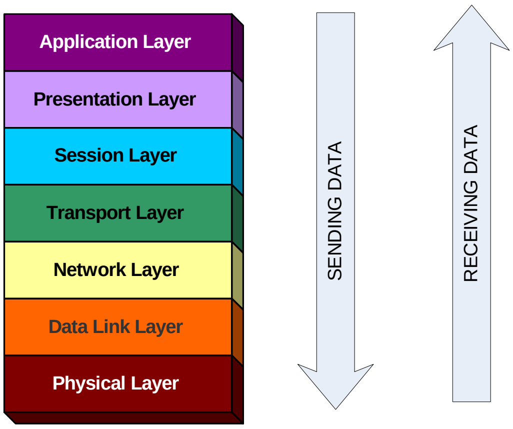
# Introduction to TCP/IP

Achmad Husni Thamrin
SOI-ASIA OW 2005

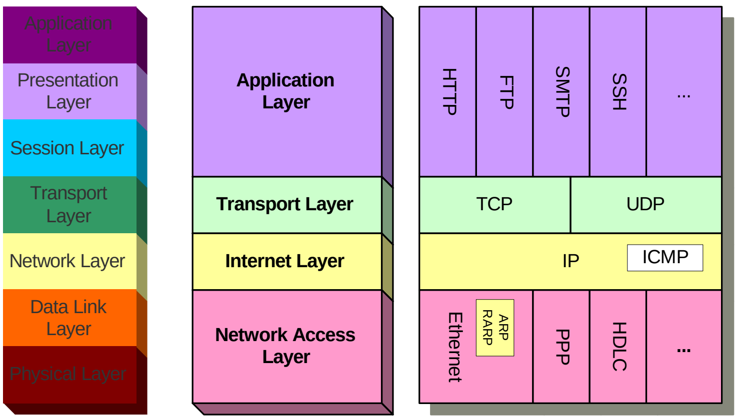
# Agenda

- TCP/IP Architecture
- IPv4 & IPv6
- Addressing
- Packet Format
- Address Resolution
- Routing
- ICMP
- Internet Service

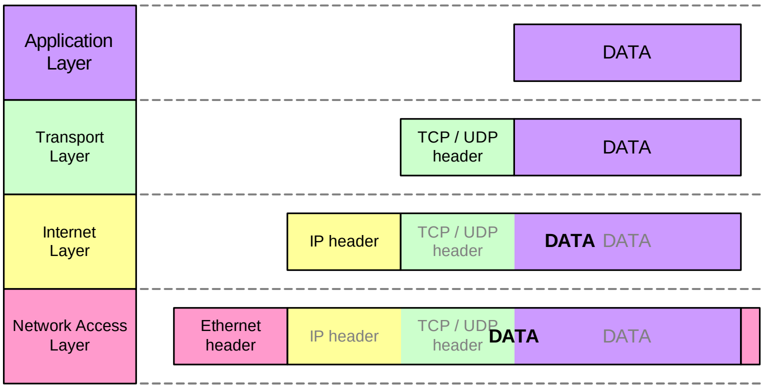
# Open System Interconnection Model




# TCP/IP Hierarchy



# Data Encapsulation




# Application Layer

- Many application protocols: standard, proprietary
- Client-server model
  - Server application runs on a host, accepting connections from clients
- Use TCP, or UDP, or both

## Transport Layer

- Provide transparent transfer of data between end users
- Use *ports* to pass data to the correct application in a host
- TCP
  - Connection oriented
  - Reliable
  - Error correction
- UDP
  - Connectionless
  - Best effort

## Internet Layer

- Provide addressing and routing
- Best effort delivery
- Two protocols: IPv4 and IPv6
- IPv4
  - Developed in 1970s
  - First standardized in 1981
- IPv6
  - First proposed in 1994

## Network Access Layer: Ethernet

- Communications between computers on the same network (local link)
- Each Ethernet interface has a globally unique 48-bit MAC address

`00:E9:4F:21:5C:32`

## A Brief History of The Internet

1969: Experimental ARPANET
1972: First public demo of ARPANET
1981: TCP/IP standardized
Jan 1, 1983: ARPANET switched to TCP/IP
1985: NSFNET connected to Internet
1990: ARPANET decommissioned
1993: Mosaic WWW browser
1994: IPv6 was proposed
1995: NSFNET became Internet backbone
1995: IPv6 basic specification was established

# Internet Layer: Addressing

# IPv4 Addressing

00001010000010100001111000101000

A 32 bit value

Representing an interface on the network

## IPv4 Address Notation

**10.20.30.40**

Dotted-decimal notation

## IPv4 Address Notation

**10.20.30.40**

00001010000101000001111000101000

32 bits

## IPv4 Address Notation

10.20.30.40

00001010 00010100 00011110 00101000

Four blocks of 8 bits separated by dots

## IPv4 Address Notation

10.20.30.40

00001010 00010100 00011110 00101000

Each block is in decimal notation

# Two parts of an IPv4 Address

10.20.30.40

Network part: 10.20.30

Host part: 40

# Two parts of an IPv4 Address

10.20.30.40

11111111111111111111111100000000

Network mask

255.255.255.0

## Shorthand Notation

- 10.20.30.40 network mask 255.255.255.0

  Host 40 on network 10.20.30.0
- Shorthand notation:

  `<address>/<prefix-length>`

  `10.20.30.40/24`
- Prefix-length = num. of `1`s in network mask

## IP Address Space

- The range of IP addresses that is limited by its network mask

Lowest Address in IP Address Space

10.20.30.40

Netmask

255.255.255.0

Lowest Address in IP Address Space

00001010000101000001111000101000

10.20.30.40

Netmask

11111111111111111111111100000000

255.255.255.0

## Lowest Address in IP Address Space

00001010000101000001111000101000

10.20.30.40

AND

Netmask

11111111111111111111111100000000

255.255.255.

00001010000101000001111000000000

## Lowest Address in IP Address Space

10.20.30.0

00001010000101000001111000000000

## Highest Address in IP Address Space

10.20.30.40

Netmask

255.255.255.0

## Highest Address in IP Address Space

00001010000101000001111000101000

10.20.30.40

Netmask

00000000000000000000000011111111

255.255.255.0

## Highest Address in IP Address Space

00001010000010100000111100101000

10.20.30.40

OR

Netmask

00000000000000000000000011111111

255.255.255.

00001010000010100000111011111111

## Highest Address in IP Address Space

10.20.30.255

00001010000010100000111011111111

## IP Address Space Notation

`<prefix>/<prefix-length>`

prefix: the lowest IP address in IP address space

`10.20.30.0/24`

Range: `10.20.30.0 – 10.20.30.255`

Size: 256 addresses

## IP Address Space Usage on a Network

- Network address = The lowest IP address in the network's IP address space
- Broadcast address = The highest IP address in the network's IP address space
- Address for hosts = other addresses

# Network 10.20.30.0/24

- Network address = `10.20.30.0`
- Broadcast address = `10.20.30.255`
- Address for hosts =

  `10.20.30. 1 – 10.20.30.254`
- Max num. of hosts = 254

# IPv4 Address Classes

| **Class** | **Prefix bits** | **Prefix-length** | **Net. Size (#hosts)** |
|---|---|---|---|
| A (unicast) | 0 | 8 | 16,777,214 |
| B (unicast) | 10 | 16 | 65,534 |
| C (unicast) | 110 | 24 | 254 |
| D (multicast) | 1110 | | |
| E (reserved) | 1111 | | |

## IP Address Space Allocation

- IP addresses were classified
- Allocated based on the IP address classes
- According to the size of an organization
- Natural netmasks
  – Class A: 255.0.0.0
  – Class B: 255.255.0.0
  – Class C: 255.255.255.0

## Classless Internet Domain Routing

- Network size differences between classes are HUGE!
- What if a network has 100k hosts?
  – Class B is too small
  – Class A is too big
- Solution: Don't use classes → CIDR
  – Allocated prefix-lengths no longer limited to 8, 16, and 24

## Subnetting

- Organizations may allocate their IP address spaces to longer prefix-lengths
- 192.168.0.0/16
  - 192.168.0.0/18
  - 192.168.64.0/18
  - 192.168.128.0/17

## IPv4 Address Space for Private Networks

- For networks using TCP/IP that are not connected to the Internet

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

The Internet is using IPv4…

but IPv4 poses problems.

# Problems of IPv4

- Exhaustion of IPv4 addresses
  - 32-bit turned out to be not enough
  - Was thought address space would be exhausted in 2003
- Proliferation of NAT
  - Networks don't have enough address space
  - Private IP address + NAT for connectivity
- Routing table explosion
  - Routers on the core of the Internet have to maintain more than 100k routing entries now

# Then came IPv6…

started in 1994

# IPv6 Features

- Larger address space (128 bits)
- New header format
- Efficient and hierarchical addressing and routing infrastructure
- Built-in security
- Better support for QoS
- Extensibility

# IPv6 Addressing

00101010000100100011010001011100
00000000000000000000000000000000
00000000011110000000100110101011
00001100000011011110000011110000

A 128 bit value

Representing an interface on the network

# IPv6 Address Notation

`2A12:3456:0:0:78:9AB:C0D:E0F0`

# IPv6 Address Notation

Eight blocks of 16 bits in hexadecimal
separated by colons (::)

`2A12:3456:0:0:78:9AB:C0D:E0F0`

```
00101010000010010001101000101100
00000000000000000000000000000000
00000000011110000000100110101011
00001100000011011110000011110000
```

# IPv6 Address Notation

Eight blocks of 16 bits in hexadecimal
separated by colons (::)

2A12:3456:0:0:78:9AB:C0D:E0F0

00101010000100100011010001011100
00000000000000000000000000000000
00000000011110000000100110101011
00001100000011011110000011110000

# IPv6 Address Notation

Eight blocks of 16 bits in hexadecimal
separated by colons (::)

2A12:3456:0:0:78:9AB:C0D:E0F0

00101010000100100011010001011100
00000000000000000000000000000000
00000000011110000000100110101011
00001100000011011110000011110000

# IPv6 Address Notation

Eight blocks of 16 bits in hexadecimal
separated by colons (::)

2A12:3456:0:0:78:9AB:C0D:E0F0

00101010000100100011010001011100
00000000000000000000000000000000
00000000011110000000100110101011
00001100000011011110000011110000

# IPv6 Address Notation

- Blocks of 0 may be shortened with double colon (::) ; only one :: is allowed

1234:5678:90AB::5678:0:CDEF
1234:5678:90AB:0:0:5678::CDEF
~~1234:5678:90AB::5678::CDEF~~

## IPv6 Address Space Notation

```
<prefix>/<prefix-length>

1234:5678::/48
1234:5678:9ABC:DEF::/64
```

## IPv6 Address Types

- Unicast
  - Single interface
- Multicast
  - Set of interfaces
  - Packets delivered to all interfaces
- Anycast
  - Set of interfaces
  - Packets delivered to one (the nearest) interface

# Address Type Identification

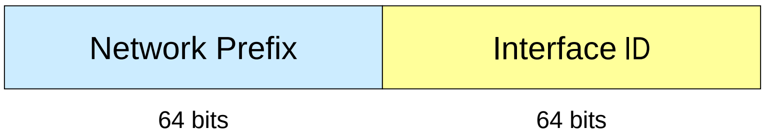
| Type | Binary Value/Prefix | IPv6 Notation |
|---|---|---|
| Unspecified | `000…0` (128bits) | `::/128` |
| Loopback | `000…1` (128bits) | `::1/128` |
| Multicast | `11111111` | `FF00::/8` |
| Link-local unicast | `1111111010` | `FE80::/10` |
| Site-local unicast | `1111111011` | `FEC0::/10` |
| Global unicast | (everything else) | |

# Global Aggregatable Unicast Address Format

| Prefix 001 | TLA ID | RES | NLA ID | SLA ID | Interface ID |
|---|---|---|---|---|---|
| 3 bits | 13 bits | 8 bits | 24 bits | 16 bits | 64 bits |

**TLA ID** Top-level aggregation identifier
**RES** Reserved for future use
**NLA ID** Next-level aggregation identifier
**SLA ID** Site-level aggregation identifier
**Interface ID** Interface identifier

## An Interface’s Unicast Address




A link’s prefix length is always 64 bit

## Allocating IP Address Space

`2001:d30:200::/48`

- 16 bits for link’s network prefixes = 65k

## Interface Identifier

- Interface ID: manual or automatic
- Automatic: Modified EUI-64 of MAC address
  - – Complement 2nd LSB of 1st byte
  - – Insert 0xfffe between 3rd and 4th bytes
- MAC: `00-12-34-56-78-9a`
- Interface ID: `212:34ff:fe56:789a`

## Link-local Address Format

`fe80::<Interface-ID>`

KAME style

```
fe80:<Interface-ID>%<ifname>
  fe80::212:34ff:fe56:789a%fxp0
```

## Multicast Address Format

| Prefix `1111 1111` | FLAGS | SCOPE | Group Identifier |
|---|---|---|---|
| 8 bits | 4 bits | 4 bits | 112 bits |

**Flags**:
LSB = 0 well-known mcast address
LSB = 1 temporary/transient mcast address

**Scope**:
1 interface-link scope
2 link-local scope
5 site-local scope
8 organization-local scope
E global scope

## Multicast Address Example

`ff02::2`

- Well-known address, link-local scope

`ff18::100`

- Temporary address, organization-local scope

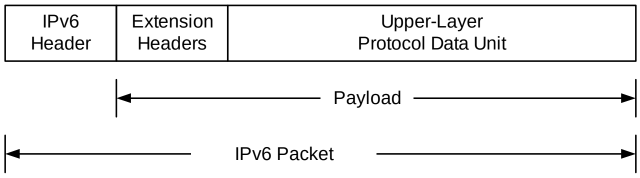
## A Node's Address

- Loopback Address
- Link-local Address for each interface
- Additional Unicast and Anycast Addresses
- All-Nodes Multicast Addresses (ff02::1)
- Solicited-Node Multicast Addresses
- Multicast Addresses of groups it joined

## A Router's Address

- A node's address
- Subnet-Router Anycast Addresses
- All other Anycast Addresses
- All-Router Multicast Addresses (ff02::2)

# IPv6 Packet Format

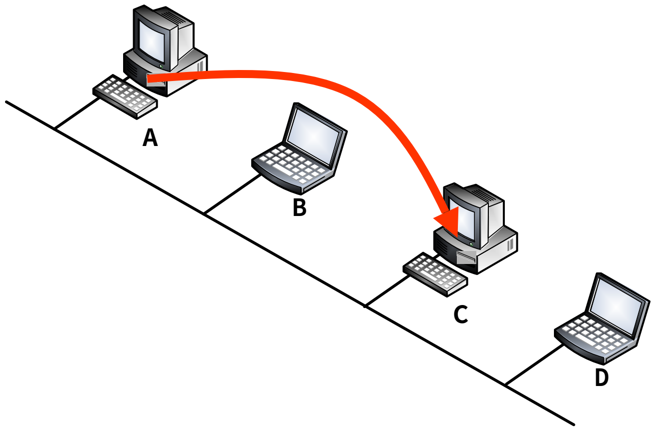



- Packet = IPv6 header + Payload
- Payload = Extension header + PDU
- What follows after a header is identified by Next Header field

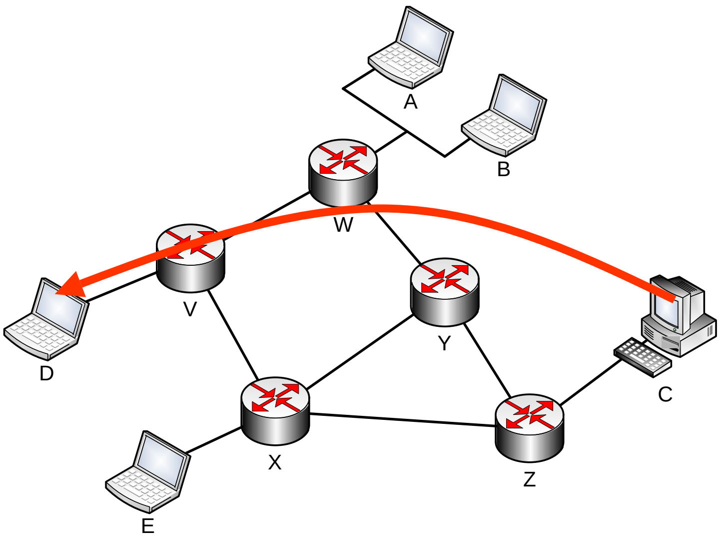
# Internet Layer: Routing

How to deliver IP packets between hosts?

## On a Link




## On Different Link



# Address Resolution

- On a LAN, an IP packet is encapsulated by an Ethernet frame
- A host should know the MAC address of another host to send Ethernet frame to that host
- Mapping between IP address and MAC address is needed
  - IPv4: Address Resolution Protocol (ARP)
  - IPv6: Neighbor Discovery Protocol (NDP)

# ARP Cache Table

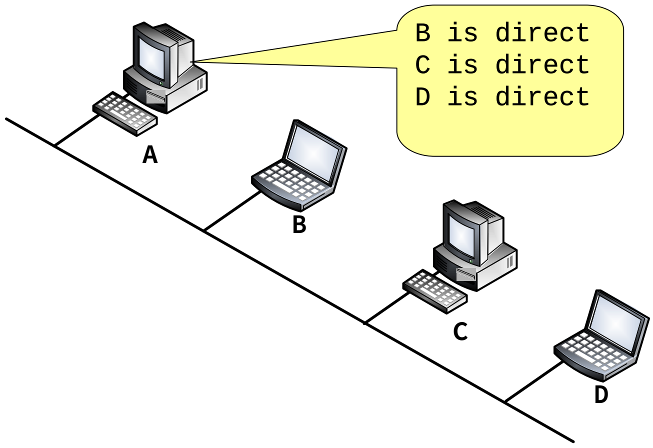
```
> arp -an
? (10.39.234.1) at 00:02:b3:ec:6c:d4 on
  fxp0 [ethernet]
? (10.39.234.121) at 00:0a:79:33:98:59 on
  fxp0 [ethernet]
```

Hostname and address

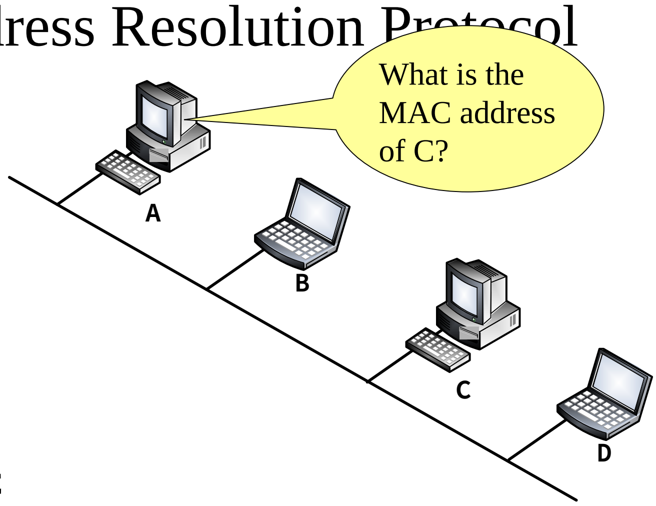
MAC address

On which interface

## Address Resolution Protocol

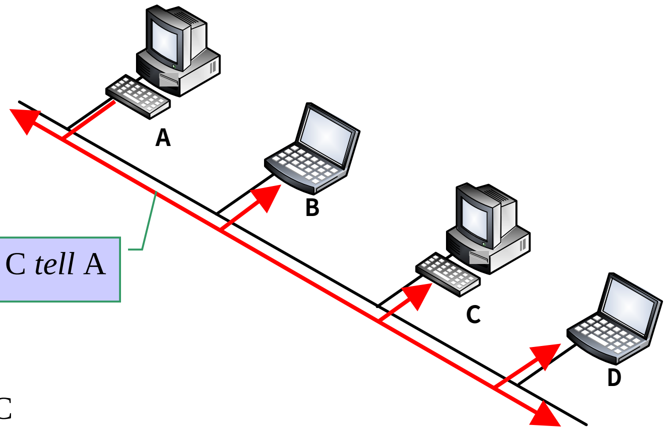



## Address Resolution Protocol




From A to C

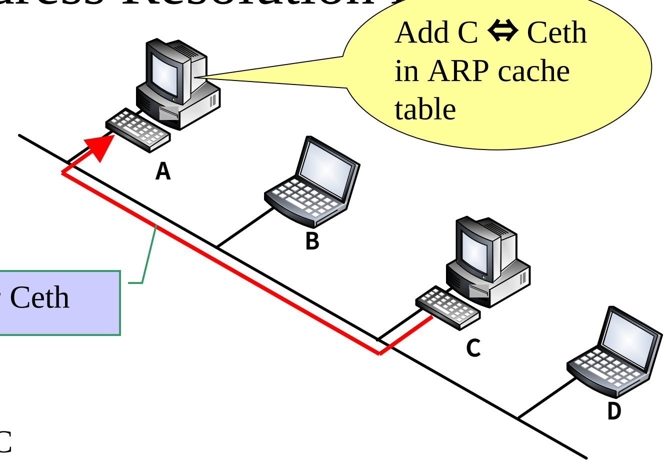
# Address Resolution Protocol




arp: *who is* C *tell* A

From A to C

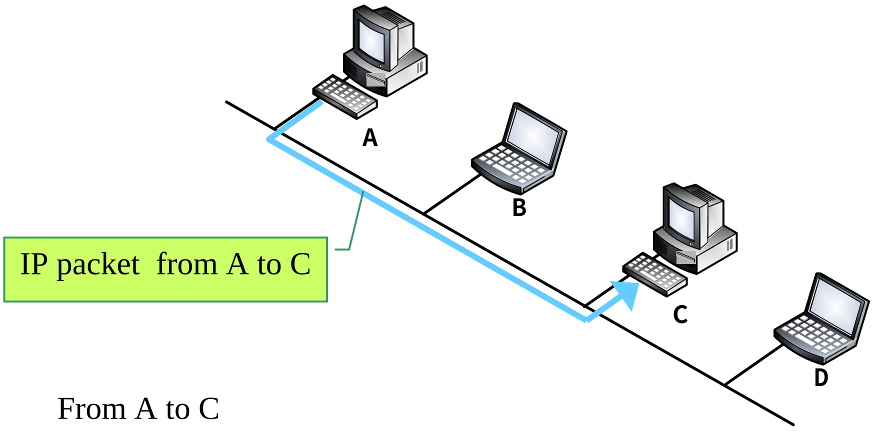
# Address Resolution Protocol




arp: C *is* Ceth

From A to C

## Address Resolution Protocol




From A to C

## Neighbor Discovery Protocol

- More functionality rather than just address resolution:
- To find neighboring routers
  - Routers advertise themselves using *NDP Router Advertisement* messages
- To keep track of reachable neighbors

# NDP Messages

- Router Advertisement
- Router Solicitation
- Neighbor Advertisement
- Neighbor Solicitation
- Redirect

# NDP Cache Table

```
> ndp -an
Neighbor                         Linklayer Address
  Netif Expire    St Flgs Prbs
3ffe:1:2:3:202:b3ff:feec:6cd4    0:2:b3:ec:6c:d4
  fxp0 16h56m31s S   R
3ffe:1:2:3:2d0:b7ff:fe9e:e5d2    0:d0:b7:9e:e5:d2
  fxp0 permanent R
fe80::202:b3ff:feec:6cd4%fxp0    0:2:b3:ec:6c:d4
  fxp0 16h56m26s S   R
fe80::2d0:b7ff:fe9e:e5d2%fxp0    0:d0:b7:9e:e5:d2
  fxp0 permanent R
fe80::1%lo0                      (incomplete)
  lo0 permanent R
```

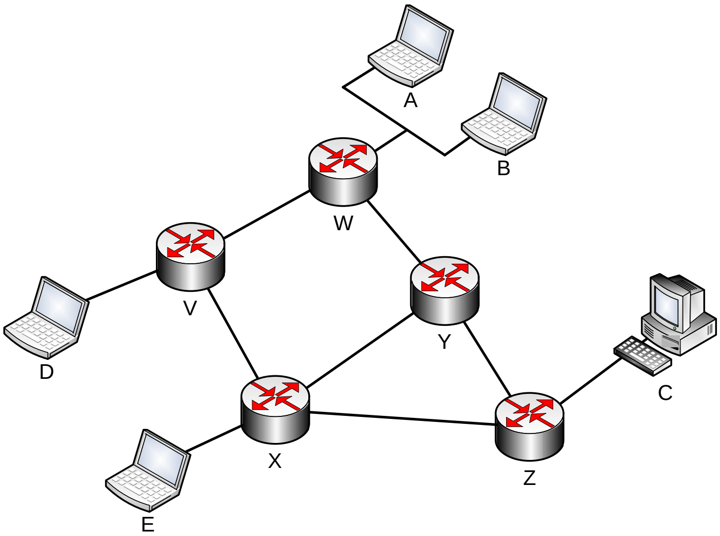
## NDP Cache Table

```
> ndp -an
Neighbor                                 Linklayer Address
   Netif Expire     St Flgs Prbs
3ffe:1:2:3:202:b3ff:feec:6cd4            0:2:b3:ec:6c:d4
   fxp0 16h56m31s  S   R
3ffe:1:2:3:2d0:b7ff:fe9e:e5d2            0:d0:b7:9e:e5:d2
   fxp0 permanent R
fe80::202:b3ff:feec:6cd4%fxp0            0:2:b3:ec:6c:d4
   fxp0 16h56m26s  S   R
fe80::2d0:b7ff:fe9e:e5d2%fxp0            0:d0:b7:9e:e5:d2
   fxp0 permanent R
fe80::1%lo0                              (incomplete)
   lo0 permanent R
```

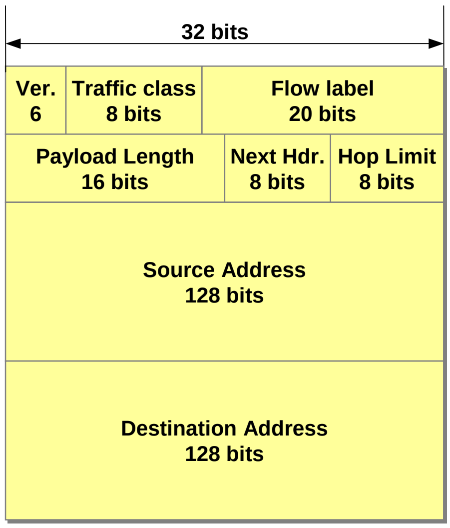
## Multicast Listener Discovery

- Enable IPv6 router to discover the presence of multicast listener on its link.
- Use ICMPv6 packets
- Messages
  - Multicast Listener Query
  - Multicast Listener Report
  - Multicast Listener Done

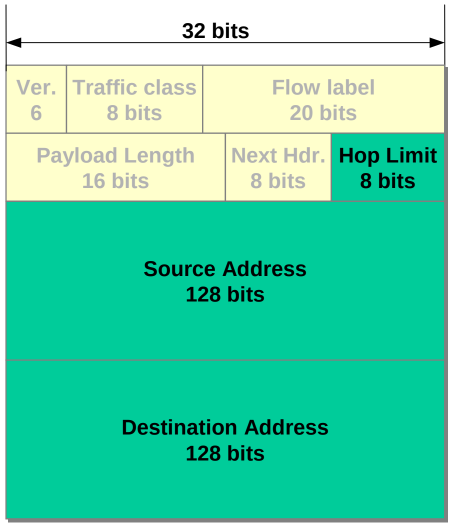
# Delivering A Packet



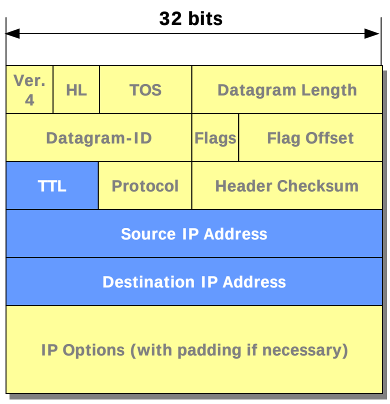

# IPv6 Header



# IPv6 Header




# IPv4 Header



# Routing Table

To reach this destination…

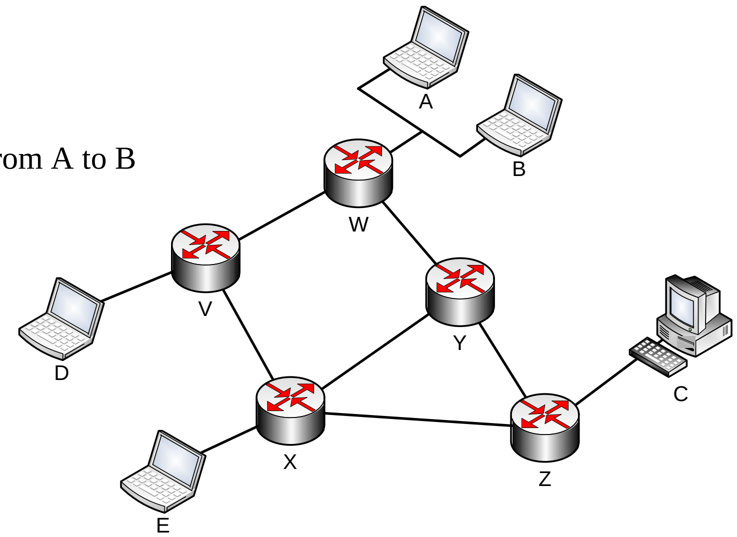
use this gateway

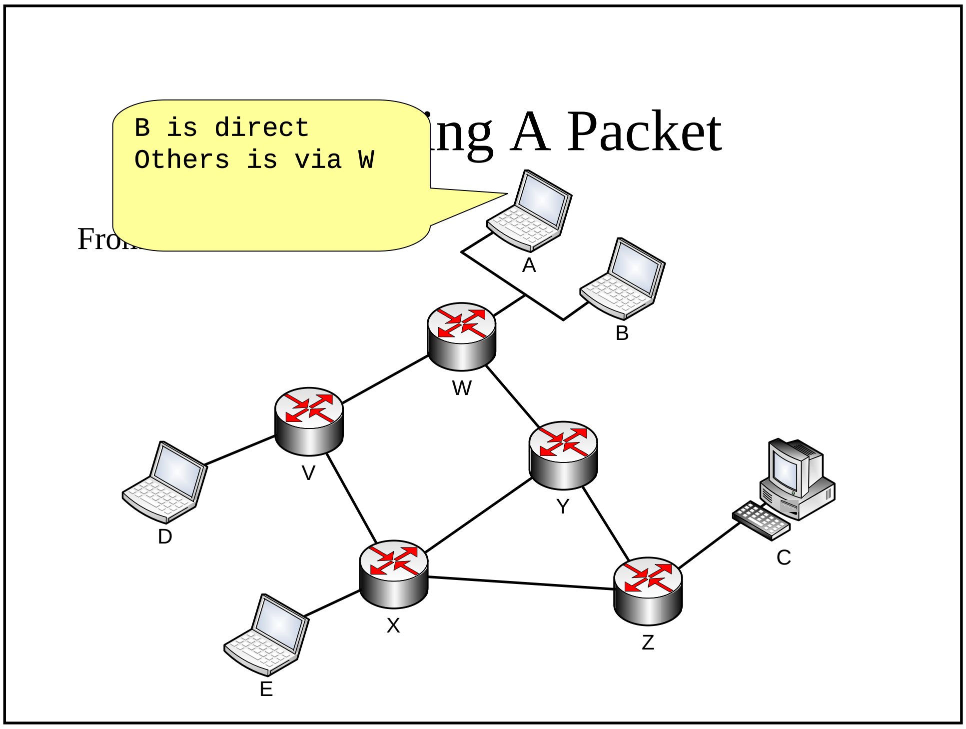
send directly to this

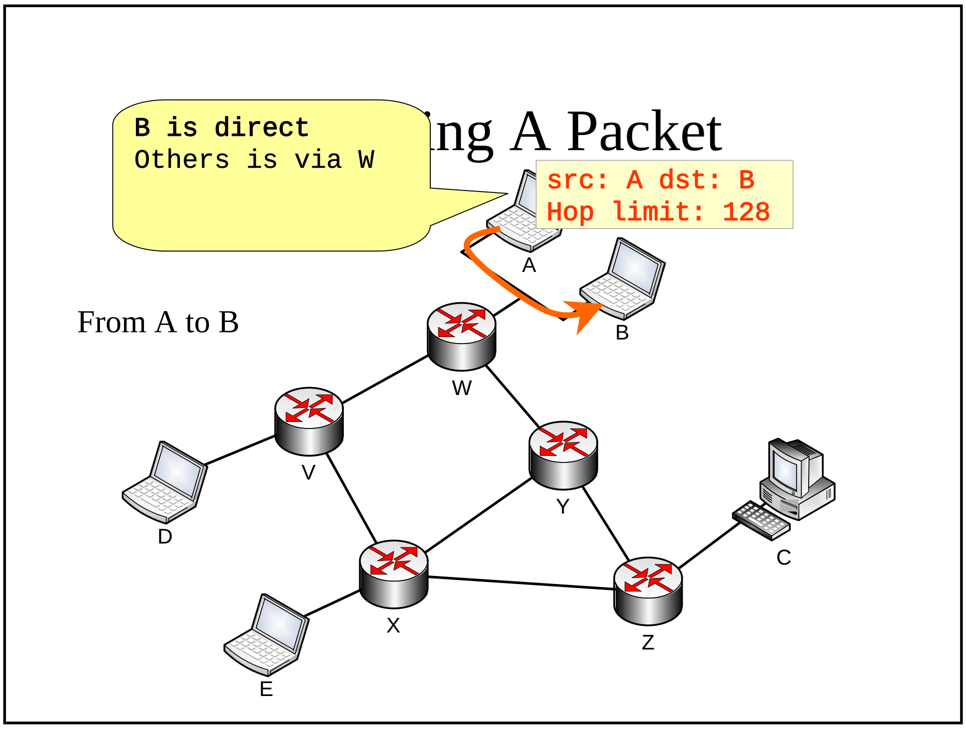
```
> netstat -nr
Routing tables

Internet:
Destination        Gateway            Flags     Refs      Use   Netif Expire
default            10.39.234.1        UGSc        11     1791    fxp0
127.0.0.1          127.0.0.1          UH           1      970     lo0
10.39.234          link#1             UC           2        0    fxp0
10.39.234.1        00:02:b3:ec:6c:d4  UHLW        12        0    fxp0   1199
10.39.234.121      127.0.0.1          UGHS         0     1463     lo0

Internet6:
Destination                       Gateway                        Flags      Netif
   Expire
::/96                             ::1                            UGRSc       lo0 =>
default                           fe80::202:b3ff:feec:6cd4%fxp0  UGc        fxp0
::1                               ::1                            UH          lo0
::ffff:0.0.0.0/96                 ::1                            UGRSc       lo0
3ffe:1:2:3::/64                   link#1                         UC         fxp0
3ffe:1:2:3:2d0:b7ff:fe9e:e5d2     00:d0:b7:9e:e5:d2              UHL         lo0
```

# Delivering A Packet

From A to B

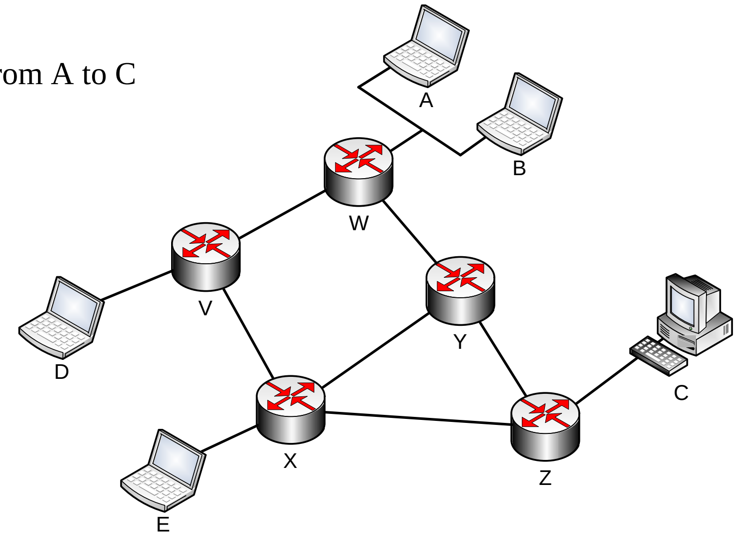



B is direct
Others is via W
ing A Packet
Fro
A
B
W
V
Y
D
C
X
Z
E


B is direct
Others is via W
ing A Packet
src: A dst: B
Hop limit: 128
A
B
From A to B
W
V
Y
D
C
X
Z
E

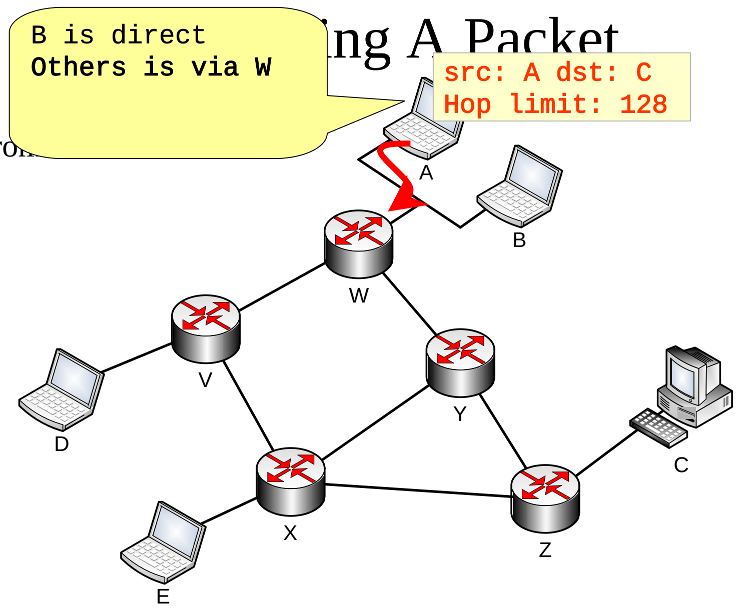
# Delivering A Packet

From A to C

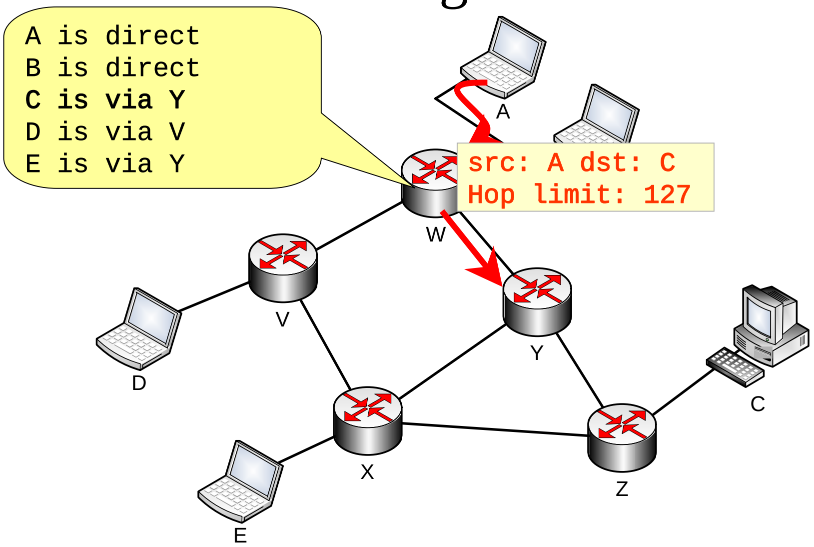



# ing A Packet

Fro

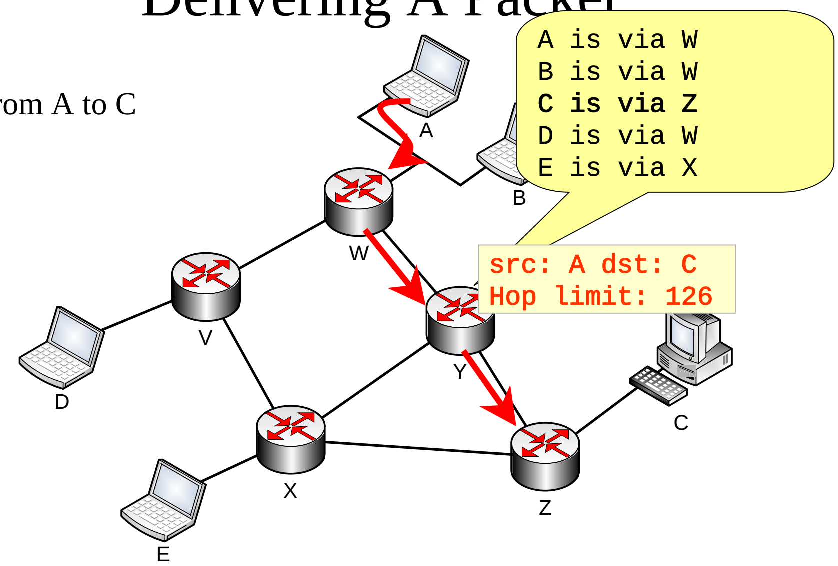


# Delivering A Packet



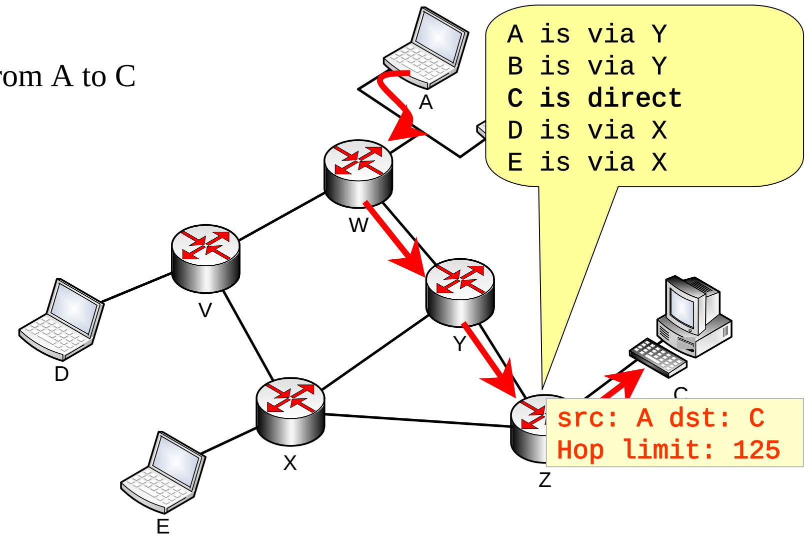

# Delivering A Packet

From A to C



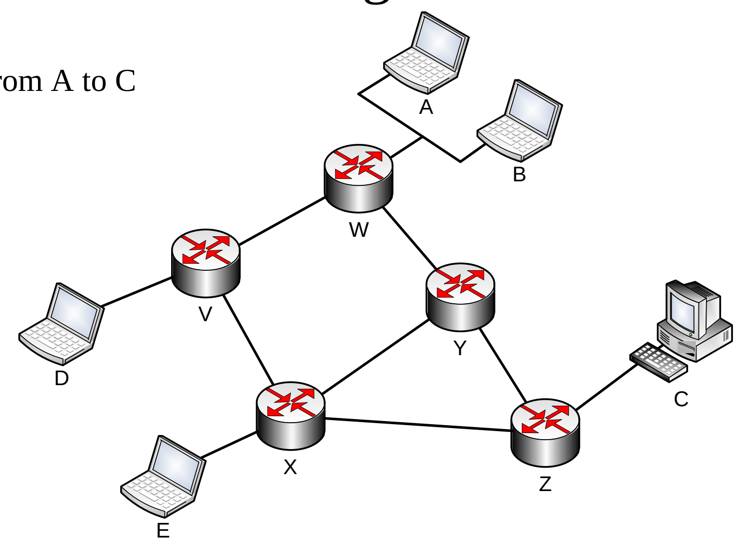
# Delivering A Packet

From A to C



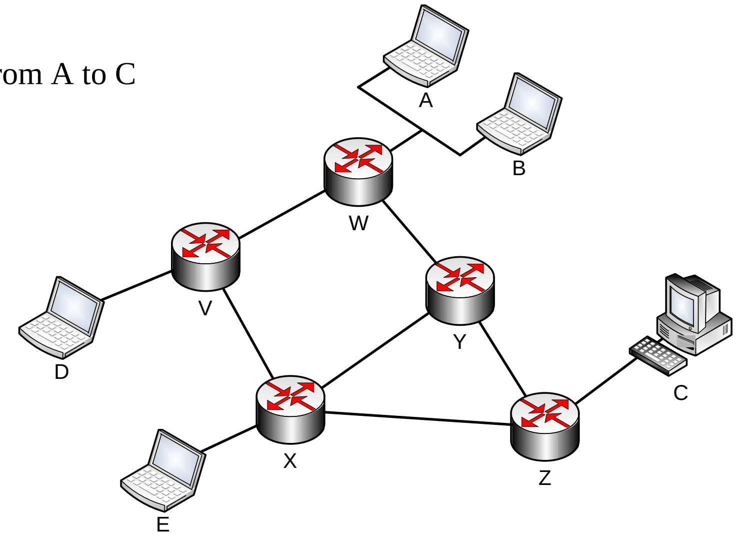

# Delivering A Packet

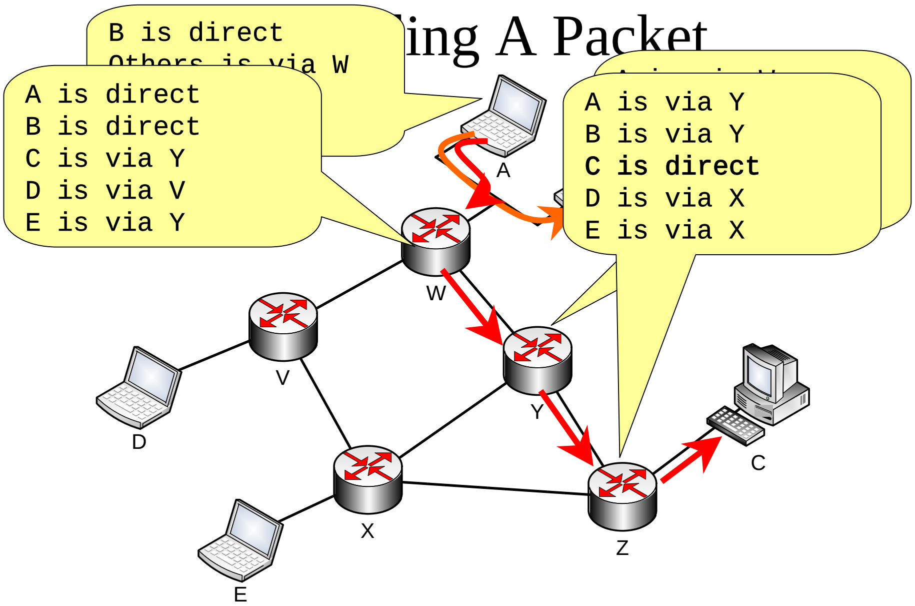
From A to C



# Delivering A Packet

From A to C



# IP Forwarding

- Mechanism to deliver IP packets across local networks
- Use Routing Table whether a packet is for local or non-local network destination
- Decrement Hop Limit (TTL in IPv4) of packet
  - Hop Limit: a field in IP header to determine the number of hops an IP packet can travel
- Issue ICMP to packet's source address if errors:
  - Hop Limit (TTL) is decremented to zero
  - Packet is larger than the Maximum Transmission Unit (MTU) of the outgoing interface

# Internet Layer: ICMP

## Internet Control Message Protocol

- IPv4 and IPv6 are only for addressing and packet delivery
- ICMP provides diagnostic and error information about the network and hosts
- ICMP for IPv4 and ICMPv6 for IPv6
- ICMPv6 includes new functionalities compared to ICMP

## ICMPv6

- Handling errors
  - Destination Unreachable
  - Packet Too Big
  - Time Exceeded
  - Parameter Problem
- Diagnostic and troubleshooting
  - Echo Request
  - Echo Reply
- Other functions
  - Neighbor discovery (NDP)
  - Multicast membership management (MLD)

## Ping:
## Use ICMP Echo Request and Reply

```
> ping6 -c 5 www.kame.net.
PING6(56=40+8+8 bytes) 2001:d30:101:1::10 -->
   2001:200:0:8002:203:47ff:fea5:3085
16 bytes from 2001:200:0:8002:203:47ff:fea5:3085, icmp_seq=0
   hlim=59 time=5.511 ms
16 bytes from 2001:200:0:8002:203:47ff:fea5:3085, icmp_seq=1
   hlim=59 time=5.836 ms
16 bytes from 2001:200:0:8002:203:47ff:fea5:3085, icmp_seq=2
   hlim=59 time=5.347 ms
16 bytes from 2001:200:0:8002:203:47ff:fea5:3085, icmp_seq=3
   hlim=59 time=4.736 ms
16 bytes from 2001:200:0:8002:203:47ff:fea5:3085, icmp_seq=4
   hlim=59 time=5.463 ms

--- www.kame.net ping6 statistics ---
5 packets transmitted, 5 packets received, 0% packet loss
round-trip min/avg/max/std-dev = 4.736/5.379/5.836/0.360 ms
```

## Traceroute:
## Use ICMP Time Exceeded

```
> traceroute6 www.kame.net.
traceroute6 to www.kame.net (2001:200:0:8002:203:47ff:fea5:3085)
   from 2001:d30:101:1::10, 30 hops max, 12 byte packets
 1  sfc-gate  0.249 ms  0.254 ms  0.187 ms
 2  n6-gate  2.905 ms  2.734 ms  2.710 ms
 3  pc6.otemachi.wide.ad.jp  3.149 ms  3.061 ms  3.139 ms
 4  hitachi1.otemachi.wide.ad.jp  3.235 ms  3.197 ms  3.201 ms
 5  pc3.yagami.wide.ad.jp  4.132 ms  3.921 ms  3.936 ms
 6  gr2000.k2.wide.ad.jp  5.363 ms  5.555 ms  5.491 ms
 7  orange.kame.net  5.037 ms  4.758 ms  4.757 ms
```

# LAB WORK

Exercise 1: IP Addressing

Page 16

# Answers #1

| | | |
|---|---|---|
| 1 | 3ffe:: | 3ffe:0:0:0:0:0:0:0 |
| 2 | 2001:d30::1234:abcd | 2001:d30:0:0:0:0:1234:abcd |
| 3 | 2001:d1:a2::698a:fc22:563b | 2001:d1:00a2:0:0:698a:fc22:563b |
| 4 | 2001:0:40c0::a0:fa32 | 2001:0:40c0:0:0:0:00a0:fa32 |

## Answers #2

| | | |
|---|---|---|
| 1 | 10.1.1.1/28 | 10.1.1.1/255.255.255.240 |
| 2 | 202.249.193.1/19 | 202.249.193.1/255.255.224.0 |
| 3 | 172.16.2.5/29 | 172.16.2.5/255.255.255.248 |
| 4 | 114.5.89.5/18 | 114.5.89.5/255.255.192.0 |

## Answers #3

| | | |
|---|---|---|
| 1 | 00:60:3e:46:e8:d9 | 260:3eff:fe46:e8d9 |
| 2 | 00:d0:b7:2c:69:14 | 2d0:b7ff:fe2c:6914 |
| 3 | 00:30:48:71:f5:8e | 230:48ff:fe71:f58e |
| 4 | 00:e0:b7:2c:22:4f | 2e0:b7ff:fe2c:224f |

## Answer #4

1. `192.168.0.0/18`
   `192.168.0.0`
   `192.168.63.255`
2. `2001:d10:00a2::/48`
   `2001:d10:a2::`
   `2001:d10:a2::ffff:ffff:ffff:ffff:ffff`
3. `3ffe:2c0::/35`
   `3ffe:2c0::`
   `3ffe:2c0::31ff:ffff:ffff:ffff:ffff:ffff`

## LAB WORK

Exercise 2: Enabling IPv6

Page 17

## Username and Password

- admin : Ow;2k5
- root : Ow;2k5

## Ex. 2: Enabling IPv6

- Ethernet interface has two IPv6 addresses
  - Link-local fe80:<Interface-ID>%<ifname>
  - Global unicast 2001:d30:XXX:<Interface-ID>
- Link-local always presents
- Global unicast address prefix is advertised by router
- Interface identifier corresponds to MAC address

# LAB WORK

Exercise 3: Local neighbors

# Ex. 3: Local neighbors

- NDP cache contains entries of your interfaces and the router

# LAB WORK

Exercise 4: ping

# Ex. 4: ping

- Ping-ing host sends ICMP Echo Request
- Ping-ed host replies with ICMP Echo Reply
- `ff02::1` is all node multicast address
- `ff02::2` is all router multicast address
- NDP cache entries will be filled with neighbor’s addresses

# LAB WORK

Exercise 5: Routing table and traceroute

## Ex. 5: routing table and traceroute

- Traceroute-ing host sends IP packets with increasing Hop Limit value
- Routers along the path to destination send back ICMP Time Exceeded messages when Hop Limit of packets is decremented to zero
- Results:
  - Route to destination
  - Round trip time to each router (RTT)

# LAB WORK

Exercise 6: tcpdump

# Ex. 6: tcpdump

- Display the packets seen by a network interface
- Use -v to see more verbose