Computer Network

Internetworking

Bridges (LAN Switches)

Bridges are link-level nodes (they forward frames from one link to another to implement an extended LAN)

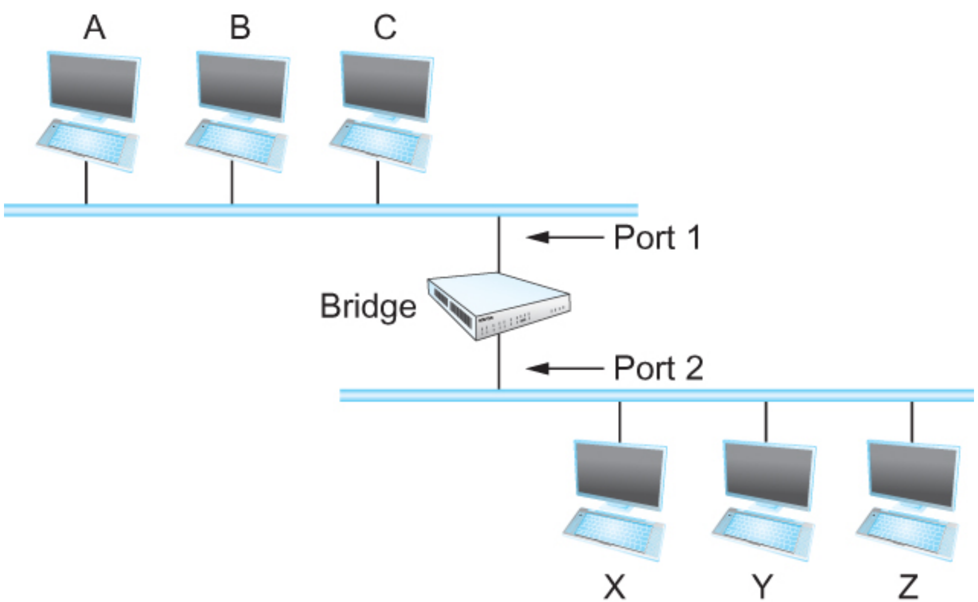
- Bridges (LAN Switches)
  - Class of switches that is used to forward packets between shared-media LANs such as Ethernets (or token ring)
  - Suppose you have a pair of Ethernets that you want to interconnect
    1. One approach is put a repeater in between them
       - It might exceed the physical limitation of the Ethernet
         - » No more than four repeaters between any pair of hosts
         - » No more than a total of 2500 m in length is allowed
    2. An alternative would be to put a bridge between the two Ethernets and have it forward frames from one Ethernet to the other
       - A collection of LANs connected by one or more bridges is usually said to form an Extended LAN

Bridges (LAN Switches)

- Simplest Strategy for Bridges (dumb bridges)
  - Accept LAN frames on their inputs and forward them out to all other outputs regardless of where the destination host resides
  - Used by early bridges
- Learning Bridges (smart bridges)
  - Observe that there is no need to forward all the frames that a bridge receives
  - Consider the example in the next slide…

Bridges (LAN Switches)

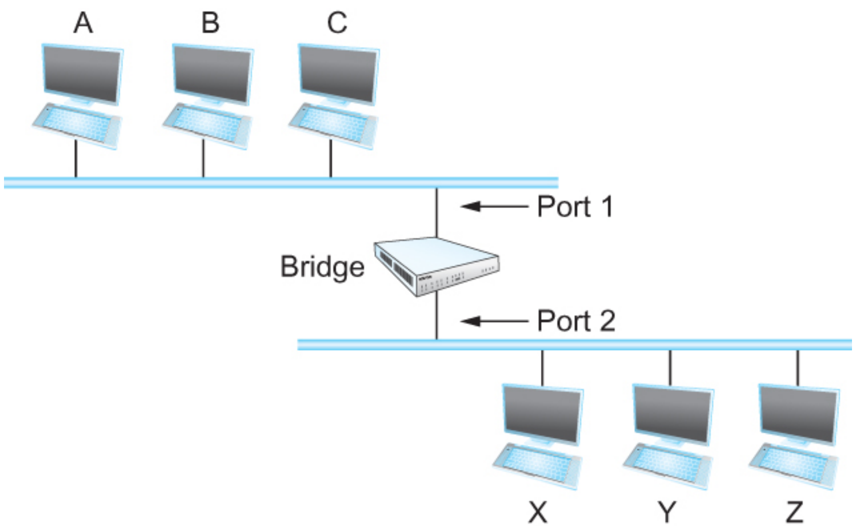
- Consider the following extended Ethernet network
  - When a frame from host A that is addressed to host B arrives on port 1, there is no need for the bridge to forward the frame out over port 2.



  - How does a bridge come to learn on which port the various hosts reside?

Bridges (LAN Switches)

- Solution
  - Download this table into the bridge



Host	Port
A	1
B	1
C	1
X	2
Y	2
Z	2

  - Who does the download?
    - Human
      - Too much work for maintenance

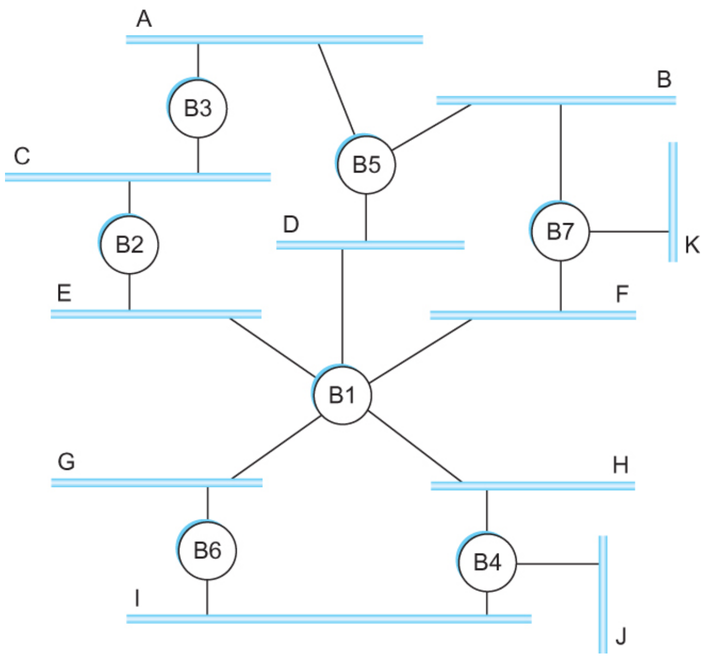
Bridges (LAN Switches)

- Can the bridge learn this information by itself?
  - Yes
- How?
  - Each bridge inspects the source address in all the frames it receives
  - Record the information at the bridge and build the table
  - When a bridge first boots, this table is empty
  - Entries are added over time
  - A timeout is associated with each entry (set a timer at creation time)
    - The bridge discards the entry after a specified period of time
    - To protect against the situation in which a host is moved from one network to another
- If the bridge receives a frame that is addressed to host not currently in the table (say right after first boot, or deleted after timeout)
  - Forward the frame out on all other ports
  - The table only optimizes performance, without it bridge acts as a hub

Bridges (LAN Switches)

- Strategy works fine if the extended LAN does not have a loop in it

- Why?
  - Frames potentially loop through the extended LAN forever



  - Bridges B1, B4, and B6 form a loop

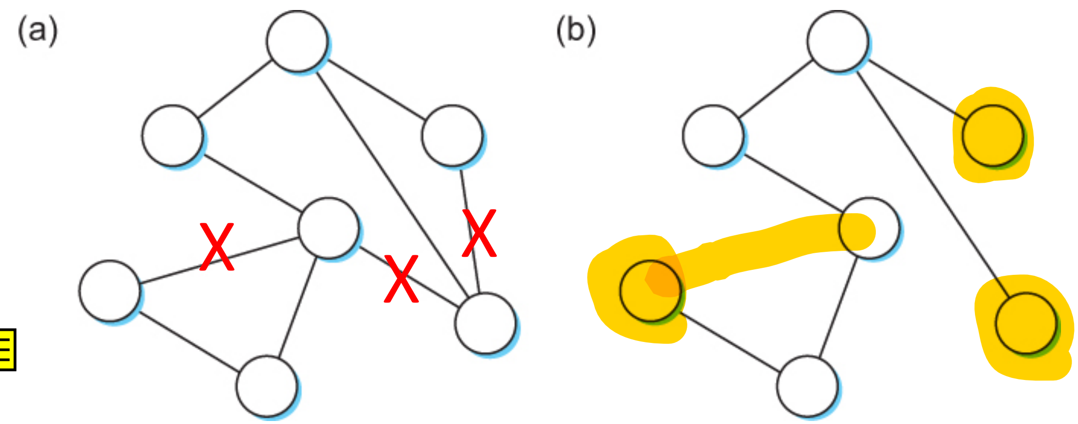
Bridges (LAN Switches)

- Network is managed by more than one administrator
  - For example, it spans multiple departments in an organization
  - It is possible that no single person knows the entire configuration of the network
    - A bridge that closes a loop might be added without anyone knowing
- Loops are built into the network to provide redundancy in case of failures
- Solution
  - Distributed Spanning Tree Algorithm

Spanning Tree Algorithm

- Think of the extended LAN as being represented by a graph that possibly has loops (cycles)

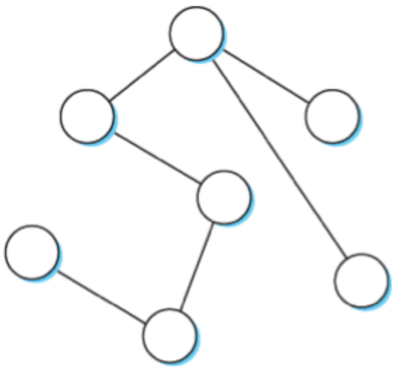
- A spanning tree is a sub-graph of this graph that covers all the vertices but contains no cycles
  - Spanning tree keeps all the vertices of the original graph but throws out some of the edges



Example of (a) a cyclic graph; (b) a corresponding spanning tree.

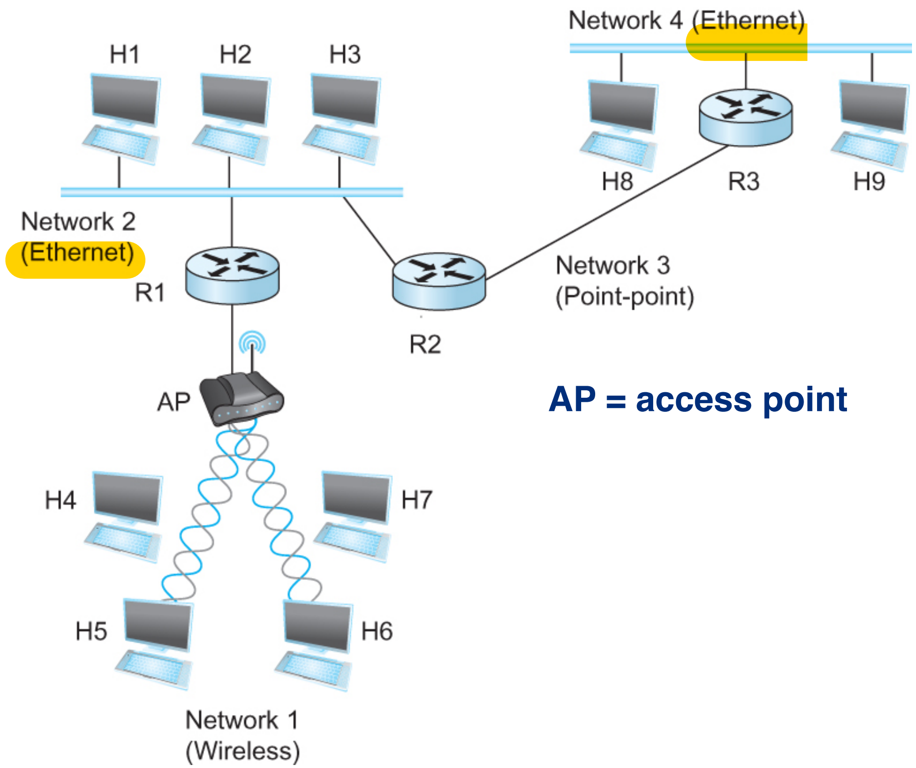
Spanning Tree Algorithm

- Developed by Radia Perlman at Digital
  - A protocol used by a set of bridges to agree upon a spanning tree for a particular extended LAN
  - IEEE 802.1 specification for LAN bridges is based on this algorithm
  - Each bridge decides the ports over which it is and is not willing to forward frames
    - In a sense, it is by removing ports from the topology that the extended LAN is reduced to an acyclic tree
    - It is even possible that an entire bridge will not participate in forwarding frames
- Algorithm is dynamic
  - The bridges are always prepared to reconfigure themselves into a new spanning tree if some bridges fail



Internetworking

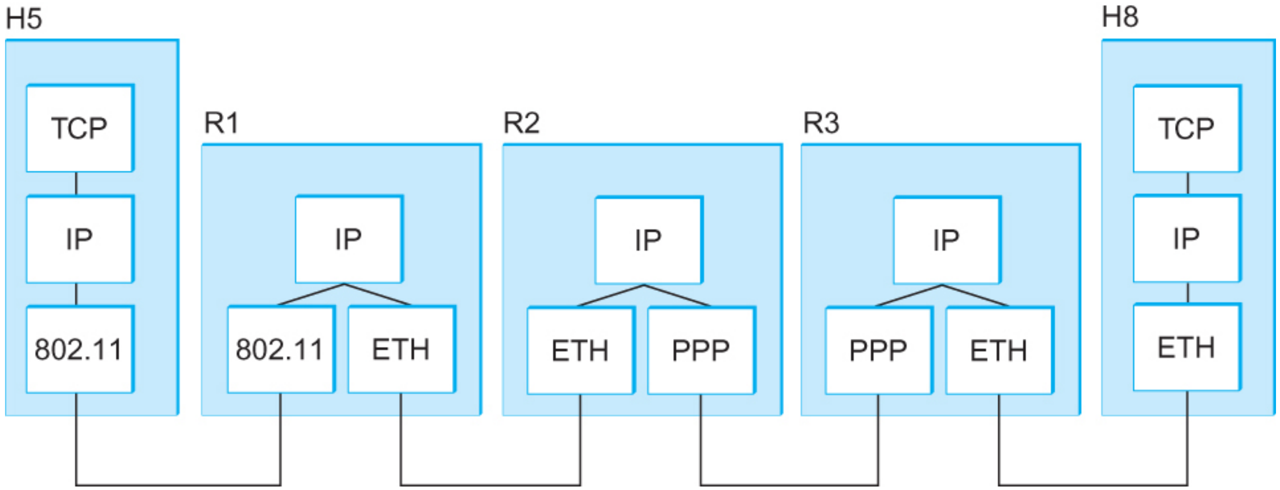
- What is internetwork?
  - A network of networks
  - An arbitrary collection of networks interconnected to provide some sort of host-to-host packet delivery service



A simple internetwork where H represents hosts and R represents routers

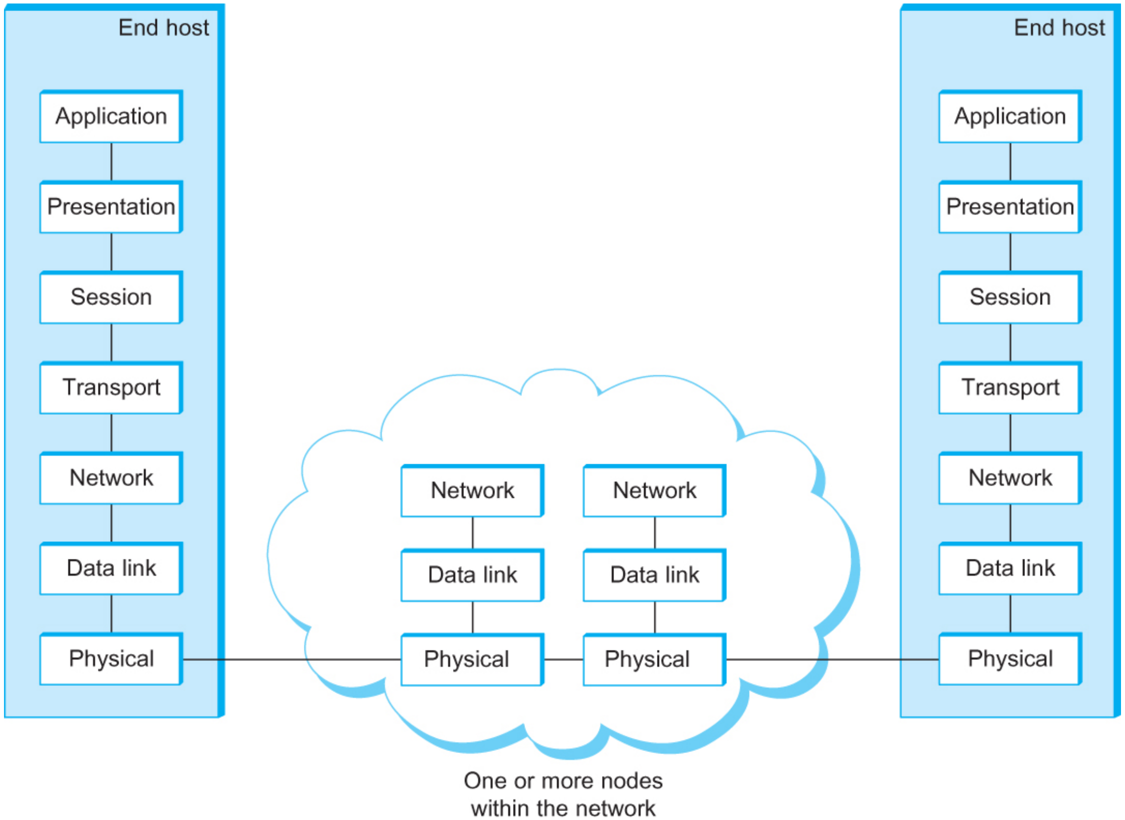
Internetworking

- What is IP?
  - IP stands for Internet Protocol
  - Key tool used today to build scalable, heterogeneous internetworks
  - It runs on all the nodes in a collection of networks and defines the infrastructure that allows these nodes and networks to function as a single logical internetwork



A simple internetwork showing the protocol layers

OSI Architecture Reminder



The OSI 7-layer Model

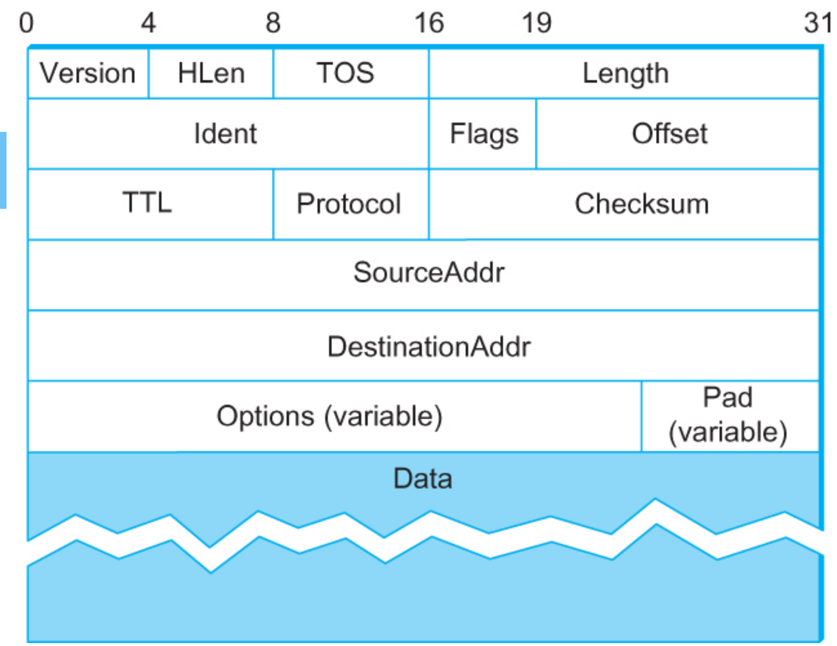
OSI – Open Systems Interconnection

IP Service Model

- Packet Delivery Model
  - Connectionless model for data delivery
    - Datagram model
  - Best-effort delivery (unreliable service)
    - Packets may be lost
    - Packets may be delivered out of order
    - Duplicate copies of a packet may be delivered
    - Packets can be delayed for a long time
    - The higher level layer may need to be aware of these possible failures at the IP layer
- Global Addressing Scheme
  - Provides a way to identify all hosts in the network

IPv4 Packet Format

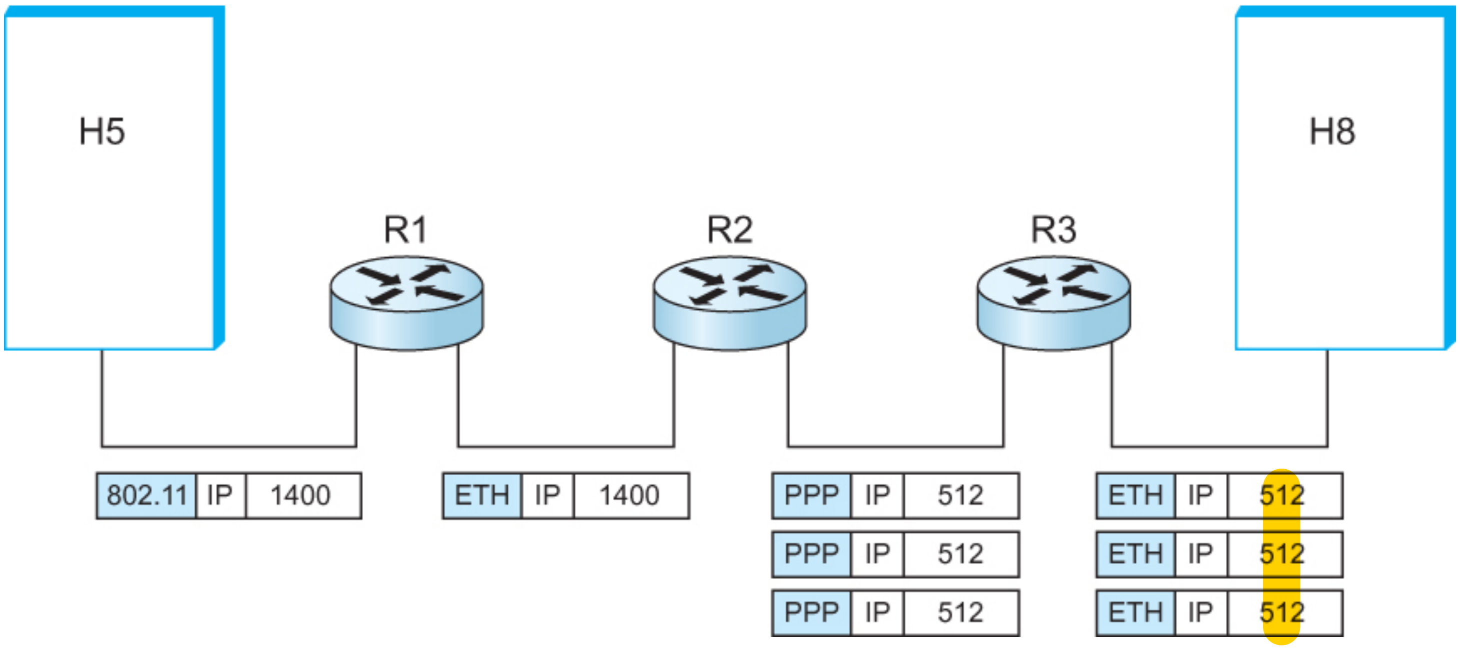
- Version (4): currently 4
- Hlen (4): number of 32-bit words in header
- TOS (8): type of service (to allow packets to be treated differently based on the application needs)
- Length (16): number of bytes in this datagram including the header (max 2^{16} = 65,535 bytes)
- Ident (16): used by fragmentation
- Flags/Offset (16): used by fragmentation
- TTL (8): number of hops this datagram has traveled
- Protocol (8): demux key (TCP=6, UDP=17)
- Checksum (16): of the header only
- DestAddr & SrcAddr (32)



IP Fragmentation and Reassembly

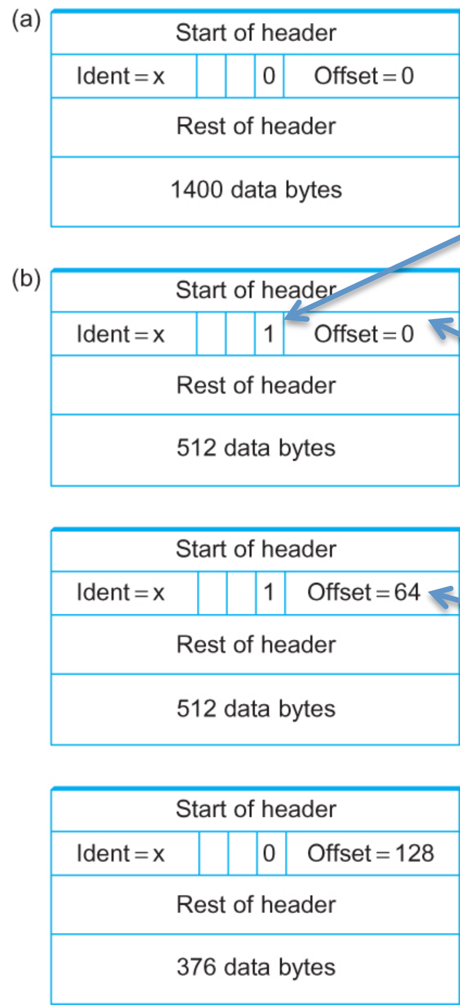
- Each network has some MTU (Maximum Transmission Unit)
  - Ethernet (1500 bytes), FDDI (4500 bytes)
- Strategy
  - Fragmentation occurs in a router when it receives a datagram that it wants to forward over a network link which has (MTU < datagram)
  - Reassembly is done at the receiving host
  - All the fragments carry the same identifier in the *Ident* field
  - Fragments are self-contained IP datagrams
  - IP does not recover from missing fragments
    - So IP fragmentation must be avoided if possible

IP Fragmentation and Reassembly



IP datagrams traversing the sequence of physical networks

IP Fragmentation and Reassembly



Flag bit set to 1 meaning more segments to follow

Byte offset of the first segment used for sequencing fragments

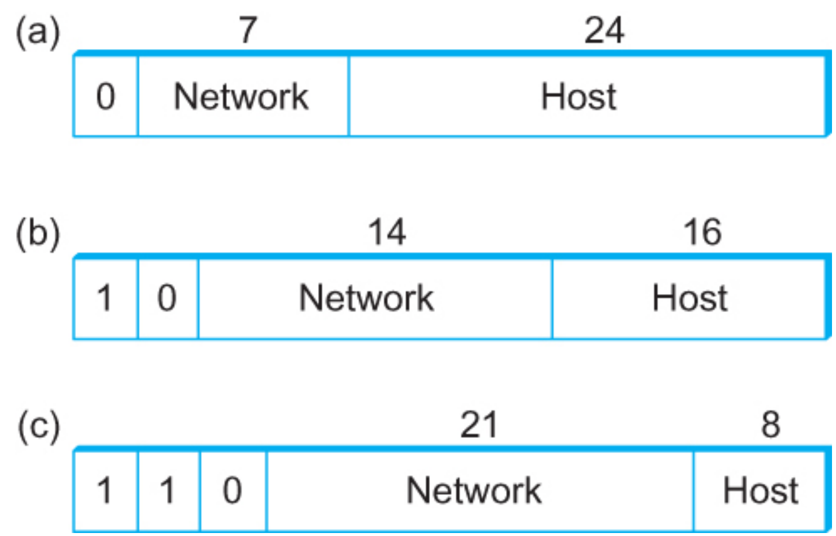
Byte offset of the next segment so 64 x 8 = 512 indicates that this segment is the second 512 bytes of data (512 to 1023)

IP Service Model

1. Packet Delivery Model
   - Connectionless model for data delivery
     - Datagram model
   - Best-effort delivery (unreliable service)
2. Global Addressing Scheme
   - Provides a way to identify all hosts in the network

Global Addresses

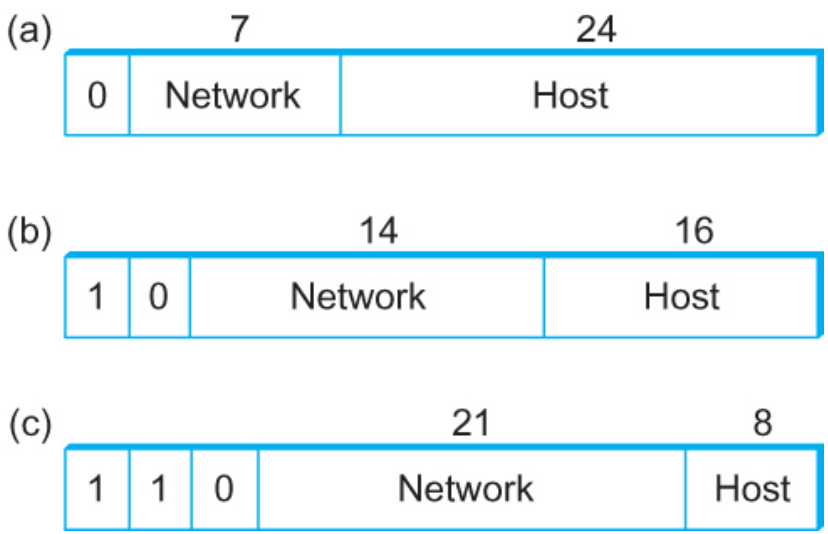
- Properties
  - globally unique
  - hierarchical: network + host
  - 4 Billion IP address (2^{32}), half are A type, ¼ is B type, and 1/8 is C type
- Format (all 32 bits)
  - Class D (multicasting)
    - Leading bits are 1110
  - Class E (research)
    - Leading bits 1111



- Dot notation
  - 10.3.2.4
  - 128.96.33.81
  - 192.168.1.1 (11000000 10101000 00000001 00000001)

Octet 1 Octet 2 Octet 3 Octet 4

Global Addresses



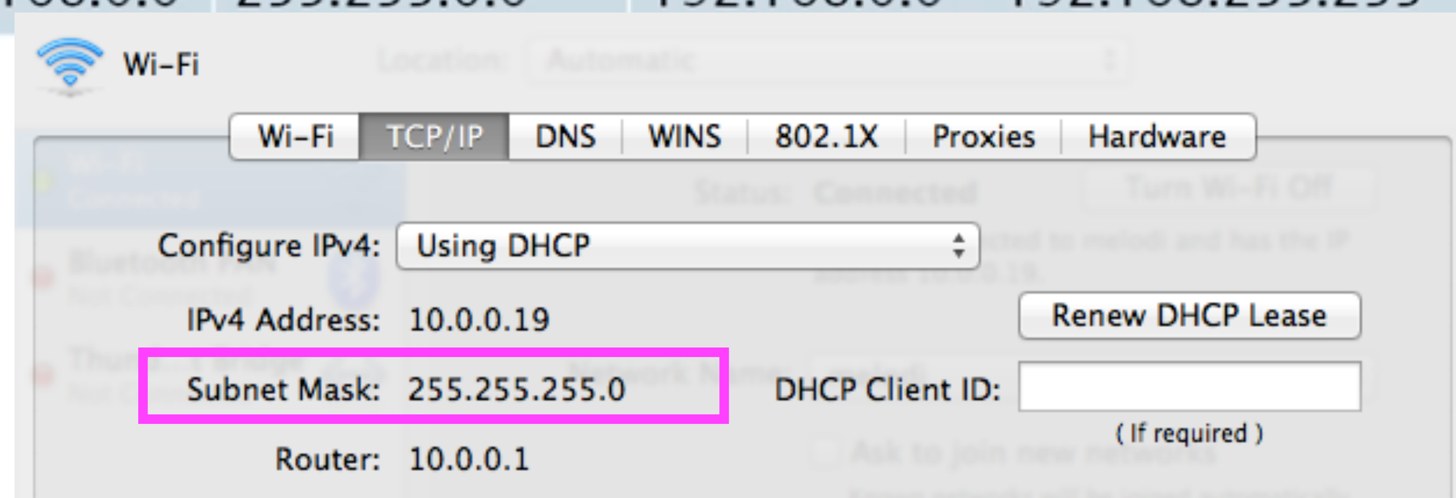
Class	1st Octet	Mask	Format
A	1 – 127	255.0.0.0	n.h.h.h
B	128 - 191	255.255.0.0	n.n.h.h
C	192 - 223	255.255.255.0	n.n.n.h
D	224 - 239	Multicast	Multicast
E	240 - 255	Experimental	Experimental

Global Addresses

Private IP Address Space

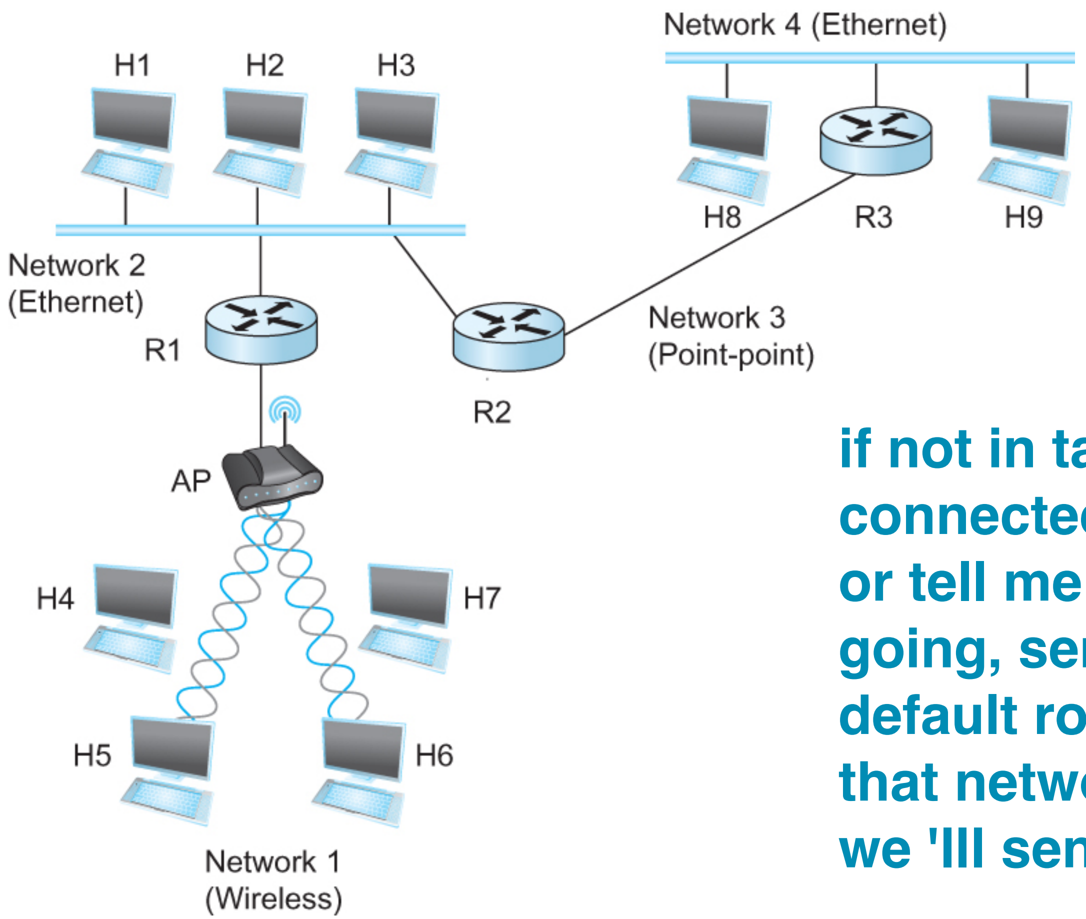
- RFC 1918 defines IP address ranges to be used on private networks.
- These IP addresses are not routable on the global internet and are used inside private networks

Network	Mask	Range
10.0.0.0	255.0.0.0	10.0.0.0 – 10.255.255.255
172.16.0.0	255.240.0.0	172.16.0.0 – 172.31.255.255
192.168.0.0	255.255.0.0	192.168.0.0 – 192.168.255.255



https://www.youtube.com/watch?v=QBqPzHEDzvo (NAT)

Global Addresses



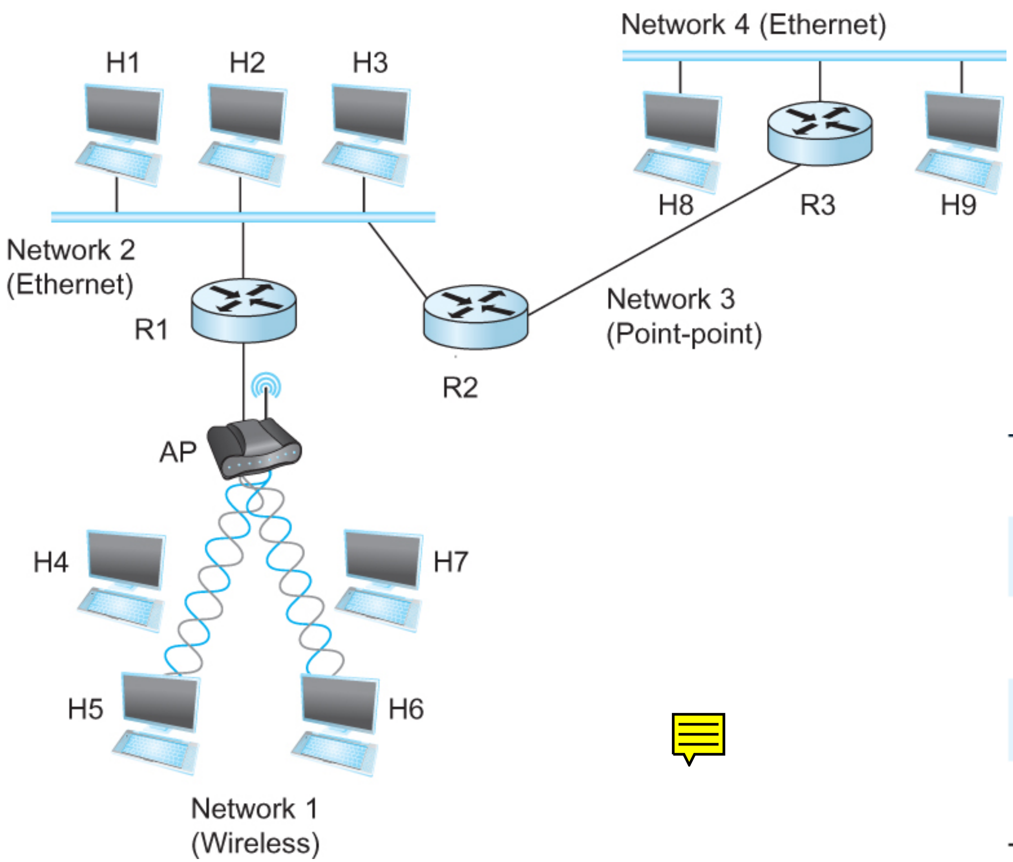
if not in table once connected directly, or tell me where I'm going, send back to default router with that network ID and we 'lll send it there.

IP Datagram Forwarding

- Strategy
  - Every datagram contains destination's address
  - If directly connected to destination network, then forward to host
  - If not directly connected to destination network, then forward to some router
    - Forwarding table maps network number into next hop
  - Each host has a default router
  - Each router maintains a forwarding table

IP Datagram Forwarding

- Example (router R2)



Forwarding Table of R2

NetworkNum	NextHop
1	R1
2	Interface 1
3	Interface 0
4	R3

~~IP Datagram Forwarding~~

- ~~Algorithm~~

~~**if** (NetworkNum of destination = NetworkNum of one of my interfaces) **then**~~

~~deliver packet to destination over that interface~~

~~**else**~~

~~**if** (NetworkNum of destination is in my forwarding table) **then**~~

~~deliver packet to NextHop router~~

~~**else**~~

~~deliver packet to default router~~

~~For a host with only one interface and only a default router in its forwarding table, this simplifies to~~

~~**if** (NetworkNum of destination = my NetworkNum) **then**~~

~~deliver packet to destination directly~~

~~**else**~~

~~deliver packet to default router~~

Classless Addressing

CIDR Notation

Classless Inter-Domain Routing Notation

- Traditionally, subnet masks were determined by the IP address class, so there were only really three subnet masks you would see – For the class A, B and C networks
- To preserve IP address space, use them more efficiently, and help decrease burdon on global routing tables classless interdomain routing was born (CIDR).
- CIDR is used for IP address aggregation and specifies the subnet mask in a different notation
- The CIDR notation lists the network followed by a "/" followed by the number of subnet mask bits
  - Example: 192.168.0.0/16 ← dotted decimal mask 255.255.0.0
  - Example: 220.140.100.0/25 ← dotted decimal mask 255.255.255.128
  - Example: 8.8.8.8/30 ← dotted decimal mask 255.255.255.252

https://www.youtube.com/watch?v=Q1U9wVXRuHA

Address Translation

- ARP (Address Resolution Protocol)
  - Table of IP to physical (MAC) address bindings
  - Broadcast request if IP address not in table
  - Target machine responds with its physical address
  - Table entries are discarded if not refreshed

ARP Packet Format

0	8	16	31
Hardware type = 1		ProtocolType = 0x0800	
HLen = 48	PLen = 32	Operation	
SourceHardwareAddr (bytes 0–3)			
SourceHardwareAddr (bytes 4–5)		SourceProtocolAddr (bytes 0–1)	
SourceProtocolAddr (bytes 2–3)		TargetHardwareAddr (bytes 0–1)	
TargetHardwareAddr (bytes 2–5)			
TargetProtocolAddr (bytes 0–3)			

https://www.youtube.com/watch?v=Ow-jESqubz4

Host Configurations

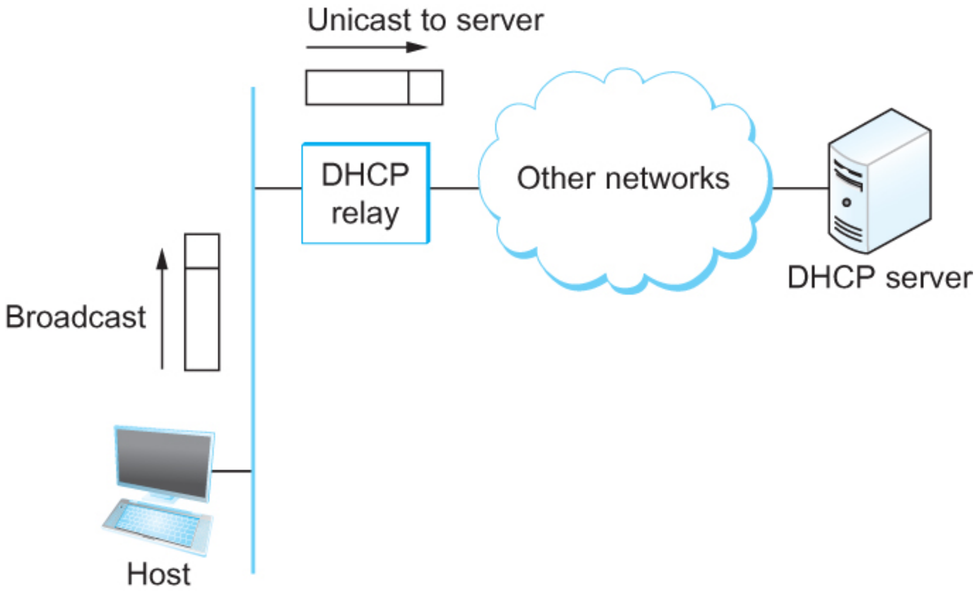
- Ethernet addresses are configured into network by manufacturer and they are unique
- IP addresses must be unique on a given internetwork but also must reflect the structure of the internetwork
- Most host Operating Systems provide a way to manually configure the IP information for the host
- Drawbacks of manual configuration
  - A lot of work to configure all the hosts in a large network
  - Configuration process is error-prone
- Automated Configuration Process is required

Dynamic Host Configuration Protocol (DHCP)

- DHCP server is responsible for providing configuration information to hosts
- There is at least one DHCP server for an administrative domain
- DHCP server maintains a pool of available addresses

DHCP

- Newly booted or attached host sends DHCPDISCOVER message to a special IP address (255.255.255.255)
  - Broadcast
- DHCP relay agent unicasts the message to DHCP server and waits for the response



https://www.youtube.com/watch?v=RUZohsAxPxQ

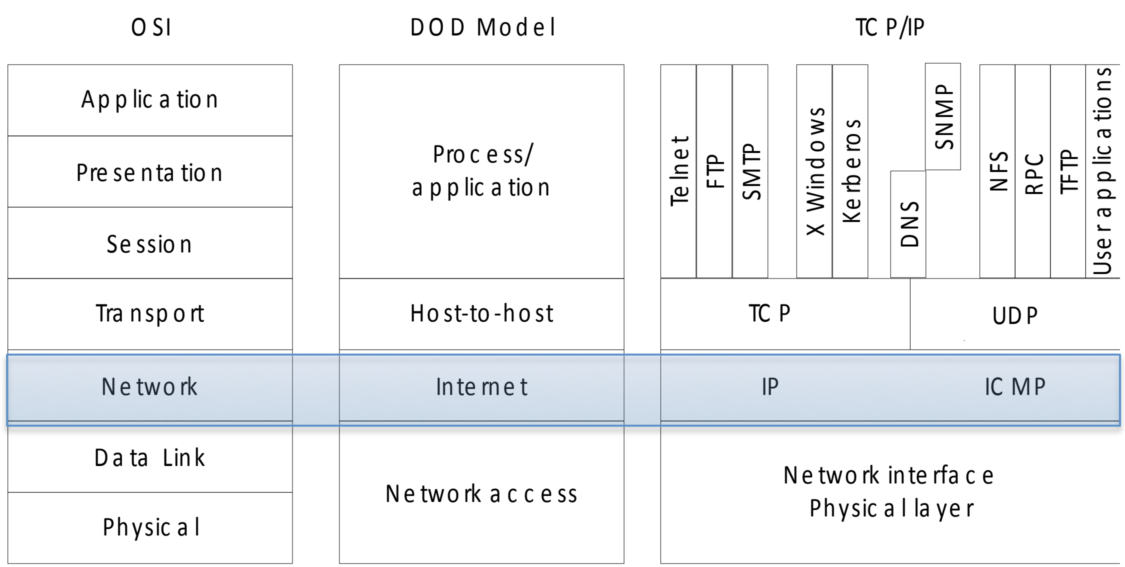
Internet Control Message Protocol (ICMP)

different ICMP for each IPv4 and IPv6

- Defines a collection of error messages that are sent back to the source host whenever a router or host is unable to process an IP datagram successfully
  - Destination host unreachable due to link /node failure
  - Reassembly process failed
  - TTL had reached 0 (so datagrams don't cycle forever)
  - IP header checksum failed

all happened in Data Link Layer

? but why this => show network layer?



Internet Control Message Protocol (ICMP)

- ICMP-Redirect (a useful ICMP control message)
  - From router to a source host
  - With a better route information (for the consequent packets)