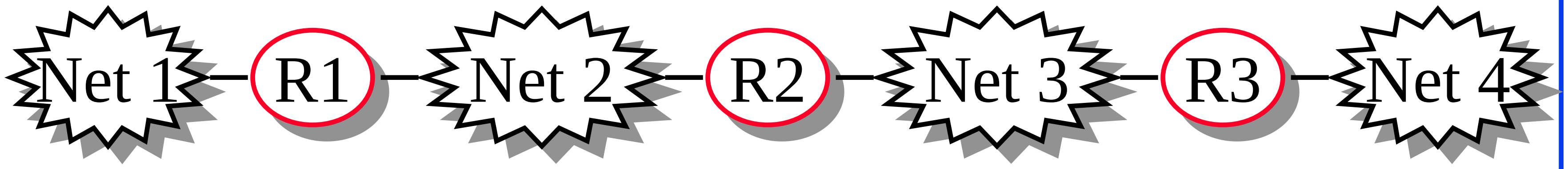
The Network Layer: Data Plane



Raj Jain
Washington University in Saint Louis
Saint Louis, MO 63130
Jain@wustl.edu

Audio/Video recordings of this lecture are available online at:

http://www.cse.wustl.edu/~jain/cse473-24/

Student Questions

1. Network Layer Basics
2. What’s inside a router?
3. Forwarding Protocols: IPv4, DHCP, NAT, IPv6
4. Software Defined Networking

Note: This class lecture is based on Chapter 4 of the textbook (Kurose and Ross) and the figures provided by the authors.

Student Questions

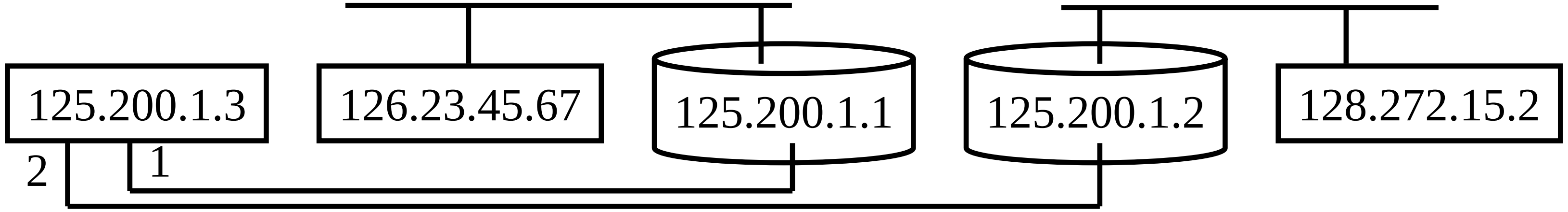
Network Layer Basics

1. Forwarding and Routing
2. Connection-Oriented Networks: ATM Networks
3. Classes of Service
4. Router Components
5. Packet Queuing and Dropping

Student Questions

Forwarding and Routing

- **Forwarding**: Input link to output link via Address prefix lookup in a table.
- **Routing**: Making the Address lookup table
- **Longest Prefix Match**



Prefix	Next Router	Interface
126.23.45.67/32	125.200.1.1	1
128.272.15/24	125.200.1.2	2
128.272/16	125.200.1.1	1

Ref: Optional Homework: R3 in the textbook

Student Questions

- Is there a limit to how long an address table can be?

No. There is no limit.

- The slides in Chapter 4 indicate optional homework R3, R4, and R5. Do we need to review all the homework problems in the textbook

Try at least those indicated.

- Can you review what prefixes match and how you get the interface numbers?

Interfaces are numbered internally in the router.

- Is the IP address lookup process done in CAM rather than software?

It can be done anywhere.

- What is the benefit of using the longest Prefix Match?

You don't need a very long table.

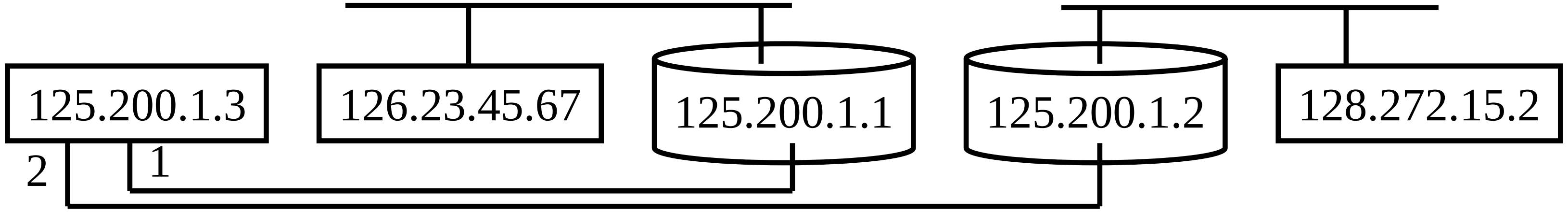
- Has generalized forwarding ever been done in the network layer, and when?

Generalized=more than destination. Yes, QoS is often used to determine the path.

- Does the cylindrical icon mean router, while the rectangle means host? *Yes*

Forwarding and Routing

- **Forwarding**: Input link to output link via Address prefix lookup in a table.
- **Routing**: Making the Address lookup table
- **Longest Prefix Match**



Prefix	Next Router	Interface
126.23.45.67/32	125.200.1.1	1
128.272.15/24	125.200.1.2	2
128.272/16	125.200.1.1	1

Ref: Optional Homework: R3 in the textbook

Student Questions

- Are CAMS comparable to Random Access Memory in computers?

CAMS=Content Addressable Memories. Have search engines built in? You give one column and get the whole row.

- How do you decide which router to use when the longest prefix match is the same for multiple routers?

Fastest speed link or round robin.

- Why is it "longest prefix match" and not the entry that matches entirely?

The table would be too long.

- Are Prefixes in the table the addresses of terminals? *Yes*
- The video quiz question states that routing is not the function of making the lookup table, but the slide states that routing makes the address lookup table.

The quiz answer has been corrected. It was announced on Piazza. No one is affected.

- What does interface here mean?

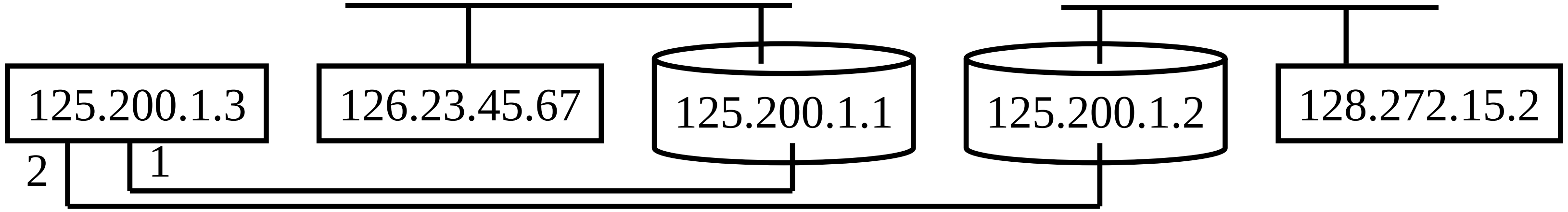
Link

- Are they matching in binary systems?

All computers are binary. Human presentation is in decimal.

Forwarding and Routing

- **Forwarding**: Input link to output link via Address prefix lookup in a table.
- **Routing**: Making the Address lookup table
- **Longest Prefix Match**



Prefix	Next Router	Interface
126.23.45.67/32	125.200.1.1	1
128.272.15/24	125.200.1.2	2
128.272/16	125.200.1.1	1

Ref: Optional Homework: R3 in the textbook

Student Questions

- What will happen if there is no match?

There is always a default entry that matches.

- How does a router determine the correct interface when multiple entries in the routing table match the destination IP address?

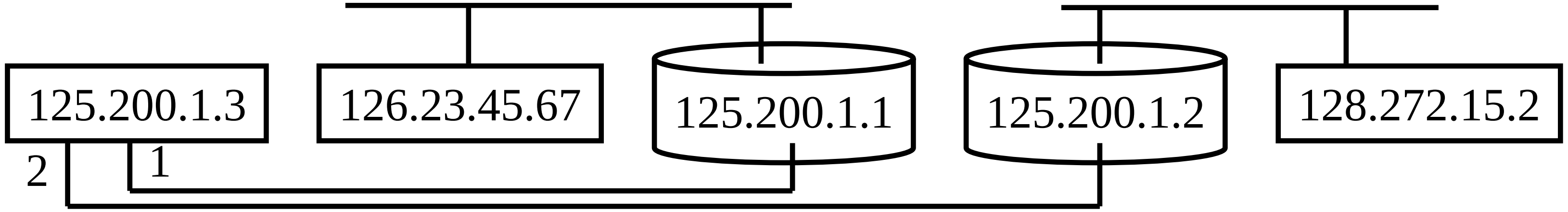
Longest prefix first

- How are content addressable memories (CAMs) utilized in the input ports for address lookup and caching?

CAM: Give content, get the address
RAM: Give the address to get the content

Forwarding and Routing

- **Forwarding**: Input link to output link via Address prefix lookup in a table.
- **Routing**: Making the Address lookup table
- **Longest Prefix Match**



Prefix	Next Router	Interface
126.23.45.67/32	125.200.1.1	1
128.272.15/24	125.200.1.2	2
128.272/16	125.200.1.1	1

Ref: Optional Homework: R3 in the textbook

Student Questions

- What exactly are we matching here?

Destination address in the datagram with our table entries.

- What is "Next Router“?

Router to send this datagram next.

- Can you explain the /32, /24, /16 notation in the prefix column of the table?

/n = Match first n bits only

- How exactly are routing tables made?

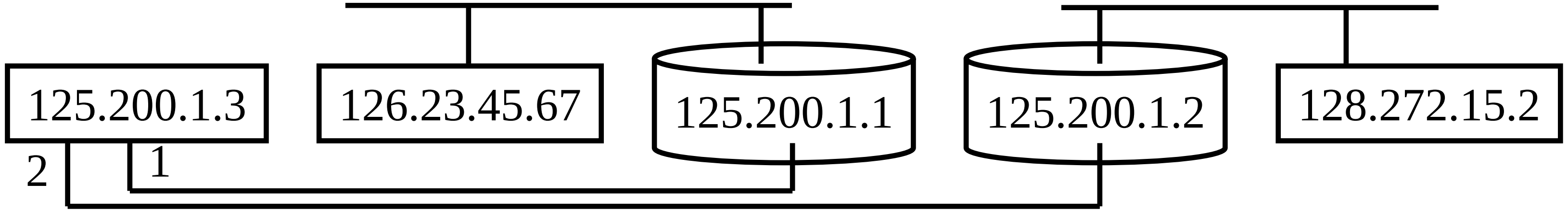
Discussed in the next chapter.

- Is the prefix match similar to a subnet mask?

Yes.

Forwarding and Routing

- **Forwarding**: Input link to output link via Address prefix lookup in a table.
- **Routing**: Making the Address lookup table
- **Longest Prefix Match**



Prefix	Next Router	Interface
126.23.45.67/32	125.200.1.1	1
128.272.15/24	125.200.1.2	2
128.272/16	125.200.1.1	1

Student Questions

- Why is port forwarding dangerous? I tried to enable it on my router, but since I don't own it. I'm not allowed to enable it.

You may deprive other users of their required ports.

Ref: Optional Homework: R3 in the textbook

Network Service Models

- Guaranteed Delivery: No packets lost
- Bounded delay: Maximum delay
- In-order packet delivery: Some packets may be missing
- Guaranteed minimal throughput
- Guaranteed maximum jitter: Delay variation
- Security Services (optional in most networks)
- ATM offered most of these
- IP offers none of these ⇒ Best effort service (Security is optional)

Optional Homework: R4, R5 in the textbook

Student Questions

- In the textbook, it uses “Guaranteed minimal bandwidth” instead of “Guaranteed minimal throughput.” Are there any differences between bandwidth and throughput?

Yes. Bandwidth relates to the frequency of the signal. Throughput is measured in the units of the output (bits). However, many people use them interchangeably.

- What are the bounded delay times, and how are these decided?

30 ms. The new information will be generated in 30 ms as in a movie with 30 frames/sec.

- Are ATM networks mainly used for telephone networks rather than internet connections?

It was designed for telephone networks. Then, the same features were incorporated into IP.

Network Service Models

- Guaranteed Delivery: No packets lost
- Bounded delay: Maximum delay
- In-order packet delivery: Some packets may be missing
- Guaranteed minimal throughput
- Guaranteed maximum jitter: Delay variation
- Security Services (optional in most networks)
- ATM offered most of these
- IP offers none of these ⇒ Best effort service (Security is optional)

Optional Homework: R4, R5 in the textbook

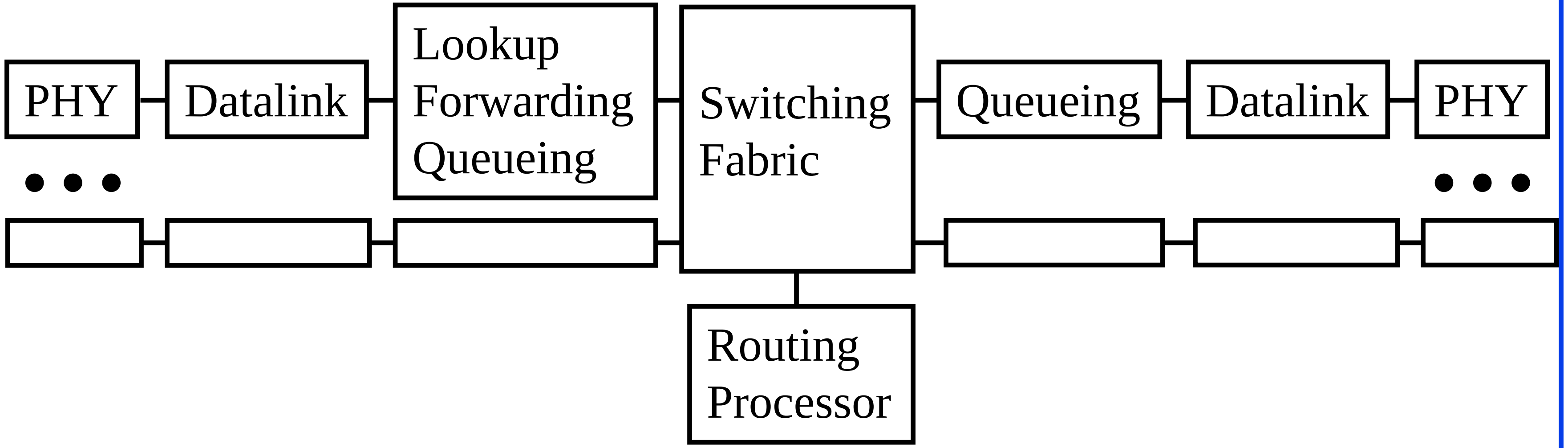
Student Questions

- What is ATM?

Asynchronous Transfer Mode (ATM) was the packet-switching technology developed by carriers. It introduced the concept of labels. Most of the concepts were later taken over by MPLS.

Next Q 4.10b

What's Inside a Router?



- **Input Ports**: receive packets, lookup address, queue
  Use **Content Addressable Memories** (CAMs) and caching
- **Switch Fabric**: Send from the input port to the output port
- **Output Ports**: Queuing, transmitting packets

Student Questions

- Does this input physical link also serve as the output physical link back to wherever the input came from?

Generally, yes. However, simplex (one-way) links are possible.

- At the beginning of this module, we learned about the layers of a router, including the physical and datalink. To be clear, we don't consider those parts of the data plane, even though we learned about it here, right?

They are part of the Layer 1 and Layer 2 data planes. Here, we are working on the Layer 3 data plane.

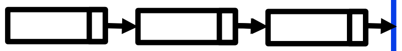
- Why is it called switch fabric?

Repetitive patterns like fabric

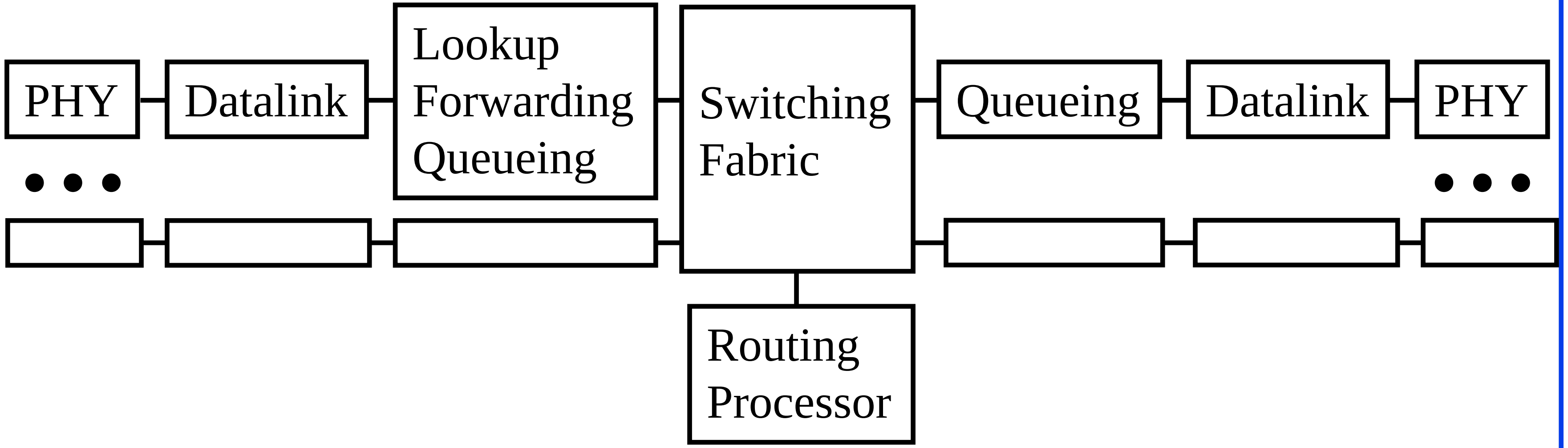
- What does a switching fabric physically consist of?

Printed circuit board

- How does the router queue the packets in the output ports?

Using a linked list 

What's Inside a Router?



- **Input Ports**: receive packets, lookup address, queue
  Use **Content Addressable Memories** (CAMs) and caching
- **Switch Fabric**: Send from the input port to the output port
- **Output Ports**: Queuing, transmitting packets

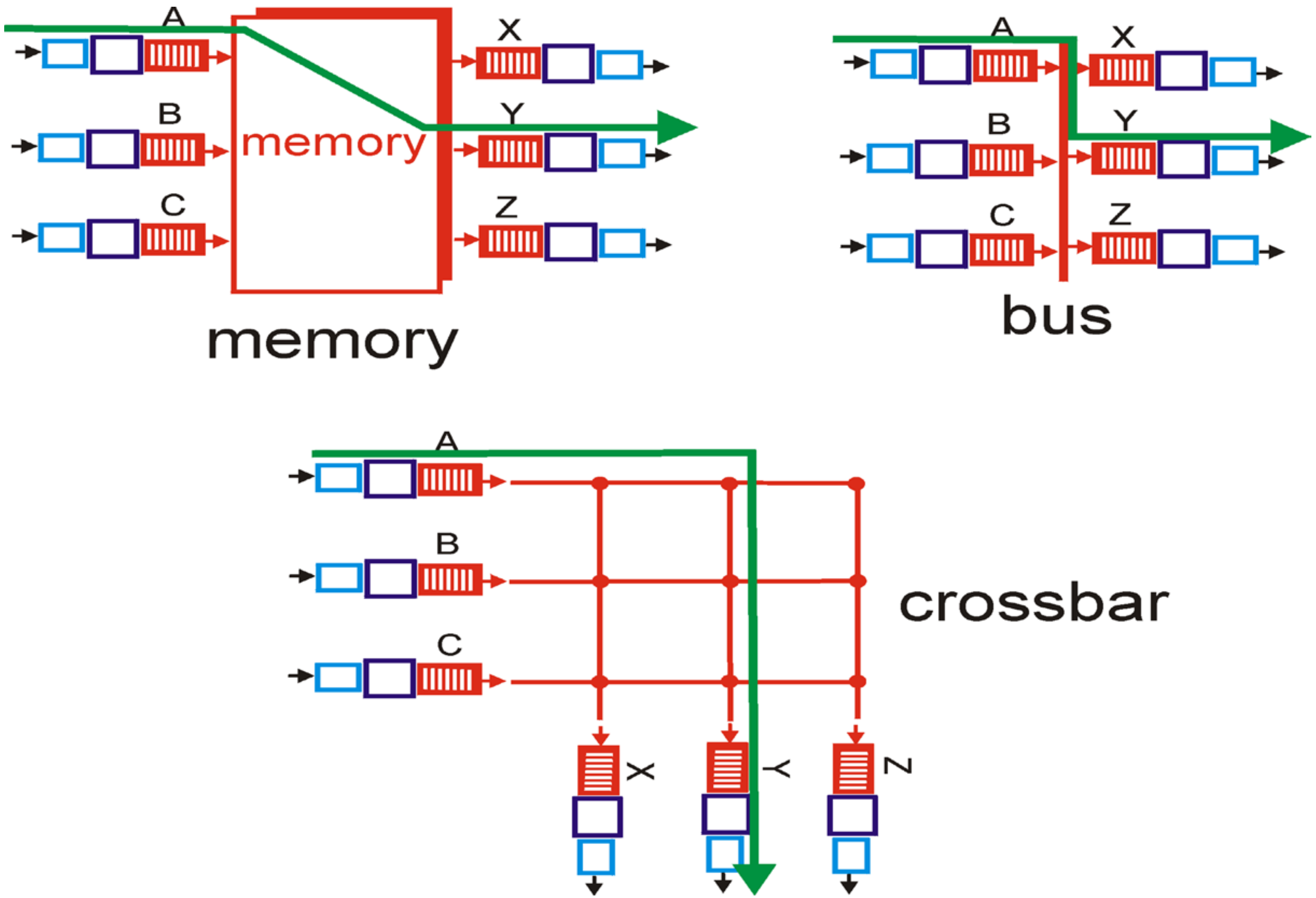
Student Questions

- Why did you say CAMs don't need addresses but use content? I'm confused about what the content is.

A table of address prefixes has many rows. You usually access a table by row number. In CAMs, you can access the table by the row's content.

128.3	3
125.7.1	2
7.37.1	3

Types of Switching Fabrics



Student Questions

- Is there an industry standard for switching, or is it at the discretion of each manufacturer?

It is at the discretion of each manufacturer.

- For switching fabrics, how in-depth do you expect us to understand the different types of switching fabrics? Would it be something like telling the difference between the 3 when given an image of each fabric type?

Whatever the book covers in this section is included, which is more than three figures.

- Do ports here refer to computer ports such as Port 80 for HTTP?

No. The port here means switch ports. E.g., a 5-port switch

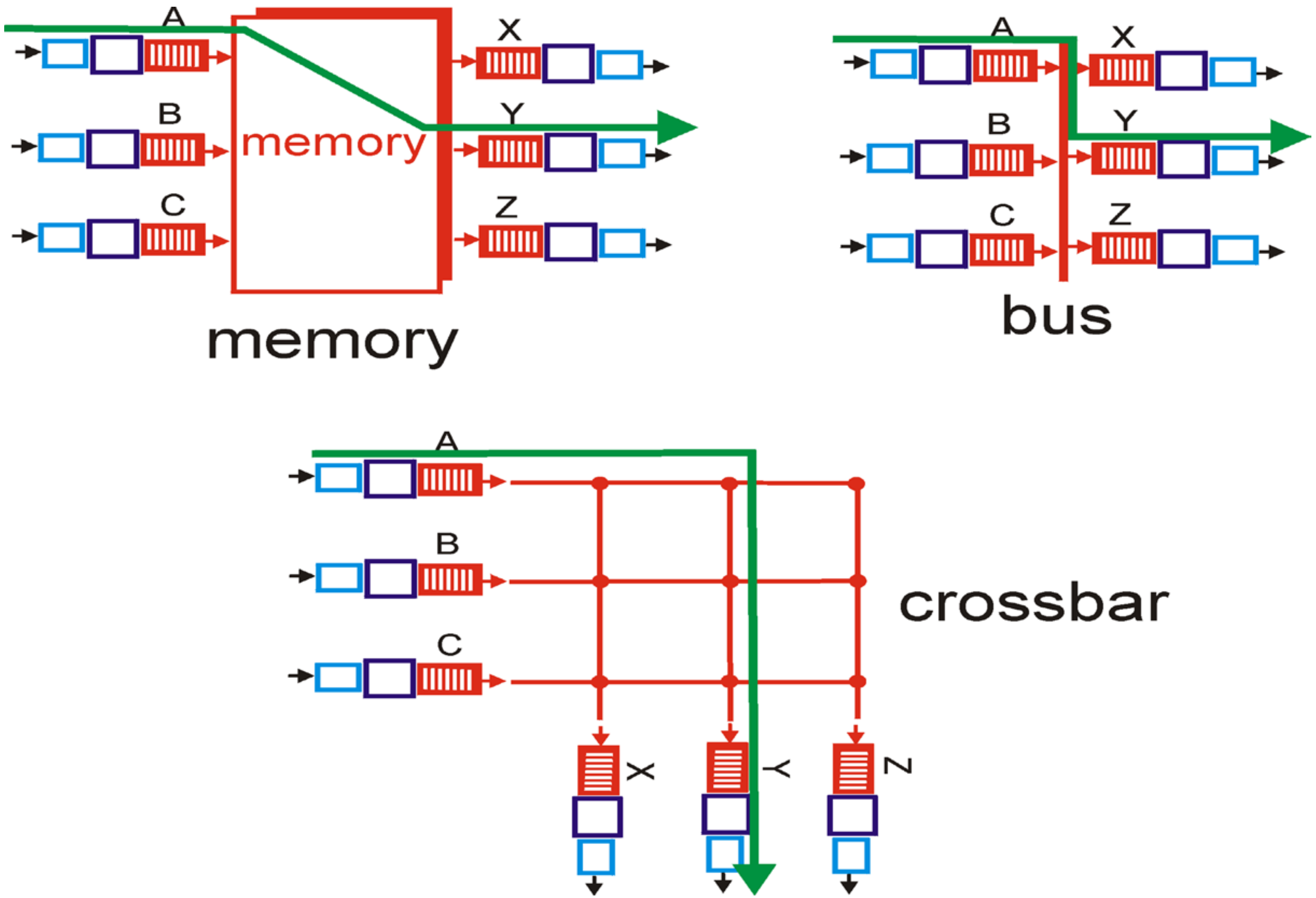
- Bus and crossbar seem very similar. ABC and XYZ share links. What is the difference?

Crossbar = many parallel busses

- How different is the performance between the different types of fiber?

Fiber? See above for fabrics.

Types of Switching Fabrics



Student Questions

- What are the pros and cons of switching fabrics

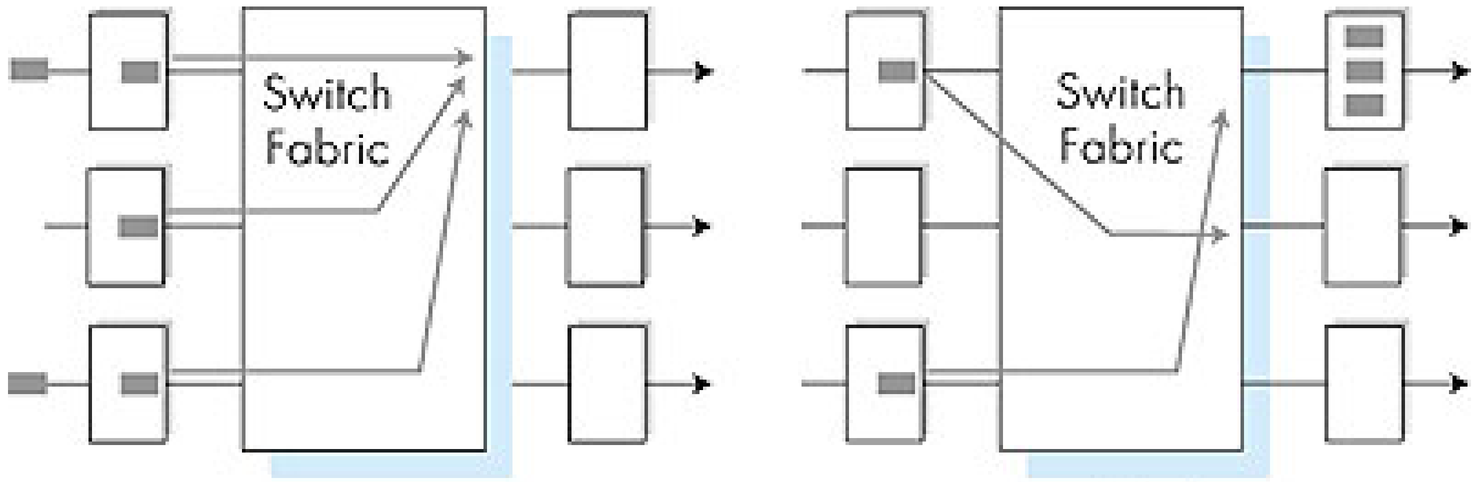
Memory is easy to implement, the bus has a queue, and crossbars need more hardware. Details in the book.

- How does crossbar switching select a particular path between inputs and outputs?

The controller programs the switches as needed for each packet.

Where Does Queuing Occur?

- If switching fabric is slow, packets wait on the input port.
- If switching fabric is fast, packets wait for the output port
  $\Rightarrow$ Queueing (Scheduling) and drop policies
- Queueing: First Come First Served (FCFS), Weighted Fair Queueing



Output Port Contention at Time t

One Packet Time Later

Student Questions

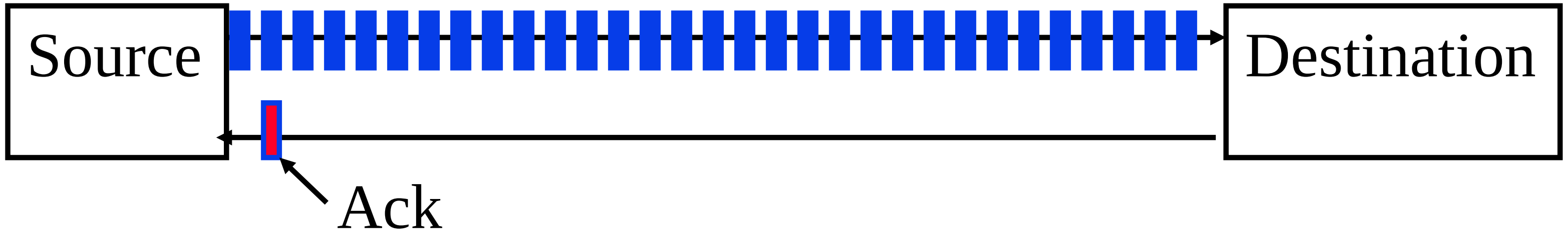
- Is the FCFS both for input port queueing and output port queueing?

Yes. Queue everywhere needs a service discipline.

- Will queueing occur at the input and output ports, or is it mutually exclusive?

Port queuing may happen because of the following switch or router. Not this one.

Ideal Buffering



- Flow Control Buffering = RTT*Transmission Rate
- Buffer = RTT*Transmission Rate/√(# of TCP flows)

Student Questions

- Can you clarify what this flow control buffering refers to? Is this the buffer for the entire link, and then when you divide by sqrt(# TCP flows), that is the buffer for what? Do input ports have a separate buffer from the entire link?

Buffers are at the destination. The buffers must be as large as the number of bits on the wire.

- The book says: "router buffers ... for buffer sizing ... the amount of buffering should be equal to the average RTT times the link capacity." Where does this fit in?

Number of bits on the wire
= Length of the link in sec × Bits/sec
=RTT Link × Capacity

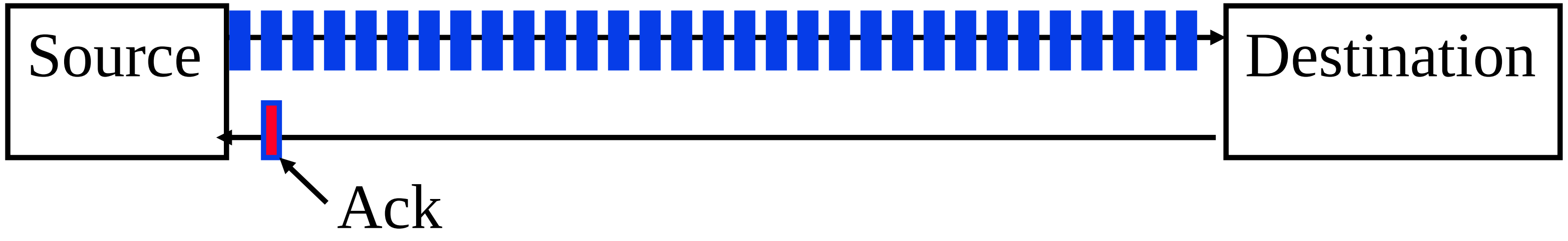
- What is the difference between flow control buffering and buffer in this slide?

This slide shows the packets on the link.

- What's the difference between "flow control buffering" and "buffer?"

Buffer = storage
Flow control buffer = storage reserved for flow control

Ideal Buffering



- Flow Control Buffering = RTT*Transmission Rate
- Buffer = RTT*Transmission Rate/√(# of TCP flows)

Student Questions

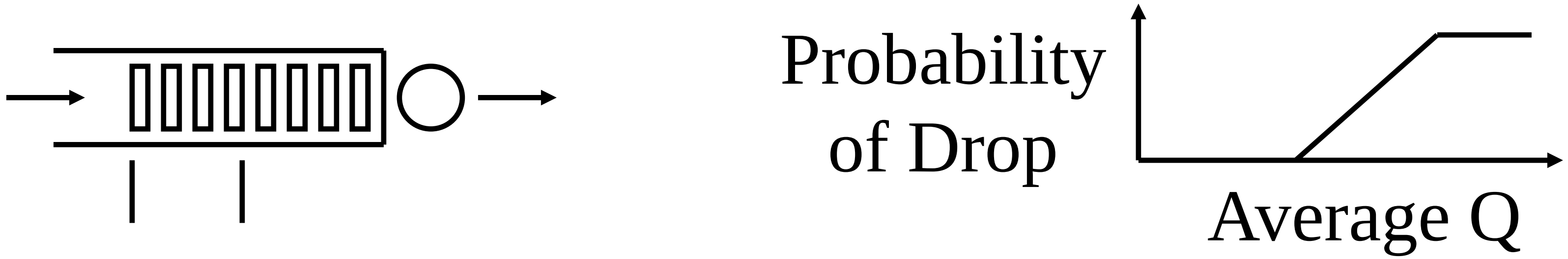
- Does the overbooking of the buffers assume that the packet can be dropped on the link?

Overbooking means packets may not find a buffer in the destination.

- How does this picture show Ideal Buffering?

Each packet occupies space on the wire, just like packets on a UPS truck.

Packet Dropping Policies



- **Drop-Tail**: Drop the arriving packet
- **Random Early Drop (RED):** Drop arriving packets even before the queue is full
  - Routers measure the average queue and drop incoming packets with a certain probability

  $\Rightarrow$ **Active Queue Management** (AQM)

Student Questions

- For both policies, packets already queued won't be dropped. The only difference is that the arriving packet might be dropped even though the queue is not complete.
- Yes.
- Is there a threshold to decide when to activate RED?

Left to the administrator

- Why not let the queues fill and only drop packets, then?

It is unfair in the presence of bursty traffic.

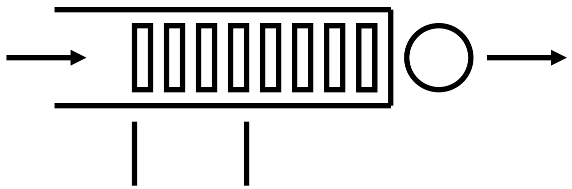
- How do we decide which one to drop

The packet that cannot get in.

- How is the packet-dropping policy chosen? For example, when do we use drop-tail versus RED?

RED is more recent.

Packet Dropping Policies



- **Drop-Tail**: Drop the arriving packet
- **Random Early Drop (RED):** Drop arriving packets even before the queue is full
  - Routers measure the average queue and drop incoming packets with a certain probability
  - ⇒ **Active Queue Management** (AQM)

Student Questions

- Is there any way to explicitly assign importance to a particular packet?

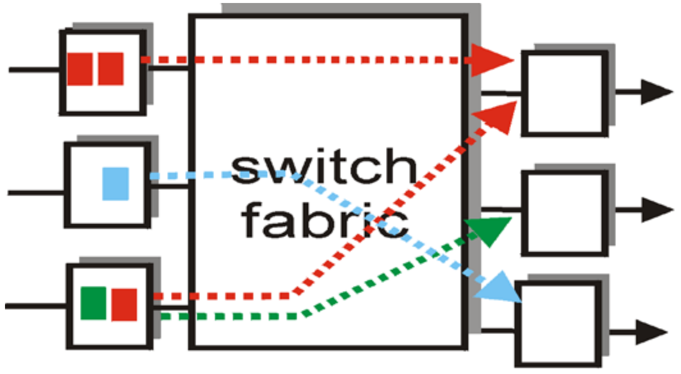
Yes. There are bits in the header that can be used. Too early to discuss those.

- Why are Drop-Tail and Random Early Drop the packet-dropping policies for the output queue? I would think it could also be applied to the input-queue.

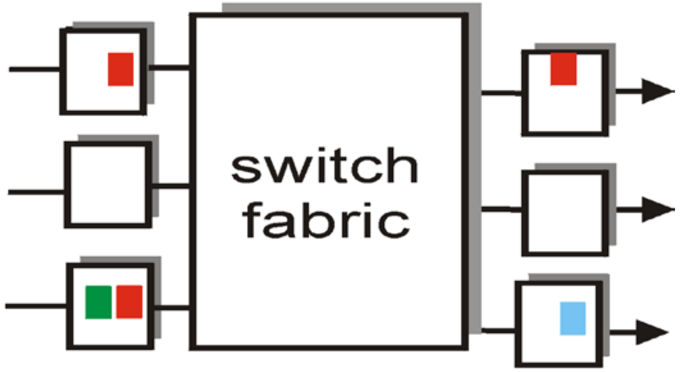
They are applied to the input queue.

Head-of-Line Blocking

- The packet at the head of the queue is waiting
  ⇒ Other packets can not be forwarded even if they go to another destination.



output port contention at time t - only one red packet can be transferred



green packet experiences HOL blocking

Student Questions

- Is HoL blocking just a pure negative, and should we always queue on the output side, or is there some tradeoff for the design?

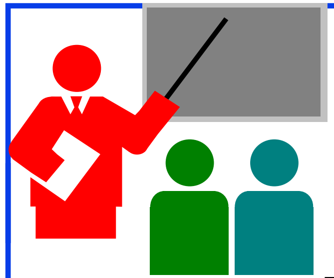
HoL is negative. Some slow fabric needs to process incoming packets faster and need an input queue.

- Why HOL blocking is an issue? Wouldn't the delay be very short?

Depends on who is blocking.

- Are there ways to avoid HOL blocking?

Yes, multiple queues by size or priority.



Network Layer Basics: Review

1. Forwarding uses a routing table to find the output port for datagrams using **the longest prefix match**. Routing protocols make the table.
2. IP provides only **the best-effort** service (KISS).
3. Routers include input/output ports, **switching fabric**, and processors.
4. Datagrams may be dropped even if the queues are not full (**Random early drop**).
5. Queueing at the input may result in **head-of-line blocking**.

Ref: Read Sections 4.1, 4.2, full, Page 303-330 of the textbook. Try R1 through R16

Student Questions

- What is the function of processors? Will it be included in the exam?

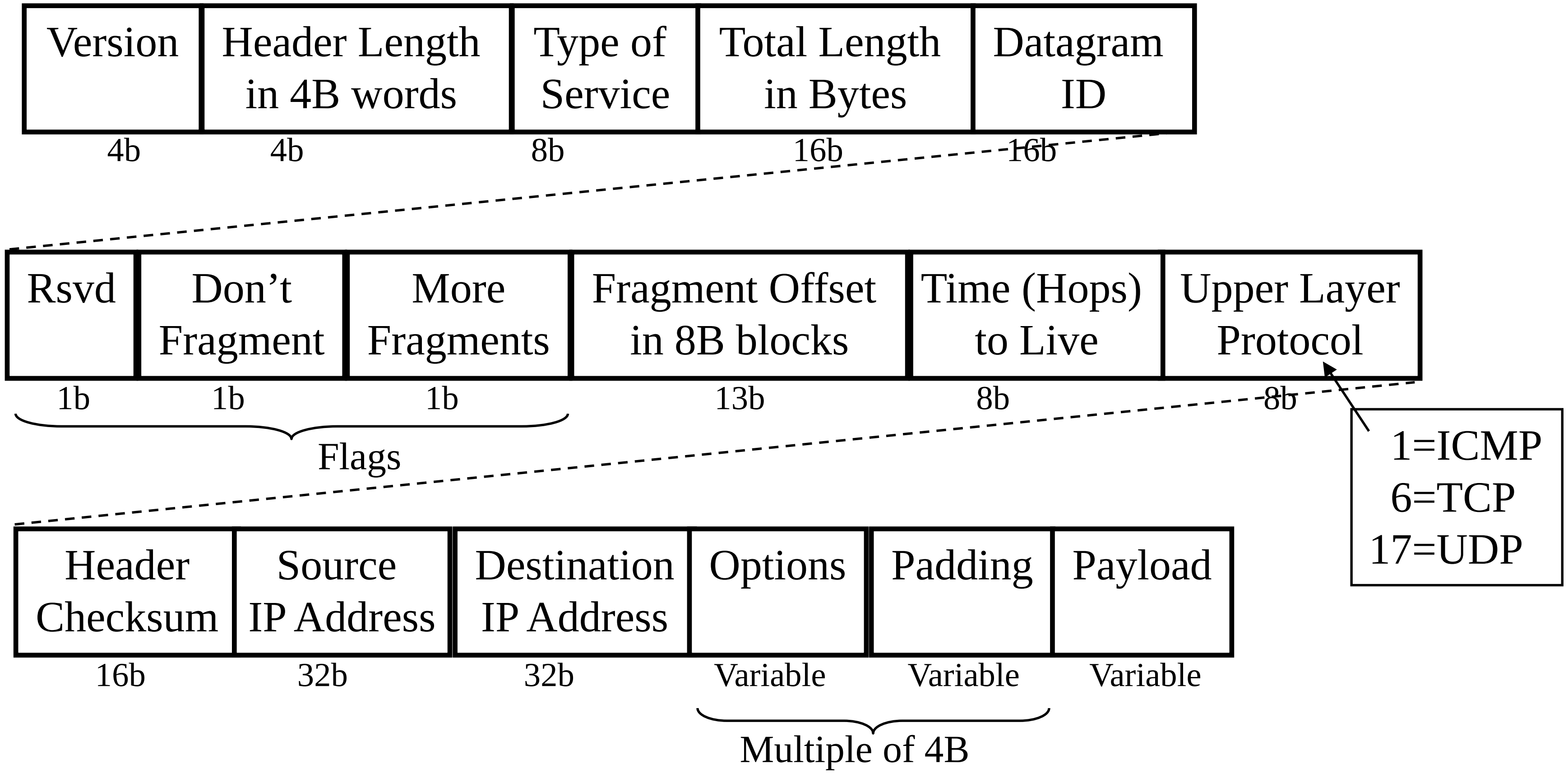
Routing processors are CPUs used for path computation. Yes, it is included.

Forwarding Protocols

1. IPv4 Datagram Format
2. IP Fragmentation and Reassembly
3. IP Addressing
4. Network Address Translation (NAT)
5. Universal Plug and Play
6. Dynamic Host Control Protocol (DHCP)
7. IPv6

Student Questions

IP Datagram Format



Student Questions

- ❑ To clarify, what type of service is not used? *It was not used for a long time. Several proposals have recently been made to use it. So it is used now.*
- ❑ Will it be possible for TTL to increase after processing?

No. TTL is the number of hops to live, specified when the packet first leaves the IP.

- ❑ What is the reserved bit used for?

Reserved for future extensions

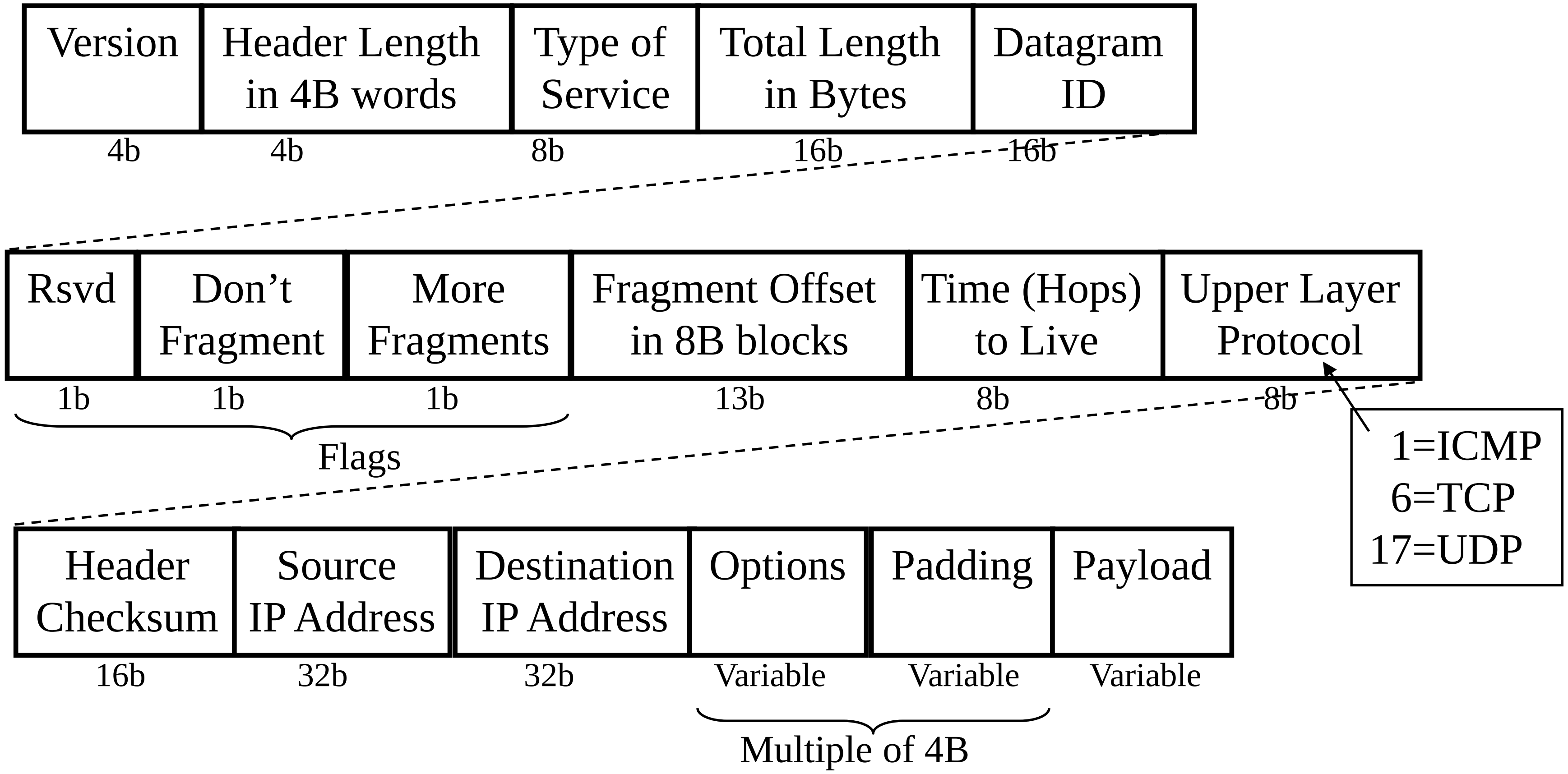
- ❑ Does the diagram show the whole length of the IP?

Yes

- ❑ Does the IP header have a fixed length?

No, options are variable length. Use 20B if not specified.

IP Datagram Format



Student Questions

- Why are source and destination IPs included again if they already exist in the UDP/TCP header?

TCP header contains port #, not an IP address. Each layer has its address. TCP addresses are called ports.

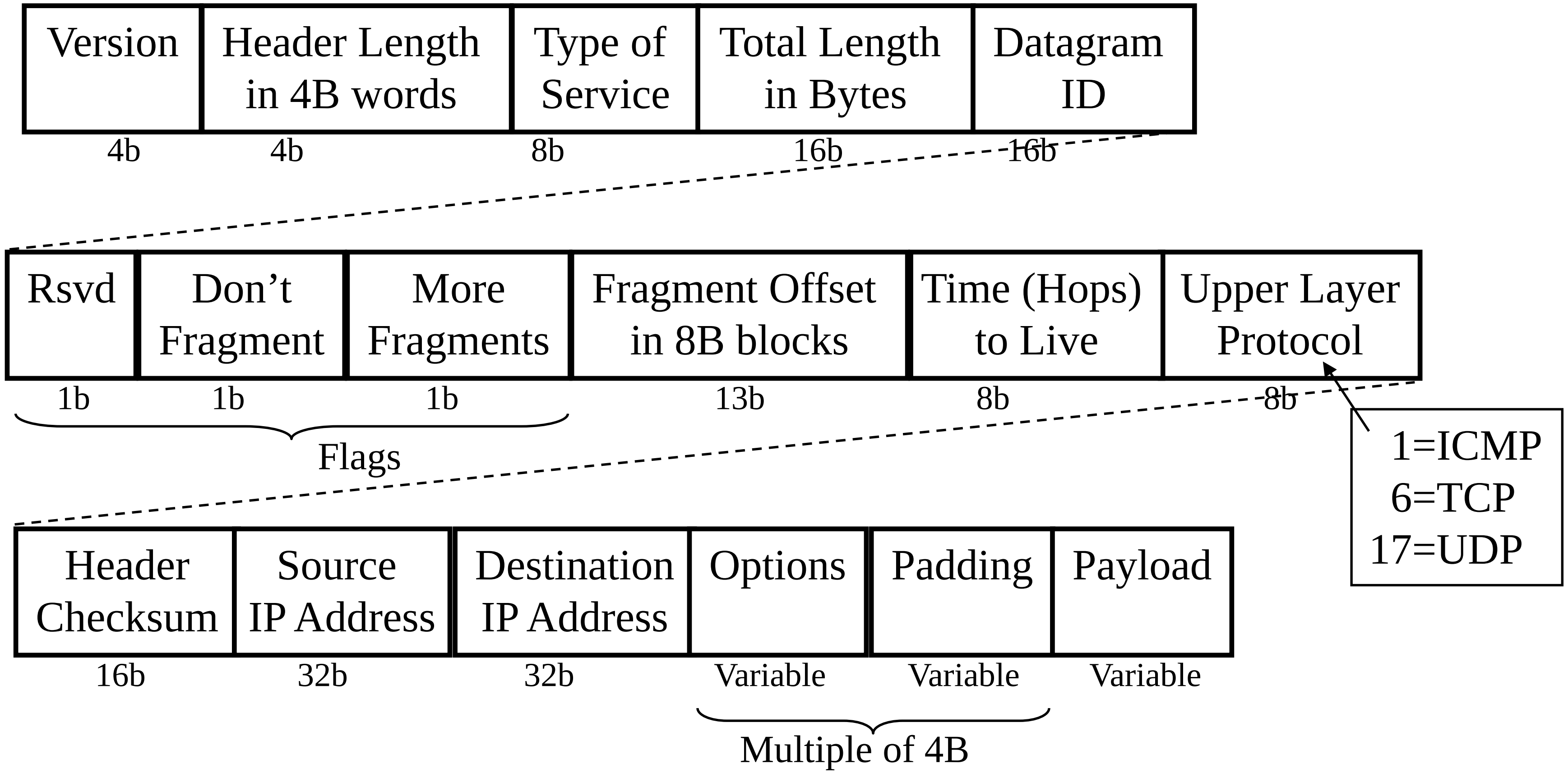
- Which block is the data from UDP/TCP in? Is it the payload portion?

Yes. Payload=L4 Header+L4 data

- Is the checksum just for the header or the header and the payload?

As labeled, it only covers the header.

IP Datagram Format



Student Questions

- Can you explain what the Header Length in the Datagram means?

Header length=Length of IP header

- What are the 4-byte words?

A word could be 16 bits, 32 bits, 64 bits, etc. 32b words are also 4-B words.

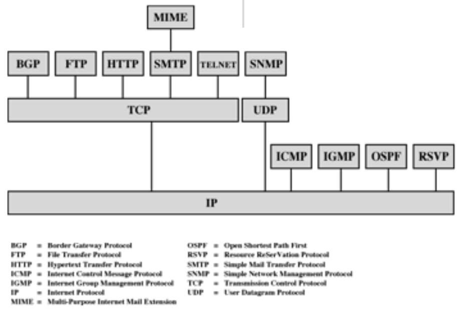
- Is it an extensive horizontal set of blocks that was too long to cover the slide?

Yes.

End of Part 1

IP Fragmentation Fields

- Header length: in units of 32-bit words
- Data Unit Identifier (ID)
  - Sending host puts an identification number in each datagram
- Total length: Length of user data plus header in bytes
- Fragment Offset - Position of a fragment in the original datagram
  - In multiples of 8-byte blocks
- *More fragments* flag
  - Indicates that this is not the last fragment
- Datagrams can be fragmented/refragmented at any router
- Datagrams are reassembled only at the destination host



Student Questions

- What are some examples of other Upper Protocol Layer numbers? How many are there?

See Slide 1.44 (Figure above)

- Does the total length include the other layers?

Higher layer headers are simply data for IP. Lower layers, it does not know.

- So, for the typical 20-byte IP header, will the header length be assigned to 5?

Yes. The header length is measured in units of 4-byte words.

- If we need to fragment a datagram, would it ever be beneficial to send two packets of roughly equal size instead of one of max size and another smaller? Could padding steps be skipped in this way? Padding can be avoided simply by sending headers and payloads of appropriate size. Not related to fragmentation.
- Does the data unit identifier change after passing each router?

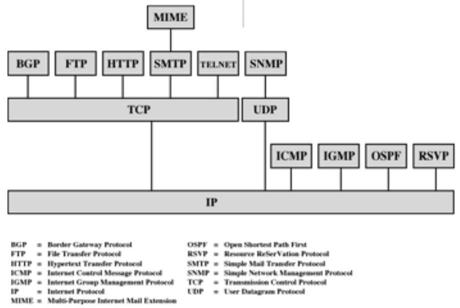
No. Set at the source.

- What are some common types of data unit identifiers?

It is just a sequence number.

IP Fragmentation Fields

- Header length: in units of 32-bit words
- Data Unit Identifier (ID)
  - Sending host puts an identification number in each datagram
- Total length: Length of user data plus header in bytes
- Fragment Offset - Position of a fragment in the original datagram
  - In multiples of 8-byte blocks
- *More fragments* flag
  - Indicates that this is not the last fragment
- Datagrams can be fragmented/refragmented at any router
- Datagrams are reassembled only at the destination host



Student Questions

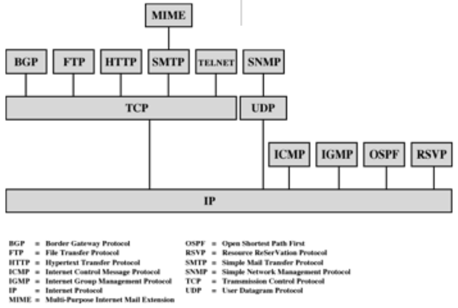
- Why perform refragment on an already fragmented datagram instead of fragmenting at once? *Every router has a different maximum datagram size.*
- Is the more fragments flag, therefore, set at pretty much all times? When/how is it un-set? *It is usually straightforward. Set only if there are "more" fragments.*
- Why does IPv6 not use fragmentation? (from the textbook)

To be discussed under IPv6

- Textbook says the unit of header length is 4 bits, but the video says it is 4 bytes. What is the unit of header length? *4 B*

IP Fragmentation Fields

- Header length: in units of 32-bit words
- Data Unit Identifier (ID)
  - Sending host puts an identification number in each datagram
- Total length: Length of user data plus header in bytes
- Fragment Offset - Position of a fragment in the original datagram
  - In multiples of 8-byte blocks
- *More fragments* flag
  - Indicates that this is not the last fragment
- Datagrams can be fragmented/refragmented at any router
- Datagrams are reassembled only at the destination host



Student Questions

- What's the purpose of refragmenting? How does a router refragment?

A fragment may be too large for another router. Reassembly is done only at the destination.

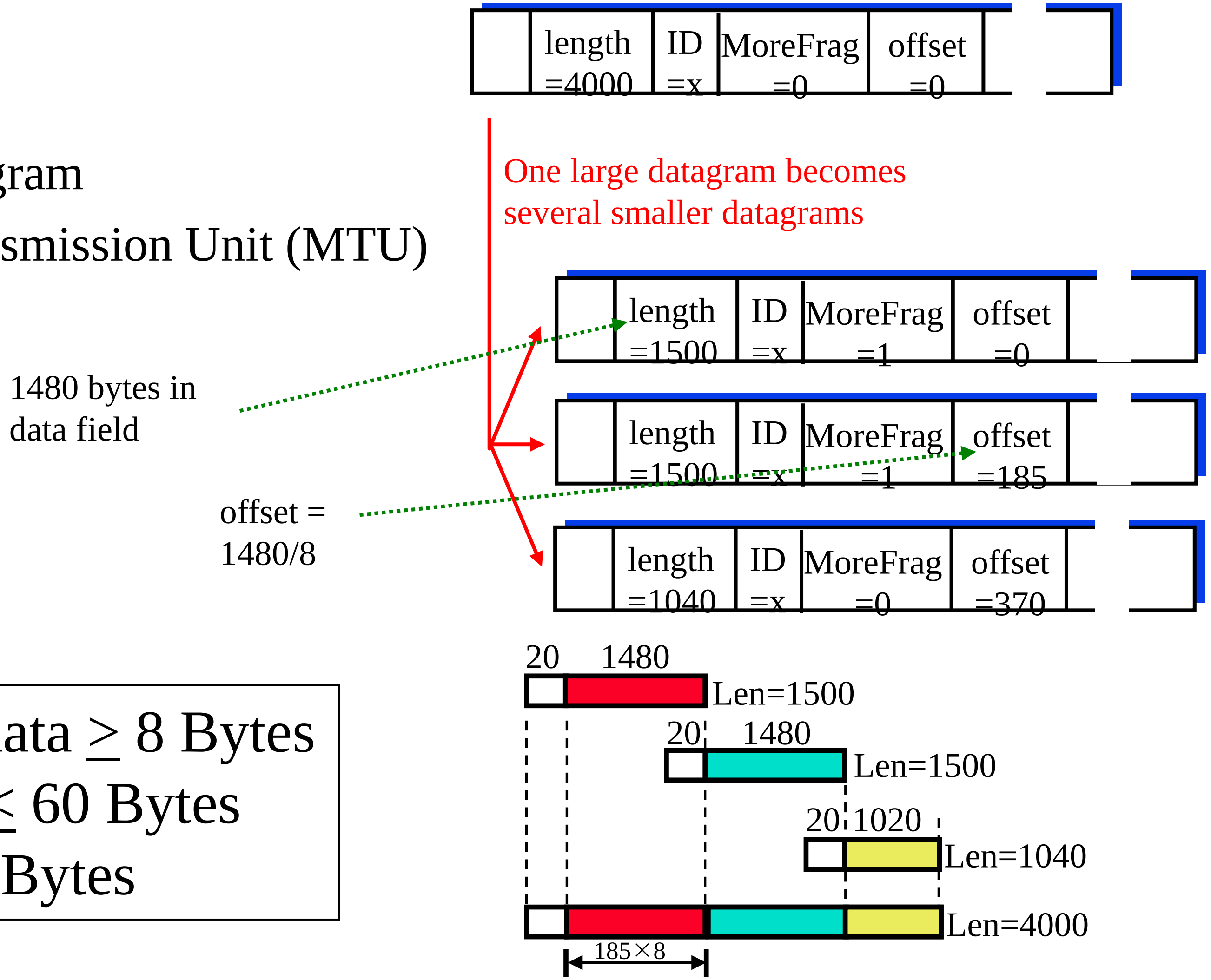
- If every datagram has an ID associated with it, why does it also store the offset? Could the receiver not just use the IDs to reassemble the datagram after fragmentation?

So that it would know where to put the fragment in the datagram?

IP Fragmentation and Reassembly

Example

- 4000 byte datagram
- Maximum Transmission Unit (MTU) = 1500 bytes



Fragment data ≥ 8 Bytes
IP Header ≤ 60 Bytes
MTU ≥ 68 Bytes

Student Questions

- Can you explain the cases for re-fragmentation and reassembling?

If a fragment cannot pass through a router, it is re-fragmented—Reassembly only at the destination.

- How do we decide the length of each segment?

TCP sends segments whenever it likes.

- Is there a length for each segment?

MSS=Max segment Size is set by TCP. To avoid fragmentation.

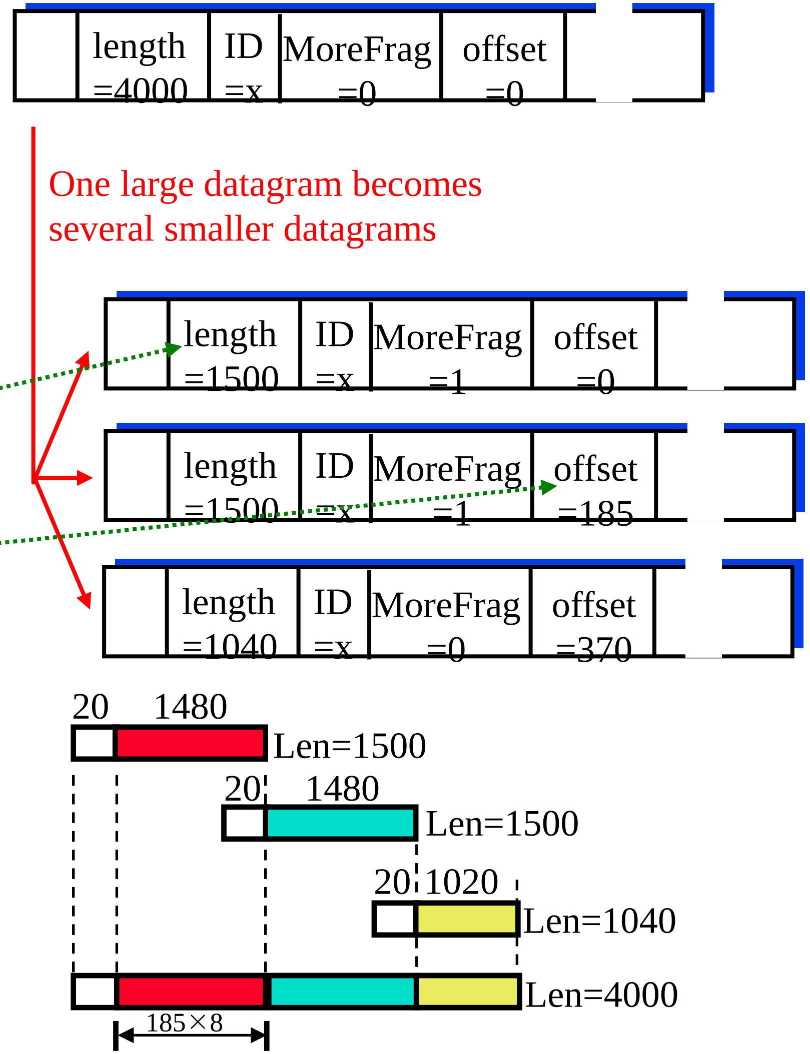
- Where does it indicate the length of the header? *See Slide 4.14*
- What is the unit of offset? *8B*
- Are fragments only reassembled at the destination? *Yes*
- Does the length have to be a multiple of 8? *Yes*
- Do we use a new header for reassembled datagrams?

Yes. It is the same as the original datagram sent from the source.

IP Fragmentation and Reassembly

Example

- 4000 byte datagram
- Maximum Transmission Unit (MTU) = 1500 bytes



Fragment data ≥ 8 Bytes
IP Header ≤ 60 Bytes
MTU ≥ 68 Bytes

Student Questions

- Why are the lengths in this example 1500 instead of 1480 again?

1480 data + 20 Header = 1500

- Can you explain this example about segments again because we can see nothing about the total length being 4000 bytes and the segment size being 1500 bytes?

The first datagram is 4000 bytes. It is broken into fragments of 1500, 1500, and 1040 bytes.

- Does this example assume that the IP header has a fixed length of 20 bytes? So that we can get a data field length of 1480 bytes.

Yes. This assumes a 20B header.

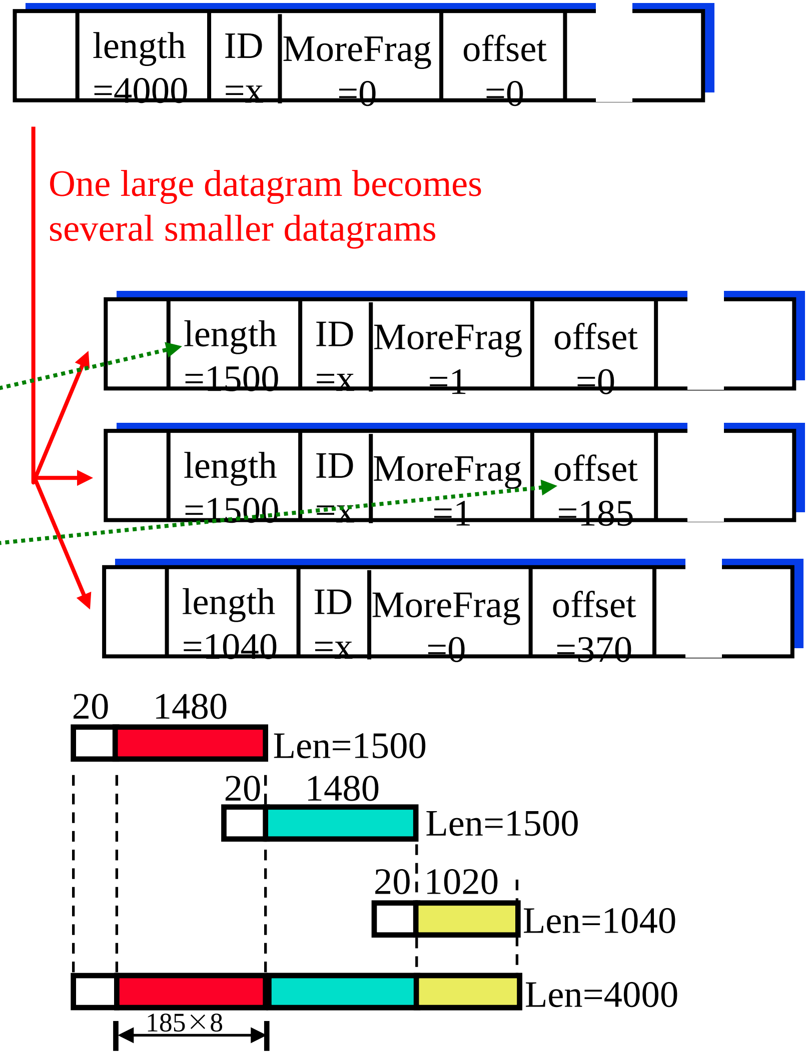
- If the offset calculated for a fragment is decimal, do we round down or up to the nearest multiple of 8?

All fragments are multiples of 8, so adjust the offset accordingly.

- Does every (non-last) datagram have to have a length equal to MTU?

Less than or equal to.

IP Fragmentation and Reassembly



Example

- 4000 byte datagram
- Maximum Transmission Unit (MTU) = 1500 bytes

Fragment data ≥ 8 Bytes
IP Header ≤ 60 Bytes
MTU ≥ 68 Bytes

Student Questions

- If the offset is not an integer, will it be rounded up or down? How would that affect the length/remaining fields?

It will be rounded down.

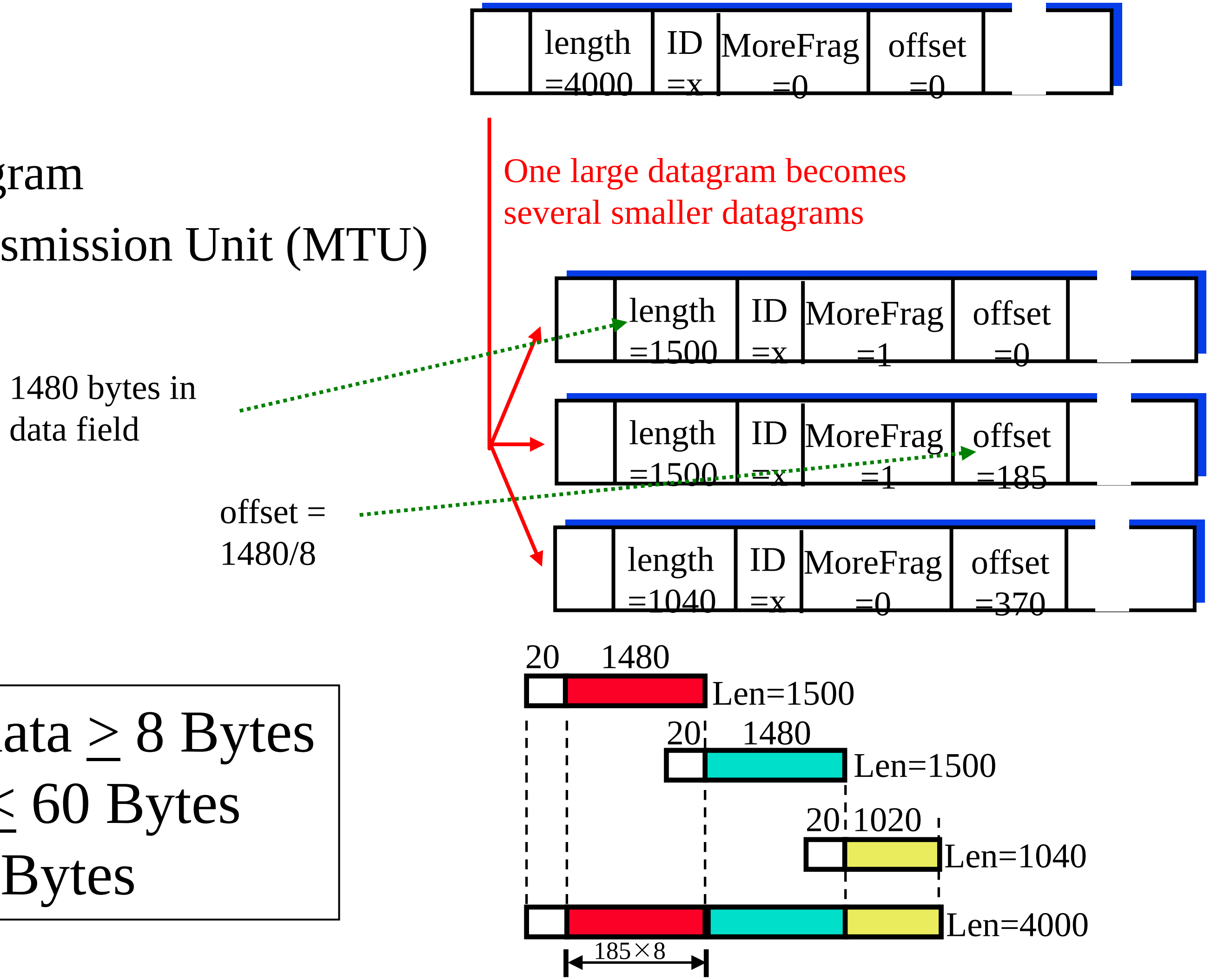
- If one large datagram becomes several smaller datagrams, do they combine later to become one datagram again at some point?

Yes, the reassembly is done at the destination before the payload is delivered to L4.

IP Fragmentation and Reassembly

Example

- 4000 byte datagram
- Maximum Transmission Unit (MTU) = 1500 bytes



Fragment data ≥ 8 Bytes
IP Header ≤ 60 Bytes
MTU ≥ 68 Bytes

Student Questions

- How can datagrams be broken down into fragments where the sum of the fragments is greater than the datagram (ex. 1500+1500+1040 > 4000)?

It is necessary because each router has a maximum size it can handle.

- How do we know how much greater than the datagram the final sum should be?

Fragment = Fragment header + Payload
ΣFragment = ΣFragment header + ΣPayload

Homework 4A: Fragmentation

- [8 points] Consider sending a 3500-byte datagram into a link that has an MTU of 800 bytes. Suppose the original datagram is stamped with the identification number 450. How many fragments are generated? What are the values in the various fields in the IP datagram(s) generated related to fragmentation?

Student Questions

- What is the identification number? Can you explain how to find the ID again? Does it go as 2^0, 2^1, 2^2, ... 2^{16}, then back to 2^0?

It is the sequence number. It goes 1, 2, 3, 4, 5, …

- Can we go over homework 4a fragmentation?

Sure.

- Can you explain how the length=796 is calculated here when MTU=800?

It is 776, not 796.

IP Address Classes

- Class A:

0	Network	Local
1	7	24 bits

- Class B:

10	Network	Local
2	14	16 bits

- Class C:

110	Network	Local
3	21	8 bits

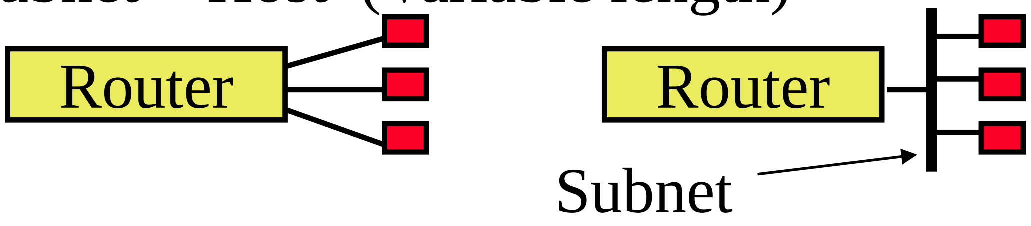
- Class D:

1110	Host Group (Multicast)
4	28 bits

- Class E:

11110	Future use
5	27 bits

- Local = Subnet + Host (Variable length)



Student Questions

- What do network and local here mean?

The Internet is a network of networks. Each home/business is a network.

- Why do I seldom see classes D and E?

Multicasts are common. Classes are now history. We use Classless Inter-Domain Routing (CIDR). See Slide 4.21

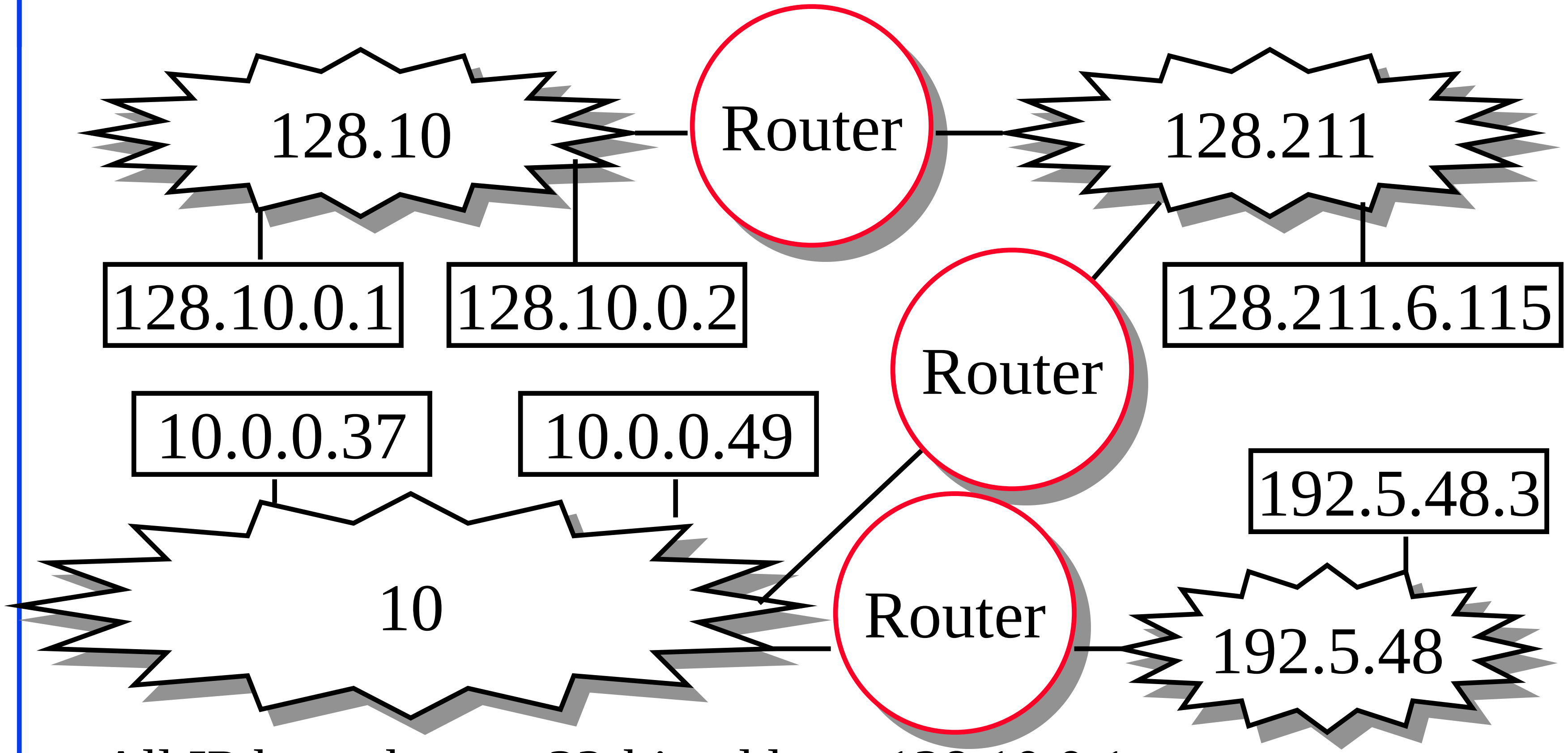
- Why are there many classes for IP addresses?

This is how it was done initially. This explains why WUSTL has more public addresses than many countries.

- What is a multicast?

A packet that needs to go to multiple nodes.

IP Addressing



- All IP hosts have a 32-bit address.128.10.0.1 =1000 0000 0000 1010 0000 0000 0000 0001
- All hosts on a network have the same network prefix

Student Questions

- Does being on the same network mean connecting to the same Wi-Fi, local ISP, or enterprise?

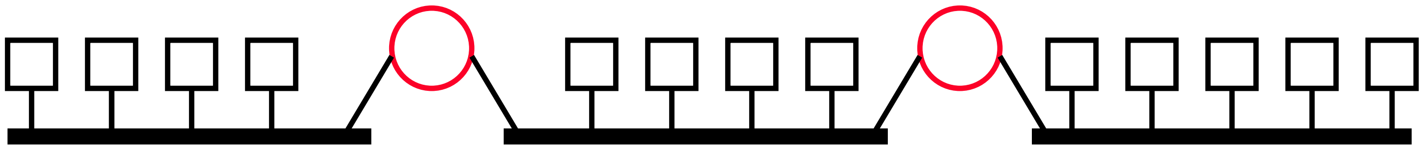
They are connected to the same subnet.
One subnet = One L2
= One Ethernet or Wi-Fi

- Can devices connecting to the WashU network have different prefixes if the campus is large enough?

Yes. There are many routers on the campus.

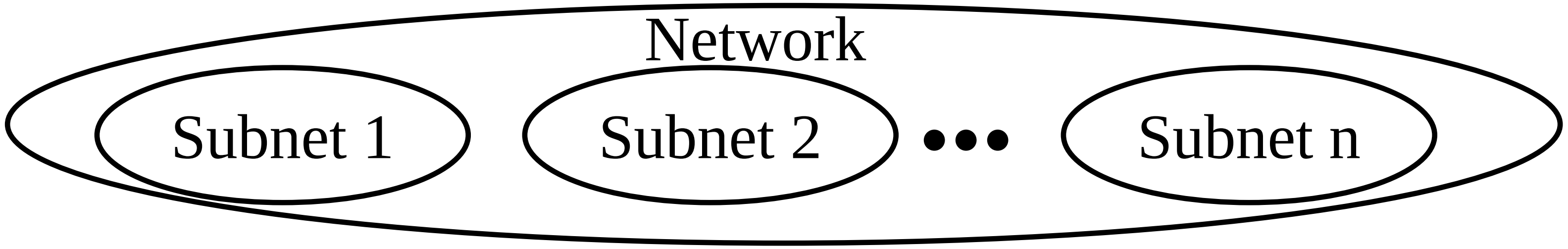
End 2/14/24

Subnetting



- All hosts on a subnetwork have the same prefix.
  The position of the prefix is indicated by a "subnet mask."
- Example: First 23 bits = subnet

Address:	10010100 10101000 00010000 11110001
Mask:	11111111 11111111 11111110 00000000
.AND.	10010100 10101000 00010000 00000000



Student Questions

- Are the subnet bits always the first n bits (a prefix)? Then why are class B's local (subnet+host) bits the last 16 bits?

Local=Host

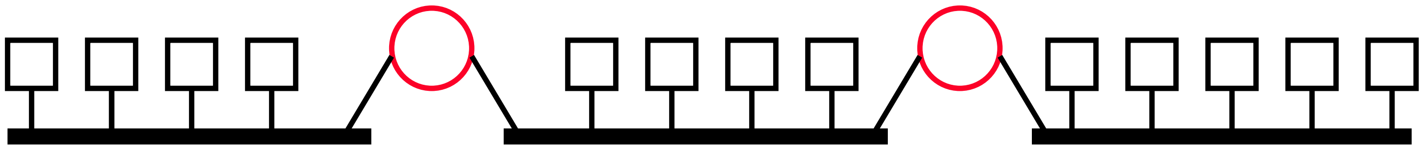
- Is there a trick other than simply increasing the number of bits that allows us to increase the number of devices on the network arbitrarily?

Private addresses. To be discussed.

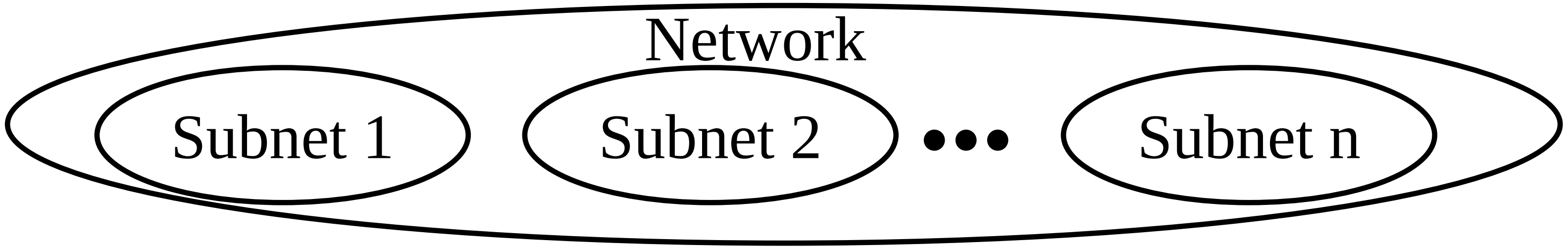
- How is the number of bits in the subnet mask decided?

By network admin

Subnetting



- All hosts on a subnetwork have the same prefix. The position of the prefix is indicated by a "subnet mask."
- Example: First 23 bits = subnet
  Address: 10010100 10101000 00010000 11110001
  Mask: 11111111 11111111 11111110 00000000
  .AND. 10010100 10101000 00010000 00000000



Student Questions

- How are packets handled differently when both hosts are within a subnet or in two different subnets?

A router is not required for forwarding within the same subnet.

- What does ".AND." mean?

Boolean AND operation

- Why should we use a subnet?

L3 forwards packets from one L2 and another. Think of L3=House, L2=Rooms in the house

- Can hosts on different subnetworks have the same IP address?

No, Not within one enterprise.

IP addressing: CIDR

- CIDR: Classless Interdomain Routing
  - Subnet portion of the address of arbitrary length
  - Address format: a.b.c.d/x, where x is # bits in the subnet portion of the address
  - All 1's in the host part is used for subnet broadcast
  - All 0s in the host part were meant as "subnet address" but not used for anything. Some implementations allow it to be used as a host address. Some don't. It's better to avoid it.

← subnet part → ← host part →

11001000 00010111 00010000 00000000

200.23.16.0/23

Student Questions

- Can you explain what is "All 1's in the host part is used for subnet broadcast"?

It is easy to detect.

- When using a computer's loopback IP address, do those datagrams get sent to the router or go direct to the local machine?

Loopback is at the local IP layer. Do not get out of the system.

- What is an example of a scenario when we would want to use multicast?

I want to say something to "all routers on this network."

- What is the difference between multicast and broadcast?

Broadcast goes to all nodes. Multicast goes to a subset.

- If a subnet has a mask of length 31 (a.b.c.d/31). Does it mean there can only be two possible hosts in this subnet?

Yes. However, some people use all 1's for subnet broadcast and all 0's for null.

Homework 4B: Subnets

- [18 points] Consider a router interconnecting three subnets: Subnet 1, Subnet 2, and Subnet 3. Suppose all of the interfaces in these three subnets must have the prefix 223.1.17/24. Also, suppose that Subnet 1 is required to support up to 60 interfaces, Subnet 2 is to support up to 80 interfaces, and Subnet 3 is to support up to 30 interfaces. Provide three network address prefixes (a.b.c.d/x) that satisfy these constraints. **Use adjacent allocations**. For each subnet, list the subnet mask to use in the hosts.

Student Questions

- Can you explain the answers to this question? *Sure.*
- Does DHCP use a greedy algorithm to arrange the subnets, or can it enumerate all the arrangements? In the example, if there is one more subnet, it may be complete, but this could be solved by putting subnet two at the beginning.

Subnet design is done manually. DHCP allocates the addresses it is given.

- The prefix for for each subnet (A, B, C) must be longer than 24 bits to narrow the address space for each subnet further, right?

Yes.

- For HW4B, why /25 network must start at either 0 or 128?

/25 leaves 7 bits for assignment.

Forwarding an IP Datagram

- Delivers **datagrams** to the destination network (subnet)
- Routers maintain a "routing table" of "next hops."
- Next Hop field does not appear in the datagram

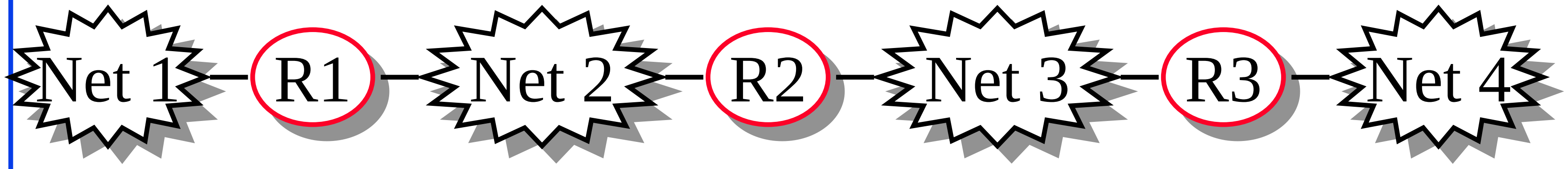


Table at R2:

Destination	Next Hop
Net 1	Forward to R1
Net 2	Deliver Direct
Net 3	Deliver Direct
Net 4	Forward to R3

Student Questions

- What is the length of the IP datagram header? Does it vary?

See Slide 4-14

- Do the other layers' headers need to get duplicated for each fragment?

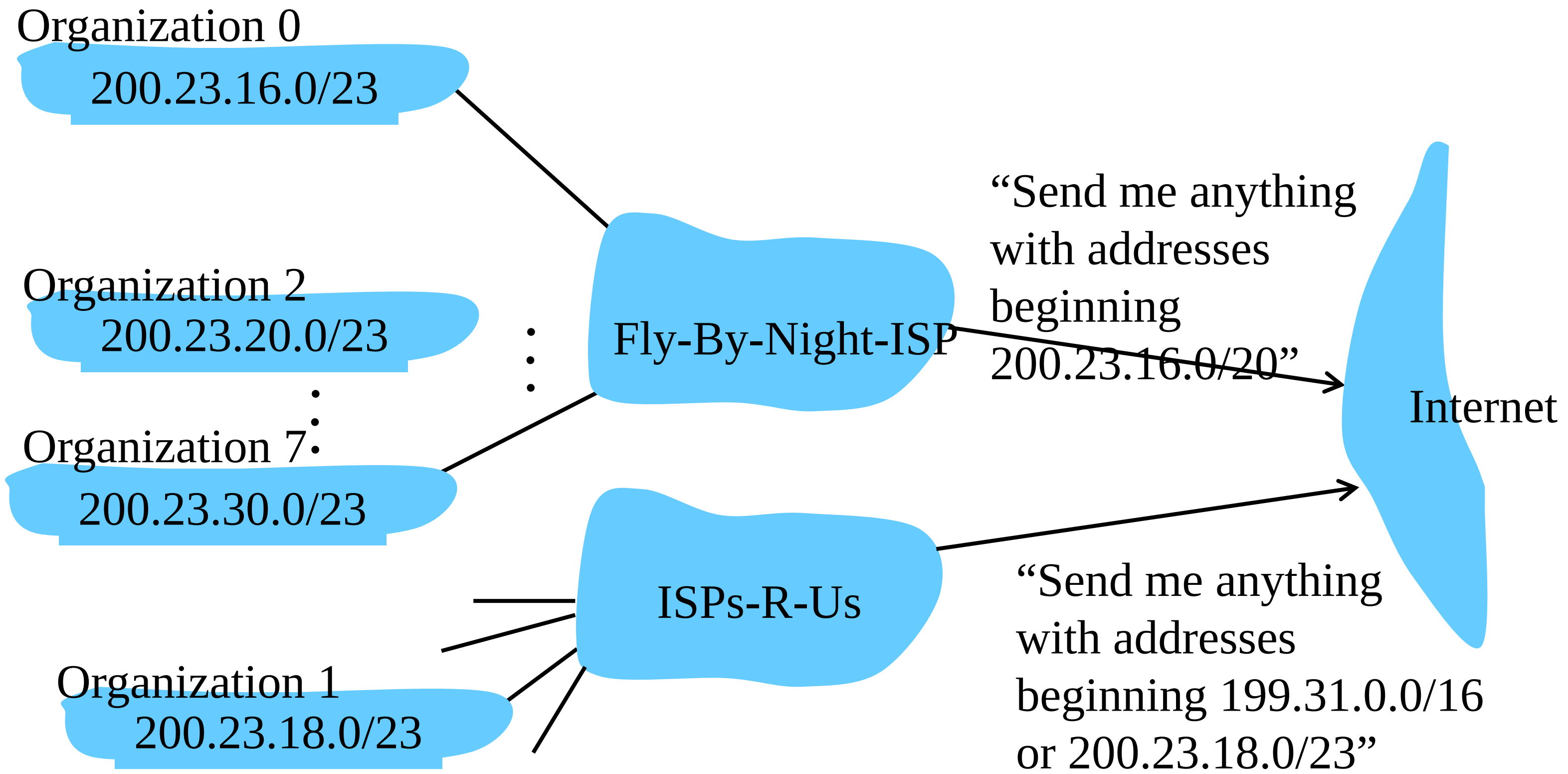
IP only cares about its headers. Its header gets duplicated. Other layers are part of the data.

- Are routing tables updated? If not, is there always the same path from source to destination?

Yes, they are updated based on time and events (link/router going up/down).

Route Aggregation

- Can combine two or more prefixes into a shorter prefix
- ISPs-R-Us has a more specific route to organization 1



Ref: Section 4.3.2. Try R27, R28

Student Questions

- Why would we want to integrate multiple subnets? *To reduce the number of entries in the routing table*
- Who decides which ISPs get which IPs?

IANA=Internet Assigned Numbers Authority

- Just to be sure, how come organizations 2 and 7 would be aggregated with organization 0 when the text specifies the third section of the address should be 16?

Orgs 0-7 are with the first ISP. The second ISP has two blocks.

"Route Print" Command in Windows

MAC: netstat -rn

```
===========================================================================
Interface List
0x1 ........................ MS TCP Loopback interface
0x2 ...00 16 eb 05 af c0 ...... Intel(R) WiFi Link 5350 - Packet Scheduler Miniport
0x3 ...00 1f 16 15 7c 41 ...... Intel(R) 82567LM Gigabit Network Connection - Packet Scheduler Miniport
0x40005 ...00 05 9a 3c 78 00 ...... Cisco Systems VPN Adapter - Packet Scheduler Miniport
===========================================================================
===========================================================================
Active Routes:
Network Destination        Netmask          Gateway       Interface  Metric
          0.0.0.0          0.0.0.0      192.168.0.1   192.168.0.108       10
          0.0.0.0          0.0.0.0      192.168.0.1   192.168.0.106       10
        127.0.0.0        255.0.0.0        127.0.0.1       127.0.0.1        1
      169.254.0.0      255.255.0.0    192.168.0.106   192.168.0.106       20
      192.168.0.0    255.255.255.0    192.168.0.106   192.168.0.106       10
      192.168.0.0    255.255.255.0    192.168.0.108   192.168.0.108       10
    192.168.0.106  255.255.255.255        127.0.0.1       127.0.0.1       10
    192.168.0.108  255.255.255.255        127.0.0.1       127.0.0.1       10
    192.168.0.255  255.255.255.255    192.168.0.106   192.168.0.106       10
    192.168.0.255  255.255.255.255    192.168.0.108   192.168.0.108       10
        224.0.0.0        240.0.0.0    192.168.0.106   192.168.0.106       10
        224.0.0.0        240.0.0.0    192.168.0.108   192.168.0.108       10
  255.255.255.255  255.255.255.255    192.168.0.106   192.168.0.106        1
  255.255.255.255  255.255.255.255    192.168.0.106           40005        1
  255.255.255.255  255.255.255.255    192.168.0.108   192.168.0.108        1
Default Gateway:       192.168.0.1
===========================================================================
Persistent Routes:
  None
```

Adr & mask = Dest
⇒ Match

Longest Prefix match is used

Metric: Lower is better

Note: 127.0.0.1 = Local Host, 224.x.y.z = Multicast on local LAN

Student Questions

- Do packets sent to 127.0.0.1 leave the computer onto the network before returning, or is it all internal?

Internal loopback.

- What is the difference between the interface and the gateway? What is network destination vs. gateway? How do you know which interface the given address specifies under that field?

Interface=Adapter
Gateway=Router
Net. Destination=Dest Adr

- What are the differences between a gateway and an interface?

Gateway=Router
Interface =Physical port address on the machine

- Is there a persistent route for unicasts?

Any change in any link may change the route.

"Route Print" Command in Windows

MAC: netstat -rn

```
===========================================================================
Interface List
0x1 ........................ MS TCP Loopback interface
0x2 ...00 16 eb 05 af c0 ...... Intel(R) WiFi Link 5350 - Packet Scheduler Miniport
0x3 ...00 1f 16 15 7c 41 ...... Intel(R) 82567LM Gigabit Network Connection - Packet Scheduler Miniport
0x40005 ...00 05 9a 3c 78 00 ...... Cisco Systems VPN Adapter - Packet Scheduler Miniport
===========================================================================
===========================================================================
Active Routes:
Network Destination        Netmask          Gateway       Interface  Metric
          0.0.0.0          0.0.0.0      192.168.0.1   192.168.0.108       10
          0.0.0.0          0.0.0.0      192.168.0.1   192.168.0.106       10
        127.0.0.0        255.0.0.0        127.0.0.1       127.0.0.1        1
      169.254.0.0      255.255.0.0    192.168.0.106   192.168.0.106       20
      192.168.0.0    255.255.255.0    192.168.0.106   192.168.0.106       10
      192.168.0.0    255.255.255.0    192.168.0.108   192.168.0.108       10
    192.168.0.106  255.255.255.255        127.0.0.1       127.0.0.1       10
    192.168.0.108  255.255.255.255        127.0.0.1       127.0.0.1       10
    192.168.0.255  255.255.255.255    192.168.0.106   192.168.0.106       10
    192.168.0.255  255.255.255.255    192.168.0.108   192.168.0.108       10
        224.0.0.0        240.0.0.0    192.168.0.106   192.168.0.106       10
        224.0.0.0        240.0.0.0    192.168.0.108   192.168.0.108       10
  255.255.255.255  255.255.255.255    192.168.0.106   192.168.0.106        1
  255.255.255.255  255.255.255.255    192.168.0.106           40005        1
  255.255.255.255  255.255.255.255    192.168.0.108   192.168.0.108        1
Default Gateway:       192.168.0.1
===========================================================================
Persistent Routes:
  None
```

Adr & mask = Dest
$\Rightarrow$ Match

Longest Prefix match is used

Metric: Lower is better

Note: 127.0.0.1 = Local Host, 224.x.y.z = Multicast on local LAN

Student Questions

- ❑ Do the routing table rules apply to incoming, outgoing, or both?

Outgoing only.

- ❑ Could you trace a forwarding example through the table rules?

Yes. See next slide (Homework)

- ❑ How does the metric value work when it is the same for multiple routes?

Either path can be taken if both metric and prefix-match lengths are identical.

Lab 4A: Routing Table

- [8 Points] Use "Route Help" in Windows (or man route in MAC) to learn the route command
- Ping www.google.com to find its address
- Ensure you have two active interfaces, preferably connected to different routers. For example, create a 2nd interface by connecting a smartphone hotspot via USB. You can also connect two smartphone hotspots to a single computer using two USB cables.
- Print route table
- Trace route to www.google.com using tracert
- Modify the routing table so that the other interface will be used.
- Note the command you used to modify the routing table
- Print the new routing table
- Trace route to the same numeric address for www.google.com as before. Submit underlined items.

Student Questions

- Don't have a phone hotspot? Could I use a non-washu VPN?

I am still determining if traceroute will work with VPN. Did you try, and did it work?

- During my traceroute, the hops differed for the first two routers but ultimately joined on the 3rd hop.

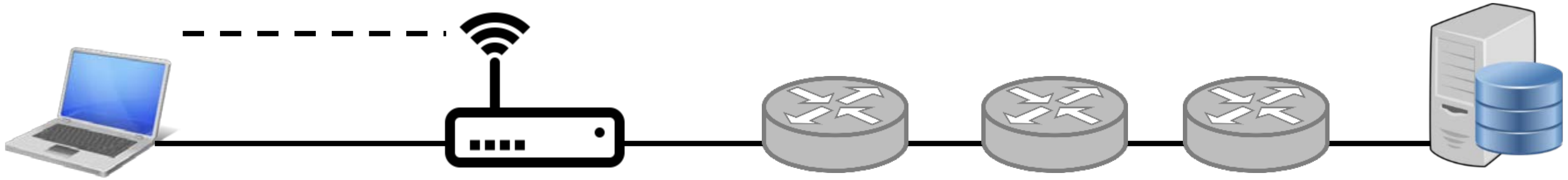
That's correct.

- Whenever I connect to my hotspot, my default gateway changes.

The computer constantly modifies the default gateway to use the fastest link. However, you can divert SOME or all of the traffic to the alternate link if there are two links. Each destination has a separate route as specified by the routing table rows. The default gateway is used only if no other rows match that destination. So you need to have two separate connections to the internet. You can still connect two smartphones to two USB ports for those without Ethernet.

Lab 4A Hints

- A host with two interfaces going to the same router:



- Trace route result will not change even if you change the interface.

```
IPv4 Route Table
===========================================================================
Active Routes:
Network Destination        Netmask          Gateway       Interface  Metric
          0.0.0.0          0.0.0.0      192.168.0.1    192.168.0.152     55
          0.0.0.0          0.0.0.0      192.168.0.1    192.168.0.151     25
```

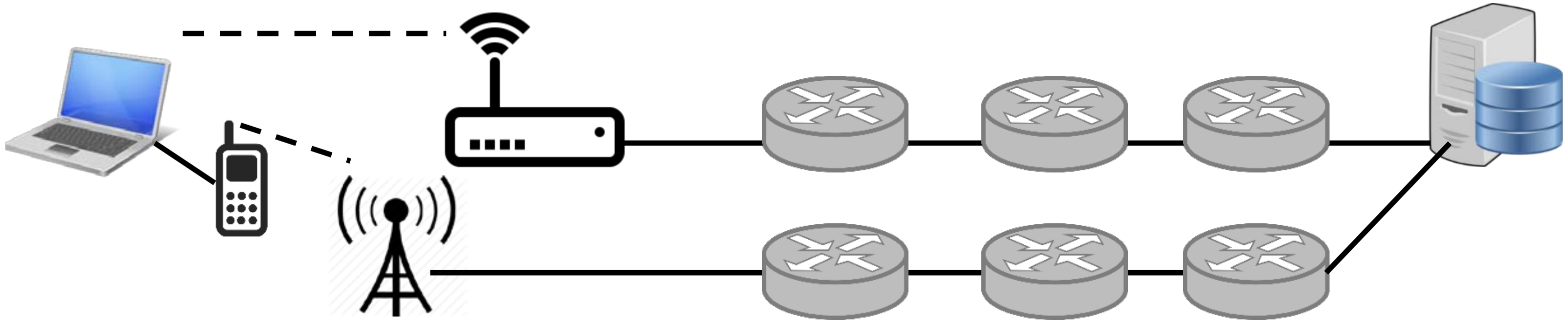
Student Questions

- What is the command for MACOS?

The command for MACOS is sudo route -n add -net <ip Address> <gateway>
For example,
"sudo route -n add -net 192.168.1.0/24 10.10.10.1."
To add two flags, e.g., n and v, write "sudo route -n -v add…"

Lab 4A Hints (Cont)

- If you have two routers, you can see the effect in trace route. One way to get two routers is to use your cell phone hot spot:



IPv4 Route Table

Active Routes:

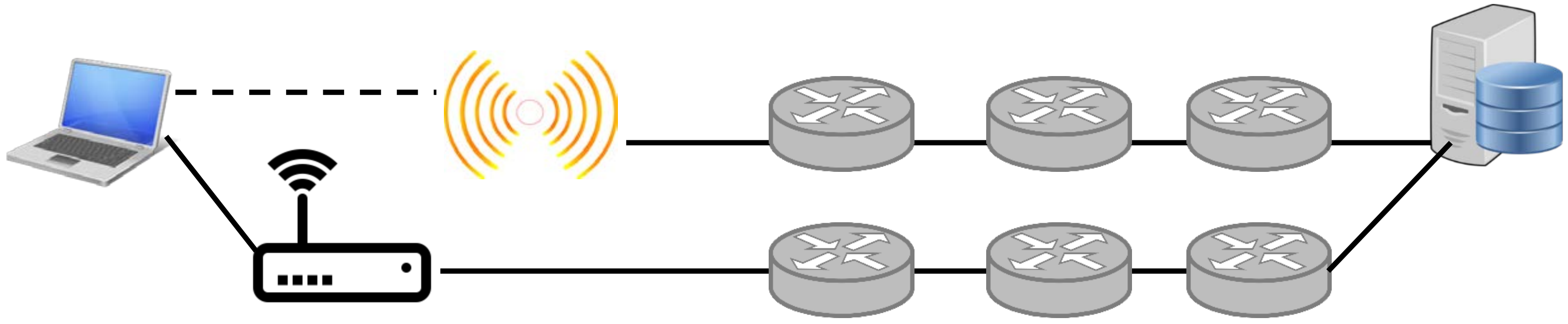
Network Destination	Netmask	Gateway	Interface	Metric
0.0.0.0	0.0.0.0	192.168.0.1	192.168.0.151	25
0.0.0.0	0.0.0.0	172.20.10.1	172.20.10.2	35

- WiFi on your phone should be disabled to ensure that it does not forward traffic to the same home router.

Student Questions

Lab 4A Hints (Cont)

- Another way to get two routers is to use another router. We have placed an extra router in our lab.



IPv4 Route Table
===
Active Routes:

Network Destination	Netmask	Gateway	Interface	Metric
0.0.0.0	0.0.0.0	192.168.0.1	192.168.0.151	25
0.0.0.0	0.0.0.0	172.20.10.1	172.20.10.2	35

Student Questions

Lab 4A Hints (Cont)

- WWW.google.com may have different IP addresses on different networks, so trace the route to the same numeric address.
- WUSTL VPN rejects all traffic not going to WUSTL.
  So, it can not be used as the 2nd interface.
- The new metric assigned by the route command may not be what you specified. So always check using route print.

Student Questions

Lab 4A Hints (Cont)

A. Use "route help" to learn the route command

- **Windows:** route help
- **Linux:** route help
- **MAC:**
  - man netstat
  - man route

B. Ping www.google.com to find its address

- ping www.google.com

C. Print the new routing table

- **Windows:**
  - route print
- **Linux:**
  - route
- **MAC:**
  - netstat -nr

D. Modify routing tables

- **Windows:**
  - route add/delete/change
- **Linux:**
  - route add/del
- **MAC:**
  - sudo route –nv add

E. Verify using tracert

- **Windows:**
  - tracert
- **Linux:**
  - traceroute
- **MAC:**
  - traceroute

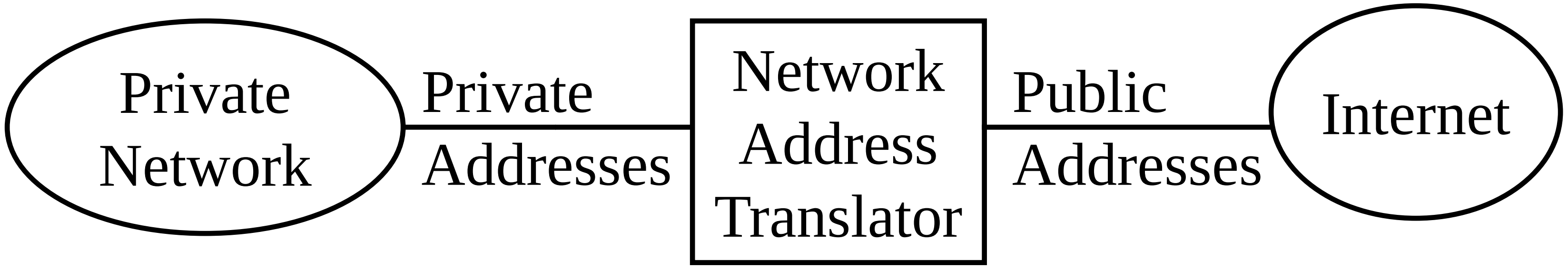
Student Questions

- Can you elaborate on what VPN does precisely?

VPN creates a "virtual link" between your computer and the company. All traffic is encapsulated and encrypted.

Private Addresses

- Any organization can use these inside their network Can't go on the internet. [RFC 1918]
- 10.0.0.0 - 10.255.255.255 (10/8 prefix)
- 172.16.0.0 - 172.31.255.255 (172.16/12 prefix)
- 192.168.0.0 - 192.168.255.255 (192.168/16 prefix)



Student Questions

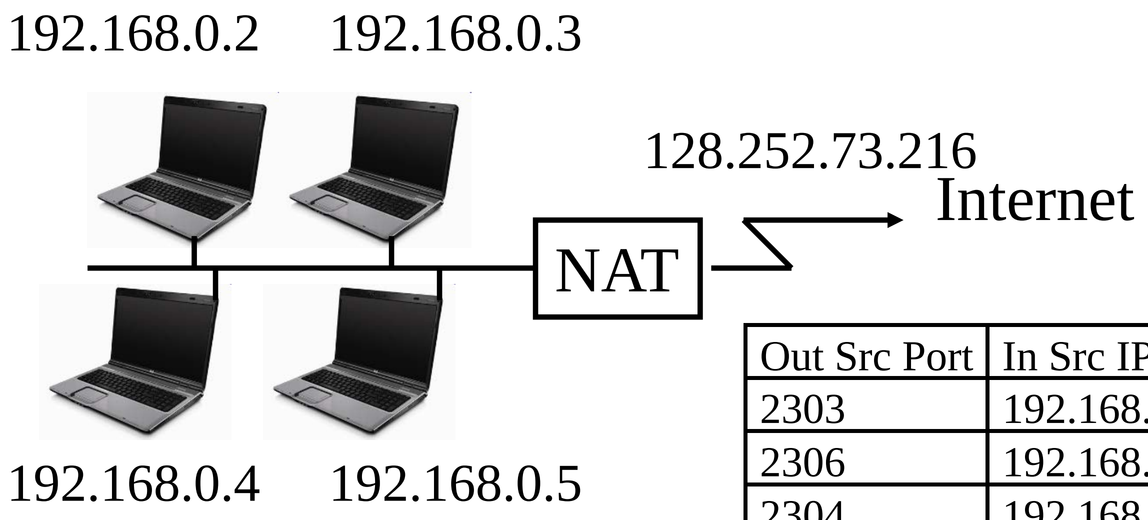
- Do we need to remember these private addresses for the exam?

Yes.

- Do ISPs ever assign private addresses for cell phones on their network?

Yes

Network Address Translation (NAT)



Out Src Port	In Src IP	In Src Port
2303	192.168.0.2	3385
2306	192.168.0.3	3385
2304	192.168.0.2	30384

- Private IP addresses 192.168.x.x: Can be used inside networks
- Cannot be used on the public Internet
- NAT overwrites the source address and source port on all outgoing packets with a unique port #. Then, it rewrites the correct destination IP and port on all incoming packets using the destination port in the packet as a key.
- Only outgoing connections are possible

Student Questions

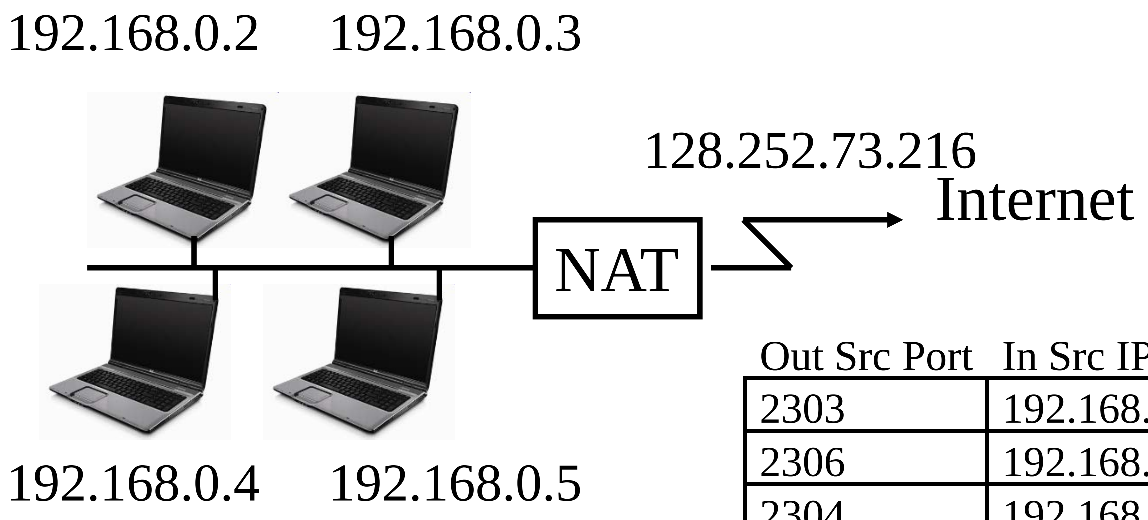
- Is incoming UDP traffic forwarded differently by NAT? *No*
- Does each subnet usually have a DHCP? Does DHCP assign private or public addresses?

Yes, but you can use statically assigned addresses and will not need a DHCP server. DHCP can assign whatever address range you give. Most companies don't have that many public addresses. Some companies do. E.g., WUSTL.

- How do hosts get a more permanent IP address? For example, a web server shouldn't constantly be changing IPs. *You can build the address in the server itself. Or ask your DHCP server (router) to assign it a fixed address.*
- What is a private IP address's purpose if it will be translated eventually? Is every private IP address translated to the same public address under the same network?

Yes. One public address can support millions of private addresses.

Network Address Translation (NAT)



Out Src Port	In Src IP	In Src Port
2303	192.168.0.2	3385
2306	192.168.0.3	3385
2304	192.168.0.2	30384

- Private IP addresses 192.168.x.x: Can be used inside networks
- Cannot be used on the public Internet
- NAT overwrites the source address and source port on all outgoing packets with a unique port #. Then, it rewrites the correct destination IP and port on all incoming packets using the destination port in the packet as a key.
- Only outgoing connections are possible

Student Questions

- P2P applications require direct connections between devices, but doesn't NAT prevent that direct connection on the internet?

No. All connections through NAT still look "direct."

- How does the NAT know which device to forward an incoming packet to if many devices share a public address?

See 4th bullet. A sample table has been added to make it more clear.

- How would two devices behind NAT find each other?

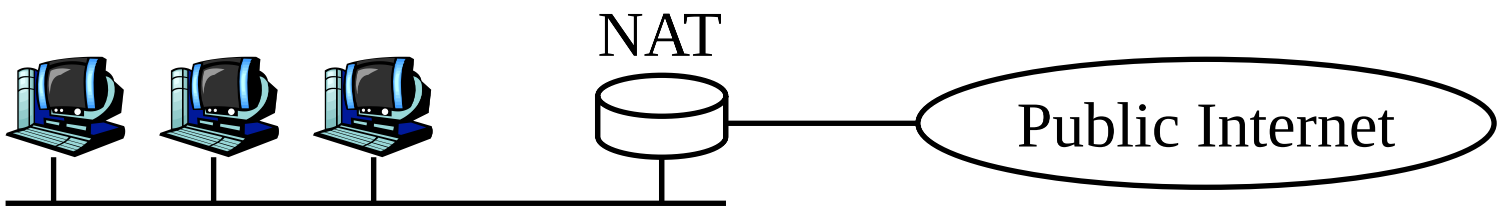
Usual IP/Ethernet address routing.

- Can DNS be used to resolve NAT specific addresses/port combinations?

Global DNS provides only the enterprise's ex.

Universal Plug and Play

- NAT needs to be manually programmed to forward external requests
- UPnP allows hosts to request port forwarding
- Both hosts and NAT should be UPnP aware
- Host requests forwarding all port xx messages to it
- NAT returns the public address and the port #.
- The host can then announce the address and port # outside
- Outside hosts can then reach the internal host (server)



Student Questions

- Could you explain what it means to UPnP aware?

UPnP = Universal Plug and Play
UPnP is relatively new, so many hosts and routers (NAT) may still need to implement it. If they have implemented it, they are UPnP aware.

- Does NAT provide additional security to hosts in the private network since it prevents incoming connection if the network admin did not manually modify its setting?

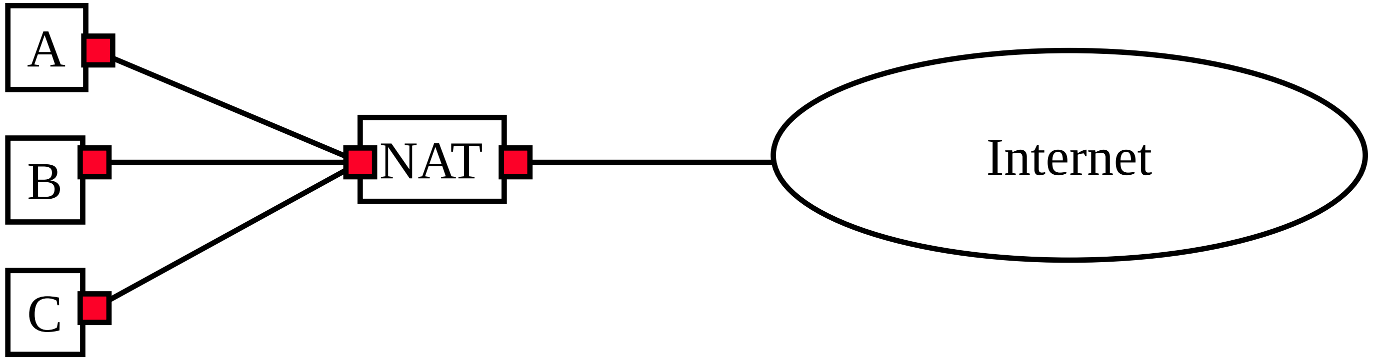
A manual setting is required only for servers that serve external hosts.

- In Universal Plug and Play, why can't the host announce the address and port number from inside?

Different hosts may conflict, so NAT rejects some requests.

Homework 4C: NAT

- [20 points] Consider a home network of 3 computers connected to the Internet via a NAT router. Suppose the ISP assigns the router the address 24.34.112.234 and that the network address of the home network is 192.168.1.0/29.
- A. Assign addresses to all interfaces in the home network, starting with the lowest possible address.
- B. What is the subnet mask for the home computers?
- C. Suppose each host has two ongoing TCP connections, all to port 80 at host 128.119.40.86. Provide the six corresponding entries in the NAT translation table. Both NAT and computers use source ports starting at 4000.



Student Questions

- Can you go through Homework 4C?

Sure.

- Can you briefly explain the solution for the NAT translation table?

Sure.

- Can you go over the correct answers for this HW?

Sure.

DHCP

- **D**ynamic **H**ost **C**ontrol **P**rotocol
- Allows hosts to get an IP address automatically from a server
- Do not need to program each host manually
- Each allocation has a limited "lease" time
- Can reuse a limited number of addresses
- Hosts broadcast "Is there a DHCP Server Here?"
  Sent to 255.255.255.255
- DHCP servers respond
- RFC 2132 defines DHCP options: DHCP Message type option is used to convey the type of the DHCP message. The code for this option is 53, and its length is 1. Legal values for this option are:

Value	Message Type	Value	Message Type
1	DHCP DISCOVER	5	DHCP ACK
2	DHCP OFFER	6	DHCP NAK
3	DHCP REQUEST	7	DHCP RELEASE
4	DHCP DECLINE	8	DHCP INFORM

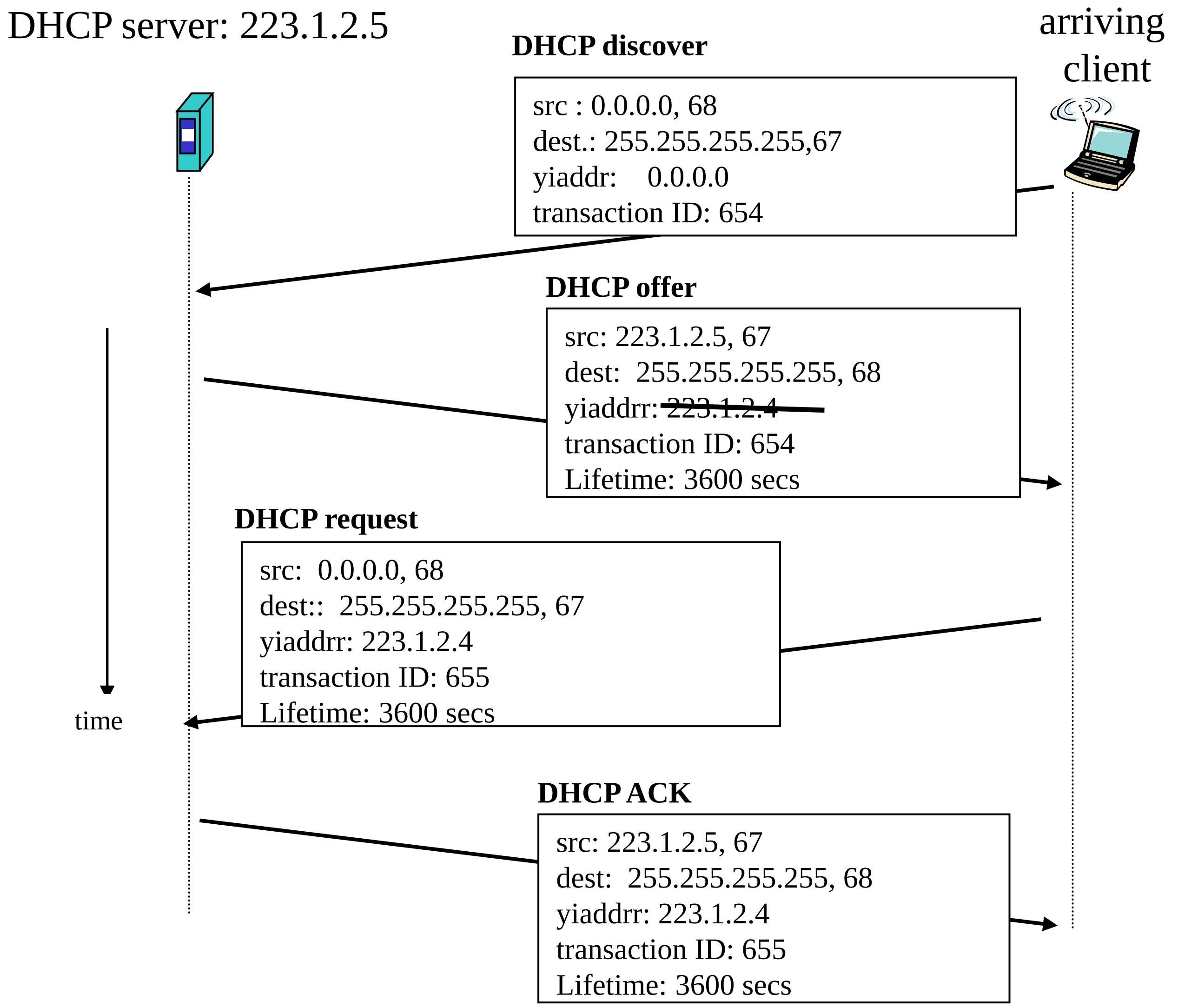
Ref: https://datatracker.ietf.org/doc/html/rfc2132

Student Questions

- Is DHCP only for private networks, or is it also used to allocate public IP addresses?

It can be used in any network but is generally used only when the broadcast is easy, such as in Wi-Fi or Ethernet.

DHCP Example



Student Questions

- Why do DHCP requests and DHCP ACK also use broadcast?

When requesting an IP address allocation, the requester does not have an IP address and does not know who can allocate it. So, it broadcasts it to everyone in the subnet. The DHCP server responds, but the destination does not know its IP address, so the response is also broadcast. The requester looks for such a broadcast, and if it finds its MAC address in the response, it knows that the allocation is for it.

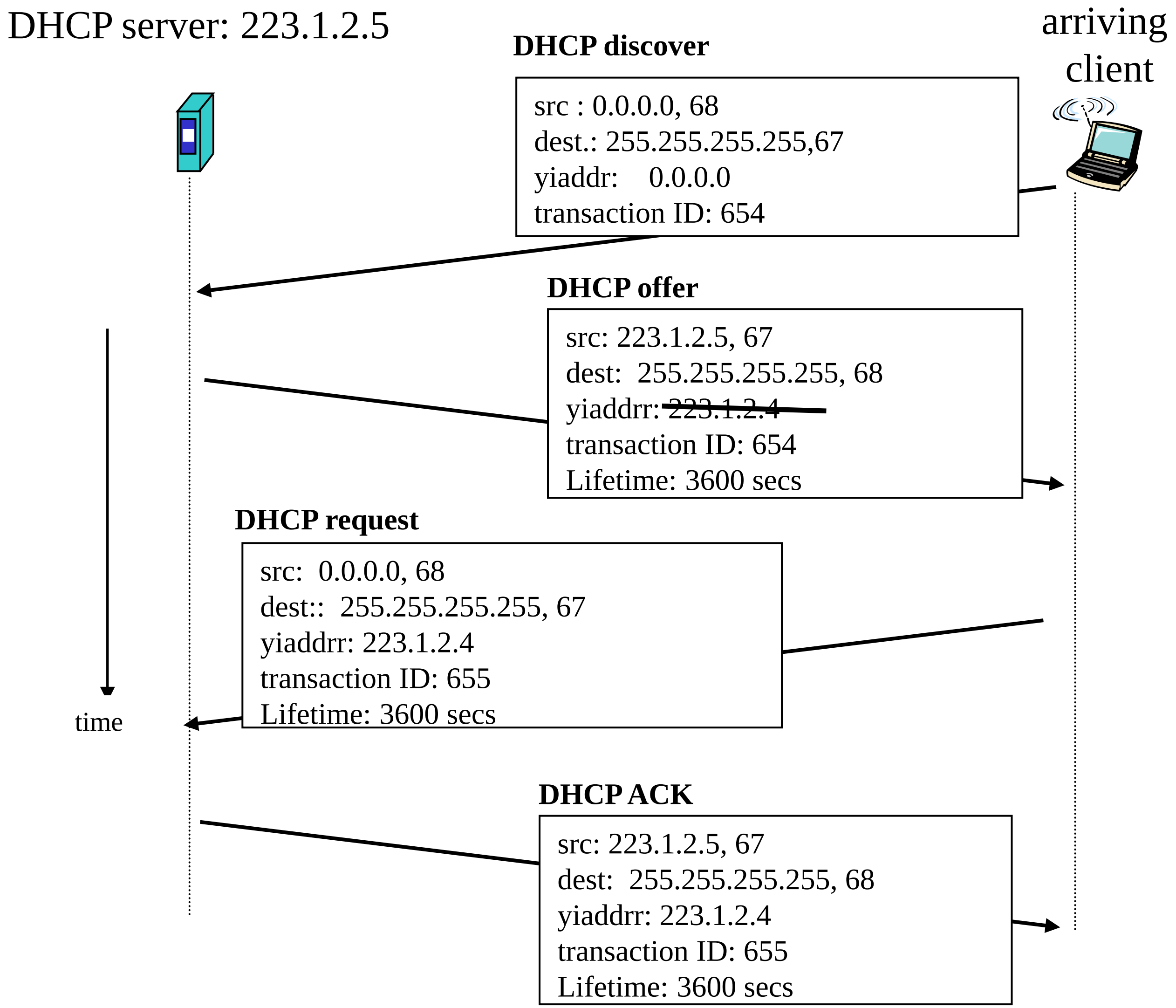
- Why are the destination addresses for DHCP requests/ack 255.255.255.255?

That's the IP broadcast address.

- Does the server only need to get IP address through DHCP for the first time or they don't need to through DHCP because their IP address is fixed?

If you mean DHCP clients, the client gets the IP address for a limited time. They need to renew it after that.

DHCP Example



Student Questions

- ❑ How are the src and dest IP addresses related?

This is a request-response protocol. The response needs to be sent to the requester.

Lab 4B: DHCP

- [15 points] Download the Wireshark traces from http://gaia.cs.umass.edu/wireshark-labs/wireshark-traces.zip
- Open *dhcp-ethereal-trace-1* in Wireshark.
  Select **View → Expand All**. Answer the following questions:
  1. Examine Frame 2 marked DHCP.
     A. What transport protocol and destination port # are used by DHCP?
     B. What are the source and destination IP addresses for this frame, and why?
     C. What is the **Code-Length-Type** for the DHCP Discover option?
  2. Examine Frames 4, 5, and 6 to find Code-Length-Type for:
     A. DHCP Offer
     B. DHCP Request
     C. DHCP Ack

Student Questions

Lab 4B: DHCP (Cont)

3. Examine Frame 4:
   A. What was the IP address assigned by the DHCP server?
   B. What IP address is this frame addressed to, and why?
   C. What was other information provided by the DHCP server?
      1. Subnet Mask:
      2. Default Gateway:
      3. DNS1:
      4. DNS2:
      5. Domain Name:
      6. Lease Time:
4. Examine Frame 5 and find what preferred IP address was requested by the client?

Student Questions

End of Part 2

IPv6

- Shortage of IPv4 addresses ⇒ Need larger addresses
- IPv6 was designed with 128-bit addresses
- $2^{128} = 3.4\times10^{38}$ addresses
  ⇒ 665×10^{21} addresses per sq. m of earth's surface
- If assigned at the rate of $10^6/\mu s$, it would take 20 years
- **Dot-Decimal**: 127.23.45.88
- **Colon-Hex:** FEDC:0000:0000:0000:3243:0000:0000:ABCD
  - Can skip leading zeros of each word
  - Can skip <u>one</u> sequence of zero words, e.g.,
    FEDC::3243:0000:0000:ABCD
    ::3243:0000:0000:ABCD
  - Can leave the last 32 bits in dot-decimal, e.g., ::127.23.45.88
  - Can specify a prefix by /length, e.g., 2345:BA23:0007::/50

Student Questions

- Could you reexplain "::" and skipping a set of zeros

There should be eight words in the address. If there is a "::" anywhere in the address. You put zeros there to bring the total number of words to 8.

- Is IPv6 a supplement to IPv4 or a replacement for IPv4? Can a website have IPv4 and IPv6 addresses at the same time?

It is a replacement. But most devices have both.

- Why have both, and is it possible to have an IPV6 address without a corresponding IPV4 version?

You can use either. To talk to the other, you must go through a tunnel (TBD).

- Why is it called IPv6 instead of IPv5?

IPv5 was there but never used.

- When do we know which 0 we could ignore and get rid of

Only one contiguous sequence of zeros.

- Is it possible to have a protocol with variable-length addresses? What are the benefits and drawbacks?

Possible. But the IP is all done now.

IPv6

- Shortage of IPv4 addresses ⇒ Need larger addresses
- IPv6 was designed with 128-bit addresses
- $2^{128} = 3.4\times10^{38}$ addresses
  ⇒ 665×10^{21} addresses per sq. m of earth's surface
- If assigned at the rate of $10^6/\mu s$, it would take 20 years
- **Dot-Decimal**: 127.23.45.88
- **Colon-Hex:** FEDC:0000:0000:0000:3243:0000:0000:ABCD
  - Can skip leading zeros of each word
  - Can skip one sequence of zero words, e.g.,
    FEDC::3243:0000:0000:ABCD
    ::3243:0000:0000:ABCD
  - Can leave the last 32 bits in dot-decimal, e.g., ::127.23.45.88
  - Can specify a prefix by /length, e.g., 2345:BA23:0007::/50

Student Questions

- Theoretically, could we use up all the IP addresses offered by IPv6, and we would have to create a new system?

Not in the near future.

- So, are most IP addresses we see IPv6 nowadays with only the last 32 bits? Or are we still mainly using IPv4?

We mainly use private IPv4 addresses.

- Must we skip the first sequence of zero words in colon-hex?

No. You can skip any one sequence of zero words. It does not have to be the first.

IPv6

- Shortage of IPv4 addresses ⇒ Need larger addresses
- IPv6 was designed with 128-bit addresses
- $2^{128} = 3.4 \times 10^{38}$ addresses
  ⇒ 665×10^{21} addresses per sq. m of earth's surface
- If assigned at the rate of $10^6/\mu s$, it would take 20 years
- **Dot-Decimal**: 127.23.45.88
- **Colon-Hex:** FEDC:0000:0000:0000:3243:0000:0000:ABCD
  - Can skip leading zeros of each word
  - Can skip one sequence of zero words, e.g.,
    FEDC::3243:0000:0000:ABCD
    ::3243:0000:0000:ABCD
  - Can leave the last 32 bits in dot-decimal, e.g., ::127.23.45.88
  - Can specify a prefix by /length, e.g., 2345:BA23:0007::/50

Student Questions

- Why can't you have more than one sequence of zeros represented by "::"

If two sequences are used, you will not know how many zeros in each sequence.

- Does DHCP nowadays assign IPv6 instead of IPv4?

IPv6 does not need DHCP. It has auto-local address generation.

- What exactly happened to IPv5?

Internet Stream Protocol (RFC 1819) was an experimental L3 protocol. It uses 5 in the version field.
https://en.wikipedia.org/wiki/Internet_Stream_Protocol

IPv6

- Shortage of IPv4 addresses ⇒ Need larger addresses
- IPv6 was designed with 128-bit addresses
- $2^{128} = 3.4\times10^{38}$ addresses
  ⇒ 665×10^{21} addresses per sq. m of earth's surface
- If assigned at the rate of $10^6/\mu s$, it would take 20 years
- **Dot-Decimal**: 127.23.45.88
- **Colon-Hex:** FEDC:0000:0000:0000:3243:0000:0000:ABCD
  - Can skip leading zeros of each word
  - Can skip one sequence of zero words, e.g.,
    FEDC::3243:0000:0000:ABCD
    ::3243:0000:0000:ABCD
  - Can leave the last 32 bits in dot-decimal, e.g., ::127.23.45.88
  - Can specify a prefix by /length, e.g., 2345:BA23:0007::/50

Student Questions

- Is IPv6 generally unnecessary because problems have been resolved by using NAT and private addresses?

Yes. If you have enough addresses for one per enterprise.

- Is there an example of backward compatible hypothetical IPv6 which can accept IPv4 data?

No.

IPv6 Header

- IPv6:

Version (4b)	Traffic Class (8b)	Flow Label (20b)	
Payload Length (16b)		Next Header (8b)	Hop Limit (8b)
Source Address (128b)			
Destination Address (128b)			

- IPv4:

Version	IHL	Type of Service	Total Length	
Identification			Flags	Fragment Offset
Time to Live		Protocol	Header Checksum	
Source Address				
Destination Address				
Options				Padding

Student Questions

- What is the benefit of dropping an IPv6 packet if the packet size is larger than the link layer MTU vs. fragmentation as in IPv6? *IPv6 nodes do not send segments more significant than the specified MTU, which is large enough for most applications. This was to keep routers simple (KISS).*
- In the textbook, on Page 349, it is said that Traffic Class in IPv6 has the same function as Type of Service in IPv4, but in these slides, you mentioned that ToS is removed and Priority is added. Why? *Both the book and I are right. The IPv4 TOS field was poorly defined, so it has many meanings different from Traffic Class. Some will argue that ToS and Traffic Class are different. You are assuming they are identical. The Traffic Class bits are well-defined and include priority and more.*
- Can the following header be the TCP header of another IPv6 header? *Yes.*
- Does the 'time to live' in the IPv4 header only change the name to 'Hop Limit' in IPv6? *Yes*

IPv6 Header

- IPv6:

Version (4b)	Traffic Class (8b)	Flow Label (20b)

Payload Length (16b)	Next Header (8b)	Hop Limit (8b)
Source Address (128b)		
Destination Address (128b)		

- IPv4:

Version	IHL	Type of Service	Total Length

Identification	Flags	Fragment Offset
Time to Live	Protocol	Header Checksum

Source Address	
Destination Address	
Options	Padding

Student Questions

- So, in IPv6, each flow id binds one port number, right?

Not necessary.

- Does IPv6 not have a checksum?

TCP checksum may protect key parts.

- Since IPv6 does not have a checksum, pass it down to TCP to do the checksum instead.

Yes.

- Why is IPv4 still commonly used if it is obsolete? Why aren't issues more common with running out of IP addresses?

We now have private addresses, which solve the problem mostly.

- Is the Hop limit the same as TTL?

Yes. It's just a more correct name.

IPv6 vs. IPv4

- 1995 vs. 1975
- IPv6 is only twice the size of the IPv4 header
- Only the version number has the same position and meaning as in IPv4
- Removed: header length, type of service, identification, flags, fragment offset, header checksum ⇒ No fragmentation
- Datagram length replaced by payload length
- Protocol type replaced by the following header
- Time to live replaced by hop limit
- Added: Priority and flow label
- All fixed-size fields.
- No optional fields. Replaced by extension headers.
- 8-bit hop limit = 255 hops max (Limits looping)
- Next Header = 6 (TCP), 17 (UDP)

Student Questions

- Wouldn't remove the fragmentation in IPv6 cause significant issues with congestion on the network?

Fragmentation is required because some routers have small memory. Congestion is caused if the link or router processing capacity is lower than the load. Compute and storage are different issues.

- Can you elaborate on the following header? What is this exactly? *Sure.*
- Is there a disadvantage to trimming the IPv6 header that allows it to be only twice as long as the IPv4 header?

The main header is twice as long. But extension headers are like options.

- Are there new versions of IP being made after IPv6?

No. But many extensions.

- What do you mean? Is IPv6 no fragmentation? Is the MTU infinite?

MTU size is large and fixed.

- Can we still use IPv4 now?

Yes.

IPv6 vs. IPv4

- 1995 vs. 1975
- IPv6 is only twice the size of the IPv4 header
- Only the version number has the same position and meaning as in IPv4
- Removed: header length, type of service, identification, flags, fragment offset, header checksum $\Rightarrow$ No fragmentation
- Datagram length replaced by payload length
- Protocol type replaced by the following header
- Time to live replaced by hop limit
- Added: Priority and flow label
- All fixed-size fields.
- No optional fields. Replaced by extension headers.
- 8-bit hop limit = 255 hops max (Limits looping)
- Next Header = 6 (TCP), 17 (UDP)

Student Questions

- Can you explain why it is called "next header?" I'm not sure what "next" refers to.

Payload type

- How does IPv6 handle address auto-configuration, and what are some best practices for ensuring secure and efficient auto-configuration in large-scale networks

There are ways

- What's the advantage of stacking multiple IPv6 headers?

Reduces header overhead if they are not needed.

- Where does an extension header reside?

Extension headers are stacked like layers – each as a payload inside the other.

IPv6 vs. IPv4

- 1995 vs. 1975
- IPv6 is only twice the size of the IPv4 header
- Only the version number has the same position and meaning as in IPv4
- Removed: header length, type of service, identification, flags, fragment offset, header checksum $\Rightarrow$ No fragmentation
- Datagram length replaced by payload length
- Protocol type replaced by the following header
- Time to live replaced by hop limit
- Added: Priority and flow label
- All fixed-size fields.
- No optional fields. Replaced by extension headers.
- 8-bit hop limit = 255 hops max (Limits looping)
- Next Header = 6 (TCP), 17 (UDP)

Student Questions

- Does IPv6 have subnets?

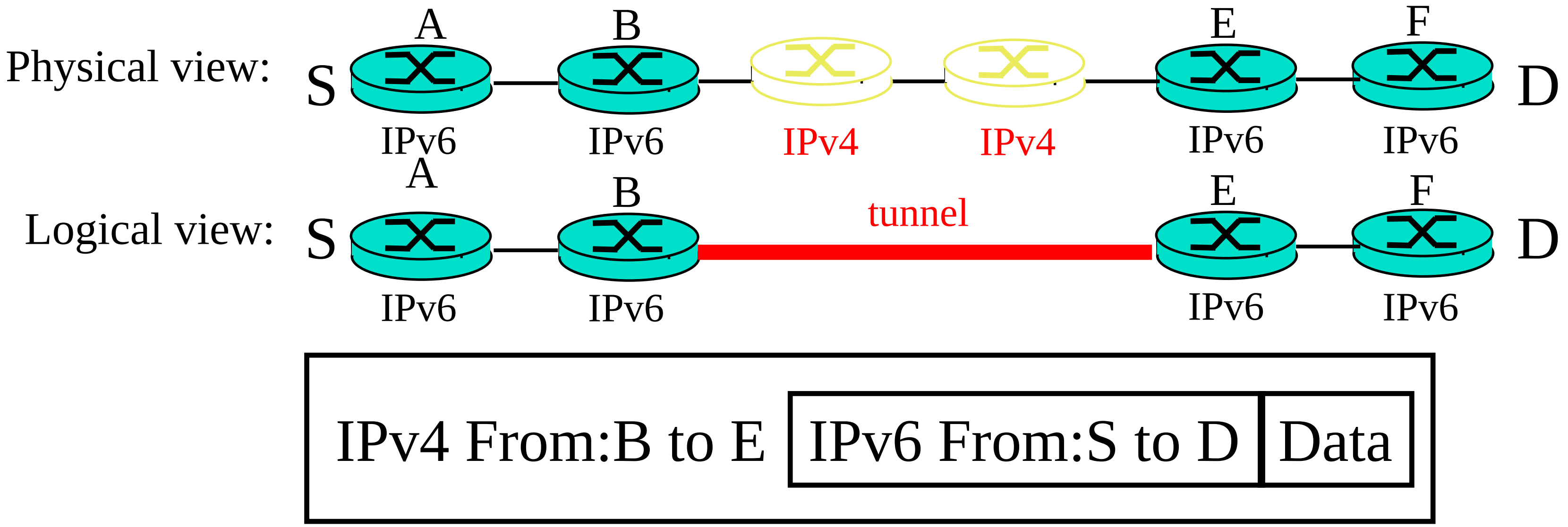
Yes. Known as Links.

- Why we need hops limit? Does it affect the performance?

An error in some nodes routing table could result in a packet circulating in the network for ever. Hop limit avoids it.

IPv4 to IPv6 Transition

- **Dual Stack**: Each IPv6 router also implements IPv4
  IPv6 is used only if the source host, destination host, and all routers on the path are IPv6 aware.
- **Tunneling**: The last IPv6 router puts the entire IPv6 datagram in a new IPv4 datagram addressed to the next IPv6 router
  = **Encapsulation**



Student Questions

- Are all new routers required to be dual-stack?

It is not required. But, yes, more and more routers are now both IPv4 and IPv6 capable.

- Will newer routers ever stop supporting older protocols like IPv4?

No. Older routers will still exist, so IPv4 needs to be supported.

- Can you further explain the tunneling? Do we attach IPv4 parts when arriving at IPv4 routers and remove IPv4 parts when arriving at IPv6 routers?

Yes.

- If we could use tunnel why we could not just use the ipv6 in ipv4 router.

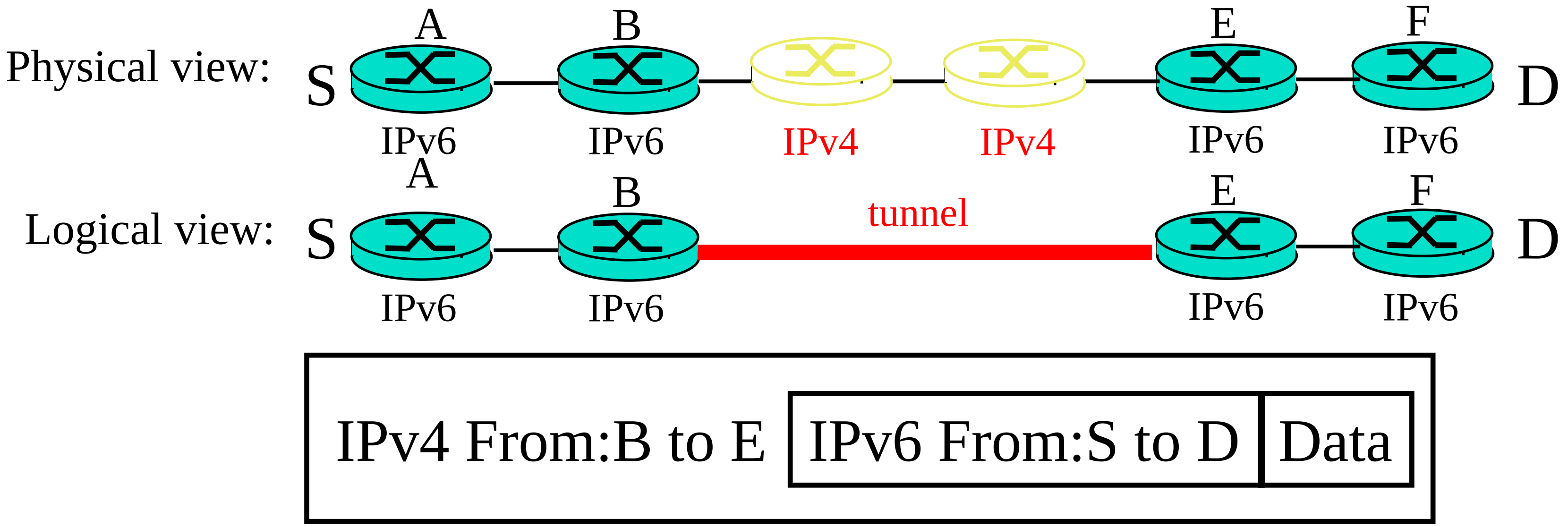
Only bi-lingual routers can tunnel.

- Why don't we do the transition on the application level? For example, we can design software to translate between IPv6 and IPv4.

We could do that if the design were strictly layered. But it is not. All applications need to know the IP addresses and their size.

IPv4 to IPv6 Transition

- **Dual Stack**: Each IPv6 router also implements IPv4
  IPv6 is used only if the source host, destination host, and all routers on the path are IPv6 aware.
- **Tunneling**: The last IPv6 router puts the entire IPv6 datagram in a new IPv4 datagram addressed to the next IPv6 router
  = **Encapsulation**



Student Questions

- What/are there any organizations that mainly use IPv6?

Most countries in the world. Only a few are in the USA.

- Why has IPv6 not wholly taken over even though IP addresses ran out long ago?

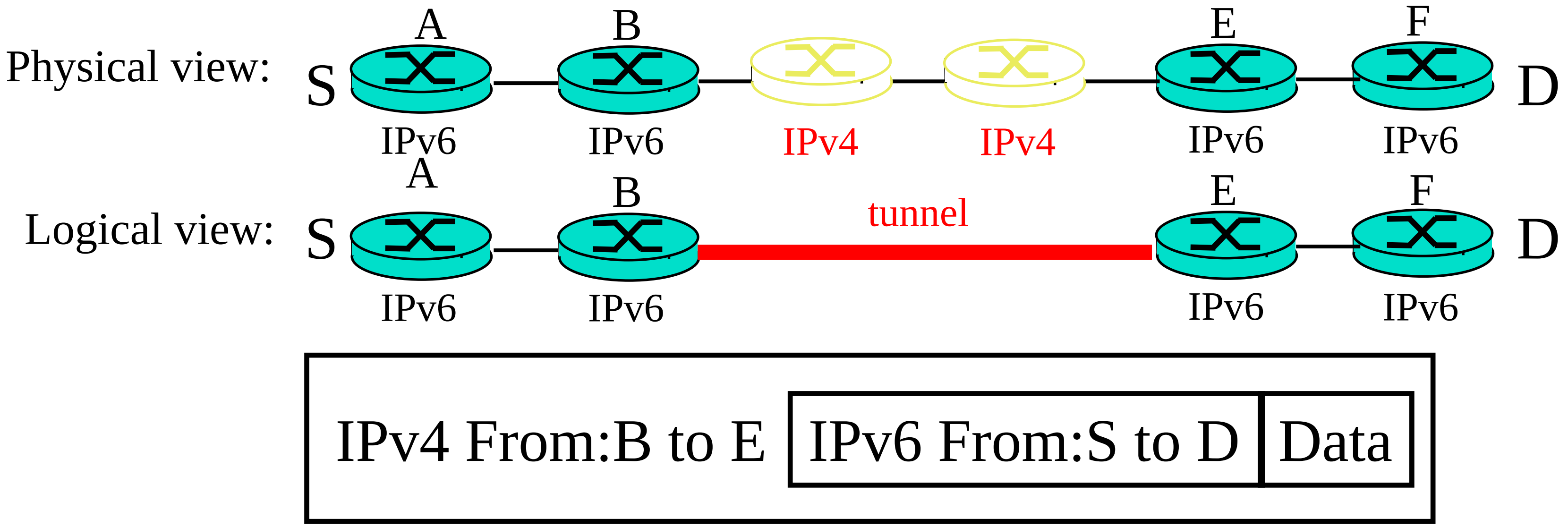
Private Addresses with NAT allow address reuse.

- Could an IPv6 packet be "downgraded" or lost because there are no IPv6 routers along the path to the destination?

Yes. It would be best if you changed IPv6 to IPv4 and vice-versa.

IPv4 to IPv6 Transition

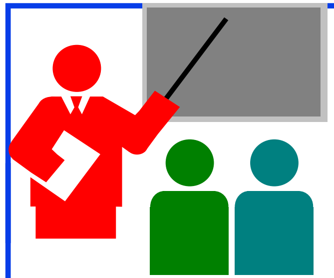
- **Dual Stack**: Each IPv6 router also implements IPv4
  IPv6 is used only if the source host, destination host, and all routers on the path are IPv6 aware.
- **Tunneling**: The last IPv6 router puts the entire IPv6 datagram in a new IPv4 datagram addressed to the next IPv6 router
  = **Encapsulation**



Student Questions

- What are the primary technical and administrative challenges organizations face transitioning from IPv4 to IPv6, and how are these being addressed within the industry?

It requires ALL new hardware. A more satisfactory transition strategy needs to be developed by the designers.



Forwarding Protocols: Review

1. IPv4 uses 32-bit addresses consisting of **subnet + host**
2. **Private addresses** can be reused
   ⇒ Helped solve the address shortage to a great extent
3. **DHCP** is used to allocate addresses to hosts automatically
4. IPv6 uses **128-bit addresses**. Requires dual-stack or **tunneling** to coexist with IPv4.

Student Questions

- will we be tested on both IPv6 and IPv4, or will questions be mainly in reference to IPv4?

Both.

- This video review's lecture and slide is mismatched.

Thanks. I checked and could not find the mismatch. Please indicate the time interval to help find the mismatch.

- Is there a cheat sheet?

For what?

Ref: Read Section 4.3 of the textbook. Try R17 through R29.

Generalized Forwarding and SDN

- Planes of Networking
- Data vs. Control Logic
- OpenFlow Protocol

Student Questions

Planes of Networking

- **Data Plane**: All activities involving as well as resulting from data packets sent by the end user, e.g.,
  - Forwarding
  - Fragmentation and reassembly
  - Replication for multicasting
- **Control Plane**: All activities that are necessary to perform data plane activities but do not involve end-user data packets
  - Making routing tables
  - Setting packet handling policies (e.g., security)
  - Base station beacons announcing the availability of services

Ref: Open Data Center Alliance Usage Model: Software Defined Networking Rev 1.0,"
http://www.opendatacenteralliance.org/docs/Software_Defined_Networking_Master_Usage_Model_Rev1.0.pdf

Student Questions

- Which layers do the various planes of networking relate to?

All layers

Planes of Networking (Cont)

- **Management Plane**: All activities related to provisioning and monitoring of the networks
  - Fault, Configuration, Accounting, Performance, and Security (**FCAPS**).
  - Instantiate new devices and protocols (Turn devices on/off)
  - Optional ⇒ May be handled manually for small networks.
- **Services Plane**: Middlebox services to improve performance or security, e.g.,
  - Load Balancers, Proxy Service, Intrusion Detection, Firewalls, SSL Off-loaders
  - Optional ⇒ Not required for small networks.

Student Questions

- What is Intrusion Detection concerning middleboxes?

Detecting security attacks

- What are the key differences between a network "layer" vs. "plane"? Are these planes that you mention only in the network layer?

There is no relationship between layers.

- Are middleboxes either management or service plane?

Service Plane.

- Which plane do NATs belong to?

Data plane

- Would the services plane be responsible for implementing QoS rules?

QoS is done in the control plane. Service boxes usually provide a service rather than enforcement. However, you could have a "policing" service that could enforce some rules.

Planes of Networking (Cont)

- **Management Plane**: All activities related to provisioning and monitoring of the networks
  - Fault, Configuration, Accounting, Performance, and Security (**FCAPS**).
  - Instantiate new devices and protocols (Turn devices on/off)
  - Optional ⇒ May be handled manually for small networks.
- **Services Plane**: Middlebox services to improve performance or security, e.g.,
  - Load Balancers, Proxy Service, Intrusion Detection, Firewalls, SSL Off-loaders
  - Optional ⇒ Not required for small networks.

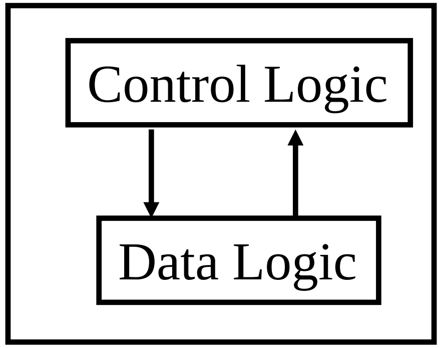
Student Questions

- Can you review the main difference between the management and services planes? It seems like they do similar things.

Management plane does FCAPS. Service plane improves performance. Both are optional.

Data vs. Control Logic

- The Data plane runs at line rate,
  e.g., 100 Gbps for 100 Gbps Ethernet ⇒ Fast Path
  ⇒ Typically implemented using special hardware,
  e.g., Ternary Content Addressable Memories (TCAMs)
- Some exceptional data plane activities are handled by the CPU in the switch ⇒ Slow path
  e.g., Broadcast, Unknown, and Multicast (BUM) traffic
- All control activities are generally handled by the CPU



Student Questions

OpenFlow: Key Ideas

1. Separation of control and data planes
2. Centralization of control
3. Flow-based control

Ref: N. McKeown, et al., ``OpenFlow: Enabling Innovation in Campus Networks," ACM SIGCOMM CCR, Vol. 38, No. 2, April 2008, pp. 69-74.

Student Questions

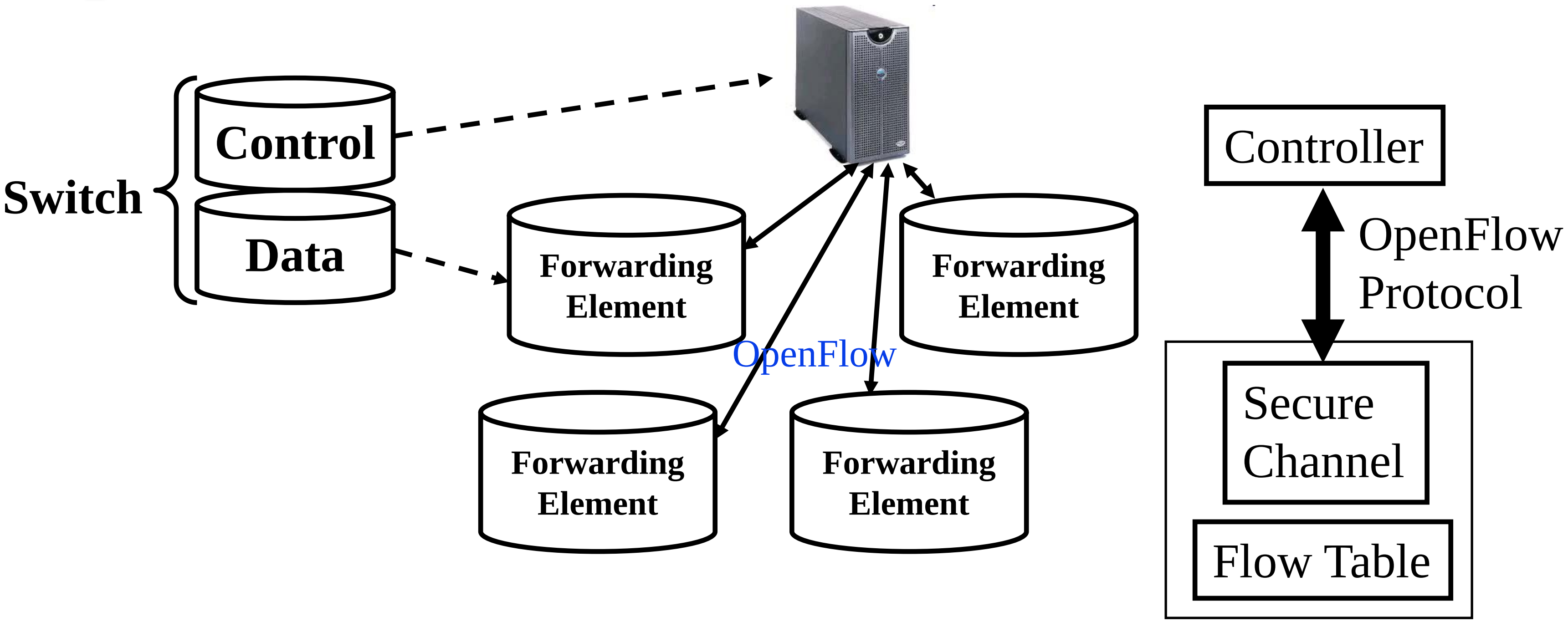
- Who were the significant entities behind OpenFlow?

OpenFlow originated from the Ph.D. thesis of Martin Casado under Prof. Nick McKeown at Stanford University.

- Why do we need OpenFlow?

It makes networks programmable.

Separation of Control and Data Plane



- Control logic is moved to a controller
- Switches only have forwarding elements
- One expensive controller with a lot of cheap switches
- OpenFlow is the protocol to send/receive forwarding rules from the controller to switches

Student Questions

- Where does the controller exist? In the server?

Yes. There is a controller for each network.

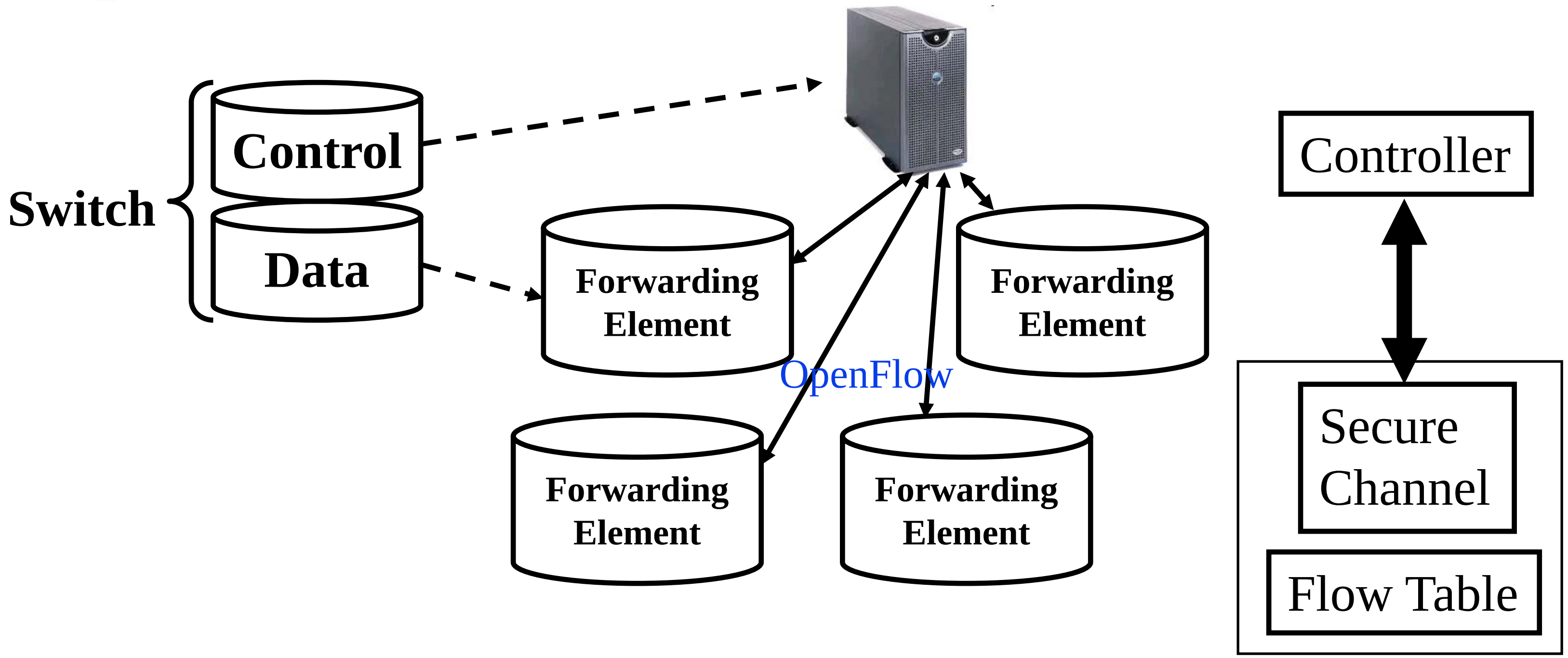
- Is a controller software running on a physical machine?

Yes.

- How does the longest prefix match principle in routing tables influence the efficiency and complexity of routing decisions in large-scale networks, especially in the context of IPv6's larger address space?

It reduces the size of the table significantly.

Separation of Control and Data Plane



- Control logic is moved to a controller
- Switches only have forwarding elements
- One expensive controller with a lot of cheap switches
- OpenFlow is the protocol to send/receive forwarding rules from the controller to switches

Student Questions

- If a controller is software, what makes it "expensive"?

It is software on fast hardware.

OpenFlow V1.0

- On packet arrival, match the header fields with flow entries in a table, if any entry matches, perform indicated actions, and update the counters indicated in that entry.

Flow Table:

Header Fields	Actions	Counters
Header Fields	Actions	Counters
…	…	…
Header Fields	Actions	Counters

Ingress Port	Ether Source	Ether Dest	VLAN ID	VLAN Priority	IP Src	IP Dst	IP Proto	IP ToS	Src L4 Port	Dst L4 Port

Ref: http://archive.openflow.org/documents/openflow-spec-v1.0.0.pdf

Student Questions

- Are most routers using OpenFlow protocol to control the traffic these days?

No.

- Was OpenFlow v1 replaced by OpenFlow v2 or an entirely different protocol?

V2 is an extension of V1.

- What's the purpose of VLAN?

We discuss this in the next chapter on LAN.

Flow Table Example

Port	Src MAC	Dst MAC	VLAN ID	Priority	EtherType	Src IP	Dst IP	IP Proto	IP ToS	Src L4 Port ICMP Type	Dst L4 Port ICMP Code	Action	Counter
*	*	0A:C8:*	*	*	*	*	*	*	*	*	*	Port 1	102
*	*	*	*	*	*	*	192.168.*.*	*	*	*	*	Port 2	202
*	*	*	*	*	*	*	*	*	*	21	21	Drop	420
*	*	*	*	*	*	*	*	0x806	*	*	*	Local	444
*	*	*	*	*	*	*	*	0x1*	*	*	*	Controller	1

- Idle timeout: Remove entry if no packets received for this time
- Hard timeout: Remove entry after this time
- If both are set, the entry is removed if either one expires.

Ref: S. Azodolmolky, "Software Defined Networking with OpenFlow," Packt Publishing, October 2013, 152 pp., ISBN:978-1-84969-872-6 (Safari Book)

Student Questions

- Do the table entries use glob-style expressions? *No. Glob is for ASCII strings. Most of these are binary strings. So, marking and matching are standard.*
- Are these counter fields denoted by the counter value (like an ID), or is the counter value the actual value being passed back of these instances? *Counters are actual counts of those rows being matched, and those actions are taken.*
- What is the purpose of the counters in OpenFlow? *Counters are used to count how many frames match that rule. For example, packets dropped could be counted to find problems in the network.*
- What do "IP ToS" and "EtherType" correspond here?

ToS = Type of Service in IPv4

EtherType=Type field in Ethernet

- What are these stars?

Wild cards.

- How is the flow table different than the routing table? Why do we need both?

Flow tables are prepared using a routing table. Flow table makes programming easy.

Flow Table Example

Port	Src MAC	Dst MAC	VLAN ID	Priority	EtherType	Src IP	Dst IP	IP Proto	IP ToS	Src L4 Port ICMP Type	Dst L4 Port ICMP Code	Action	Counter
*	*	0A:C8:*	*	*	*	*	*	*	*	*	*	Port 1	102
*	*	*	*	*	*	*	192.168.*.*	*	*	*	*	Port 2	202
*	*	*	*	*	*	*	*	*	*	21	21	Drop	420
*	*	*	*	*	*	*	*	0x806	*	*	*	Local	444
*	*	*	*	*	*	*	*	0x1*	*	*	*	Controller	1

- Idle timeout: Remove entry if no packets received for this time
- Hard timeout: Remove entry after this time
- If both are set, the entry is removed if either one expires.

Ref: S. Azodolmolky, "Software Defined Networking with OpenFlow," Packt Publishing, October 2013, 152 pp., ISBN:978-1-84969-872-6 (Safari Book)

Student Questions

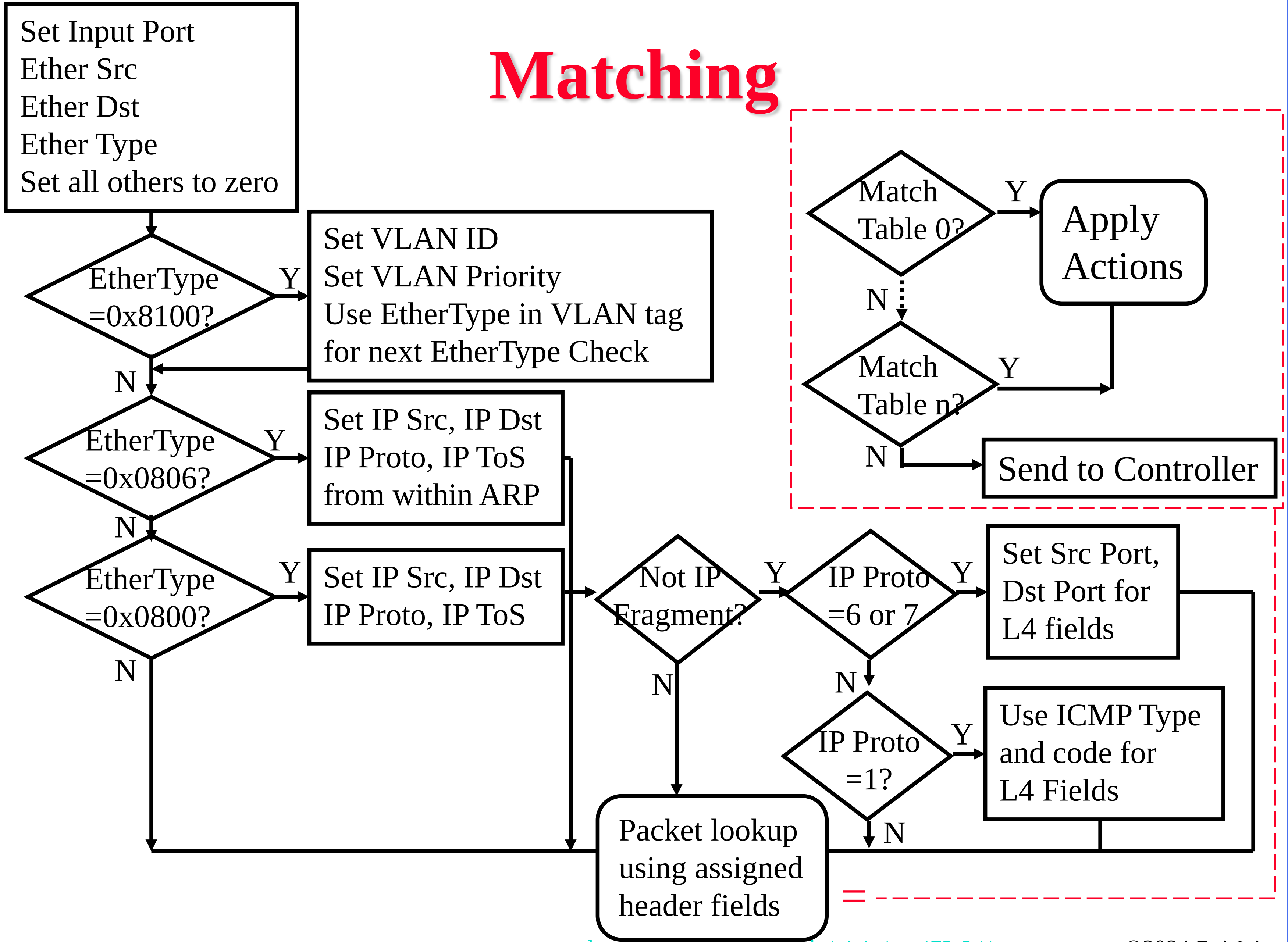
- Are flow and forwarding tables used together? Since each time is a different type of forwarding

Yes. If the flow table tells you to forward it to x, then you use the forwarding table to get it there.

- Please go over the flow table again.

Sure.

Matching



Student Questions

- To clarify, are only the fields necessary for the EtherType command set, and are others left blank?

No. The top box indicates the fields used in the three decision boxes on the left.

- What are actions the controller could take after receiving a packet?

Drop it, forward it to interface x, or give it to a layer protocol.

- Will a packet ever match multiple actions? Can a packet, for example, match a modify field action followed by a forwarding action?

Yes. You can have multiple tables and do multiple actions.

Counters

Per Table	Per Flow	Per Port	Per Queue
Active Entries	Received Packets	Received Packets	Transmit Packets
Packet Lookups	Received Bytes	Transmitted Packets	Transmit Bytes
Packet Matches	Duration (Secs)	Received Bytes	Transmit overrun errors
	Duration (nanosecs)	Transmitted Bytes	
		Receive Drops	
		Transmit Drops	
		Receive Errors	
		Transmit Errors	
		Receive Frame Alignment Errors	
		Receive Overrun erorrs	
		Receive CRC Errors	
		Collisions	

Student Questions

- Are these counters accumulated on the controller in real-time or periodically aggregated from the switches?

These counters are kept on the switches and collected periodically by the management.

Actions

- Forward to Physical/**Virtual Port** *i*
- Enqueue: To a particular **queue** in the port ⇒ QoS
- Drop
- Modify Field: E.g., add/remove VLAN tags, ToS bits, Change TTL.
- Masking allows matching only selected fields, e.g., Dest. IP, Dest. MAC, etc.
- If the header matches an entry, corresponding actions are performed, and counters are updated.
- If no header matches, the packet is queued and The **header is sent to the controller**, which sends a new rule. Subsequent packets of the flow are handled by this rule.
- Secure Channel: Between the controller and the switch using TLS

Student Questions

- Were there ever attacks on OpenFlow networks by generating and sending lots of distinct packets with distinct headers to force queries of the controller?

No. Even if these were to happen, rate control could easily overcome these.

- Would you elaborate on the TLS mechanism?

Transport layer security (TLS) will be discussed in Chapter 8.

- Does forwarding here mean forwarding IP datagrams?

No. Whatever is received from the link. This is generally a datalink frame (Ethernet or Wi-Fi frame).

- Is the virtualization of ports similar to the virtualization of memory?

Virtual = You can see it, but it is not there.

- So, when a packet does not meet any rules in the matching table, we essentially get a new rule for this packet?

Yes.

Actions

- Forward to Physical/**Virtual Port** *i*
- Enqueue: To a particular **queue** in the port ⇒ QoS
- Drop
- Modify Field: E.g., add/remove VLAN tags, ToS bits, Change TTL.
- Masking allows matching only selected fields, e.g., Dest. IP, Dest. MAC, etc.
- If the header matches an entry, corresponding actions are performed, and counters are updated.
- If no header matches, the packet is queued and the **header is sent to the controller**, which sends a new rule. Subsequent packets of the flow are handled by this rule.
- Secure Channel: Between the controller and the switch using TLS

Student Questions

- Does the modified field mean that when a packet comes in that matches a particular row, the values of that row are modified?

No. Some field in the packet is modified.

- Will the unknown packet be held in the switch while waiting for a new rule?

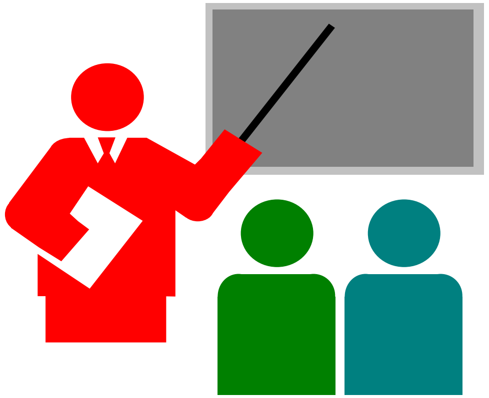
Yes.

Actions (Cont)

- Modern switches already implement flow tables, typically using Ternary Content Addressable Memories (TCAMs)
- A controller can change the forwarding rules if a client moves.
  $\Rightarrow$ Packets for mobile clients are forwarded correctly
- A controller can send flow table entries beforehand (**Proactive**) or Send them on demand (**Reactive**). OpenFlow allows both models.

Student Questions

SDN Data Plane: Summary



1. **The Data plane** consists of packets sent by the users
2. OpenFlow separates the data plane from the **control plane** and centralizes the control plane.
3. The **controller** makes rules for forwarding and sends them to switches
4. Switches match the rules and take specified actions

Student Questions

- Unsure, but the second to last question may have selected the wrong answer.

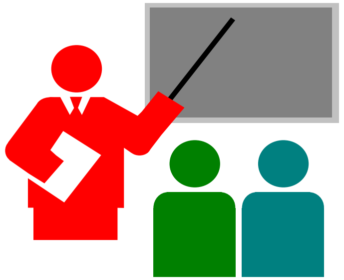
In OpenFlow, forwarding decisions are made by matching flow table entries with packet headers.

- Are the duties of the control plane ever carried out by end systems?

End systems also have a control plane; data plane division applies to all systems.

- Will new grade rankings from the exam be released? *Sure.*

Network Layer Data Plane: Summary



1. **Forwarding** consists of matching the destination address to a list of entries in a table. **Routing** consists of making that table.
2. IP is a forwarding protocol. IPv4 uses 32-bit addresses in **dot-decimal notation**. IPv6 uses 128-bit addresses in **Hex-Colon notation**.
3. **DHCP** is used to assign addresses dynamically.
4. **Private addresses** are used inside an enterprise network. **NAT** allows a single public address to be used by many internal hosts with private addresses.
5. **OpenFlow** separates the data plane from the control plane and centralizes the control plane.

Student Questions

- Can you explain the last quiz again? If no header matches, data will be sent to the controller. Does that mean that packets initially directly engage with OpenFlow without the controller?

No. The controller has yet to consider that possibility.

Acronyms

- ACK Acknowledgement
- ACM Automatic Computing Machinery
- AQM Active Queue Management
- ARP Address Resolution Protocol
- ATM Asynchronous Transfer Mode
- BGP Border Gateway Protocol
- BUM Broadcast, Unknown, and Multicast
- CAMs Content Addressable Memories
- CBR Constant bit rate
- CCR Computer Communications Review
- CIDR Classless Inter-Domain Routing
- CPU Central Processing Unit
- DHCP Dynamic Host Control Protocol
- DNS Domain Name Service
- FCAPS Fault, Configuration, Accounting, Performance and Security
- FCFS First Come First Served

Student Questions

Acronyms (Cont)

- FTP File Transfer Protocol
- GFR Guaranteed Frame Rate
- HTTP Hyper-Text Transfer Protocol
- ICMP IP Control Message Protocol
- ID Identifier
- IP Inter-Network Protocol
- IPv4 IP Version 4
- IPv6 IP Version 6
- ISP Internet Service Provider
- KISS Keep it simple stupid
- LAN Local Area Network
- MAC Media Access Control
- MS Microsoft
- MTU Maximum Transmission Unit
- NAT Network Address Translation
- PBX Private Branch Exchange

Student Questions

Acronyms (Cont)

- PHY   Physical Layer
- QoS   Quality of Service
- RED   Random Early Drop
- RFC   Request for Comment
- RIP   Routing Information Protocol
- RTT   Round Trip Time
- SDN   Software Defined Networking
- SMTP   Simple Mail Transfer Protocol
- SSL   Secure Socket Layer
- TCAM   Ternary Content Addressable Memory
- TCP   Transmission Control Protocol
- TLS   Transport Level Security
- ToS   Type of Service
- TTL   Time to live
- UBR   Unspecified bit rate
- UPnP   Universal Plug and Play

Student Questions

Acronyms (Cont)

❑	VBR	Variable bit rate
❑	VCI	Virtual Circuit Identifiers
❑	VLAN	Virtual Local Area Network
❑	VPN	Virtual Private Network
❑	WAN	Wide Area Network
❑	WiFi	Wireless Fidelity

Student Questions

Scan This to Download These Slides

Raj Jain

http://rajjain.com

http://www.cse.wustl.edu/~jain/cse473-24/i_4nld.htm

Student Questions

- Is the router in my home that the ISP gave me a gateway router that can do BGP broadcasting?

No. It sends all packets to the carrier router.

Related Modules

CSE 567: The Art of Computer Systems Performance Analysis
https://www.youtube.com/playlist?list=PLjGG94etKypJEKjNAa1n_1X0bWWNyZcof

CSE473S: Introduction to Computer Networks (Fall 2011),
https://www.youtube.com/playlist?list=PLjGG94etKypJWOSPMh8Azcgy5e_10TiDw

CSE 570: Recent Advances in Networking (Spring 2013)
https://www.youtube.com/playlist?list=PLjGG94etKypLHyBN8mOgwJLHD2FFIMGq5

CSE571S: Network Security (Spring 2011),
https://www.youtube.com/playlist?list=PLjGG94etKypKvzfVtutHcPFJXumyyg93u

Video Podcasts of Prof. Raj Jain's Lectures,
https://www.youtube.com/channel/UCN4-5wzNP9-ruOzQMs-8NUw

Student Questions