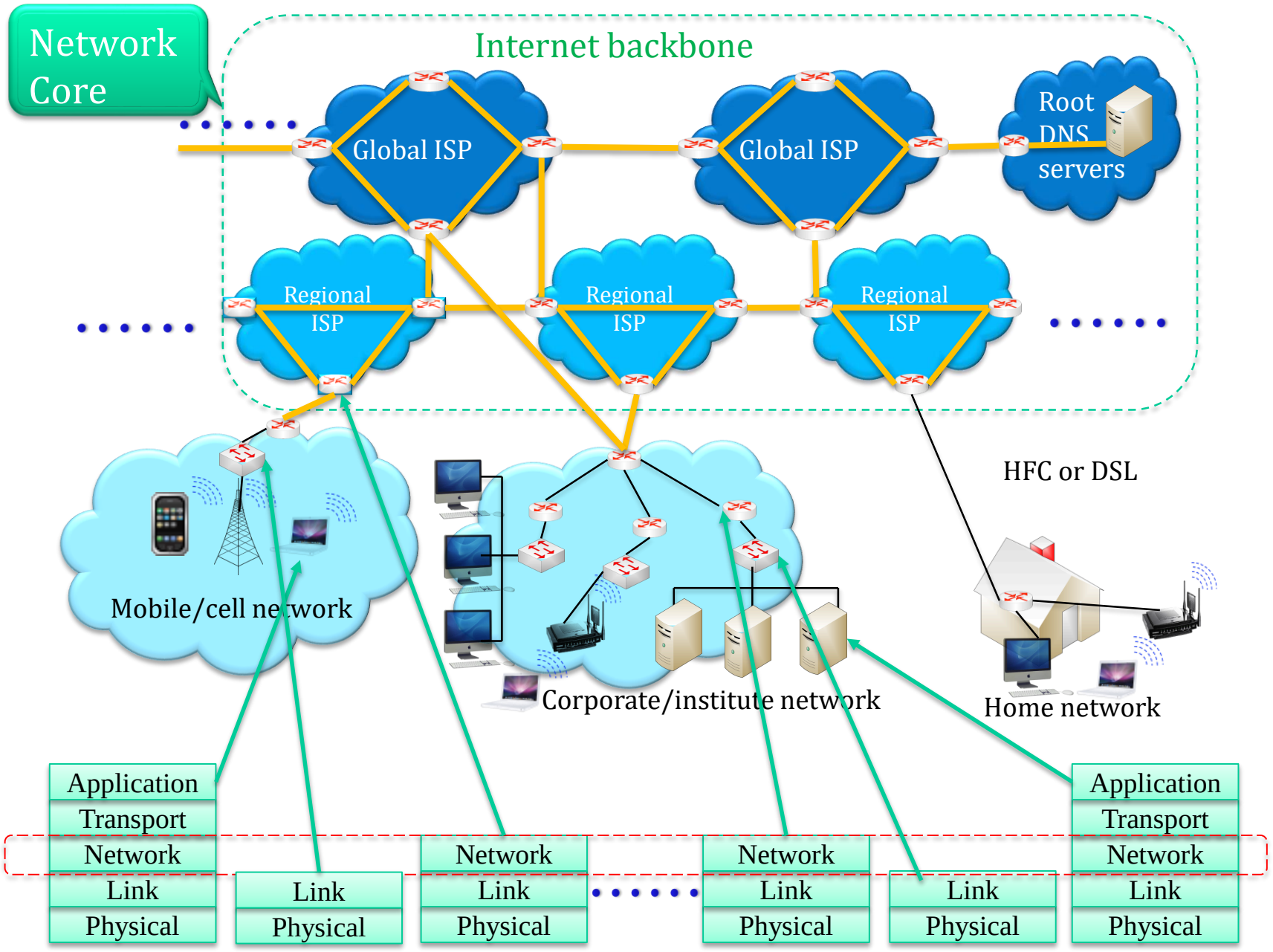
Lecture 4

Network Layer Overview

Outline

Part 3

- Network layer
- Connection-oriented network
- Network layer functions in the protocol stack
- IPv4 header
- IP address
- DHCP
- Routing between LANs
- Network address translation (NAT)
- The Internet control message protocol (ICMP)


Network Core
Internet backbone
Global ISP
Global ISP
Root DNS servers
Regional ISP
Regional ISP
Regional ISP
HFC or DSL
Mobile/cell network
Corporate/institute network
Home network
Application
Transport
Network
Link
Physical
Network
Link
Physical
Network
Link
Physical
Network
Link
Physical
Application
Transport
Network
Link
Physical

Network Layer

- Route datagram from sending to receiving host
- Source host encapsulates segments (passed down by transport layer) into datagrams
- Destination host delivers segments up to transport layer
- Network layer protocols are built in every host, and router
  - The majority of hosts do not know how to route
  - Client OS does not have routing module
- Router examines header fields in all IP datagrams from one interface and forwards to another interface in accordance with the routing table
  - Routers work together to generate routing tables
- Router understands the network, link and physical layers

Two Key Router Functions

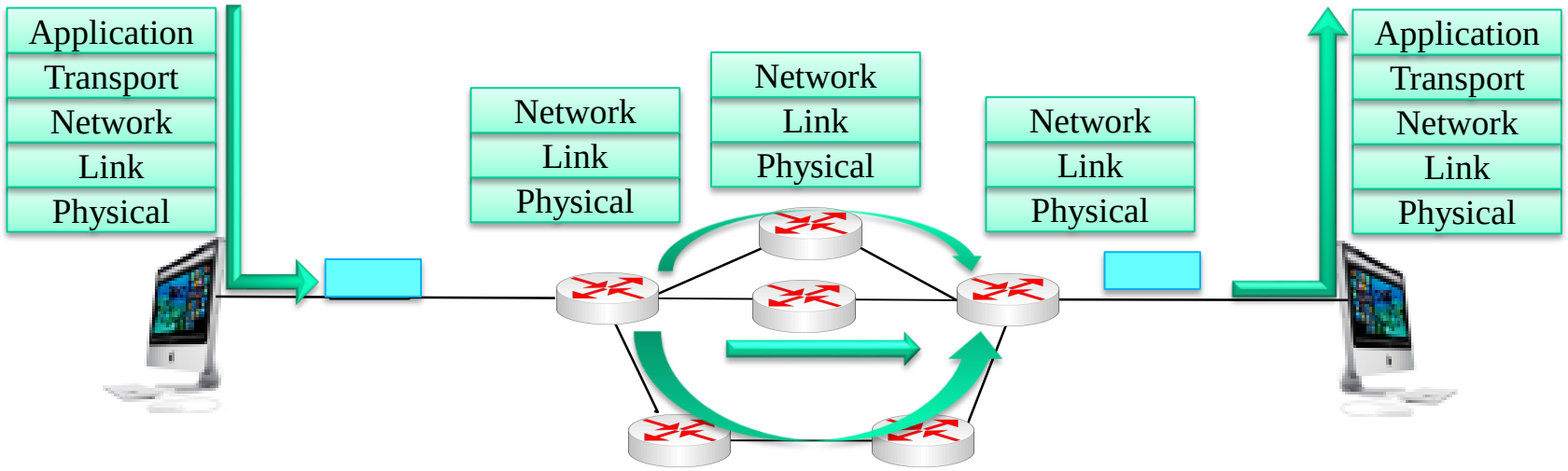
- Routing
  - Generate and maintain routing table
  - Routing algorithms: routers work together to find routes from a subnet to the other subnets
  - Distributed processing motivated by the cold war
  - If there exists a path, routers will put it in the routing table automatically
- Forwarding:
  - A router/layer 3 switch moves a packets from router's input port to appropriate router output in accordance with the routing table
- Human analogy:
  - Routing: airline timetables contain flight schedules, fleet, in-flight entertainment, and food menu. A traveler can reserve flights using airline timetables
  - Forwarding: a traveler moves from one flight to another flight in an airport using flight tables displayed on monitors

Layer2 switch

Layer 3 router

Datagram Networks

- Connectionless: no call setup at network layer
- Routers: no state information about end-to-end connections
- Packets forwarded using destination host IP address
  - Packets between same source-destination pair may take different paths
  - Packets may not arrive in the original order



Routing Table

Destination Address Range	Router Interface
11001000 11010111 00000000 00000000 through 11001000 11010111 00001111 11111111	0
11001000 11010111 00001000 00000000 through 11001000 11010111 00001000 11111111	1
11001000 11010111 00010000 00000000 through 11001000 11010111 00011111 11111111	2
Otherwise	3

Prefix, or subnet part (network ID + subnet ID)

Host ID

Use Longest Prefix Matching

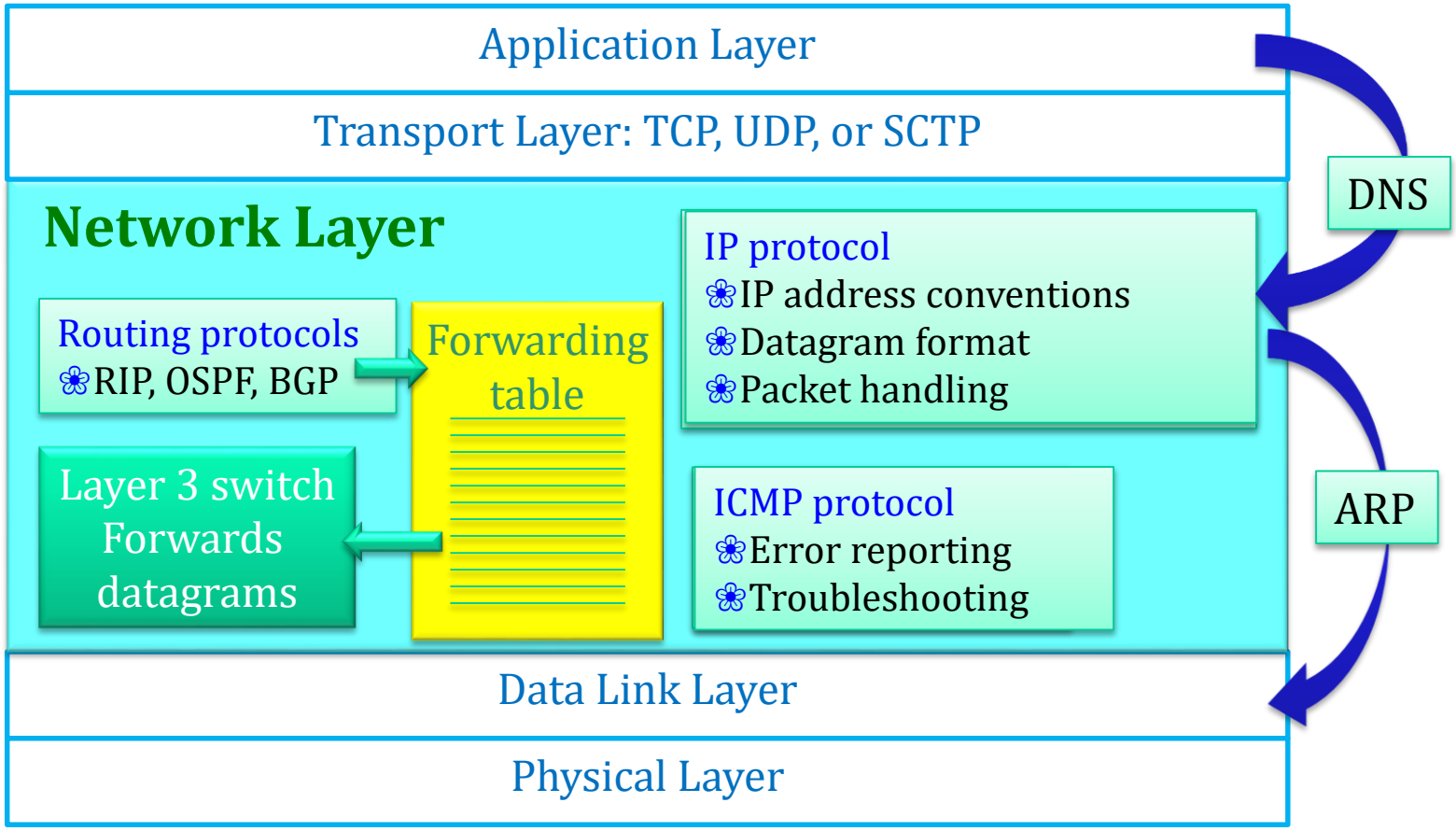
IP address prefix: 200.215.x.x (C8.D7.x.x)

Prefix for router Interface	Router Interface
11001000 11010111 0000	0
11001000 11010111 00001000	1
11001000 11010111 0001	2
otherwise	3

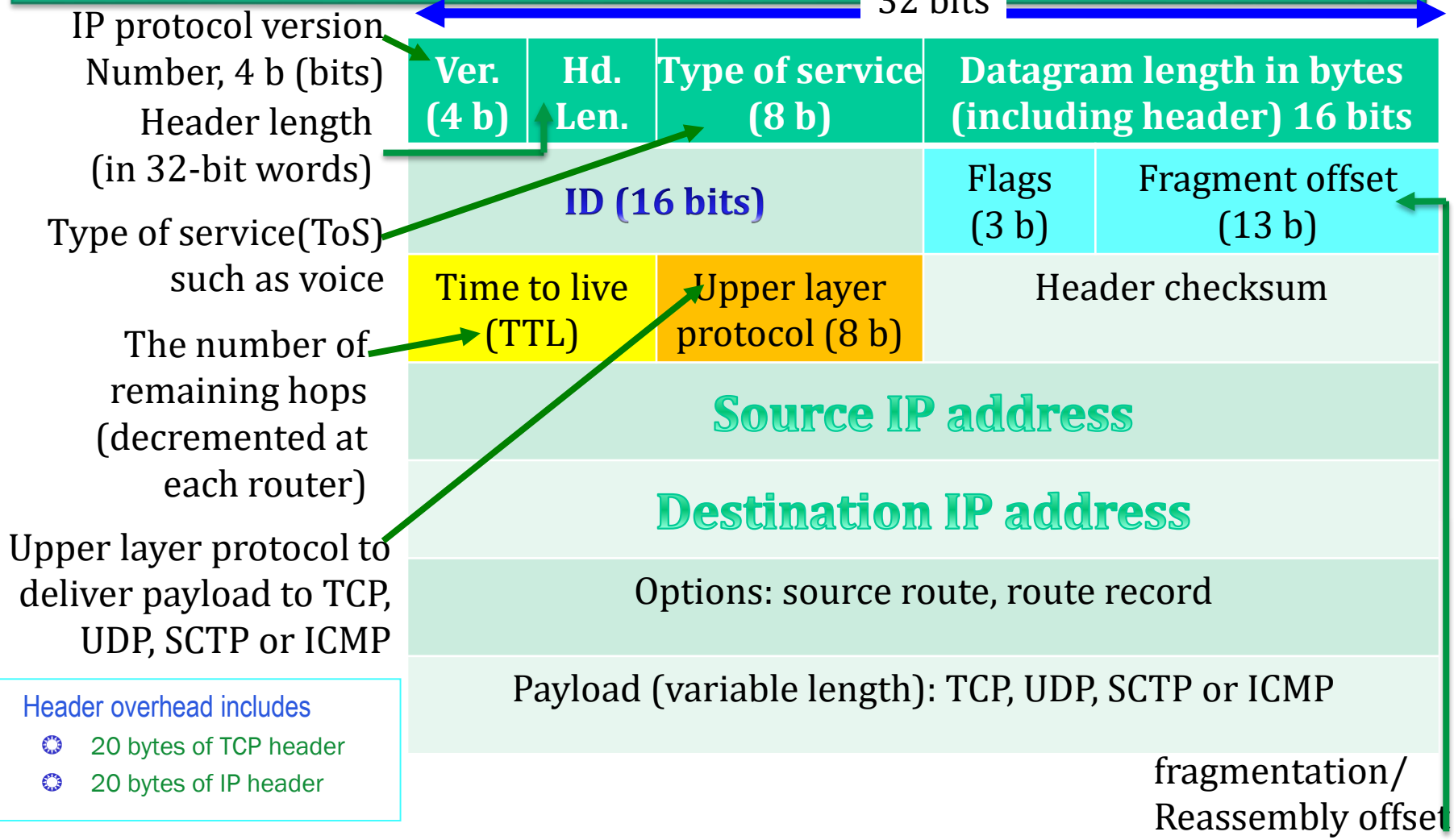
- Examples
  - Router uses the longest matching prefix in the routing table to select the output interface
    - Destination IP Address: 11001000 11010111 00001000 10100001
      - Route to Interface 1
    - Destination IP Address: 11001000 11010111 00000000 10101010
      - Route to Interface 0

The Internet Network Layer

- Router and network layer functions:



IPv4 Datagram Format



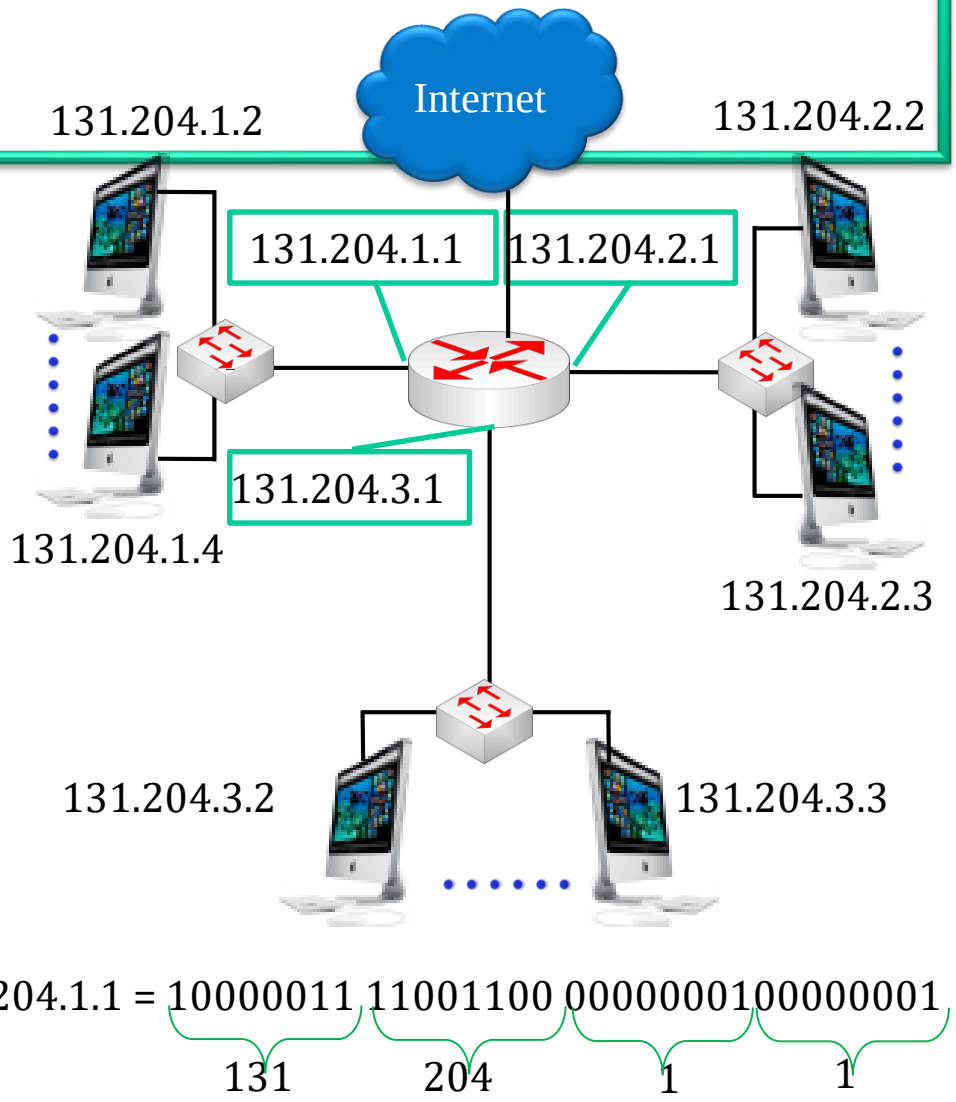
- Header overhead includes
  - 20 bytes of TCP header
  - 20 bytes of IP header

The purpose for each IP header field

IP header field	Description
The *IP protocol version number*	It is 4 bits long and 0100 for IPv"4".
The 4-bit *header length*	Specifying the number of 32-bit words in the header.
8-bit *type of service*	e.g. voice over IP in order to prioritize traffic.
The *datagram length in the number of bytes*	Including header and data, comprises the remaining 16 bits.
The *ID* and *Flag* fields	They are used to identify and control fragments.
The *fragment offset*	It is used for fragmentation/reassembly.
The *time to live*	It is the allowed maximum number of remaining hops, decremented at each router.
The *upper layer protocol*	It is the one used in the payload of the IP datagram, e.g. TCP has a value of 6 as shown in Table
The *header checksum*	It does error checking on the header.
The options	It follows the *source* and *destination addresses*. The options can be used for source route, i.e. specifying the IP address of each hop, and route record, i.e. recording the IP address of each hop. But they are rarely used for security reasons.

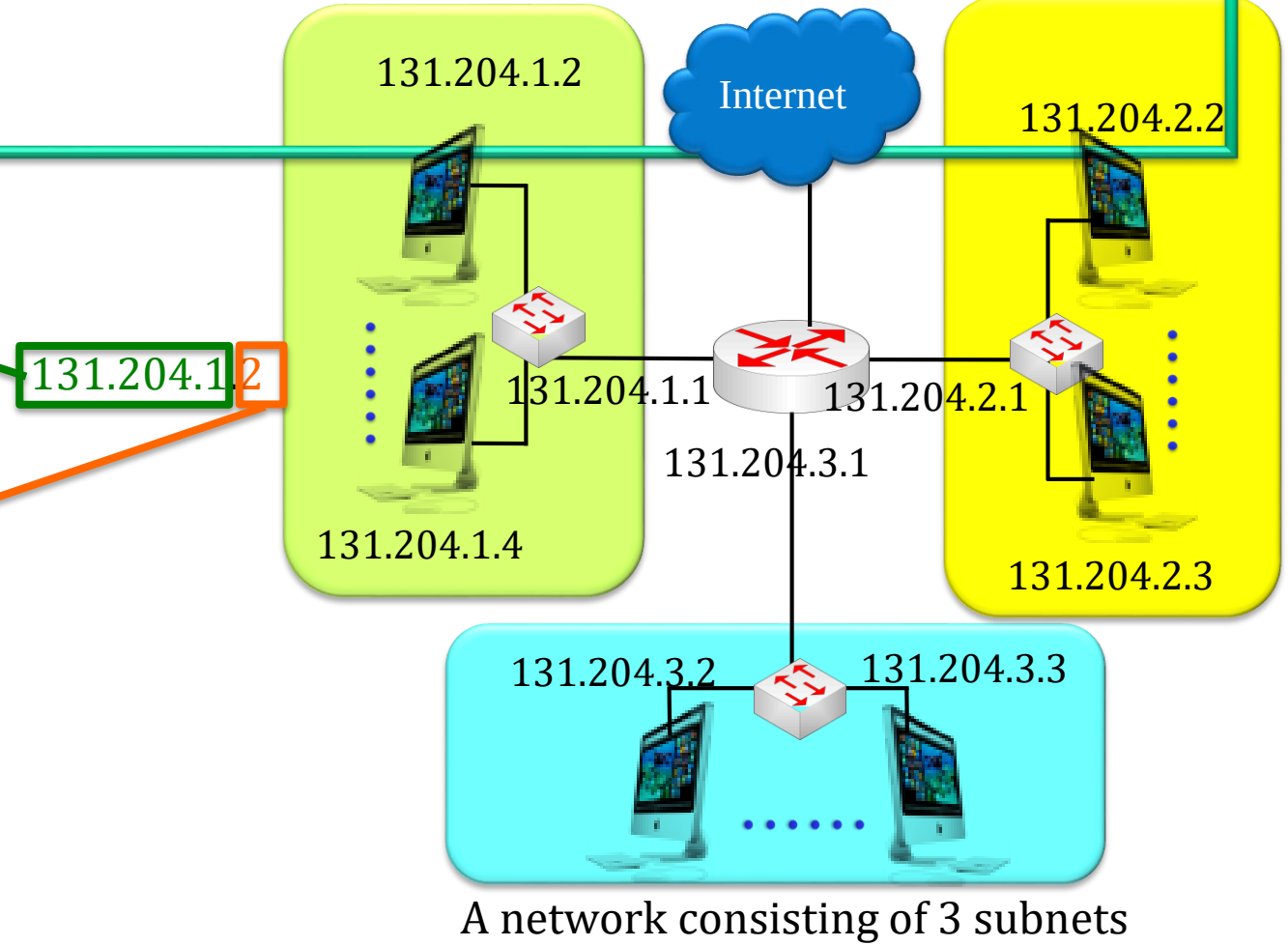
IPv4 Addressing

- IP address:
  - 32-bit identifier (IPv4) for host, or router interface
- Interface
  - A network module that has one physical link
  - Router typically has multiple interfaces (at least 2 interfaces)
  - Host typically has one interface
  - One IP address and one MAC address associated with each interface
  - Note: a layer 2 switch has no interface



Subnet

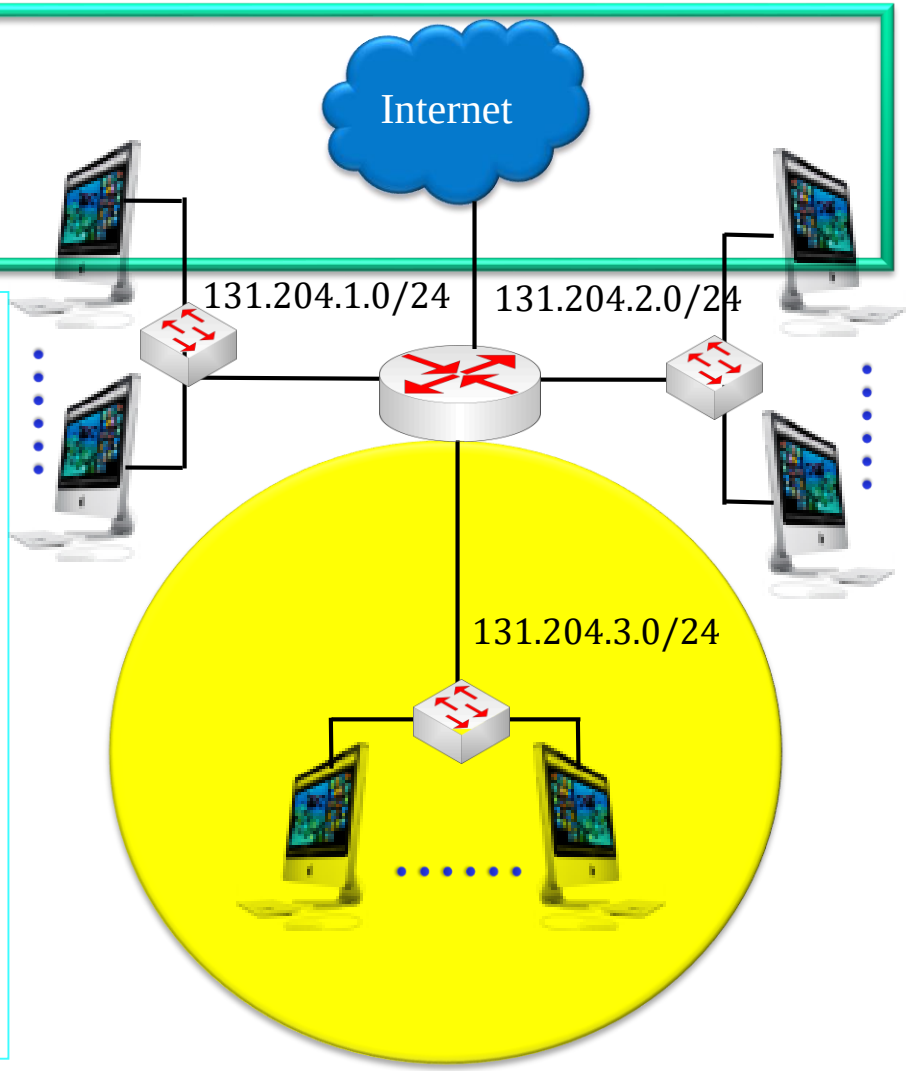
- Each IP address has:
  - Subnet part (high order bits): aka prefix
  - Host part (low order bits)
- Subnet
  - Hosts that have interfaces with the same subnet part of IP address
  - Hosts can communicate with each other without router



A network consisting of 3 subnets

Subnet

- Each subnet contains at least one gateway interface
- To determine a subnet:
  - Detach each interface from its hosts and router
  - Make layer 2 switch/hub transparent
    - They have no interface
  - Those interfaces form one island in isolated networks
  - Each isolated network is called a subnet
- Subnet mask: Replace host ID portion by 0's and subnet part by 1's



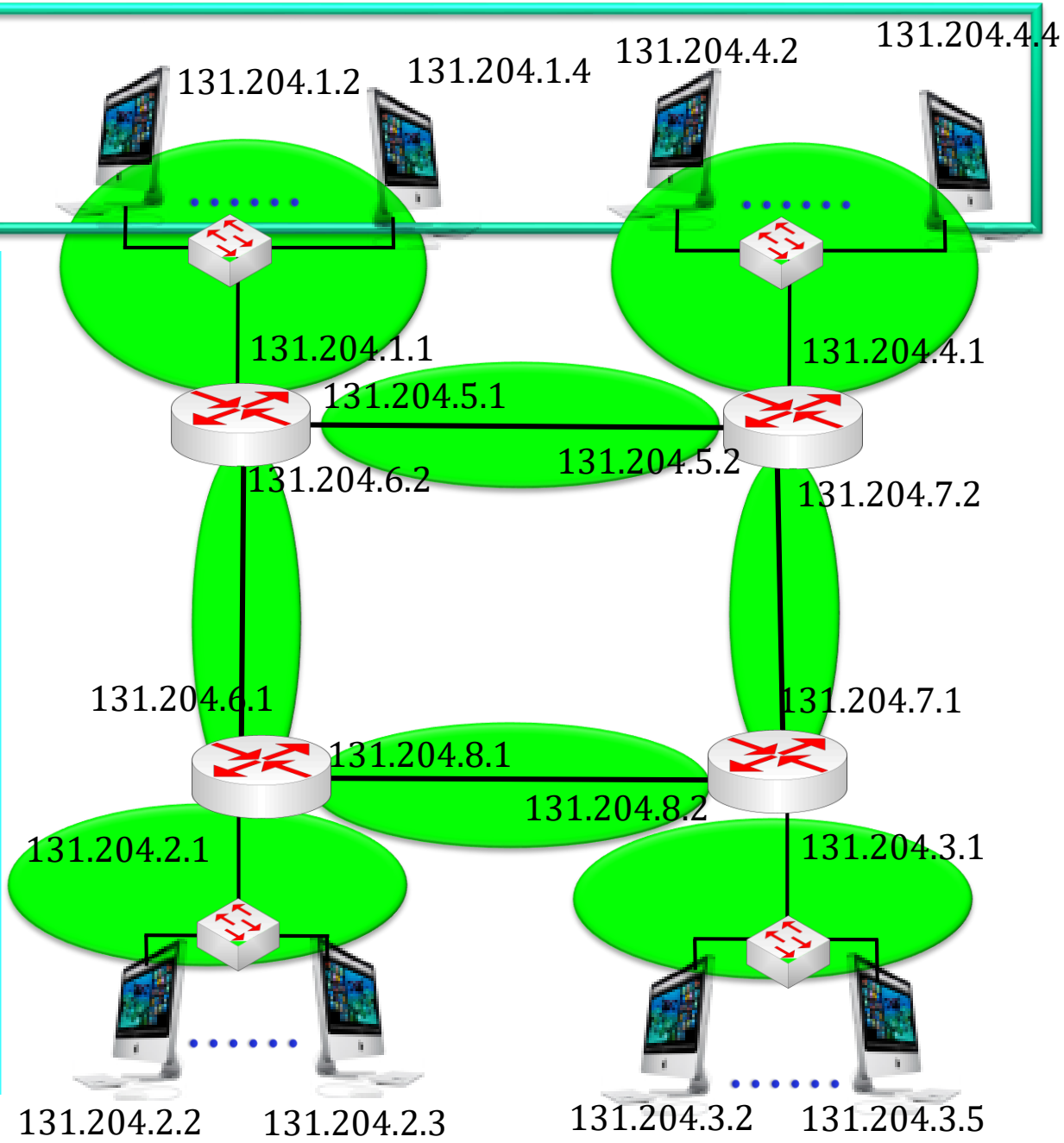
Subnet mask: 255.255.255.0

10000011 11001100 00000011 00000000

131.204.2.0/24

Subnets

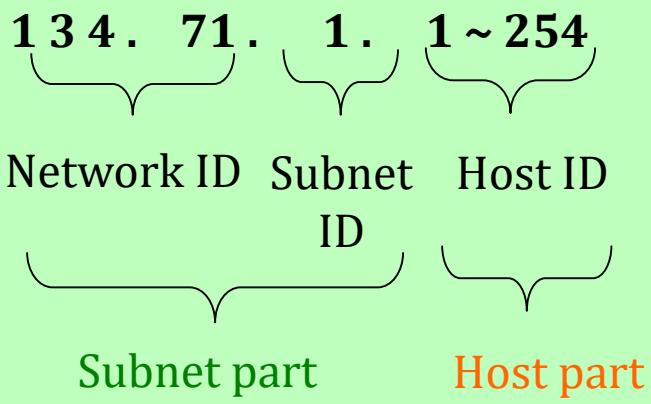
- 8 subnets
  - 131.204.1.0/24
  - 131.204.2.0/24
  - 131.204.3.0/24
  - 131.204.4.0/24
  - 131.204.5.0/30
    - 2 gateways
  - 131.204.6.0/30
    - 2 gateways
  - 131.204.7.0/30
    - 2 gateways
  - 131.204.8.0/30
    - 2 gateways



Class A, B, C, and D

- Class A
  - Network ID: 1 byte
  - MSBit = 0
  - 0<MSByte<127
- Class B
  - Network ID: 2 byte
  - MSBs = 10
  - 128≤MSByte≤191
- Class C
  - Network ID: 3 byte
  - MSBs = 110
  - 192≤MSByte≤223
- Class D
  - MSBs= 1110
  - 224≤MSByte≤239 (Multicast)

CPP IP address:



CPP:

- 134.71.255.255 (broadcast address to every IP address in cpp.edu)
- 134.71.0.0 (refer to CPP network)

Special Address

- 127.0.0.0 through 127.255.255.255 for loopback purposes
  - Localhost
  - The adapter/NIC intercepts all loopback messages and returns them to the sending application
  - A true story: A hacker emailed a religious organization (the Church of Scientology) and told the administrator that the server was hacked and all info was stored at 127.0.0.1. The admin ftp to 127.0.0.1, found out everything was there and panicked at the loss
  - Here is the actual court deposition (quite funny but long) http://www.whyaretheydead.net/krasel/biased/biased.2.10.html
- Zero Addresses
  - As with the loopback range, the address range from 0.0.0.0 through 0.255.255.255 should not be considered part of the normal Class A range.
  - 0.x.x.x addresses serve no particular function in IP, but nodes attempting to use them will be unable to communicate properly on the Internet
  - 0.0.0.0 means any IP address

Private IP Addresses

- Private Addresses
  - Relieve the shortage of IPv4 addresses
  - The IP standard defines specific address ranges within Class A, Class B, and Class C reserved for use by private networks
  - The table below lists these reserved ranges of the IP address space
- Hosts are effectively free to use addresses in the private ranges if they are not connected to the Internet
- If they reside behind firewalls or other gateways that use Network Address Translation (NAT)
  - Network Address Port Translation (NAPT) Operation (RFC 3022)
- Private IP addresses are blocked by an ISP router's firewall

Class	Private start address	Private finish Address
A	10.0.0.0	10.255.255.255
B	172.16.0.0	172.31.255.255
C	192.168.0.0	192.168.255.255

Classless Inter Domain Routing (CIDR)

- Eliminating class limitations of a network ID
- Subnet portion of the address of an arbitrary length
- Address format: a.b.c.d/x, where x is # of bits in subnet part of IP address
- Subnet mask: use all 1's for the subnet part and 0's for the host part
- The number of hosts that can be in this subnet: $2^9 - 2$ hosts

131.204.2.0/23

CIDR Benefits

- Dramatically reduce the size of routing tables in Internet core routers
  - Most organizations received multiple class C addresses, which are continuous
  - Multiple entries become a single entry in routing tables for one organization
- The CIDR is the representation used for configuring network equipment, such as a router and firewall

Aggregate Multiple Class C Nets

- Continuous Class C: 193.1.0.0/24 to 193.1.3.0/24
  - 0: 00000000
  - 3: 00000011

```
193.1. 00000000.00000001
.....................................................
193.1. 00000011.
```

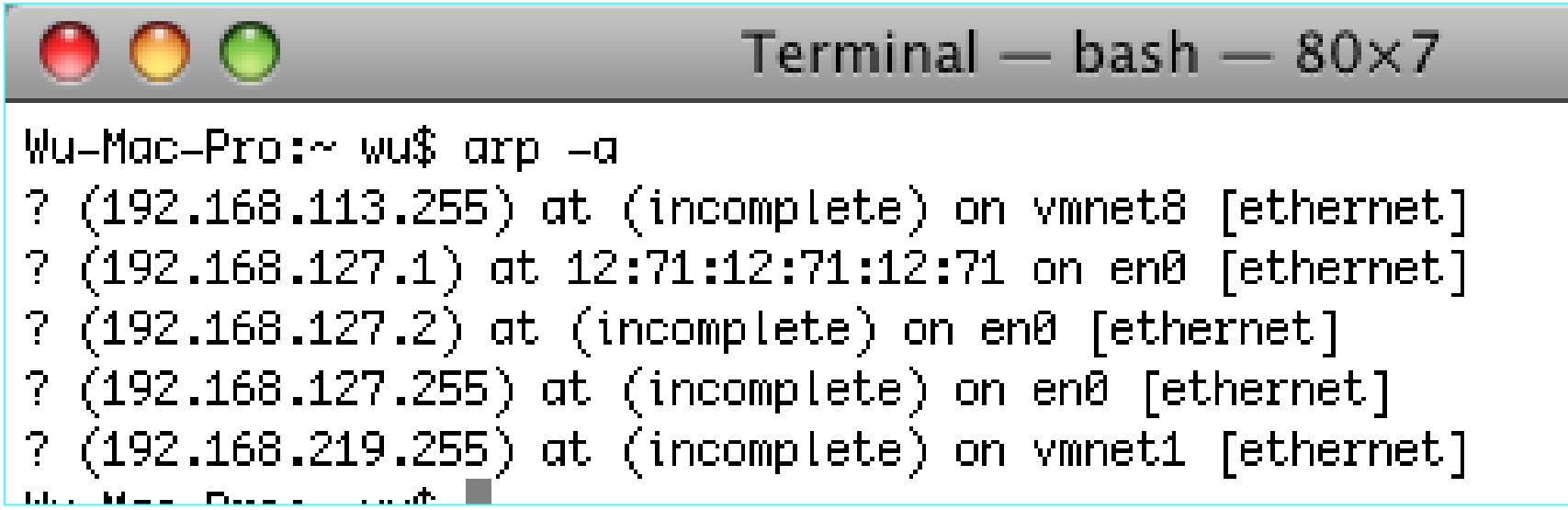
- CIDR: 193.1.0.0/22: 10-bit host ID (freedom bits)
  - 2^{10} – 2 hosts (can be used to assign subnet ID and host ID)
  - Only one entry in the routing table
- Subnet mask 255.255.11111100.0 (255.255.252.0) (replace host ID portion by 0's and subnet part by 1's)
- Broadcast address
  - 193.1.00000011.11111111 = 193.1.3.255 (replace host ID part by 1's)

Use of Subnet Mask in Host

- A host uses subnet mask to extract the subnet part of its IP address
- A host uses subnet mask to extract the subnet part of the destination IP address
- Compare to see if both subnet parts are the same
  - If they are the same, use the ARP to obtain the destination MAC address
  - If they are different, then packet must be sent to the gateway (router), and use ARP to obtain the router's MAC address

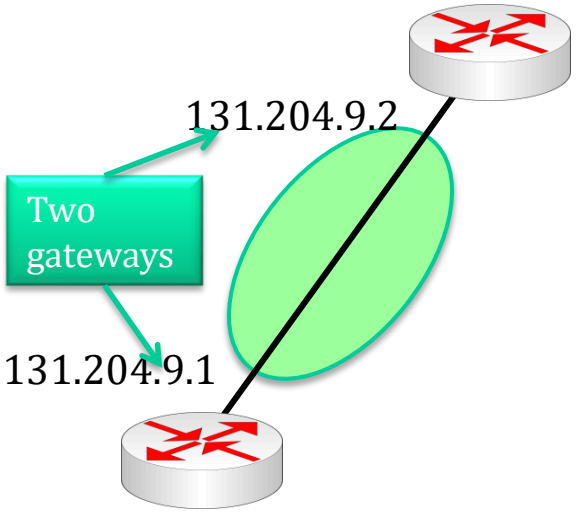
ARP

- Each host has an ARP cache
- 192.168.127.1 is the gateway to the Internet and its MAC address is 12:71:12:71:12:71



Example: Assign Subnet Mask for Routers Forming a Subnet

- The number of interfaces: 2
- 2 bits for host IDs
- $2^2 - 2 = 2$ interfaces
- Subnet mask = 255.255.255.252
- CIDR: 131.204.9.0/30
- Two gateways:
  - 131.204.9.1
  - 131.204.9.2



Example: CIDR and Subnet

131.204.2.00010000: IP address representing the subnet

- Gateway:131.204.2.00010001
- Hosts:

131.204.2.0001 0010

.

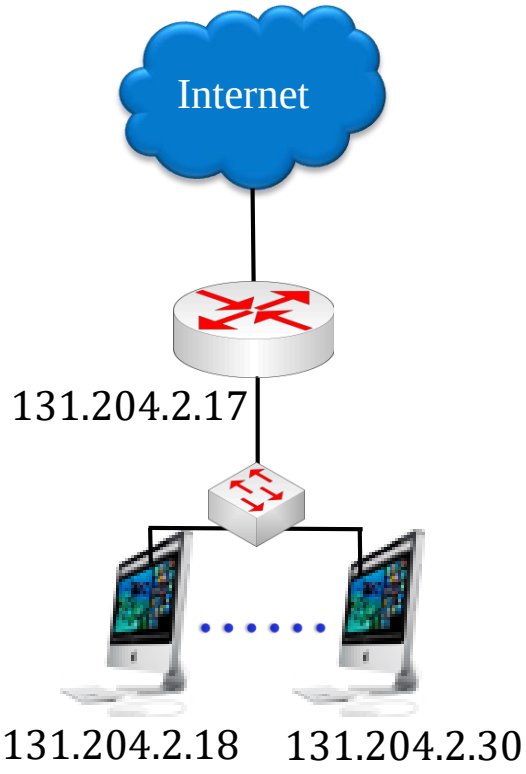
.

.

131.204.2.0001 1110

13 hosts

- CIDR: 131.204.2.16/28

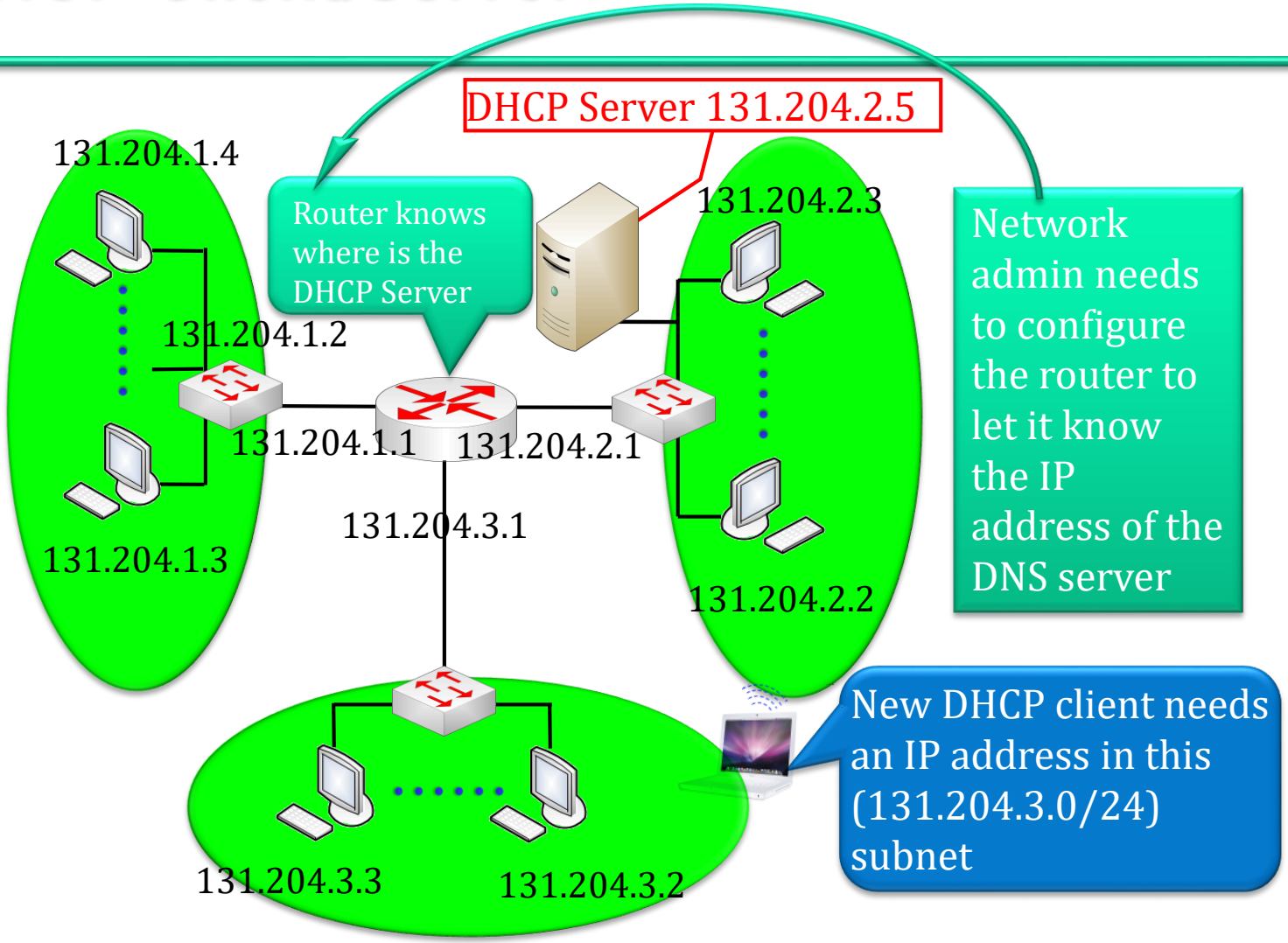


131.204.2.00011111: broadcast

DHCP: Dynamic Host Configuration Protocol

- Function:
  - Allow host to dynamically obtain its IP address from DHCP server when joining network
  - Allow IP address reuse
- Support wired/mobile stations joining network
- Host holds an IP address only while actively connected
- Renew IP address already in use
- DHCP provides:
  - IP address
  - Subnet mask
  - Gateway IP address
  - DNS server IP address

DHCP Client/Server

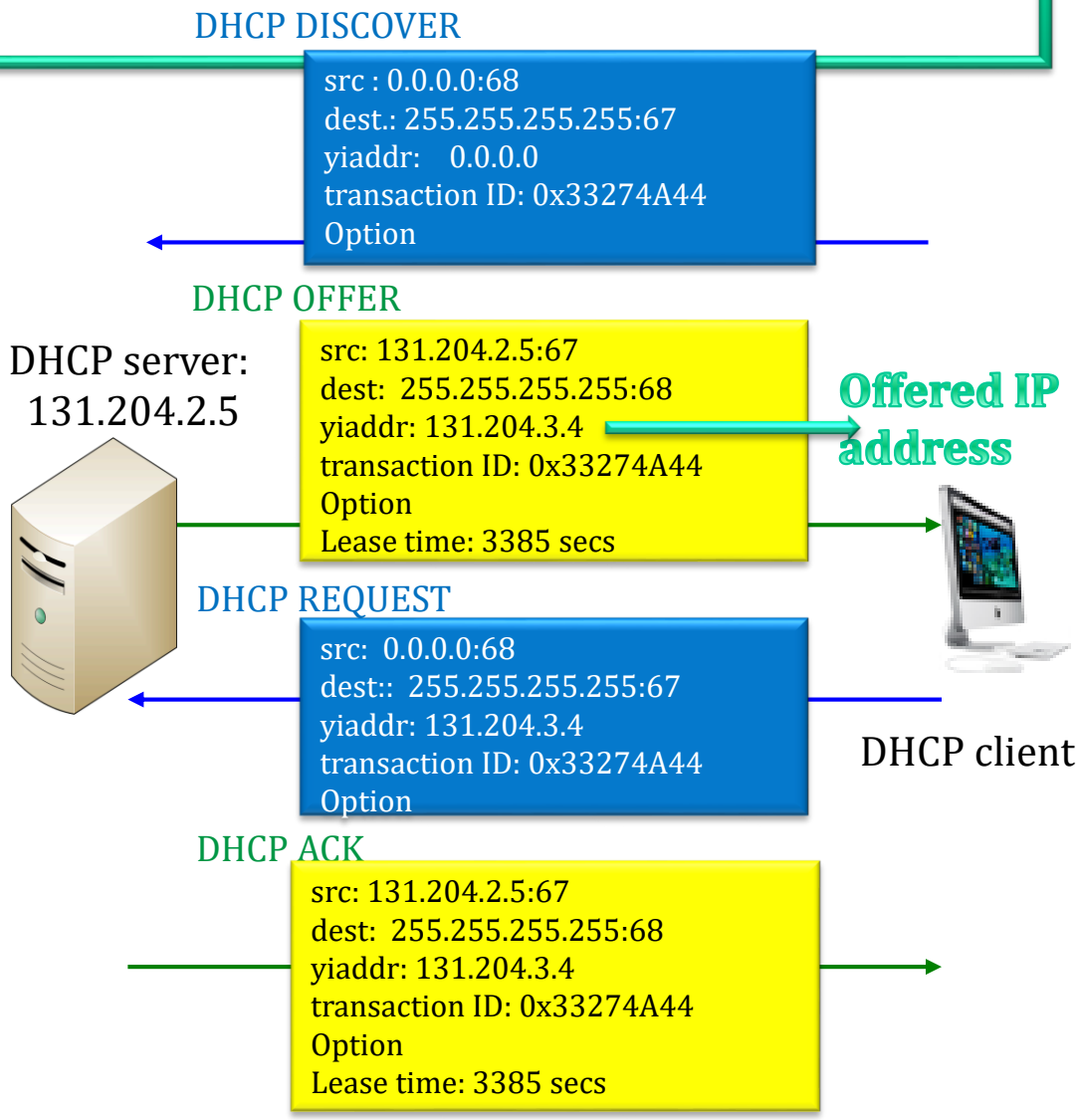


DHCP Client/Server

DHCP procedure:

1. Host broadcasts "DHCP Discover" message
2. DHCP server responds with "DHCP offer" message
3. Host requests IP address: "DHCP request" message
4. DHCP server sends address: "DHCP ACK" message

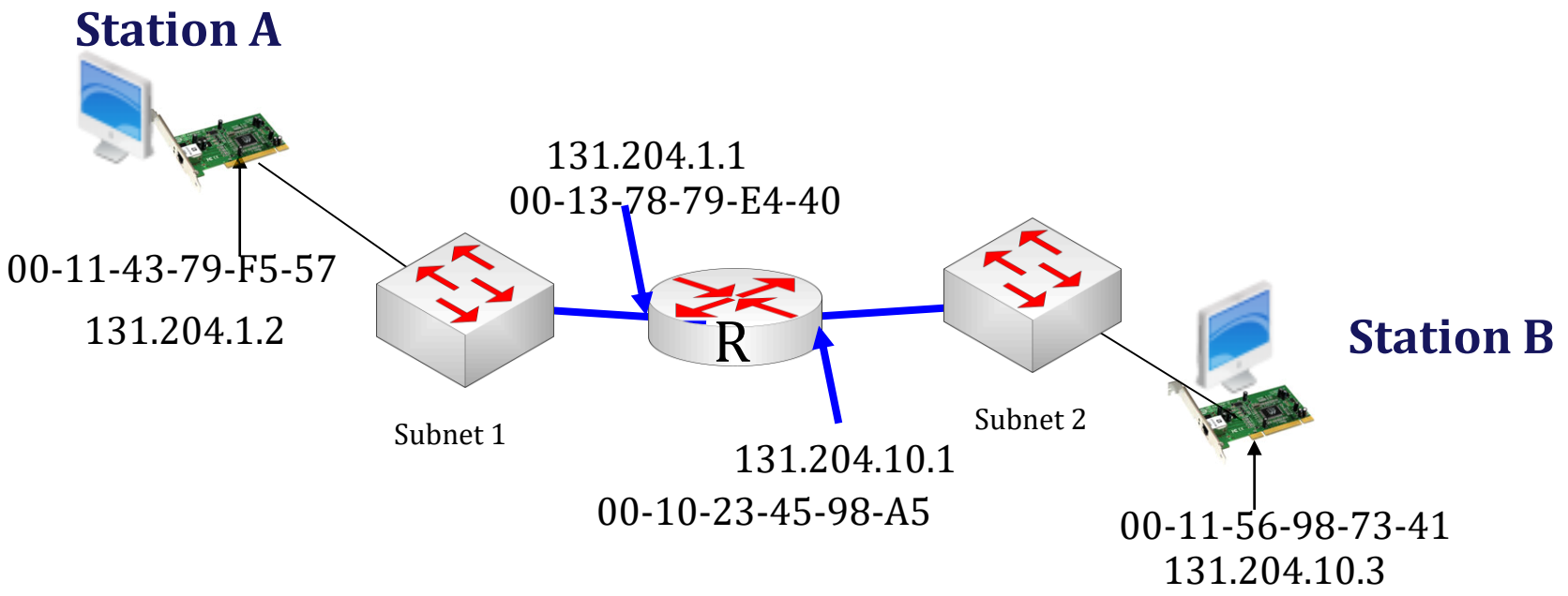
yiaddr: (offered) your IP address



Reuse a previously allocated network address (1)

- Most DHCP servers are configured to let a client reuse a previously allocated network address
  - This can reduce the amount of broadcast traffic resulting from a DHCP DISCOVER message and a DHCP OFFER message
  - If a client remembers and wishes to reuse a previously allocated network address, a client may choose to omit the DHCP discover message
  - The client broadcasts a DHCPREQUEST message on its local subnet. The message includes the client's network address in the 'requested IP address' option
- If the client used a 'client identifier' to obtain its address, the client must use the same 'client identifier' in the DHCPREQUEST message
  - chaddr may be used both as a hardware address for transmission of DHCP reply messages and as a client identifier
- Servers with knowledge of the client's configuration parameters respond with a DHCPACK message to the client

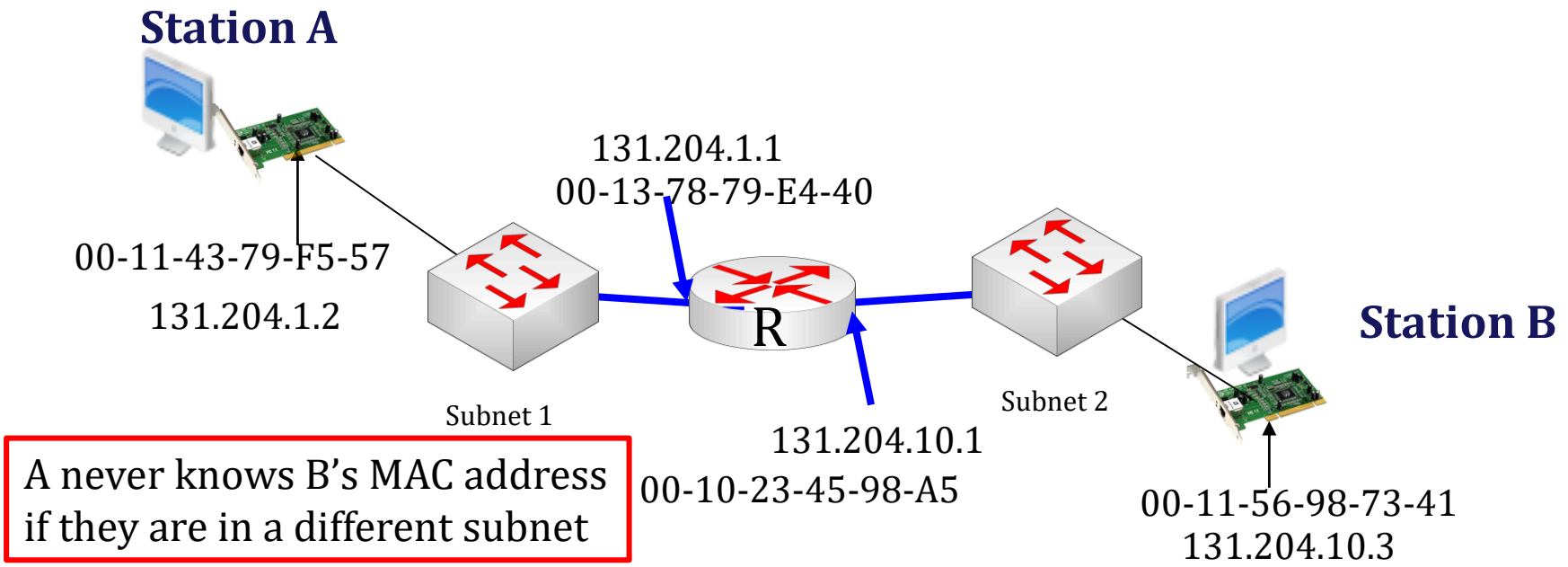
Routing to Another LAN



- Walkthrough:
  - Send datagram from Station A to Station B via router R
  - Station A knows B's IP address
- Two ARP tables in router R
  - One for each IP network (LAN)

- Station A creates an IP datagram with source A, destination B
- Station A uses ARP to get R's MAC address for 131.204.1.1 that is obtained by DHCP
- Station A creates a link-layer frame with R's MAC address as dest MAC address, and frame contains A-to-B IP datagram
- A's NIC sends frame
- R's NIC receives frame
- R removes IP datagram from Ethernet frame, sees its destined to Station B
- R forwards the packet to Interface 131.204.10.1 interface based on routing table
- R uses ARP to get B's MAC address
- R creates frame containing A-to-B IP datagram and sends to B using B's MAC address as the destination MAC address

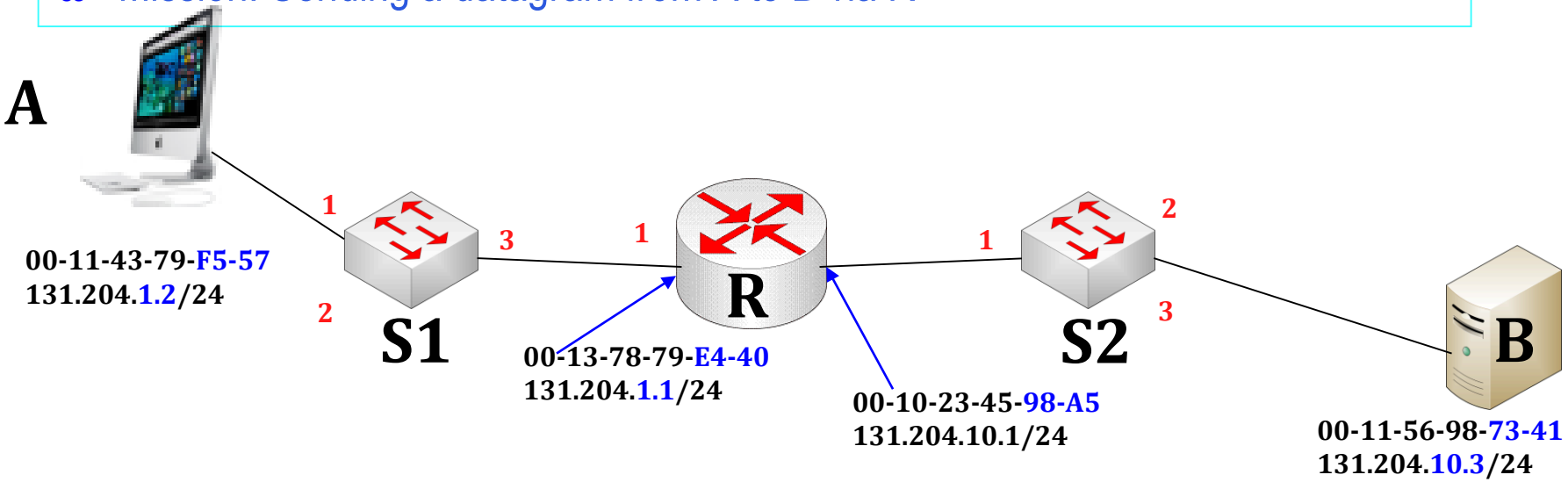
This is a really important example

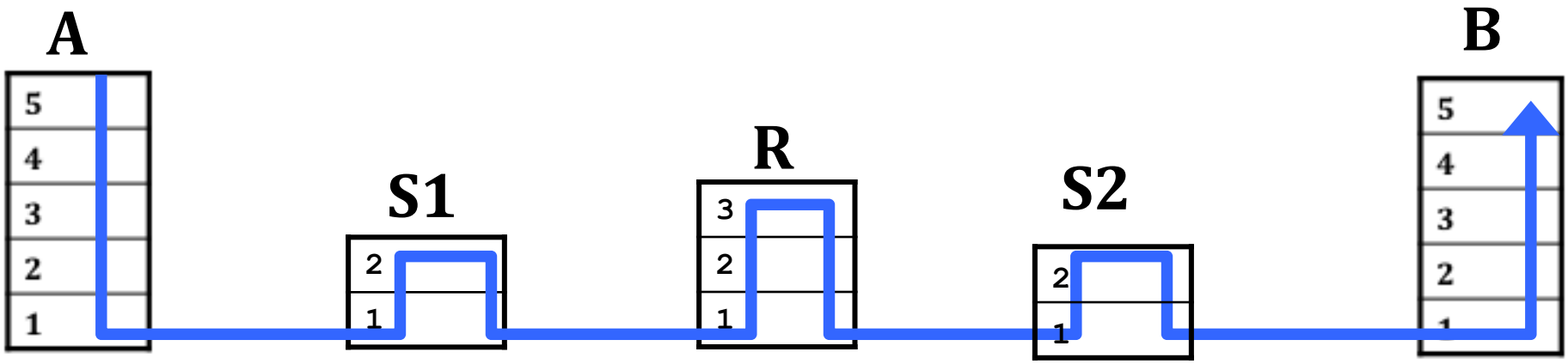


A never knows B's MAC address if they are in a different subnet

Sending Datagram to Another Subnet/LAN

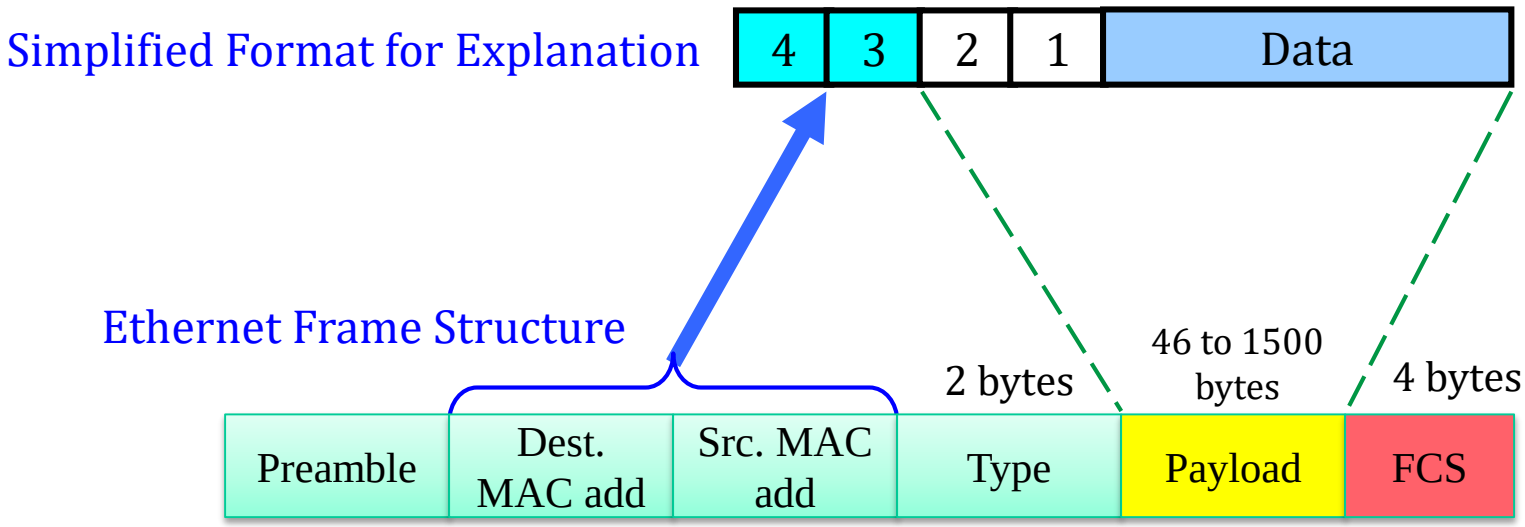
- Mission: Sending a datagram from A to B via R



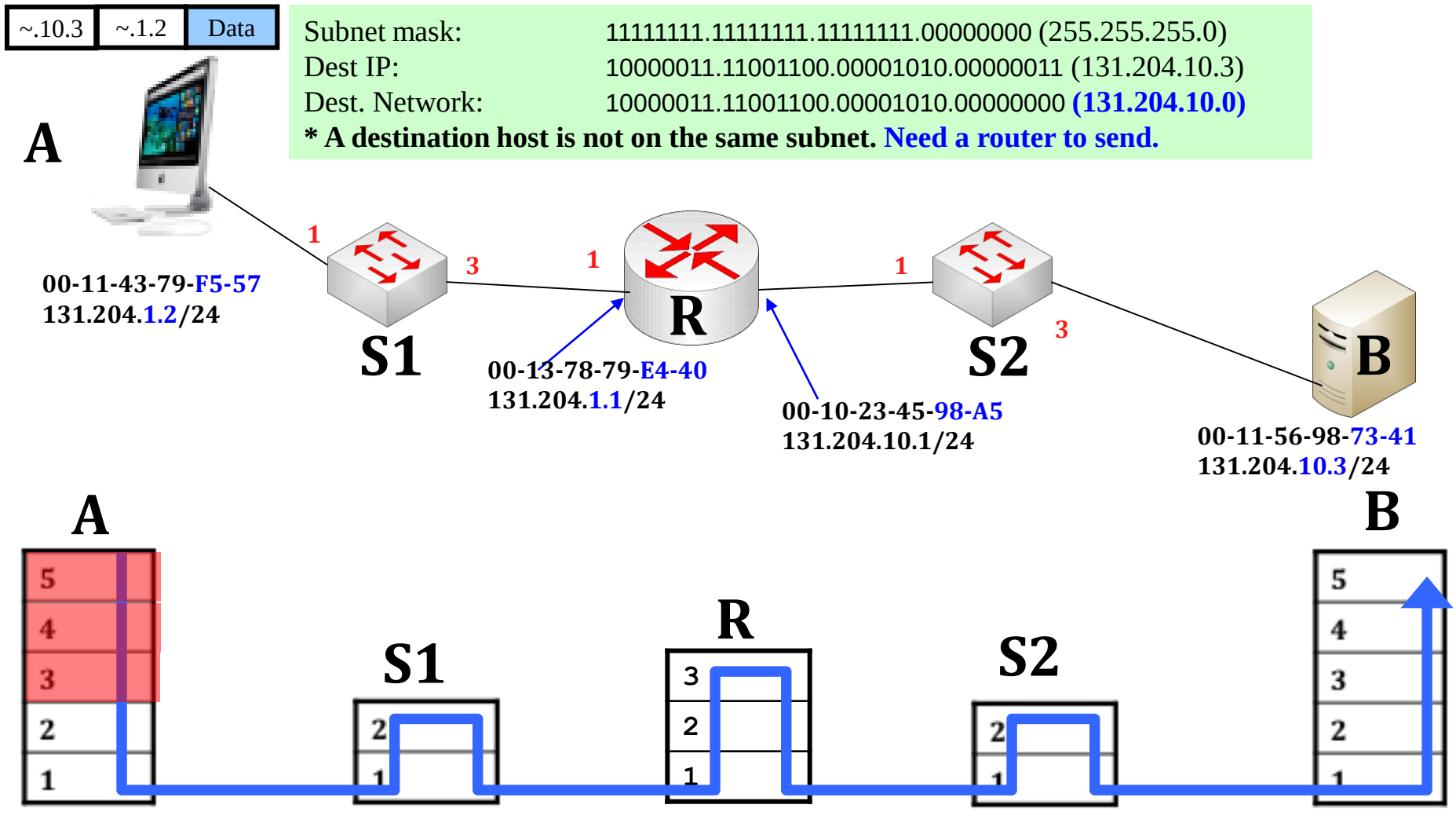


Simplified Data Format

- Source IP
- Destination IP
- Source Mac Address
- Destination Mac Address

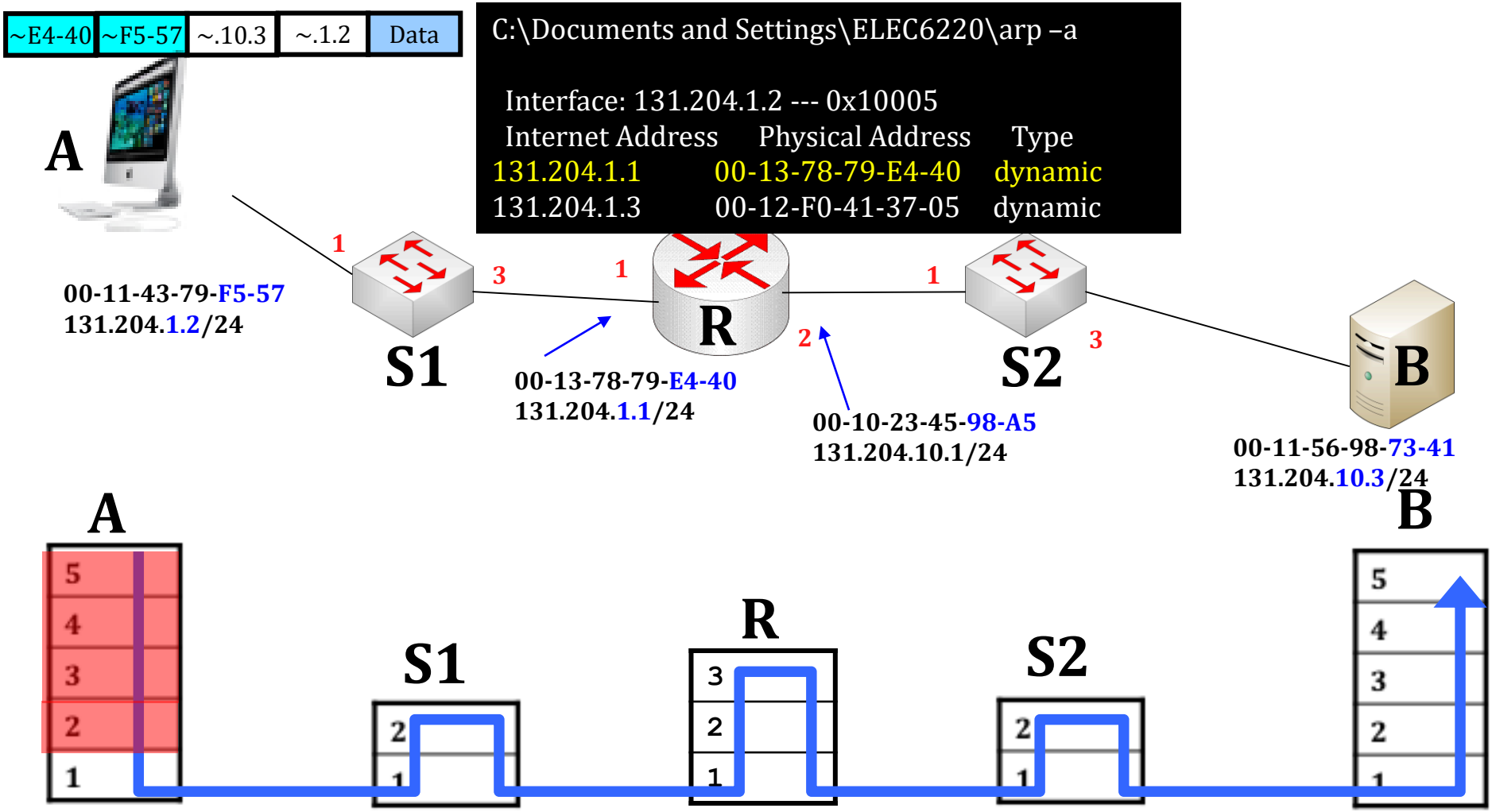


Host A (Network Layer)
- Set ***source & destination IP*** in datagram format
- A datagram will be forwarded to a gateway (router) using the subnet portion of the network ID since the destination host is not on the same subnet



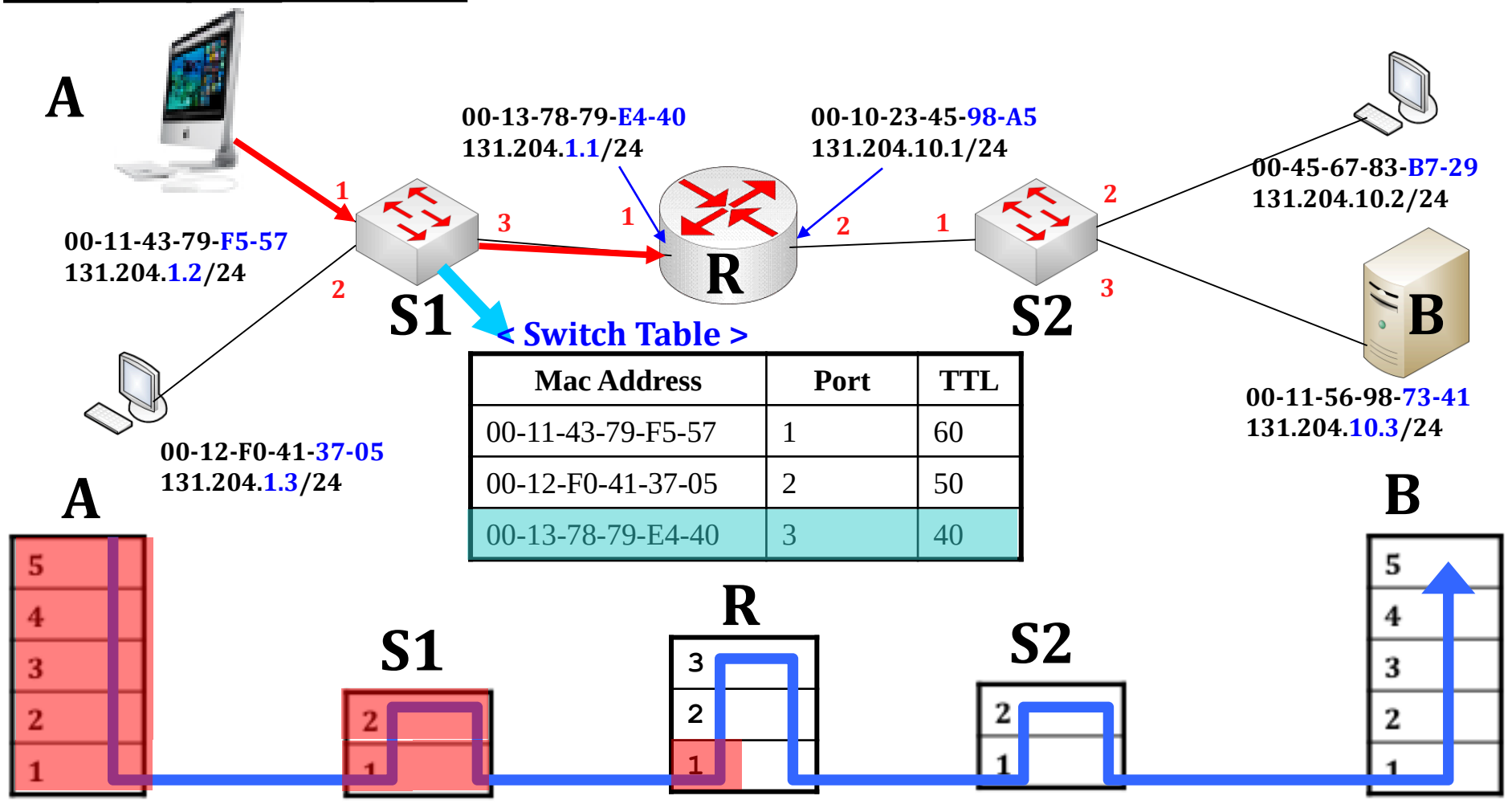
Host A (Data link layer)
- Set ***source & destination Mac address*** in an Ethernet frame by means of an ARP table

< ARP Table Information>



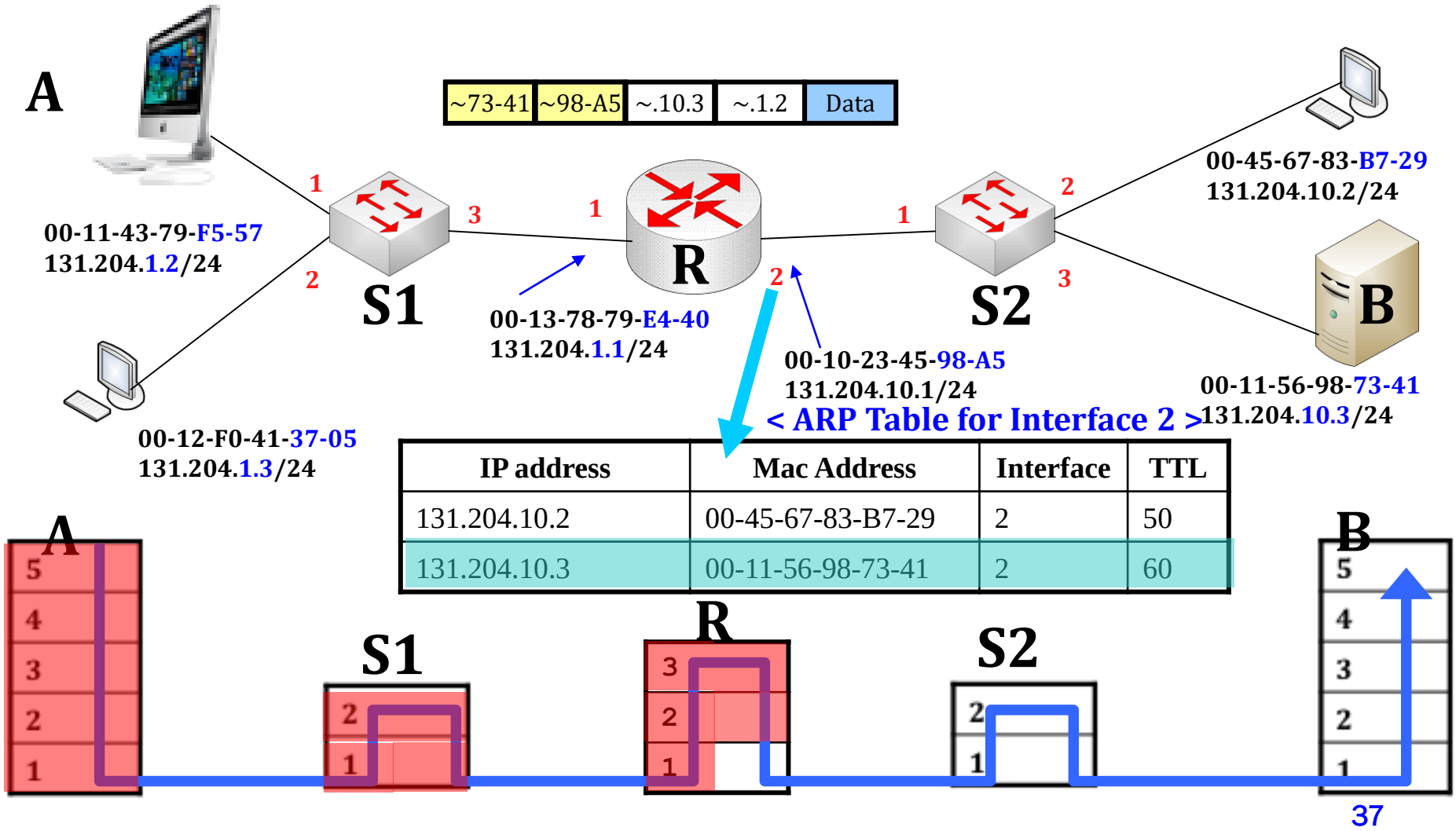
Host A : A's NIC sends an Ethernet frame
Switch S1
- S1's NIC of interface 1 receives an Ethernet frame
- ***Look up a switch table*** & ***Forward a frame*** into a corresponding port

~E4-40	~F5-57	~.10.3	~.1.2	Data

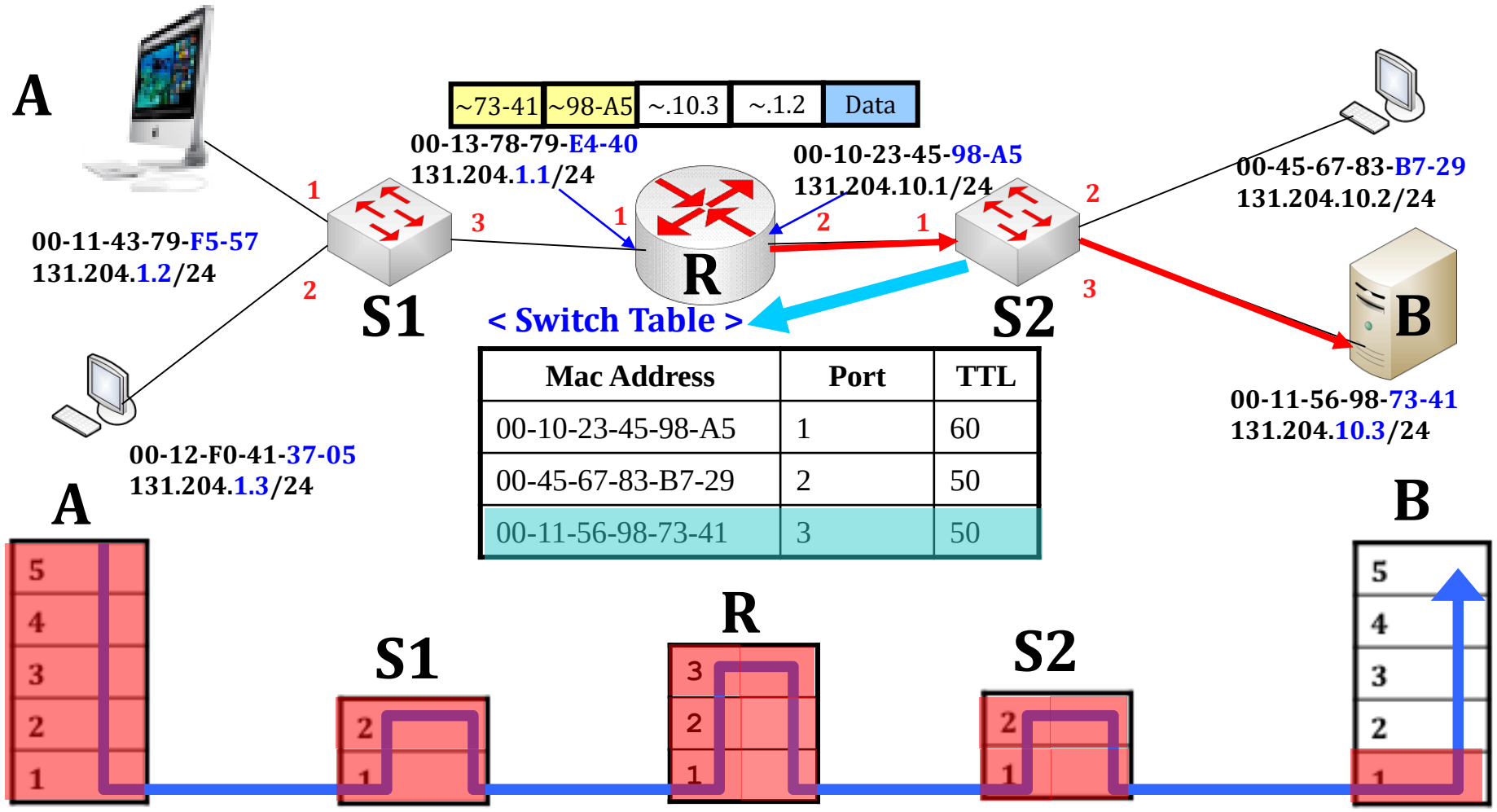


Mac Address	Port	TTL
00-11-43-79-F5-57	1	60
00-12-F0-41-37-05	2	50
00-13-78-79-E4-40	3	40

Router R
- Look up a routing table & determine a next forwarding link (Layer 3)
- Change MAC addresses for next hop (Layer 2)

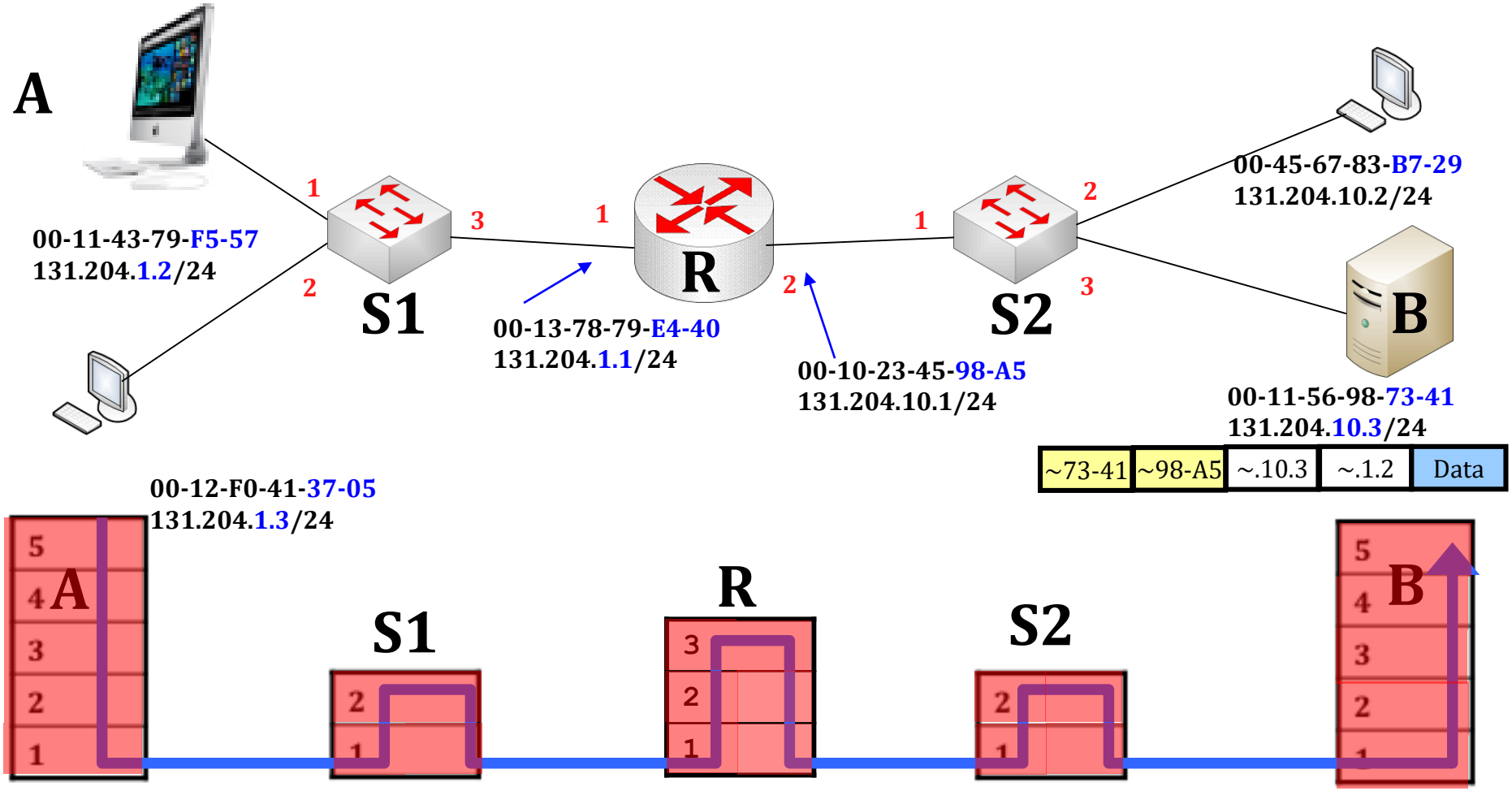


Router R : R's interface sends an Ethernet frame
Switch S2
- S2's NIC for port 1 receives an Ethernet frame.
- ***Look up a switch table*** & ***Forward a frame*** into corresponding port 3



Mac Address	Port	TTL
00-10-23-45-98-A5	1	60
00-45-67-83-B7-29	2	50
00-11-56-98-73-41	3	50

Host B (From layer 1 to layer 5)
- Check if destination addresses are the same as those of itself.
  (Data link layer – MAC Address, Network layer – IP)
- Each layer drops its own header and trailer.



Network Address Translation (NAT)

- Goals
  - A private network uses just one IP address provided by ISP to connect to the Internet
  - Private networks use private IP addresses provided by IETF
  - Can change address of devices in private network without notifying outside world
  - Can change ISP without changing the address of devices in private network
  - Devices inside private network are not explicitly addressable by external network, or visible by outside world (a security plus)
- A NAT device would not advertise private networks to the external/public network; however, the external/public network services may be advertised within the private network

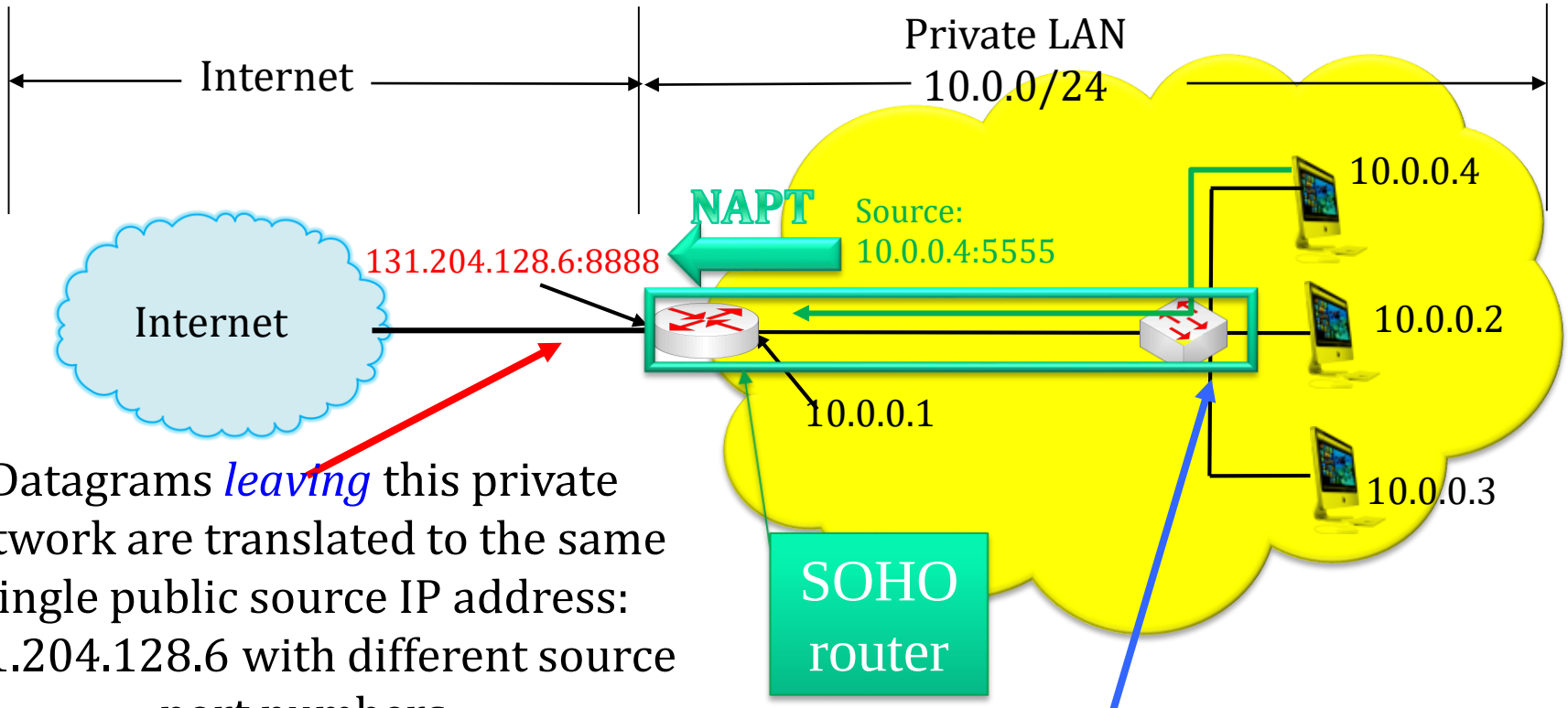
NAT and NAPT

- Basic NAT
  - A block of external/public IP addresses are set aside for translating the addresses of hosts within a private domain as they originate sessions to the external domain
    - For packets outbound from the private network, the source IP address and related fields such as IP, TCP, UDP and ICMP header checksums are translated
    - For inbound packets, the destination IP address and the checksums as listed above are translated
  - However, multiple external/public IP addresses are difficult to obtain due to the shortage of IPv4 addresses
- Network Address Port Translation (NAPT)
  - The NAPT also translates transport identifiers, e.g., TCP and UDP port numbers as well as ICMP query identifiers

NAPT: Network Address Port Translation

- This permits the transport identifiers of a number of private hosts to be multiplexed into the transport identifier of a single external/public IP address
  - The IP address binding extends to transport level identifiers such as TCP/UDP ports
  - Address binding is done at the start of a session
- For most of the small office and home (SOHO) routers
  - The private network usually relies on a single IP address, supplied by the ISP to connect to the Internet
  - SOHO can change ISPs without changing the private IP addresses of the devices within the network
- The terms NAT and NAPT are used interchangeably in the literature
  - However the RFCs, such as RFC 3022, use the term NAPT when port numbers are involved in translation
  - Cisco refers to NAPT as PAT, i.e. Port Address Translation

NAPT (1)



Datagrams *leaving* this private network are translated to the same single public source IP address: 131.204.128.6 with different source port numbers

No NAPT for datagrams within this private network that have 10.0.0.0/24 IP address as a destination

NAPT (2)

- A NAPT router performs the following for every outgoing datagram
  - Replace (source IP address, port #) with (Public IP address, new port #)
  - Remote clients/servers will respond using (Public IP address, new port #) as destination addr
  - Remember (in NAPT translation table) every (source IP address, port #) to (Public IP address, new port #) translation pair
- A NAPT router performs the following for every incoming datagram
  - Replace (Public IP address, new port #) in destination fields of every incoming datagram with corresponding (source IP address, port #) stored in NAPT table
- 10.0.0.4:5555 represents, for a host, an IP address and port number separated by a ":"
  - 10.0.0.4:5555 is referred as a transport address in the RFCs

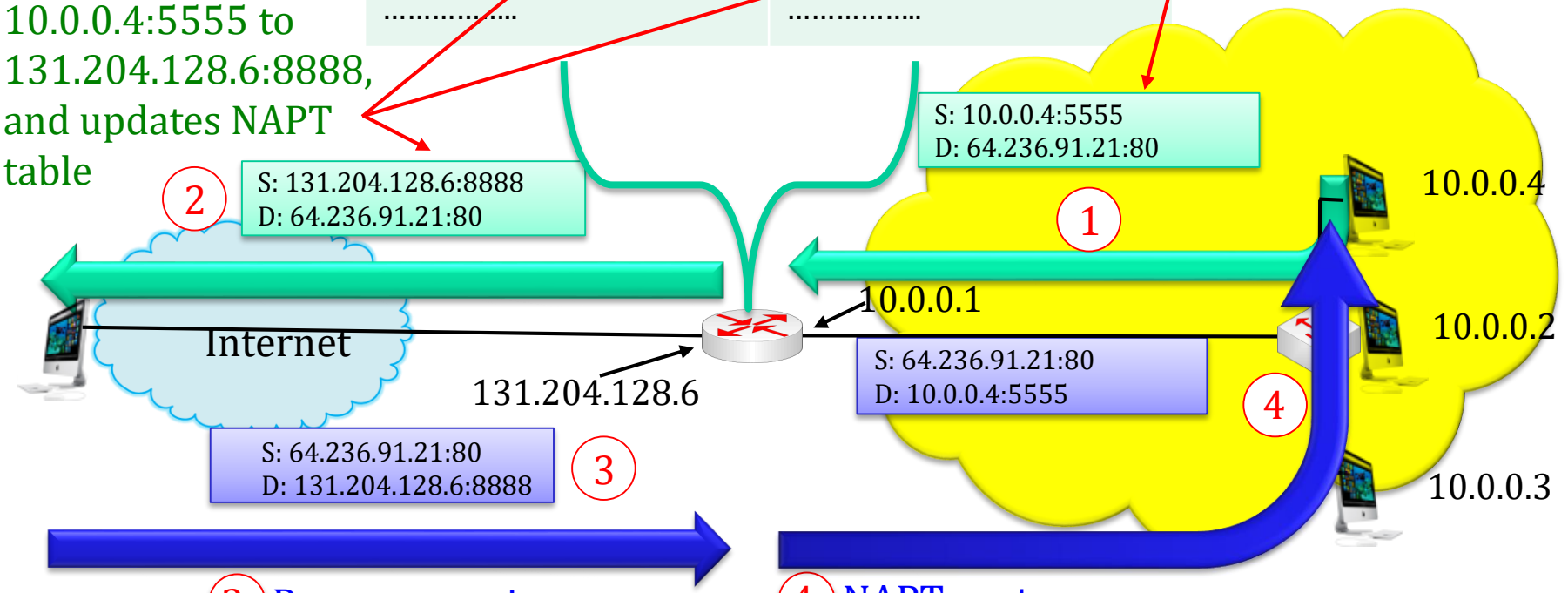
NAPT (3)

NAPT translation table

Public IP Address/Port	Private IP Address/Port
131.204.128.6:8888	10.0.0.4:5555
……	……

2 NAPT router translates datagram source from 10.0.0.4:5555 to 131.204.128.6:8888, and updates NAPT table

1 Host 10.0.0.4:5555 sends datagram to 64.236.91.21:80



3 Response arrives dest. address: 131.204.128.6:8888

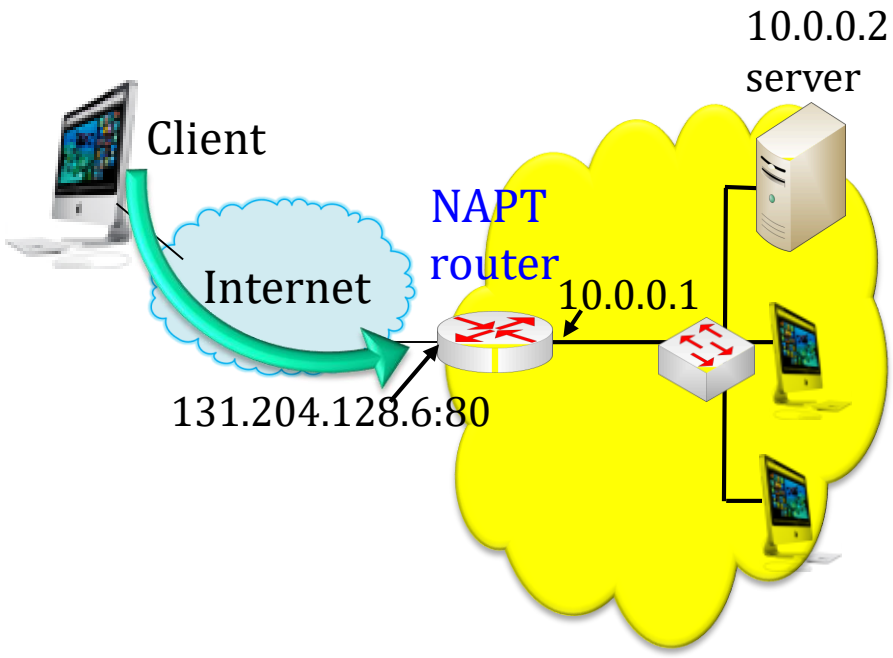
4 NAPT router translates datagram Dest from 131.204.128.6:8888 to 10.0.0.4:5555

NAPT (4)

- 16-bit port-number field:
  - 64,000 simultaneous connections with a single public IP address
- NAPT is controversial:
  - Routers should only process up to layer 3
  - Violates layer 3 limit
    - NAPT modification of port number must be taken into account by app designers, e.g., P2P applications
    - Security protocols must take care of NAPT
  - Address shortage should instead be solved by IPv6

NAPT for Incoming Requests

- NAPT router blocks all incoming ports by default
- Many applications have had problems with NAPT in the past in their handling of incoming requests
- Four major methods
  - 1. Application Level Gateways (ALGs)
  - 2. Static port forwarding
  - 3. Universal Plug and Play (UPnP) Internet Gateway Device (IGD) protocol
  - 4. Traversal Using Relays around NAT (TURN)

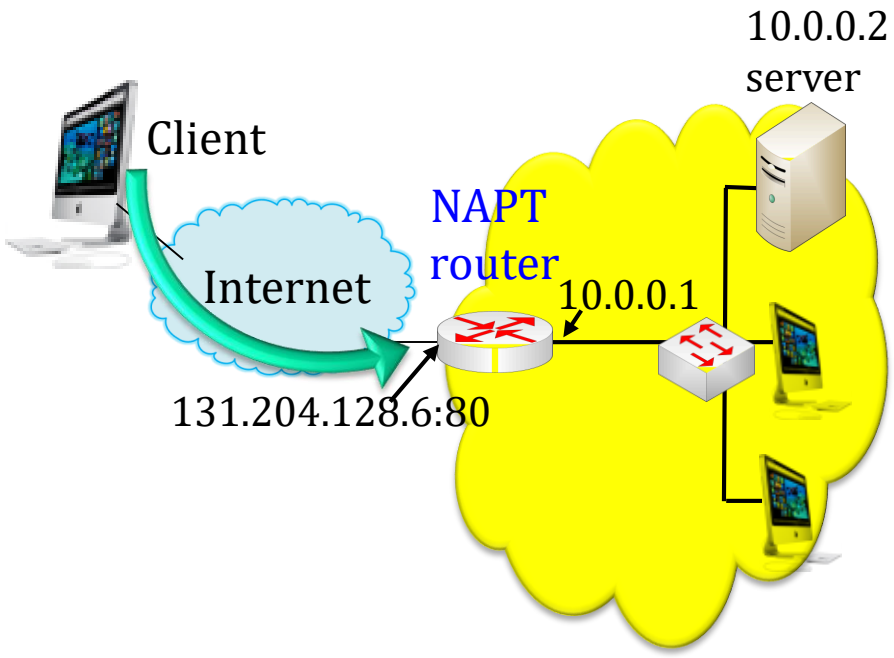


The application Level Gateways or Application Layer Gateways (ALGs)

- The application level gateways or application layer gateways (ALGs) have been embedded in NAT firewall/router products to mitigate the NAPT problem
  - ALGs perform the application layer functions required for a particular protocol to traverse a NAT device
  - The ALGs are application specific translation agents that allow an application on a host in one address realm to transparently connect to its counterpart running on a host in different realm
    - Typically, this involves rewriting application layer messages in the packet payload to contain translated IP addresses/port numbers, rather than the ones inserted by the sender
    - An ALG may interact with the NAT device to set up state, use the NAT state information, modify the application specific payload and perform all the necessary operations required for the application running across address realms

The Static Port Forwarding

- Punch a hole for incoming requests to servers
- A client wants to connect to server with address 10.0.0.2
  - Server address 10.0.0.2 is private for the LAN so client cannot use it as destination address
  - Only one public IP address: 131.204.128.6
- Static port mapping:
  - Statically configure NAPT to forward incoming connection requests at given port to server
  - e.g., (131.204.128.6, port 80) always forwarded to 10.0.0.2 port 80



Error Reporting and Diagnosis

- Examples of errors a router may encounter
  - Router does not know where to forward a packet
  - Packet's time-to-live field expires
- Best effort
  - Router just silently drop packets
- Network diagnosis
  - IP includes basic test and feedback for solving network problems
  - Internet Control Message Protocol (ICMP)
    - RFC 792: Internet Control Message Protocol
    - RFC 1122: Requirements for Internet Hosts -- Communication Layers
- ICMP runs on top of IP
  - In parallel with TCP, UDP and SCTP

Internet Control Message Protocol

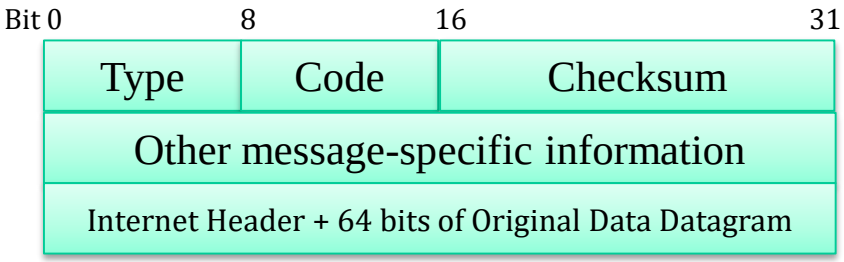
- Diagnostics
  - Triggered when an IP packet encounters a problem
    - E.g., time exceeded or destination unreachable
  - ICMP packet sent back to the source IP address
    - Includes the error information (e.g., type and code)
    - Excerpt of the original data packet for identification
  - Source host receives the ICMP packet
    - Inspects the excerpt of the packet (e.g., protocol and ports)
    - Inform the socket that should receive the error information

ICMP: Internet Control Message Protocol

Type	Code	Description
0	0	echo reply (ping)
3	0	Destination network unreachable
3	1	Destination host unreachable
3	2	Destination protocol unreachable (the designated transport protocol is not supported)
3	3	Destination port unreachable (the designated protocol is unable to inform the host of the incoming message)
3	6	Destination network unknown
3	7	Destination host unknown
4	0	source quench (congestion control - not used)
8	0	echo request (ping)
9	0	route advertisement
10	0	router discovery
11	0	TTL expired
12	0	bad IP header

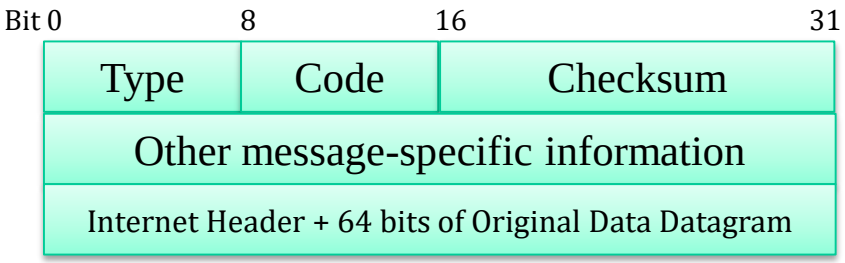
ICMP: Internet Control Message Protocol Packet Format (1)

- For hosts & routers to communicate network-level information
  - Error reporting: unreachable host, network, port, protocol
  - Echo request/reply (used by ping)
- ICMP message is carried in IP datagram as payload
- ICMP message: type, code plus first 8 bytes of IP datagram causing error



ICMP: Internet Control Message Protocol Packet Format (2)

- 2nd word (bit 32 to 63):
  - Echo or Echo Reply Message, Information Request or Information Reply Message: identifier (bit 32 to 47), Sequence Number (bit 48 to 63)
  - Redirect Message: Gateway Internet Address
  - Source Quench Message, Time Exceeded Message, Destination Unreachable Message: unused
  - Parameter Problem Message: Pointer (bit 32 to 39), unused (bit 40 to 63)



Echoes and Replies

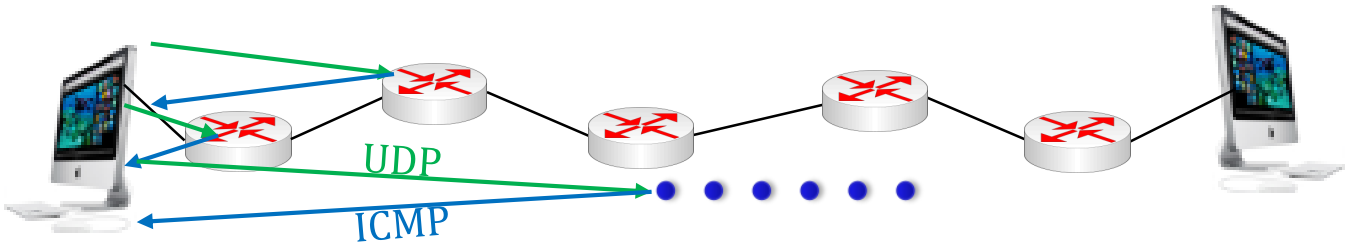
- The data received in the echo message must be returned in the echo reply message
  - The identifier and sequence number may be used by the echo sender to aid in matching the replies with the echo requests
  - For example, the identifier might be used like a port in TCP or UDP to identify a session, and the sequence number might be incremented on each echo request sent
  - The echoer returns these same values in the echo reply
- Code 0 may be received from a gateway or a host
  - If code = 0, an identifier to aid in matching echos and replies
    - Identifier may be zero
  - If code = 0, a sequence number to aid in matching echos and replies,
    - Sequence number may be zero

Traceroute Based on ICMP

- Source host sends a series of UDP packets in UNIX OS to destination host
  - Send the first UDP with TTL =1 for three times
  - Send the second UDP with TTL=2 for three times
  - .....
  - Send the ith UDP with TTL=i for three times
  - All UDP packets use a port number that is not opened as a service at the destination host
- When ith datagram arrives at the ith router:
  - The router discards datagram
  - Sends an ICMP message to source host (type 11, code 0: TTL expired)
  - Message includes name of router & IP address

Traceroute Based on ICMP

- When ICMP message arrives, source host calculates RTT (round trip time)
- UDP packet eventually arrives at destination host
- Destination host returns ICMP port unreachable message (type 3, code 3)
- When source host gets this ICMP message, it stops the sending of UDP packets



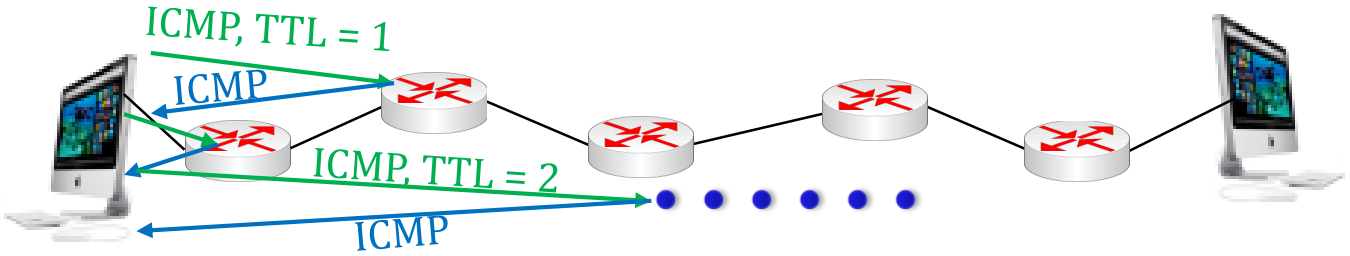
Traceroute is a security risk and not allowed in most networks

ICMP Risks

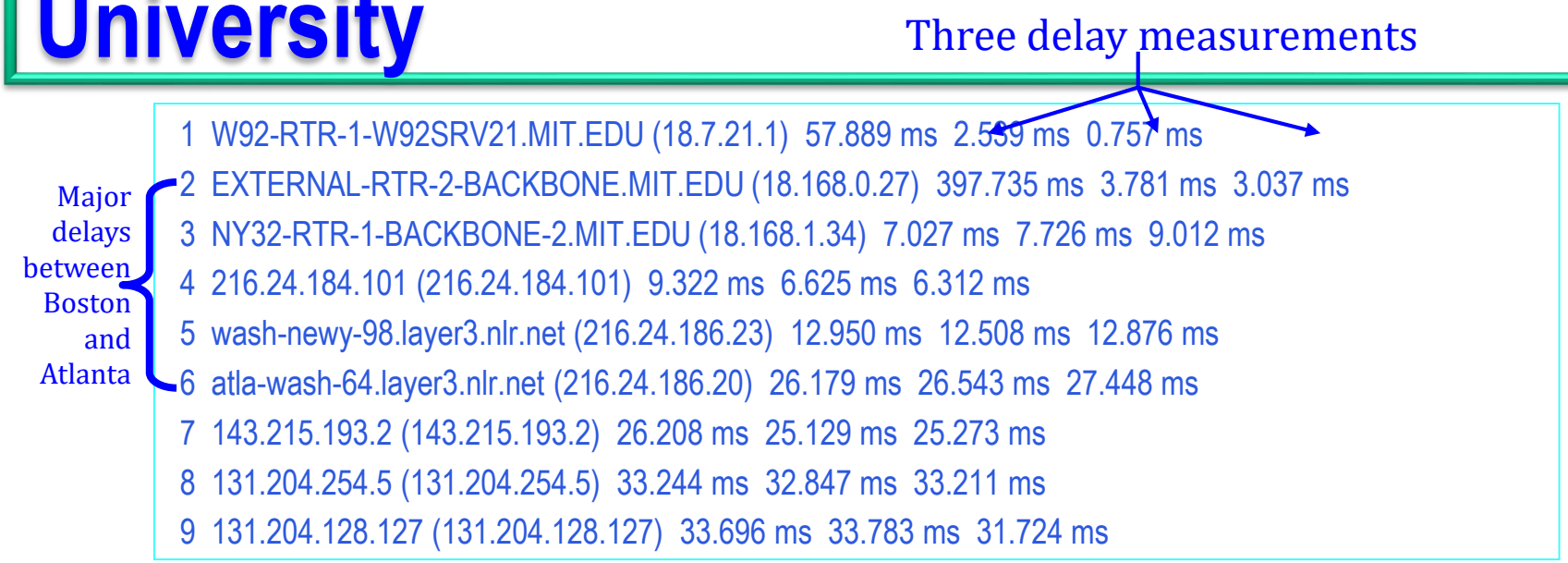
- ICMP can be used to scan hosts and available ports (services)
  - Turn off ICMP in Firewalls, including host firewalls
- Traceroute can be used to discover the router interface IP addresses
  - Attacks can be directed toward a particular interface
- To protect hosts and routers, ICMP responses must be turned off
  - Reconnaissance is useless if a host does not respond

Windows Traceroute

- Use ICMP packets



Traceroute from MIT to Auburn University



Using the site at http://bs.mit.edu:8001/cgi-bin/traceroute?JIGSAW-SESSION-ID=J-1556502962-2546

Try it by yourself