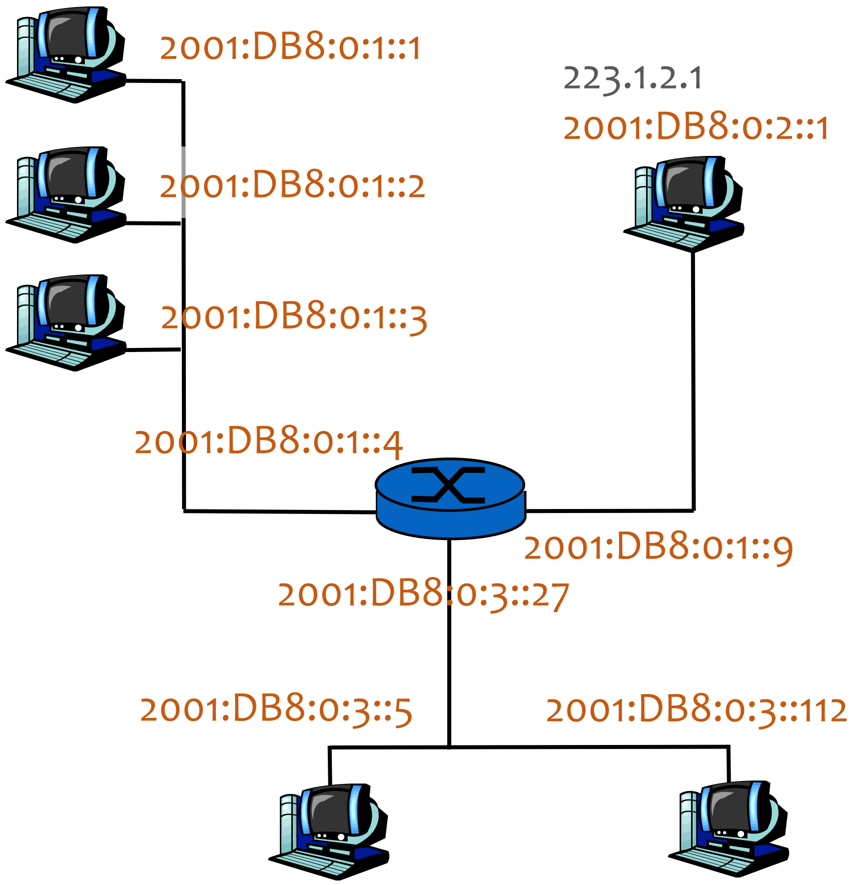
Source: https://www.google.com/intl/en/ipv6/statistics.html

IPv6: Addressing

IP address:
Identifier for host or router interface

IPv6 address:

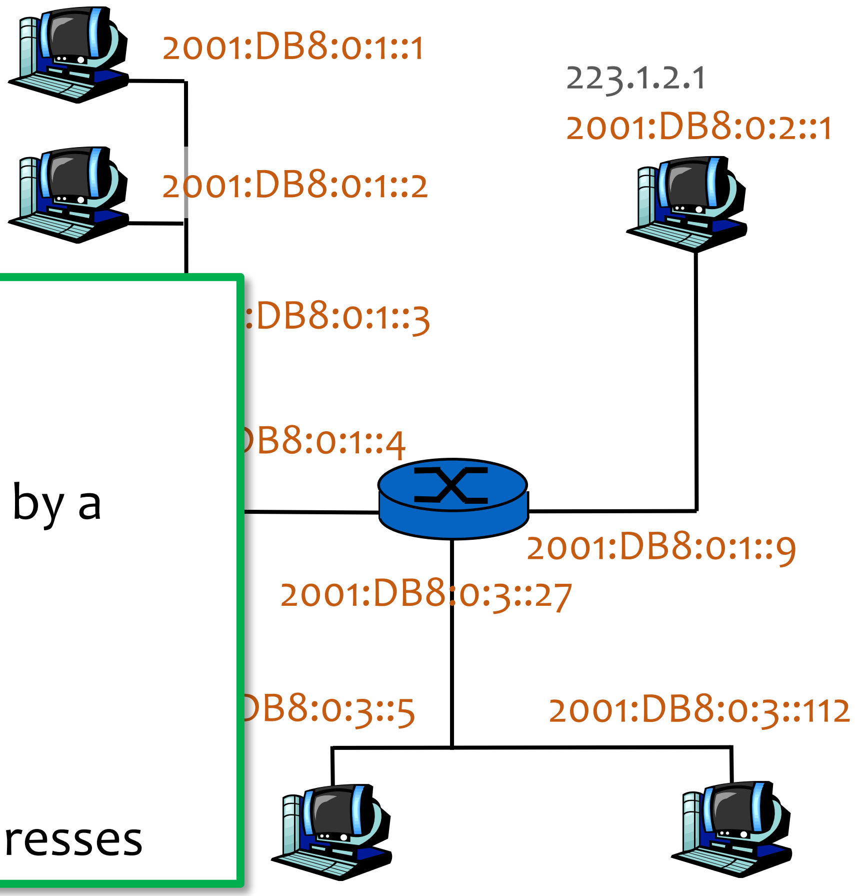
- 128 bits written as eight 16-bit groups in hexadecimal
- Hex-tets are separated by colons



IPv6: Addressing

IP address:

Identifier for host or router interface



2001:0DB8:0000:0001:0000:0000:0000:0004

- Leading zeros in hex-tets can be left out
- Multiple "empty" (all zero) hex-tets can be abbreviated by a double-colon. This can only be done at one position.

2001:DB8:0:1::4

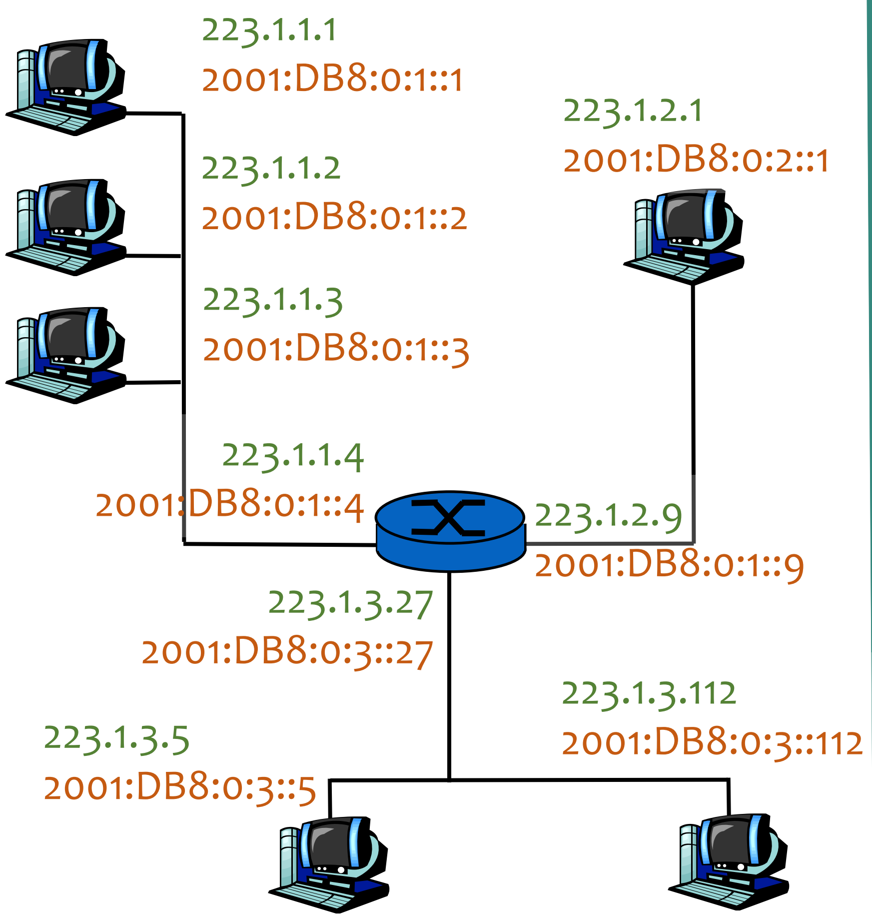
- Leading zeros can be left out
- Shortening multiple times would lead to *ambiguous* addresses

Dual Stack Addressing

IPv4 and IPv6 identifiers per host or router interface

Host decides whether to use IPv4 or IPv6

- Routers may support both
- IPv4 (IPv6) traffic stays IPv4 (IPv6)
- IPv4/IPv6 addresses on the same host are not linked at all at the network layer

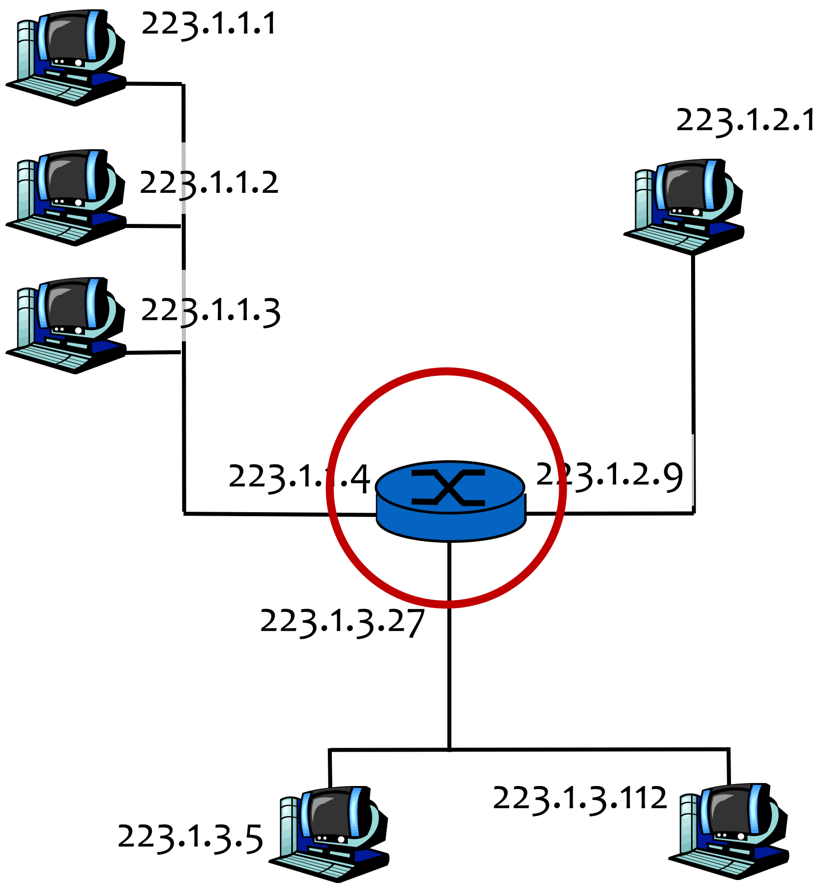


IP Network

What is a network?

(from IP address perspective)

- Can physically reach each other without intervening router

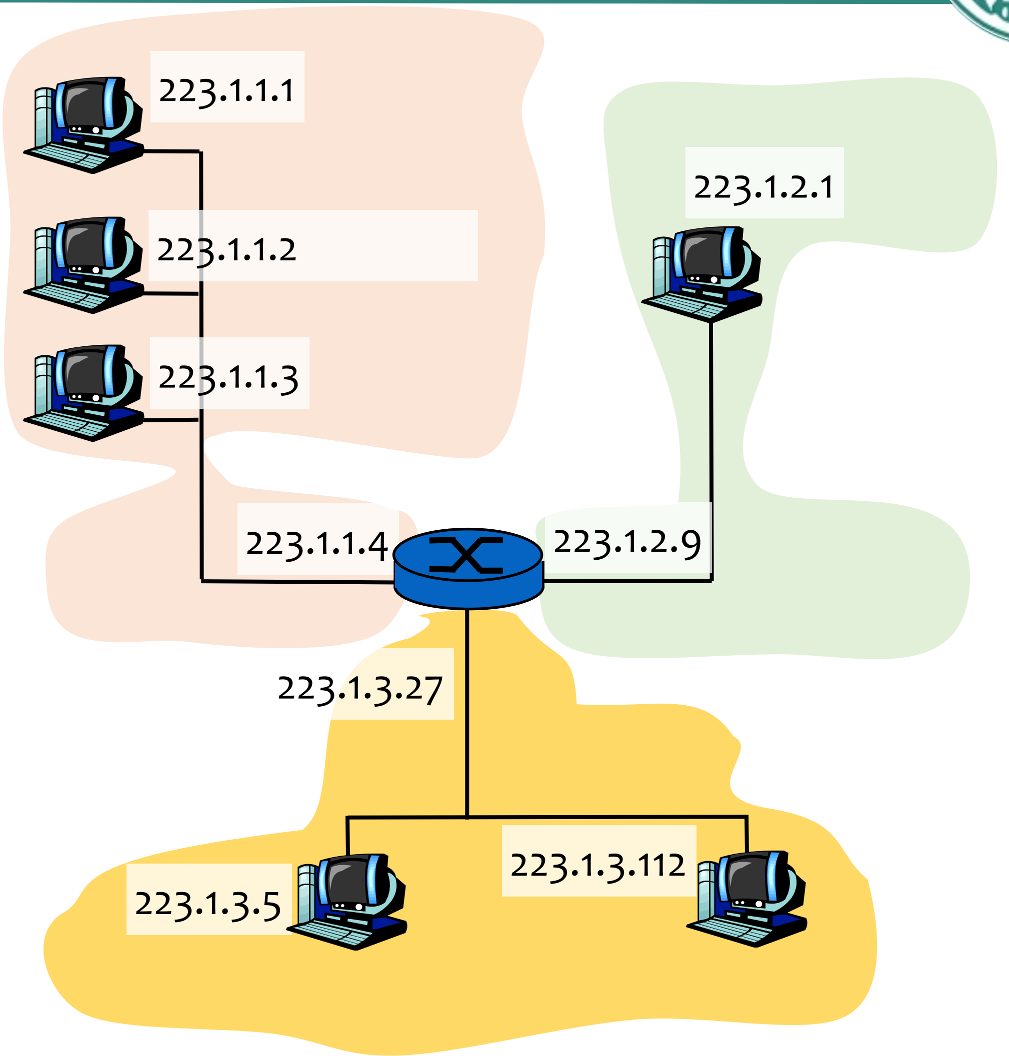


IP Network

What is a network?

(from IP address perspective)

- Can physically reach each other without intervening router
- Device interfaces with same high order bits in their IP address



Example: Network consisting of 3 IPv4 networks

IPv6 Network

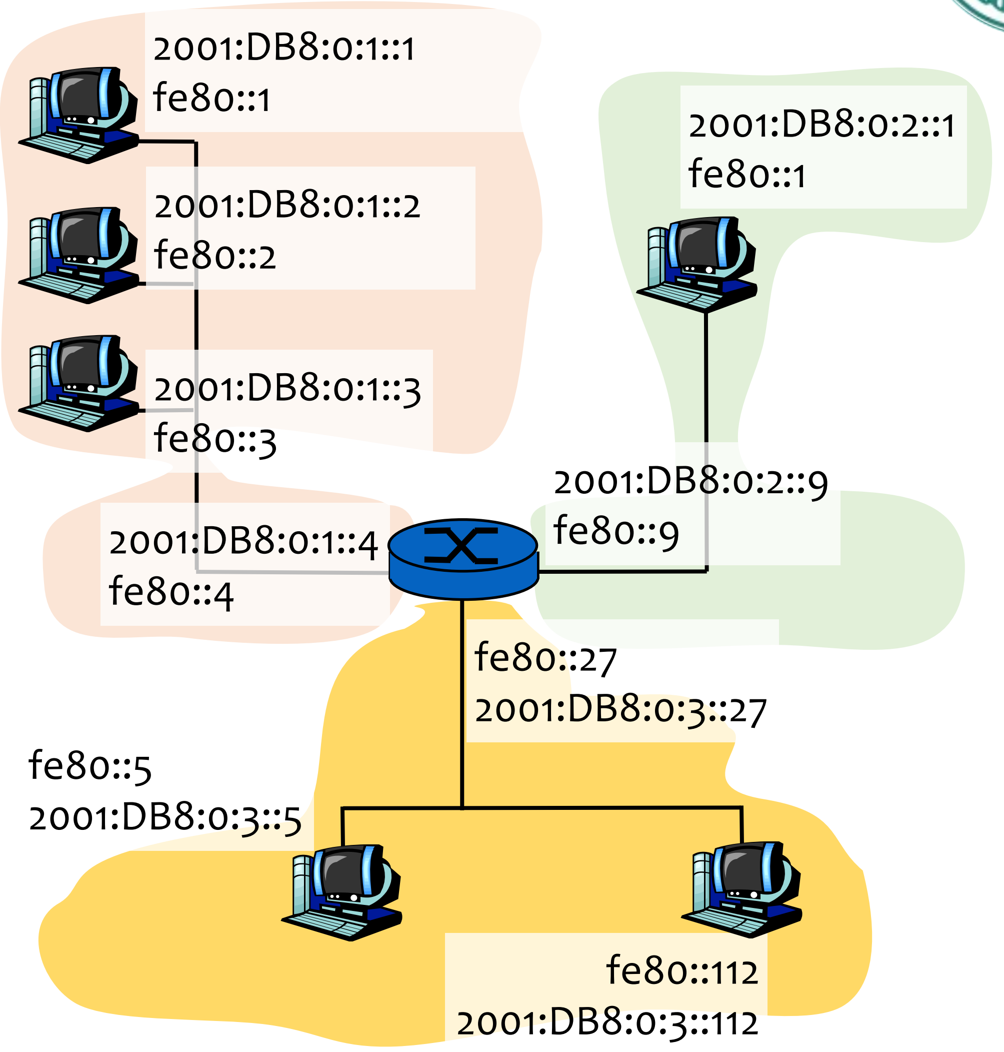
What is a network?

(from IP address perspective)

- Can physically reach each other without intervening router
- Device interfaces with same high order bits in their IP address

What's different in IPv6?

- Usually more than one IPv6 address per host
- Special link-local network



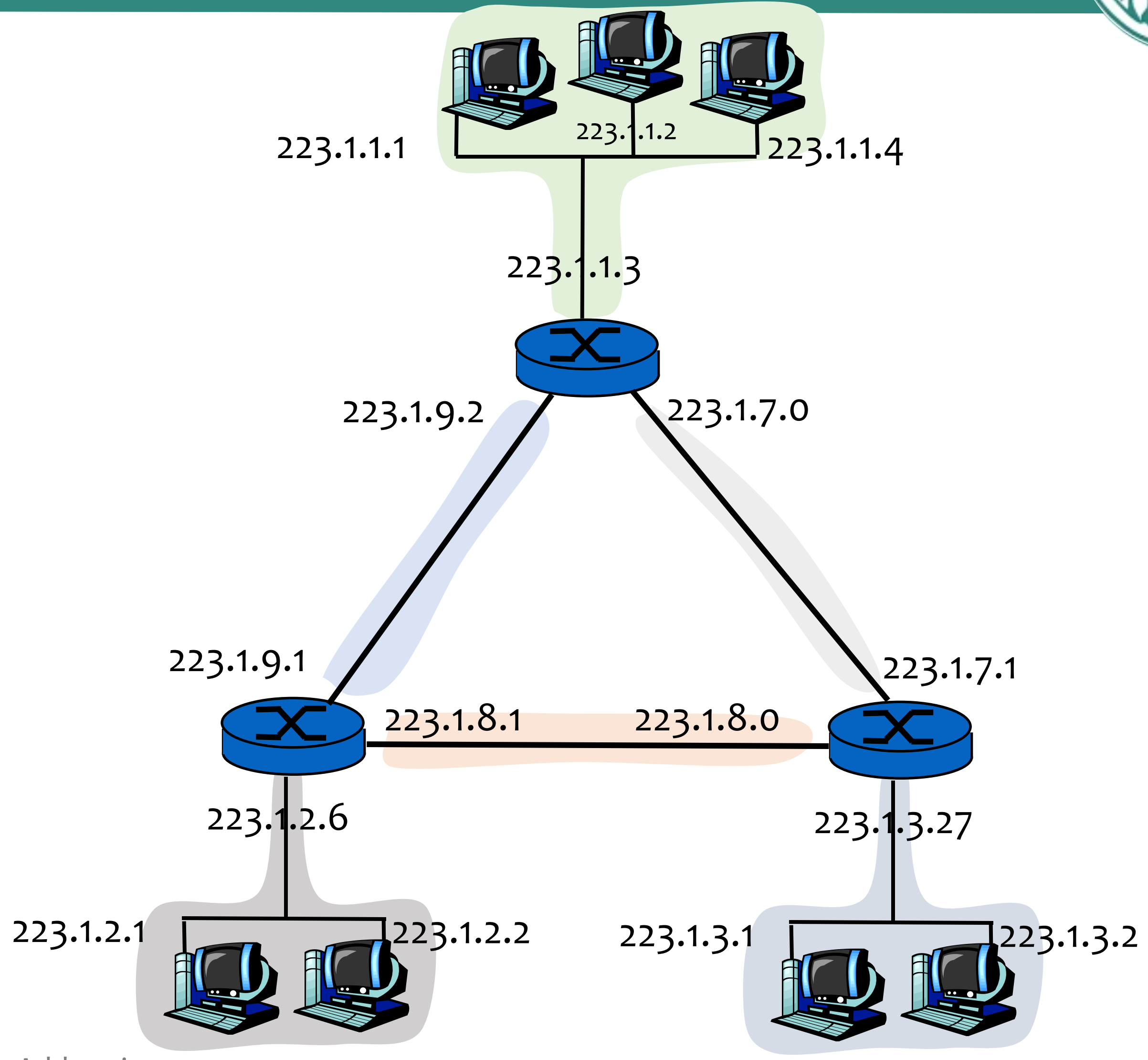
Example: Network consisting of 3 IPv6 networks

IP Networks: Top-down View

How to find the networks?

- Detach each interface from router, host
- Create "islands" of isolated networks

Interconnected system consisting of six networks



IP Subnetting

Subnetting divides address space into

- A network part referred to as prefix
- Host address

network part | *host part*

11001000 00010111 00010000 00000000

200.23.16.0/24

Address format (CIDR)

- IPv4: **a.b.c.d/m** **200.23.16.0/24**
- IPv6: **x:y:z::/m** **2001:DB8:0:3::/64**
- Network part of the address in number of bits; referred to as **prefix length** (bit representation == **netmask**)

IPv4 Classful Subnetting (Deprecated)

Class	Address format (32 bits)	Range
A	0 network | host	0.0.0.0 to 127.255.255.255
B	10 network | host	128.0.0.0 to 191.255.255.255
C	110 network | host	192.0.0.0 to 223.255.255.255
D	1110 multicast address	224.0.0.0 to 239.255.255.255

Problem: Wastes IP address space

- If you need more addresses than a class C network, e.g., 256, you need to get at least a class B network (65536)

CIDR: Motivation

Motivation

- Classful class A is too large, classful class C is too small
- Everyone had a Class B network!!! ⟹ running out of networks!

CIDR to the rescue

- **Flexible** network prefix length
  - Sites are given contiguous blocks of class C addresses and a mask, or
  - Parts of former class A/B networks

CIDR (Current Standard)

CIDR: Classless Inter-Domain Routing

- Subnet portion of address of arbitrary length
- Address format: **a.b.c.d/x**,
  where **x** is number of bits in the subnet portion of address

subnet part | host part

11001000 00010111 00010000 00000000

200.23.16.0/**23**

Special IPv4 Address Ranges

Loopback 127.0.0.0/8

Multicast 224.0.0.0/4

Private ranges 10.0.0.0/8
172.16.0.0/12
192.168.0.0/16

Link-Local 169.254.0.0/16

Special IPv6 Address Ranges

Loopback	::1/128
Global Unicast	2000::/3
Unique Local	FC00::/7
Link-Local Unicast	FE80::/10
Multicast	FF00::/8

Addresses for use in the Internet are **Global Unicast** and parts of Multicast

Link-Local addresses are limited to a physical link (RFC 3513)

Multicast Addresses

Addresses a group of hosts at once

- Useful for streaming and conferencing applications
- Heavily used in IPv6 for signaling

Only certain ranges are usable for multicast

- IPv4: 224.0.0.0/16
- IPv6: FF00::/8

Link-local Addresses

Non-routable addresses

- Can only be used within a network
- Addresses not unique(!)
- Heavily used in IPv6 for local signaling

Reserved address ranges:

- 169.254.0.1/16 RFC 3927
- FE80::/10 RFC 4291

Private IP Addresses

For local use only; not routable in the Internet; used for NAT

Private IPv4 addresses (RFC 1918)

- 10.0.0.0/8
- 172.16.0.0/12
- 192.168.0.0/16

Unique local IPv6 unicast addresses (RFC 4193)

- FC00::/7

How does a host get an IP addresses?

- Hard-coded by system administrator
- DHCP/DHCPv6 (Dynamic Host Configuration Protocol)
  - Host requests the address from a DHCP server
- IPv6 SLAAC (Stateless Address Autoconfiguration)
  - Router advertises the IPv6 prefix
  - Host adds an interface identifier as host part

IP Address Allocation Process

1. ICANN (Internet Corporation for Assigned Names and Numbers) gives **IP address blocks** to RIRs

2. RIRs (Regional Internet Registries)—RIPE, ARIN, APNIC, LACNIC, AFRINIC—assign addresses to LIRs

3. LIRs (Local Internet Registries) are larger providers that assign addresses or address blocks to their customers

IP Address Allocation Process

1. ICANN (Internet Corporation for Assigned Names and Numbers) gives **IP address blocks** to RIRs
2. RIRs (Regional Internet Registries)—RIPE, ARIN, APNIC, LACNIC, AFRINIC—assign addresses to LIRs
3. LIRs (Local Internet Registries) are larger providers that assign addresses or address blocks to their customers

IPv4 address space

- None left in the ICANN pool since January 31, 2011
- All RIRs also exhausted since November 25, 2019; still allocate recovered address blocks

IPv6 address space

- Typical allocation for an LIR: /32
- Typical allocation for a site: /48 – /56

IP Address Allocation Process

What do I do if I do not have a public address?

- Recall private IP addresses
  - 10/8 RFC 1918
  - 172.16/12
  - 192.168/16
  - FC00::/7 RFC 4193
- Private use only—not routable in the Internet

- Recall link local addresses
  - 169.254.0.1/16 RFC 3927
  - FE80::/10 RFC 4291
- Local or single network use only—not routable in the Internet

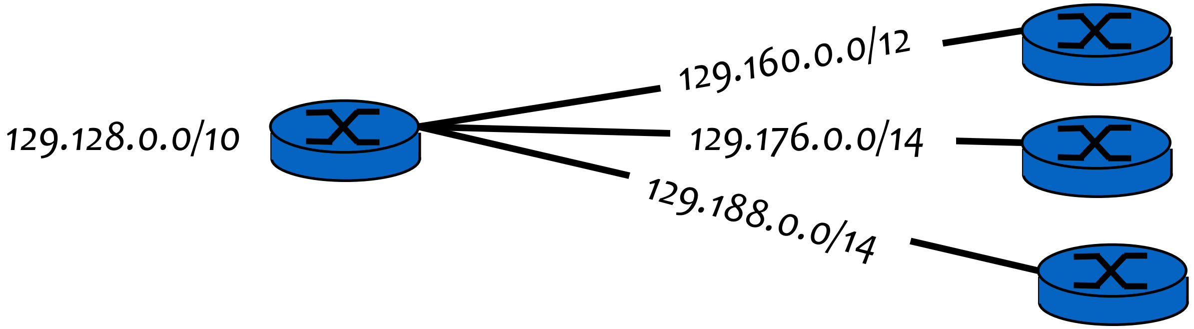
Hierarchical Address Structure

Recall **CIDR**

18 relevant bits

- 128.119.48.12/18 = 10000000 01110111 00110000 00001100
- High order bits form the prefix
- Once inside the network, can subnet: divide remaining bits

Subnet example:



Picture shows prefixes, not interface addresses!

Forwarding decision according to longest prefix match

Forwarding vs. Routing

Forwarding: Process of moving packets from input to output

- The forwarding table
- Information in the packet

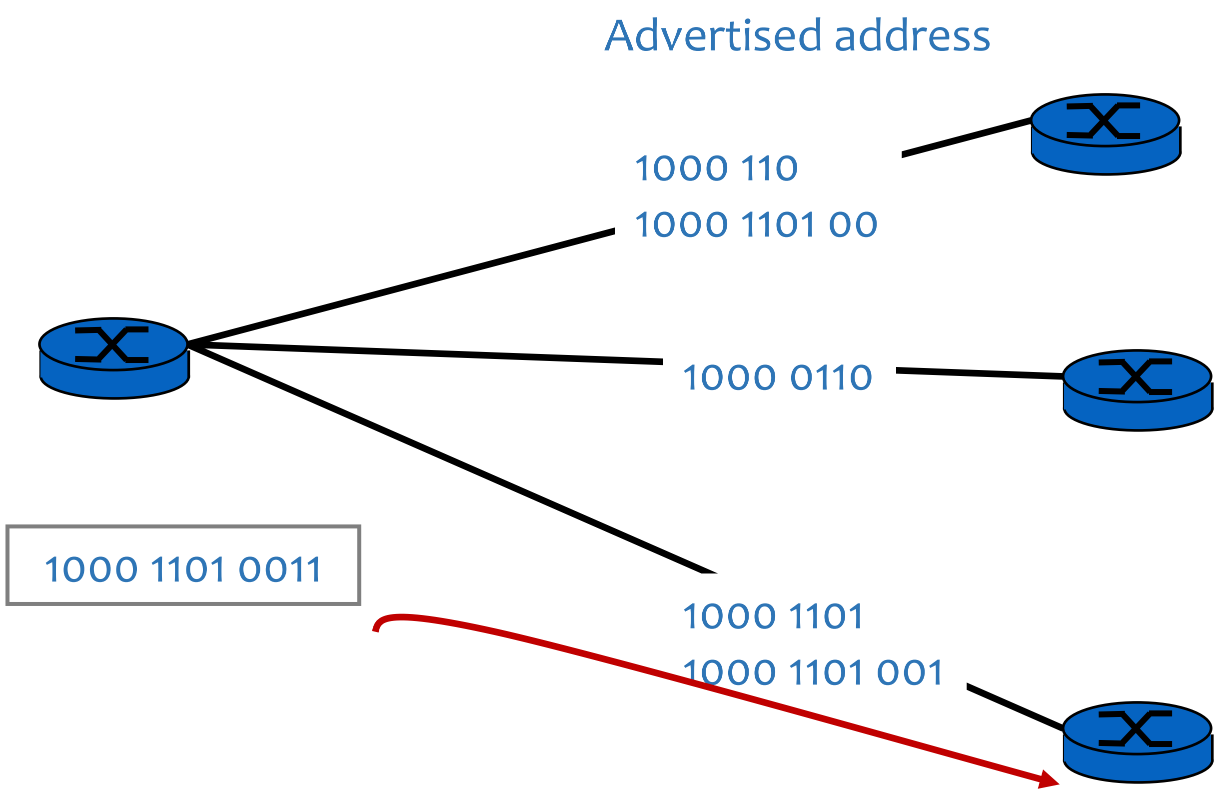
Routing: Process by which the forwarding table is built and maintained

- One or more routing protocols
- Procedures (algorithms) to convert routing data to forwarding table

(more later …)

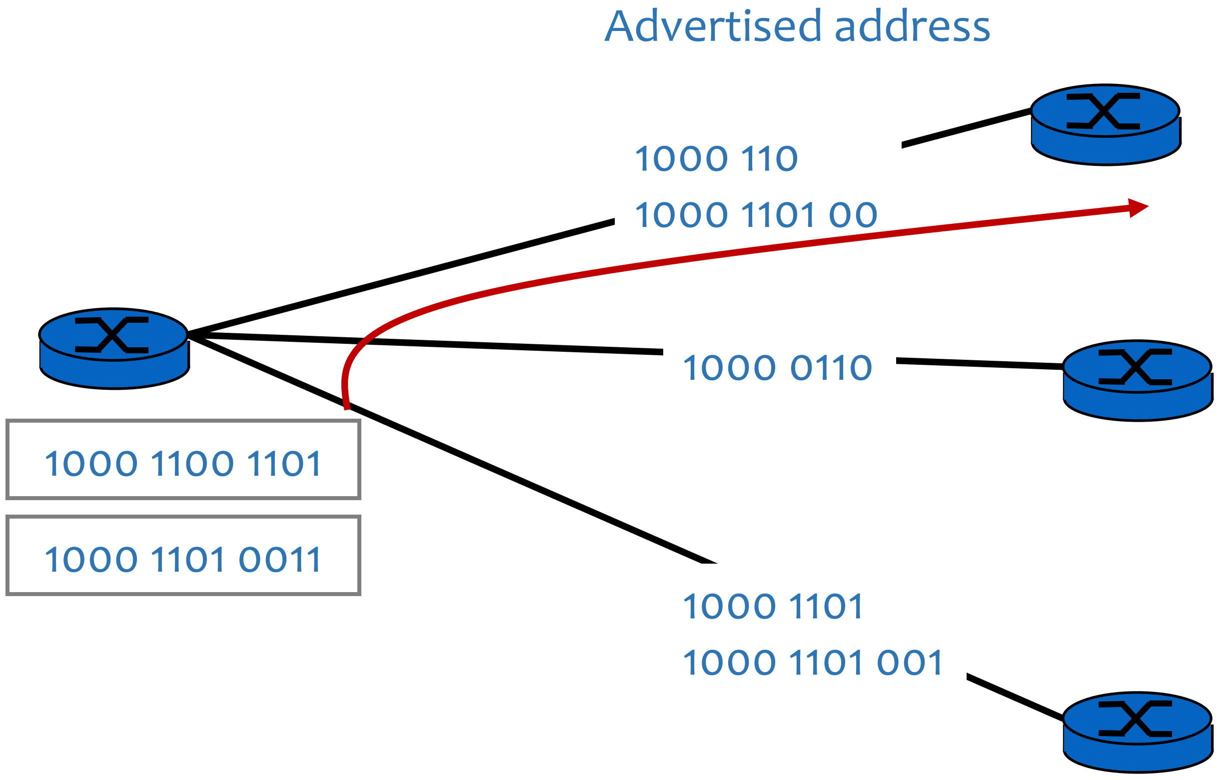
Forwarding with CIDR

Packet should be sent toward the interface with the **longest matching prefix**



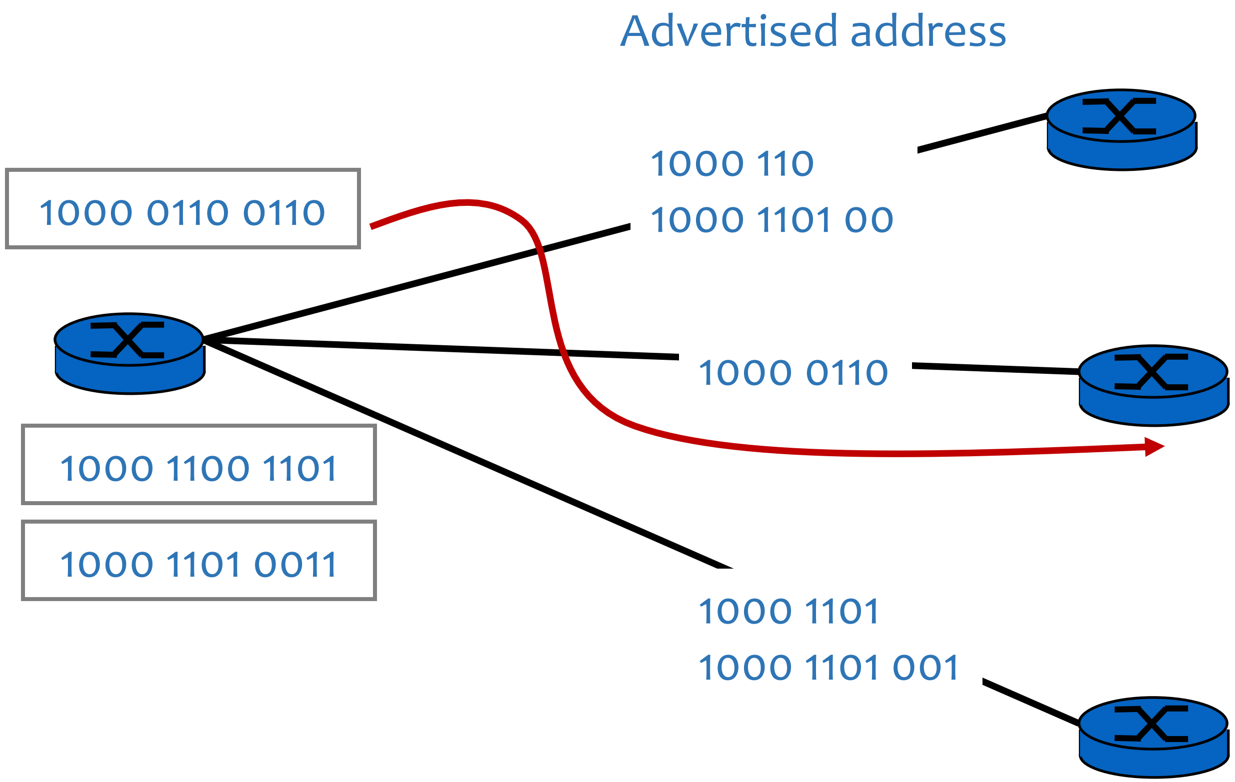
Forwarding with CIDR

Packet should be sent toward the interface with the **longest matching prefix**



Forwarding with CIDR

Packet should be sent toward the interface with the **longest matching prefix**



Lookup: Longest Prefix Match

Forwarding table:

<network>/<mask> <next-hop>

IP Packets: Destination IP address

- Find next-hop via longest prefix match

Example (IPv4):

Forwarding table

134.96.252.0/24	A
134.96.0.0/16	C
134.96.240.0/20	B
134.96.252.192/28	B
134.96.252.196/30	A

Packets

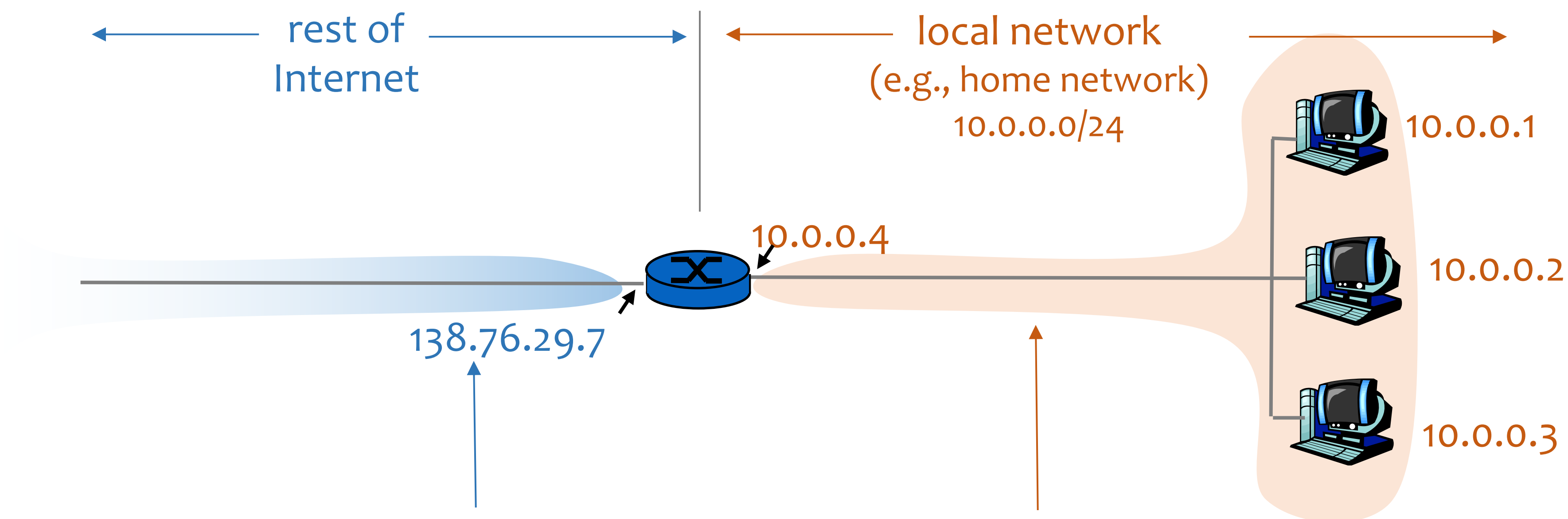
134.96.254.2	B
134.96.240.200	B
134.96.239.200	C
134.96.252.191	A
134.96.252.200	B

Network Address Translation (NAT)

Motivation: Local network has just one IP address as far as outside world is concerned

- Just one IP address for all devices
- No need for range of addresses from ISP
- Workaround for IPv4 exhaustion (carrier-grade NAT)

Network Address Translation (NAT)



All packets **leaving** local network have the **same** single source NAT IP address: 138.76.29.7, different source port numbers

Packets with source or destination in this network have 10.0.0.0/24 address as source or destination (as usual)

Network Address Translation (NAT)

Motivation: Local network uses just one IP address as far as outside world is concerned

- Range of addresses not needed from ISP:
  Just one IP address for all devices
- Can change addresses of devices in local network without notifying outside world
- Can change ISP without changing addresses of devices in local network
- Devices inside local network not explicitly addressable by (or visible to) outside world!

NAT Implementation

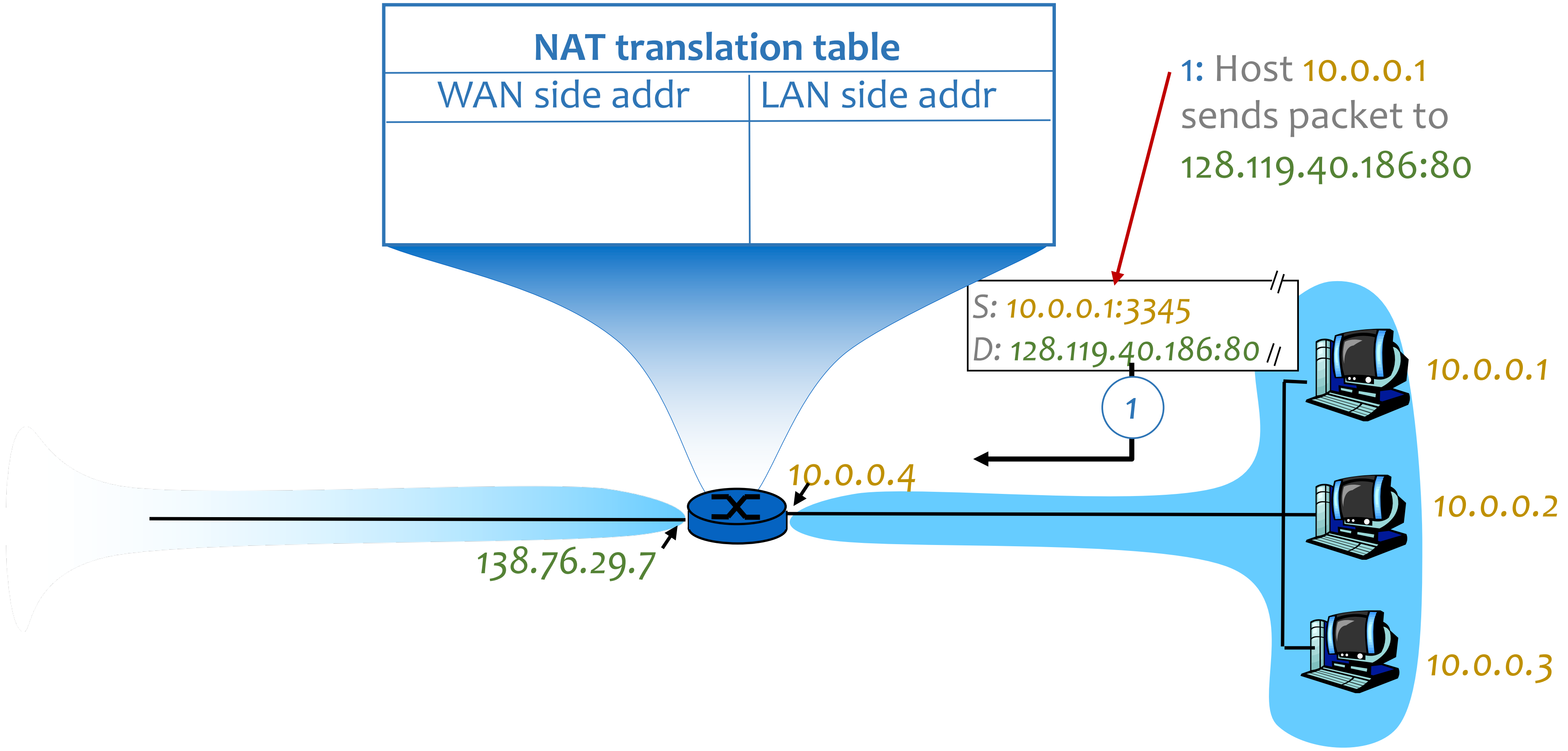
NAT router must

- **Replace** (source IP address, port number) of every outgoing packet to (NAT IP address, new port number)
  - Remote clients and servers will respond using (NAT IP address, new port number) as destination address

- **Remember** (in NAT translation table) every (source IP address, port number) to (NAT IP address, new port number) translation pair

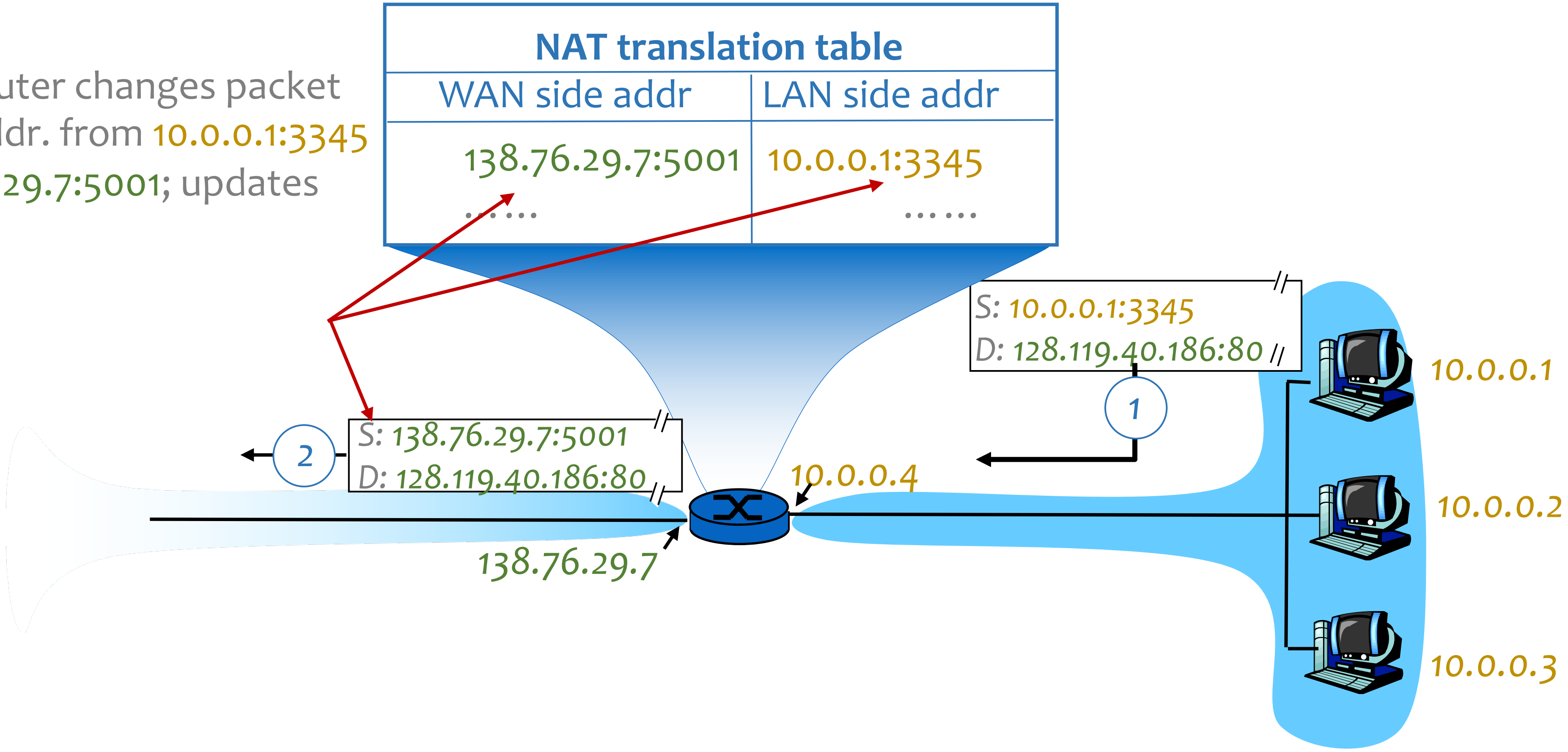
- **Replace** (NAT IP address, new port number) in destination fields of every incoming packet with corresponding (source IP address, port number) stored in the NAT table

NAT Implementation: Example



NAT Implementation: Example

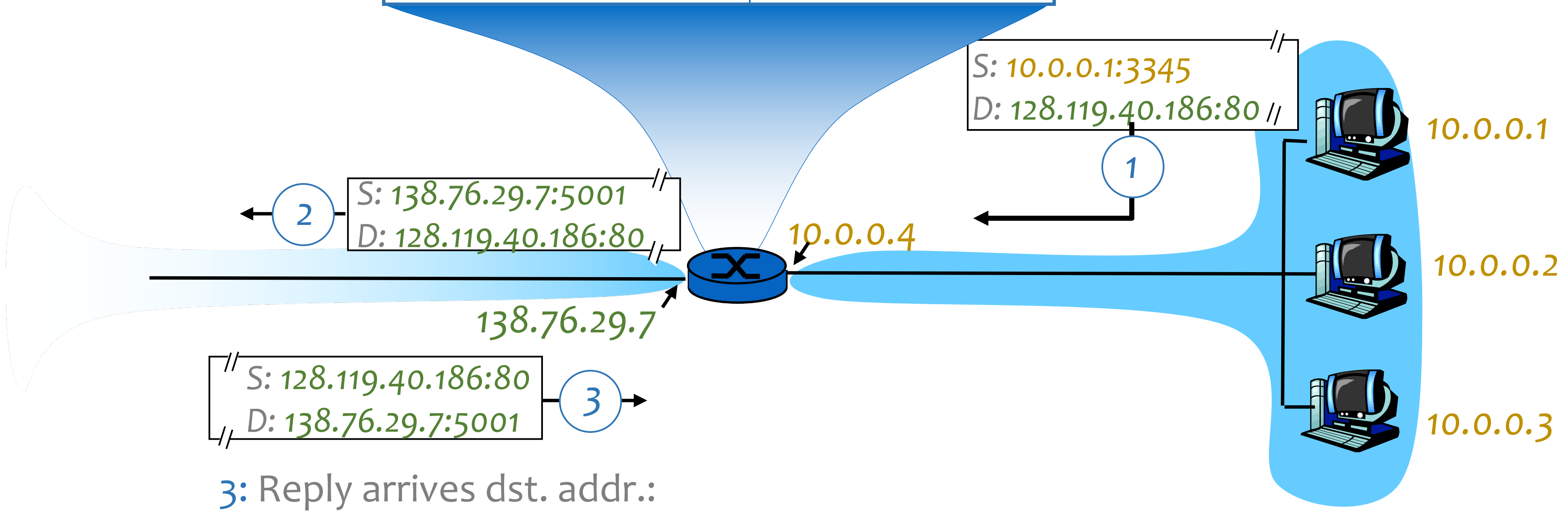
2: NAT router changes packet source addr. from 10.0.0.1:3345 to 138.76.29.7:5001; updates table



NAT Implementation: Example

NAT translation table

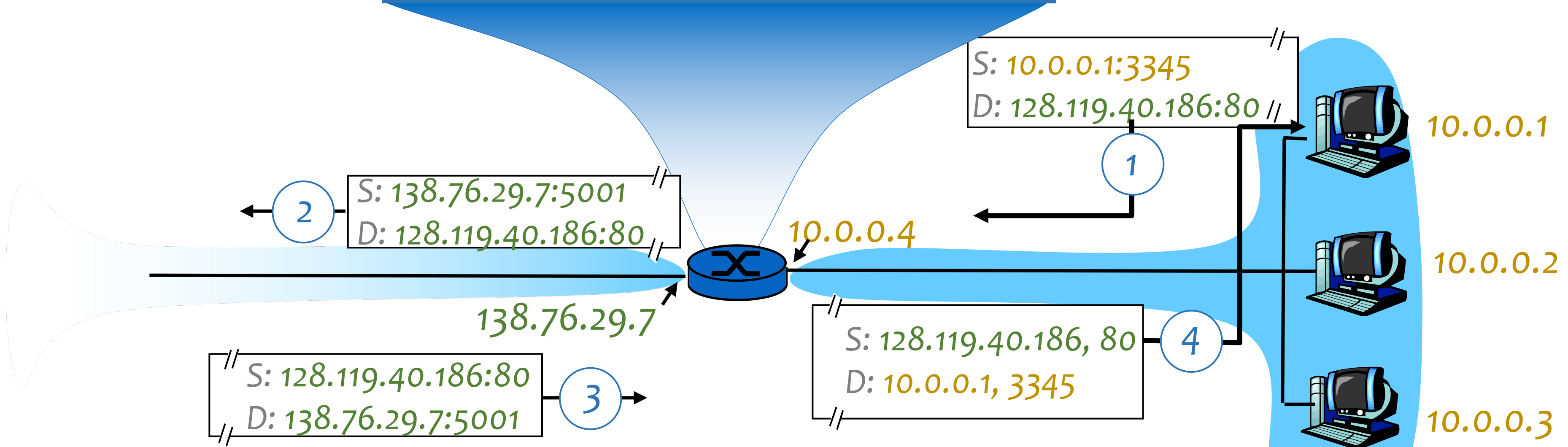
WAN side addr	LAN side addr
138.76.29.7:5001	10.0.0.1:3345
……	……



3: Reply arrives dst. addr.:
138.76.29.7:5001

NAT Implementation: Example

NAT translation table	
WAN side addr	LAN side addr
138.76.29.7:5001	10.0.0.1:3345
……	……



4: NAT router
changes packet dst. addr. from
138.76.29.7:5001 to 10.0.0.1:3345

NAT: Parting Thoughts

16-bit port-number field

- > 60,000 simultaneous connections with a single WAN-side address!

NAT is controversial

- Routers should only process up to layer 3
- Violates end-to-end argument
- NAT possibility must be taken into account by application designers (e.g., P2P applications)
- Address shortage should instead be solved by IPv6

Recap

- IPv4 and IPv6 addressing
  - Subnetting
  - Allocation
  - Special address ranges

- Routing and forwarding
  - Longest prefix matching

- Network address translation