Network Forwarding and Link Access

Qiao Xiang

https://qiaoxiang.me/courses/cnns-xmuf21/index.shtml

12/14/2021

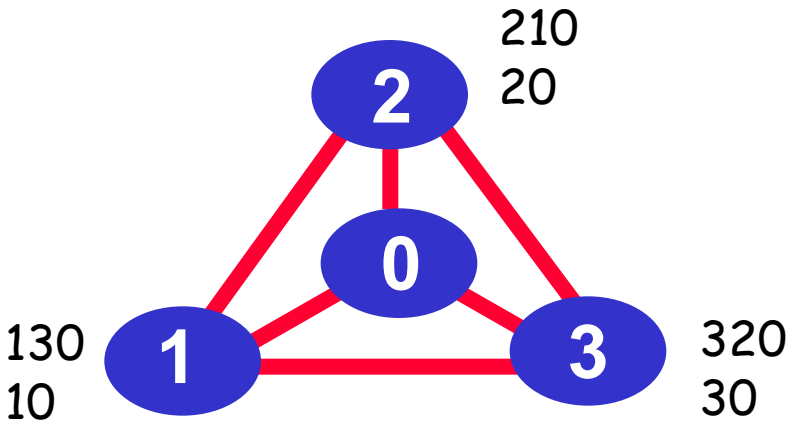
This deck of slides are heavily based on CPSC 433/533 at Yale University, by courtesy of Dr. Y. Richard Yang.

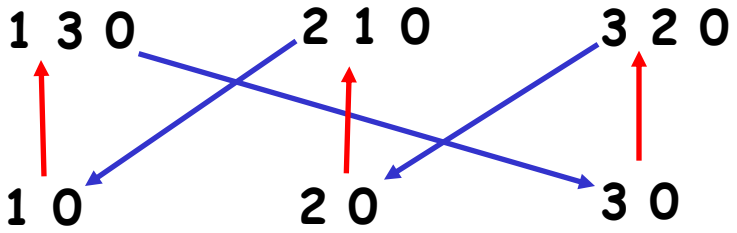
Outline

- Admin and recap
- Network layer
  - Overview
  - Routing
    - Link weights assignment
    - Routing computation
      - Basic routing computation protocols
      - Global Internet routing
        - Basic architecture
        - BGP (Border Gateway Protocol): The de facto Inter-domain routing standard
          - Basic operations
          - BGP as a policy routing framework (control interdomain routes)
          - Policy/interdomain routing analysis
            - Global preference aggregation and Arrow's Theorem
            - Local preference aggregation
            - Economics and interdomain routing patterns
          - IP addresses for interdomain routing
  - Forwarding
- Link layer

Recap: Policy Routing Analysis

- Local preference introduces dependency
- Complete dependency can be captured by a structure called P-graph
- If the P-graph of the networks has no loop, then policy routing converges.





Recap: Interdomain Routing and Economics

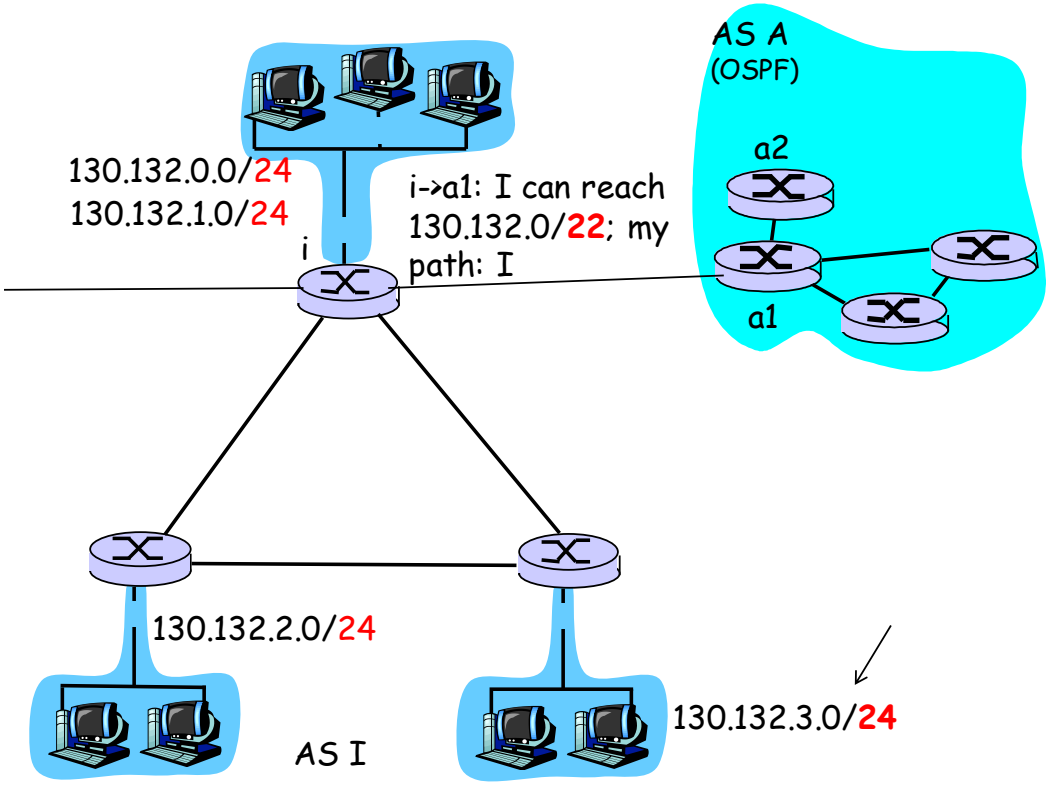
- Economics => typical routing selection policy and export policies
- Typical export policies => routes have patterns
  - e.g., Valley free routing
- Assumptions on economical relationship => a proof of no loop in P-graph.

Recap: Interdomain Routing and IP Addressing

- Requirements
  - Uniqueness
  - Allow aggregation

=>

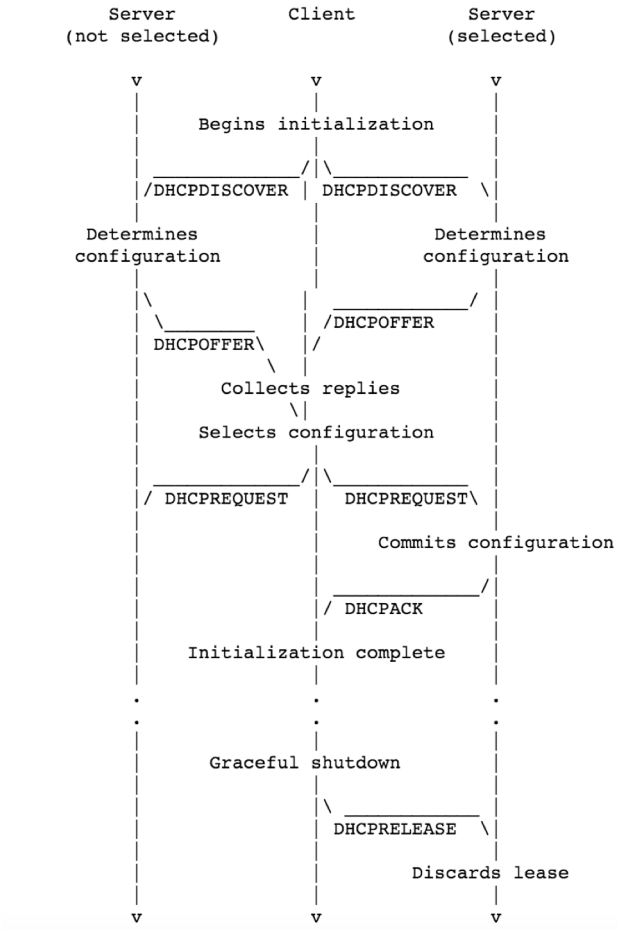
- Classless InterDomain Routing (CIDR) addressing adopted in Internet



Recap: DHCP: Dynamic Host Configuration Protocol

The often used DORA model (4 messages)

- host broadcasts "DHCP discover" msg
- DHCP server responds with "DHCP offer" msg
- host requests IP address: "DHCP request" msg
- DHCP server sends address: "DHCP ack" msg



igure 3: Timeline diagram of messages exchanged between DHCP client and servers when allocating a new network address

Outline

- Admin and recap
- Network layer
  - Overview
  - Routing
  - Forwarding (put it together)

Network Forwarding: Putting it Together

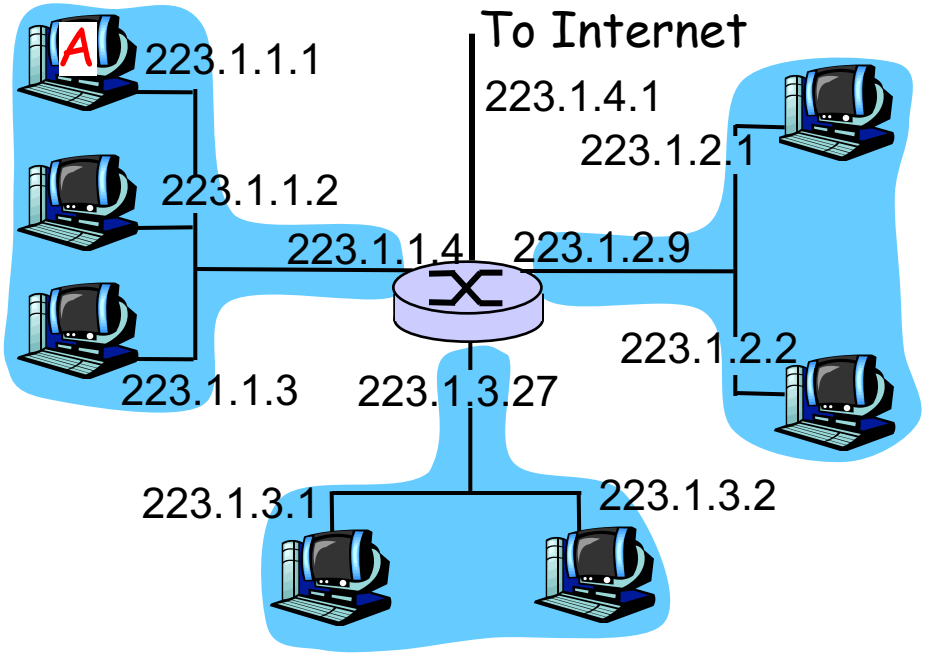
- Forwarding is also called the fast path (upon receiving each packet)

- Slow path: not per packet
  - Get IP address (DHCP, or static)
  - Setup/compute routing table

Forwarding: Example 1

	src	dst	
misc fields	223.1.1.1	223.1.1.3	data

- Setting: Host A network layer receives a packet above.
- Action:
  - Host A looks up destination in routing table
    - Exercise: Suppose A uses DHCP to obtain its address, how can A construct its routing table (routing information base, RIB)?



Host Routing Table Example: my Mac

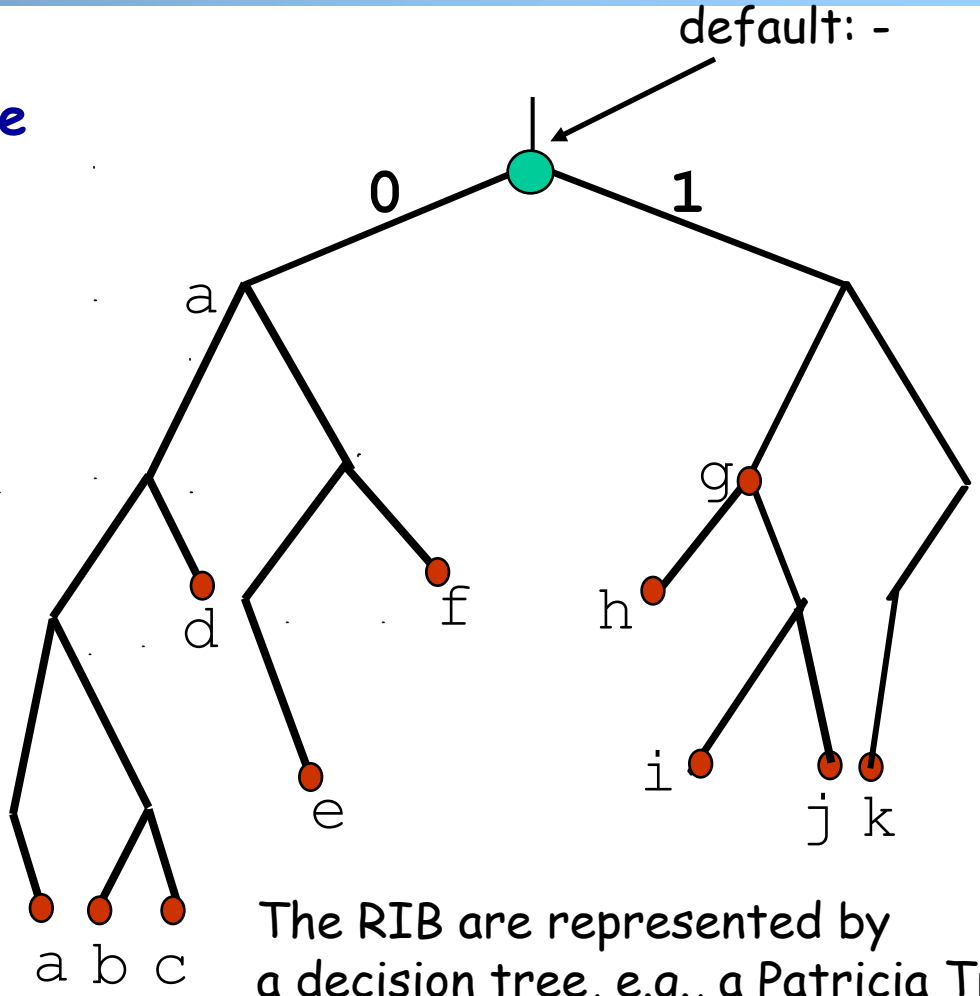
- Mac
  - ifconfig –a
  - netstat –rn (man netstat to see description)

```
Routing tables

Internet:
Destination        Gateway            Flags        Refs      Use   Netif Expire
default            172.27.16.1        UGSc         1470        0     en0
127                127.0.0.1          UCS             1        0     lo0
127.0.0.1          127.0.0.1          UH              4     3788     lo0
169.254            link#4             UCS           106        0     en0
169.254.1.229      link#4             UHLSW           1        0     en0
169.254.5.209      f0:99:bf:1e:6f:de  UHLSW           1        0     en0    989
169.254.8.254      link#4             UHLSW           1        0     en0
169.254.11.96      0:cd:fe:75:59:75   UHLSW           1        0     en0   1009
169.254.13.89      64:9a:be:af:34:53  UHLSW           1        0     en0   1145
169.254.16.49      link#4             UHLSW           1        0     en0
169.254.19.58      link#4             UHLSW           1        0     en0
169.254.19.82      link#4             UHLSW           1        0     en0
169.254.21.198     link#4             UHLSW           1        0     en0
169.254.22.67      0:23:12:12:bc:39   UHLSW           1        0     en0     31
169.254.23.4       link#4             UHLSW           1        0     en0
...
```

CIDR Forwarding Look Up: Software

#	prefix	interface
a)	00001	
b)	00010	
c)	00011	
d)	001	
e)	0101	
f)	011	
g)	10	
h)	100	
i)	1010	
j)	1011	
k)	1100	



The RIB are represented by a decision tree, e.g., a Patricia Trie to look for the longest match of the destination address

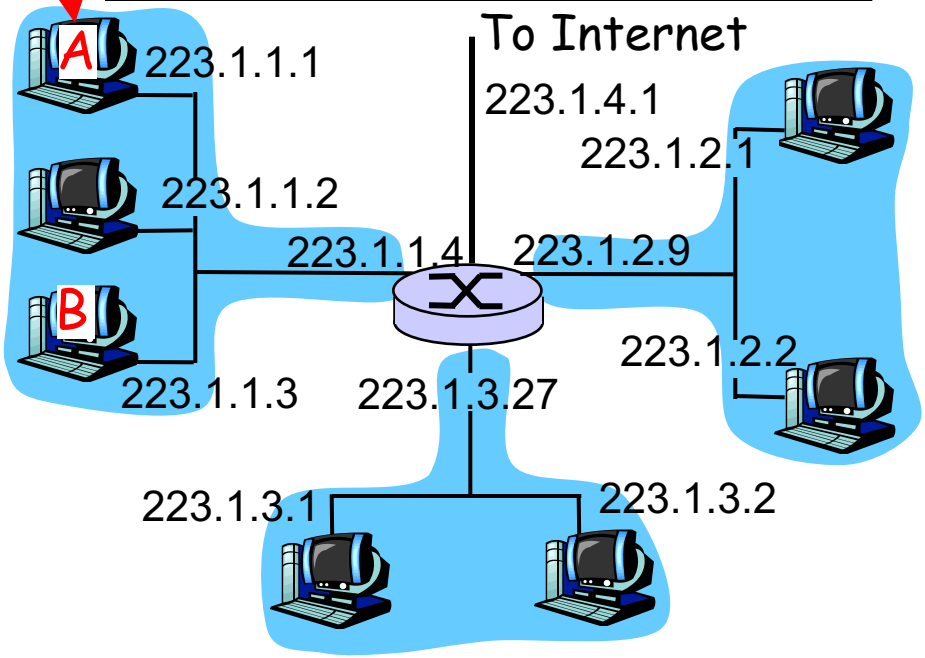
Putting it Together: Example 1: A->B

	src	dst	
misc fields	223.1.1.1	223.1.1.3	data

- Setting: Host A network layer receives a packet above.
- Action:
  - Host A looks up destination in routing table (on same subnet)
  - Hand datagram to link layer to send inside a link-layer frame
    - Key step: need to map B's IP address 223.1.1.3 to B's MAC address

forwarding table in A

Dest. Net.	next router	Nhops
223.1.1/24		1
223.1.2/24	223.1.1.4	2
223.1.3/24	223.1.1.4	2
0.0.0.0/0	223.1.1.4	-



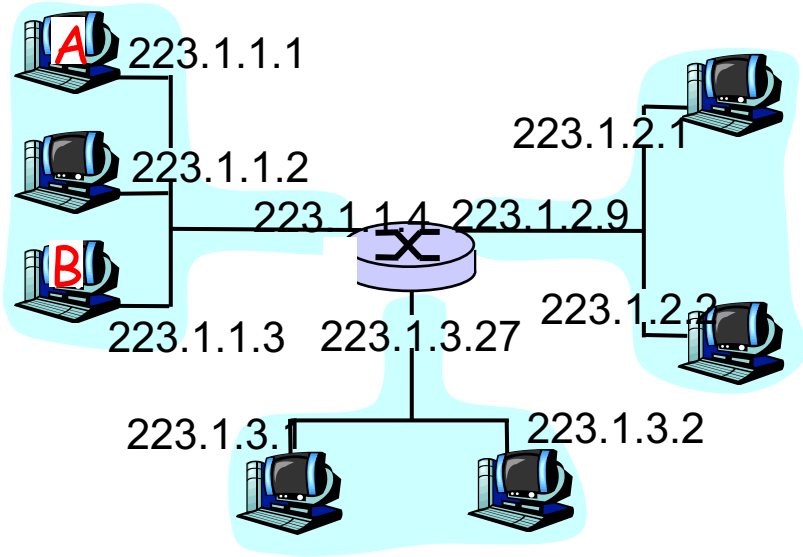
Comparison of IP address and MAC Address

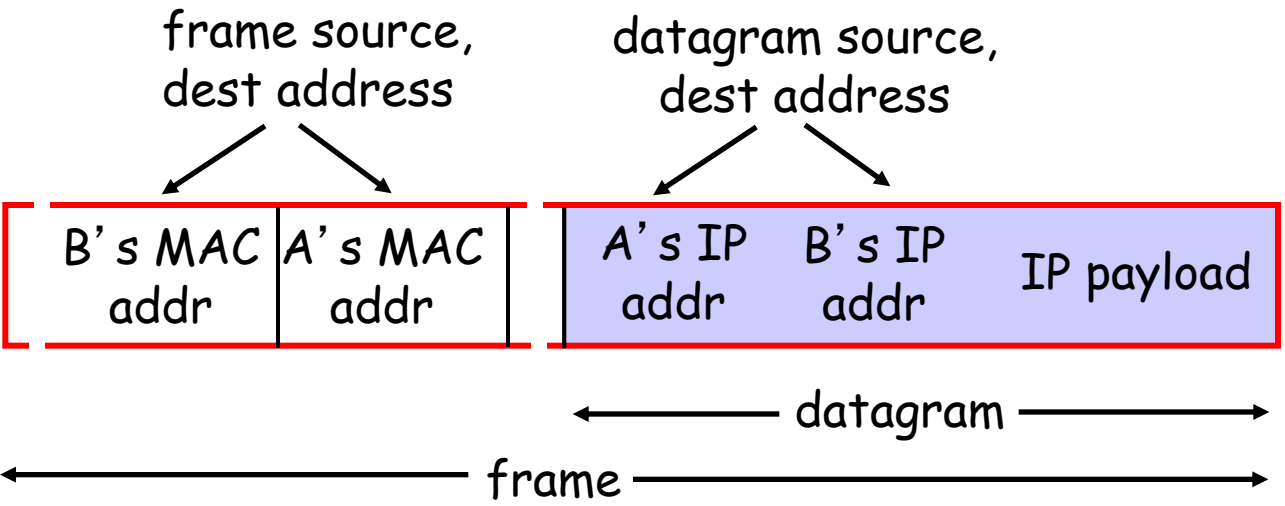
- IP address is locator
  - address depends on network to which an interface is attached
    - NOT portable
  - introduces features (e.g., CIDR) for routing scalability
- IP address needs to be globally unique (if no NAT)

- MAC address is an identifier
  - dedicated to a device
    - portable
  - flat
- MAC address does not need to be globally unique, but the current assignment ensures uniqueness

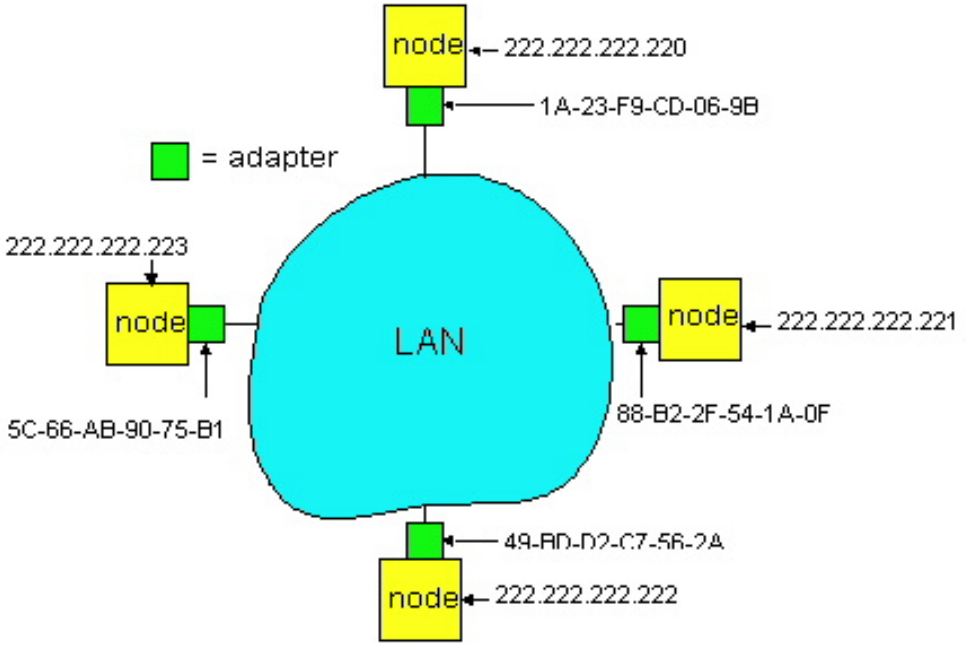
Issue

A finds the MAC address of B to construct





Recall: Address Resolution Table



- Each IP node (Host, Router) on LAN has ARP table
- ARP Table: IP/MAC address mappings for some LAN nodes

  < IP address; MAC address; TTL>
  - TTL (Time To Live): time after which address mapping will be forgotten (typically 20 min)

```
[yry3@cicada yry3]$ /sbin/arp
Address                  HWtype  HWaddress           Flags Mask            Iface
zoo-gatew.cs.yale.edu    ether   AA:00:04:00:20:D4   C                     eth0
artemis.zoo.cs.yale.edu  ether   00:06:5B:3F:6E:21   C                     eth0
lab.zoo.cs.yale.edu      ether   00:B0:D0:F3:C7:A5   C                     eth0
```

Recall: ARP Protocol

- ARP table by the ARP Protocol, which is a "plug-and-play" protocol
  - nodes create their ARP tables without intervention from net administrator

- A broadcast protocol:
  - source broadcasts query frame, containing queried IP address
    - all machines on LAN receive ARP query
  - destination D receives ARP frame, replies
    - frame sent to A's MAC address (unicast)

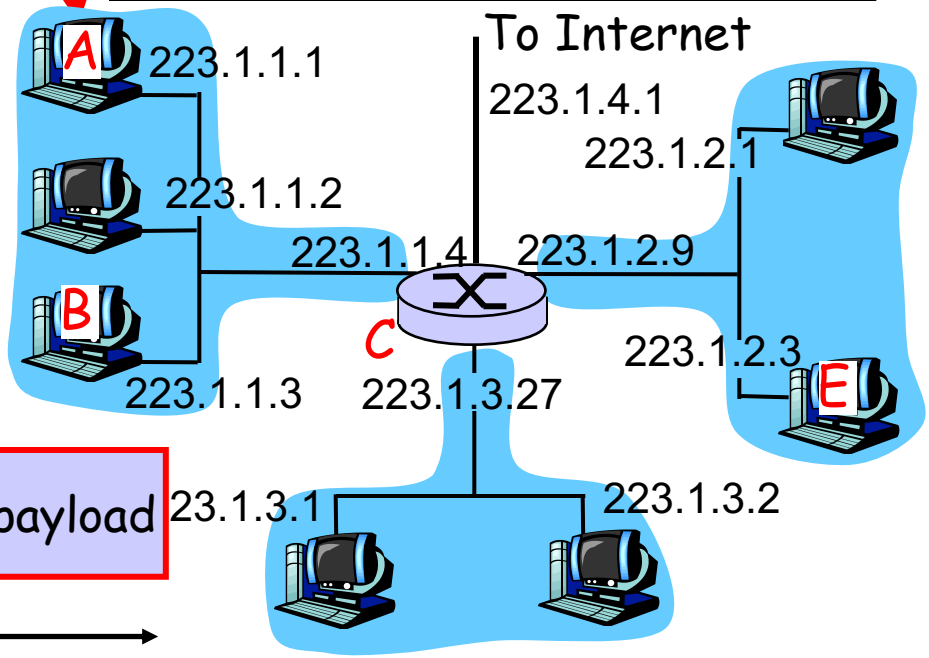
Putting it Together: Example 2 (Different Networks): A-> E

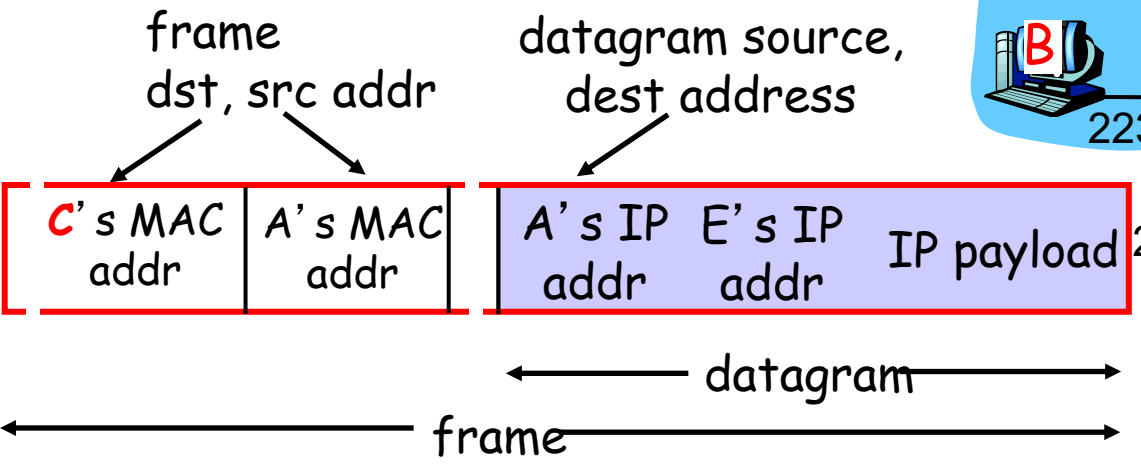
misc fields	223.1.1.1	223.1.2.3	data

- Setting: Host A network layer receives a packet above.
- Action:
  - Host A looks up destination in routing table
    - Find next hop should be 223.1.1.4
  - Hand datagram to link layer to send inside a link-layer frame

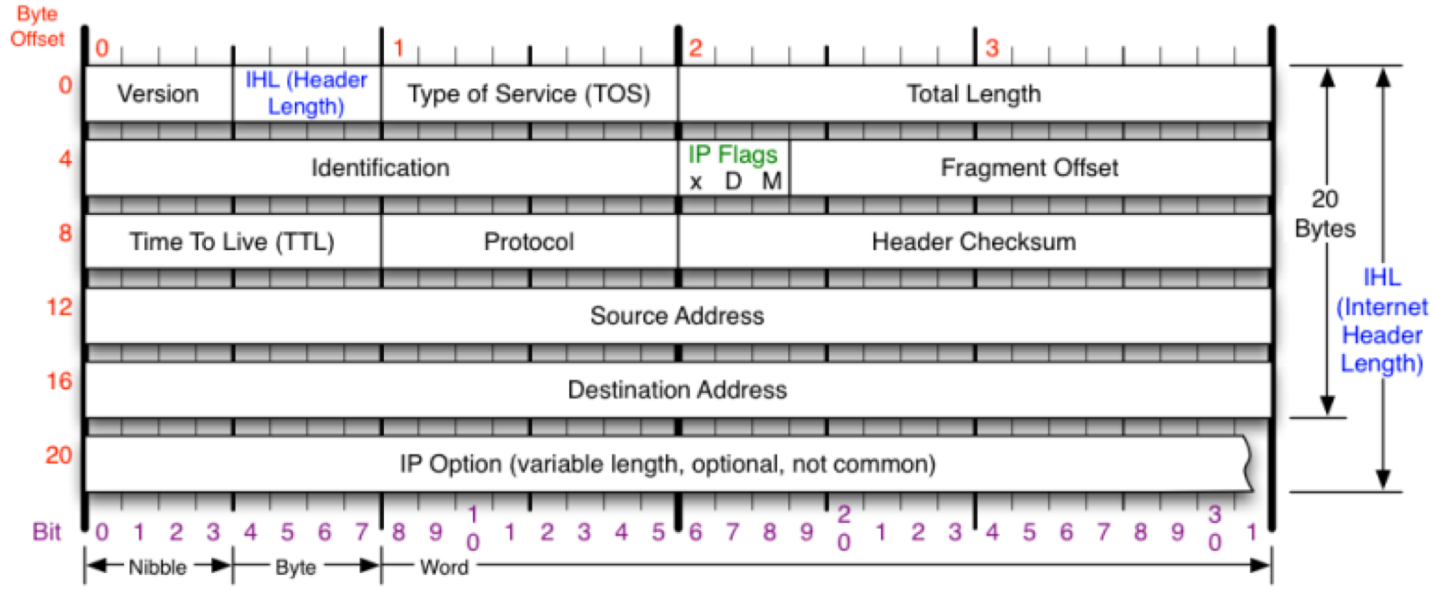
forwarding table in A

Dest. Net.	next router	Nhops
223.1.1/24		1
223.1.2/24	223.1.1.4	2
223.1.3/24	223.1.1.4	2
0.0.0.0/0	223.1.1.4	-





Exercise: Actions at a Router



Version

Version of IP Protocol. 4 and 6 are valid. This diagram represents version 4 structure only.

Header Length

Number of 32-bit words in TCP header, minimum value of 5. Multiply by 4 to get byte count.

Protocol

IP Protocol ID. Including (but not limited to):

1 ICMP	17 UDP	57 SKIP
2 IGMP	47 GRE	88 EIGRP
6 TCP	50 ESP	89 OSPF
9 IGRP	51 AH	115 L2TP

Total Length

Total length of IP datagram, or IP fragment if fragmented. Measured in Bytes.

Fragment Offset

Fragment offset from start of IP datagram. Measured in 8 byte (2 words, 64 bits) increments. If IP datagram is fragmented, fragment size (Total Length) must be a multiple of 8 bytes.

Header Checksum

Checksum of entire IP header

IP Flags

x D M

x 0x80 reserved (evil bit)
D 0x40 Do Not Fragment
M 0x20 More Fragments follow

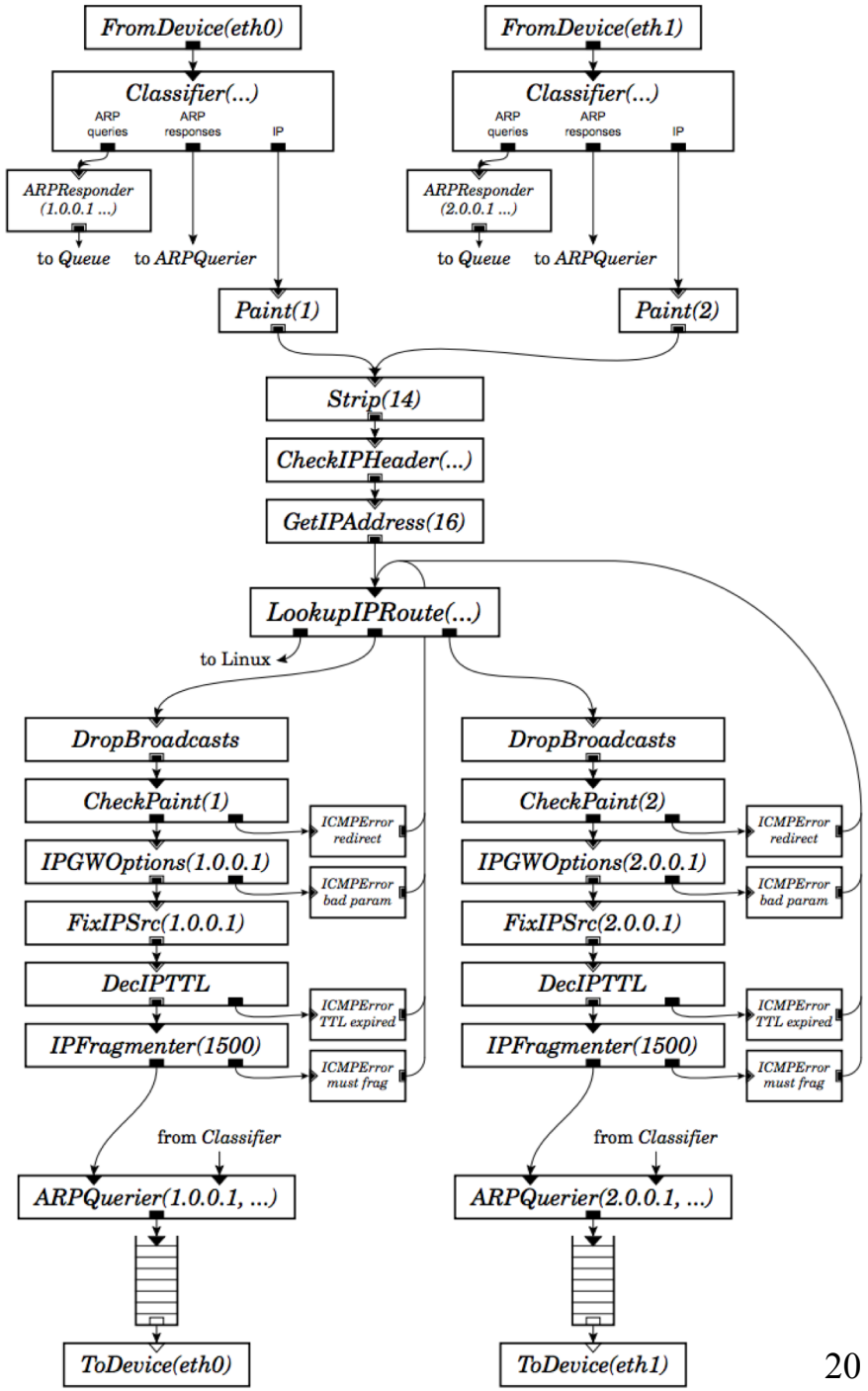
RFC 791

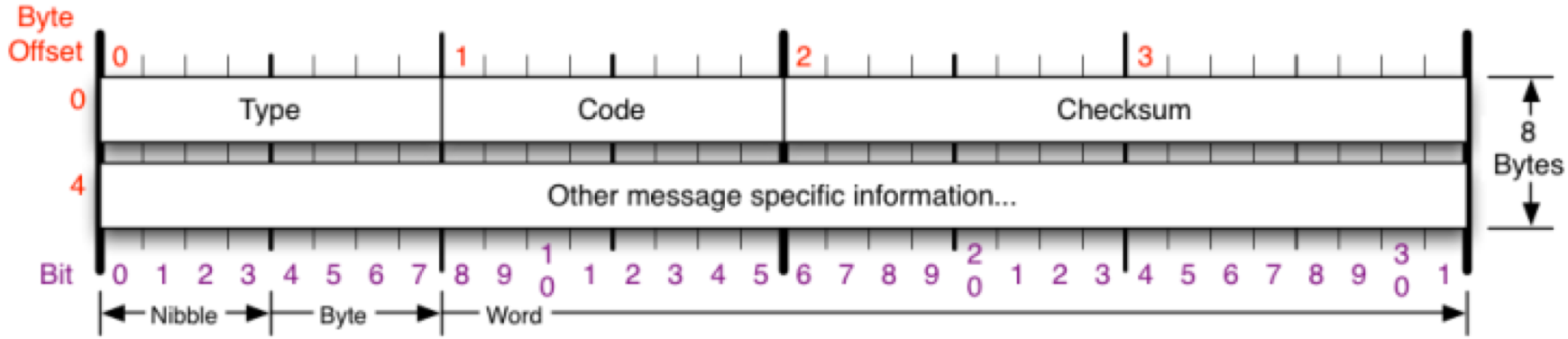
Please refer to RFC 791 for the complete Internet Protocol (IP) Specification.

Source: https://nmap.org/book/tcpip-ref.html

Router Actions

- Routing
  - Look up
  - Prevent loops (TTL--)
  - Notify host better routers (ICMP redirect)
- Handle IPGW options
  - (e.g., record routes)
- Handle fragmentation
- Error checking and reporting using ICMP
- Scheduling (e.g., linux tc)





ICMP Message Types

Type	Code/Name
0	Echo Reply
3	Destination Unreachable
	0 Net Unreachable
	1 Host Unreachable
	2 Protocol Unreachable
	3 Port Unreachable
	4 Fragmentation required, and DF set
	5 Source Route Failed
	6 Destination Network Unknown
	7 Destination Host Unknown
	8 Source Host Isolated
	9 Network Administratively Prohibited
	10 Host Administratively Prohibited
	11 Network Unreachable for TOS

Type	Code/Name
3	Destination Unreachable (continued)
	12 Host Unreachable for TOS
	13 Communication Administratively Prohibited
4	Source Quench
5	Redirect
	0 Redirect Datagram for the Network
	1 Redirect Datagram for the Host
	2 Redirect Datagram for the TOS & Network
	3 Redirect Datagram for the TOS & Host
8	Echo
9	Router Advertisement
10	Router Selection

Type	Code/Name
11	Time Exceded
	0 TTL Exceeded
	1 Fragment Reassembly Time Exceeded
12	Parameter Problem
	0 Pointer Problem
	1 Missing a Required Operand
	2 Bad Length
13	Timestamp
14	Timestamp Reply
15	Information Request
16	Information Reply
17	Address Mask Request
18	Address Mask Reply
30	Traceroute

Checksum

Checksum of ICMP header

RFC 792

Please refer to RFC 792 for the Internet Control Message protocol (ICMP) specification.

What A Router Looks Like: Outside

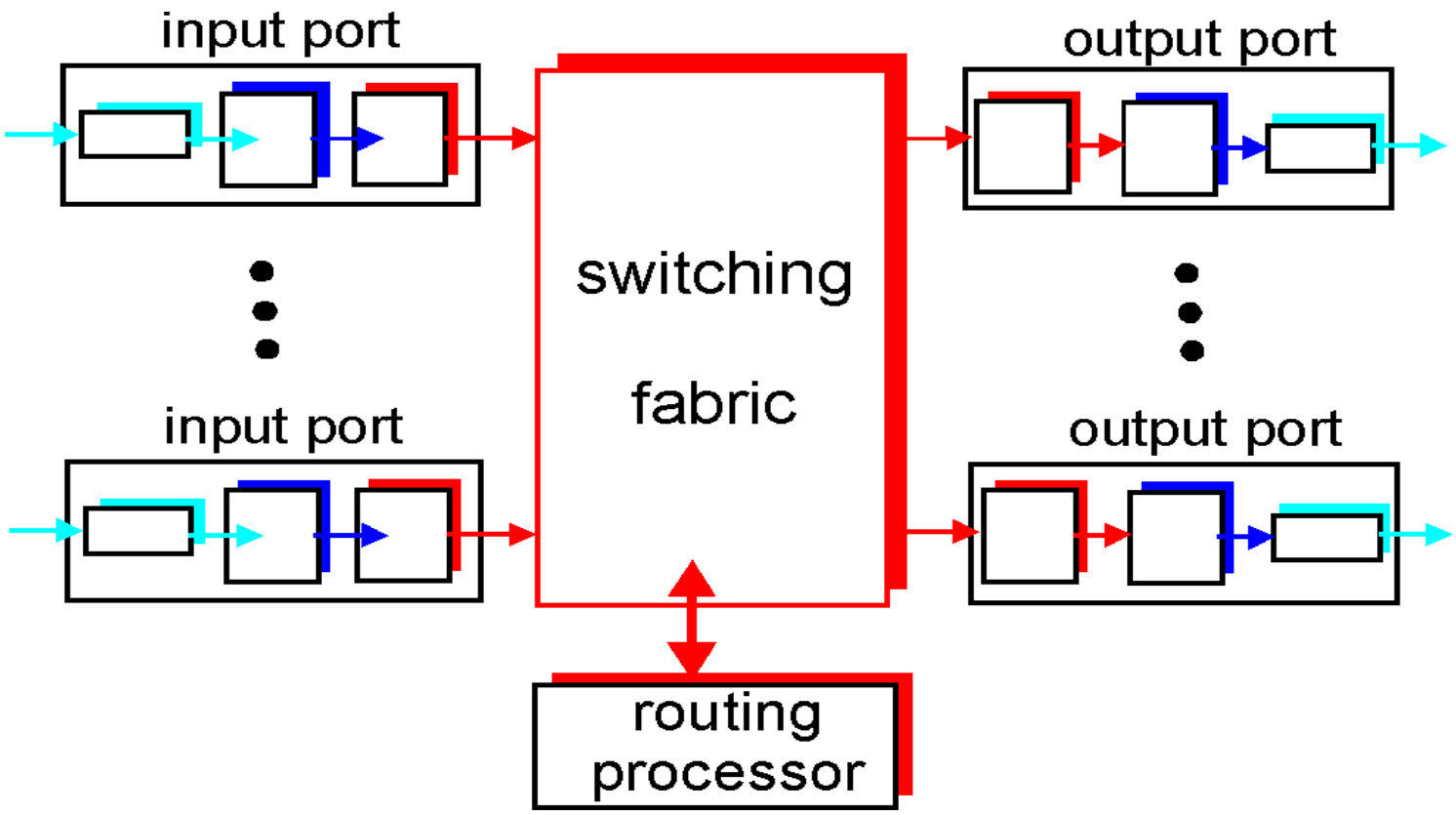
Cisco GSR 12416

Capacity: 160Gb/s
Power: 4.2kW

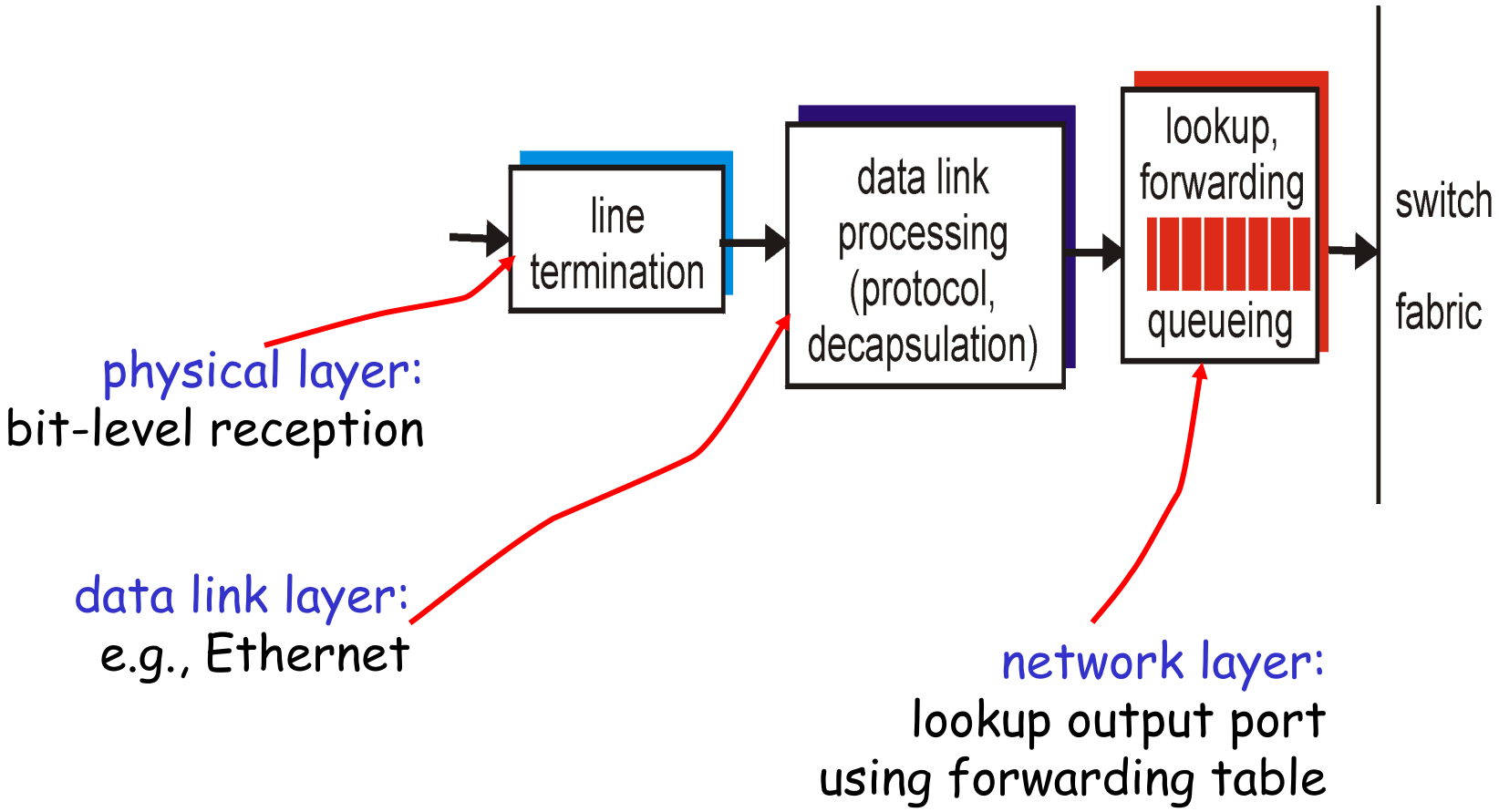
Juniper M160

Capacity: 80Gb/s
Power: 2.6kW

Look Inside a Router

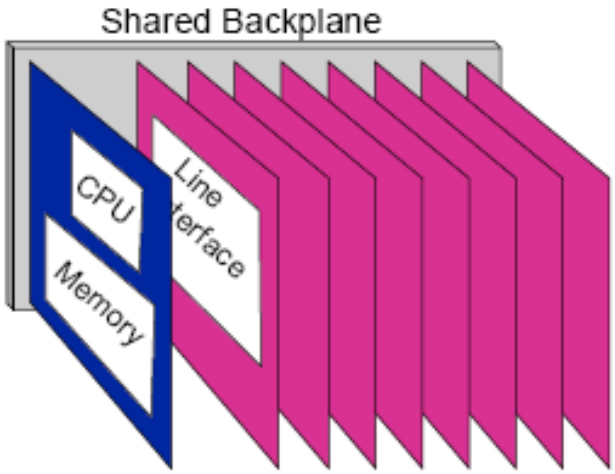


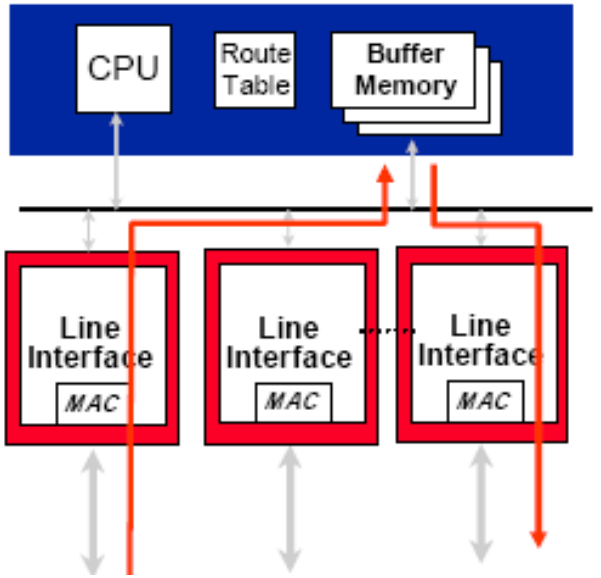
Look Inside a Router: Input Port



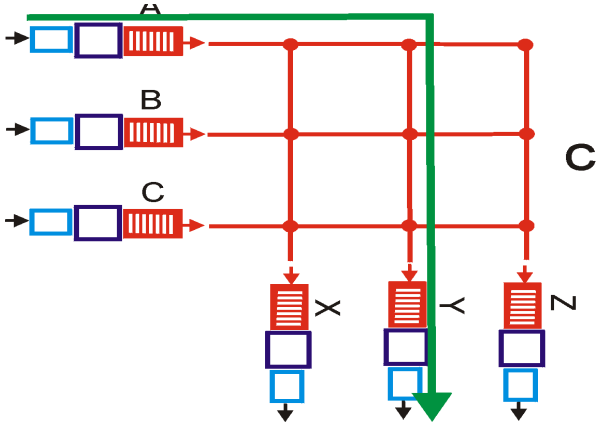
Look Inside a Router: Switching Fabric

Low End

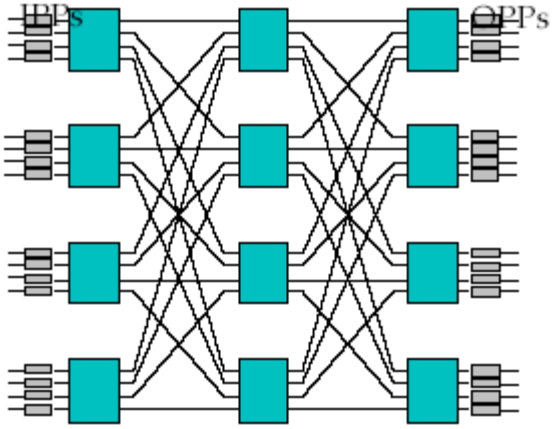


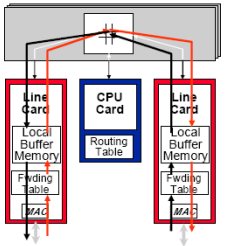


High End



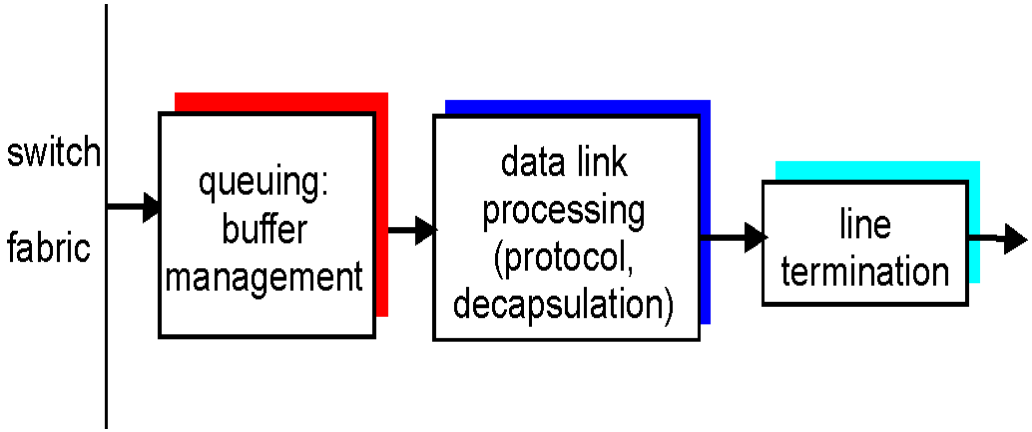
crossbar





Banyan

Look Inside a Router: Output Port



- *Buffering* required when datagrams arrive from fabric faster than the transmission rate
- *Queueing (delay) and loss due to output port buffer overflow !*
- *Scheduling and queue/buffer management* choose among queued datagrams for transmission

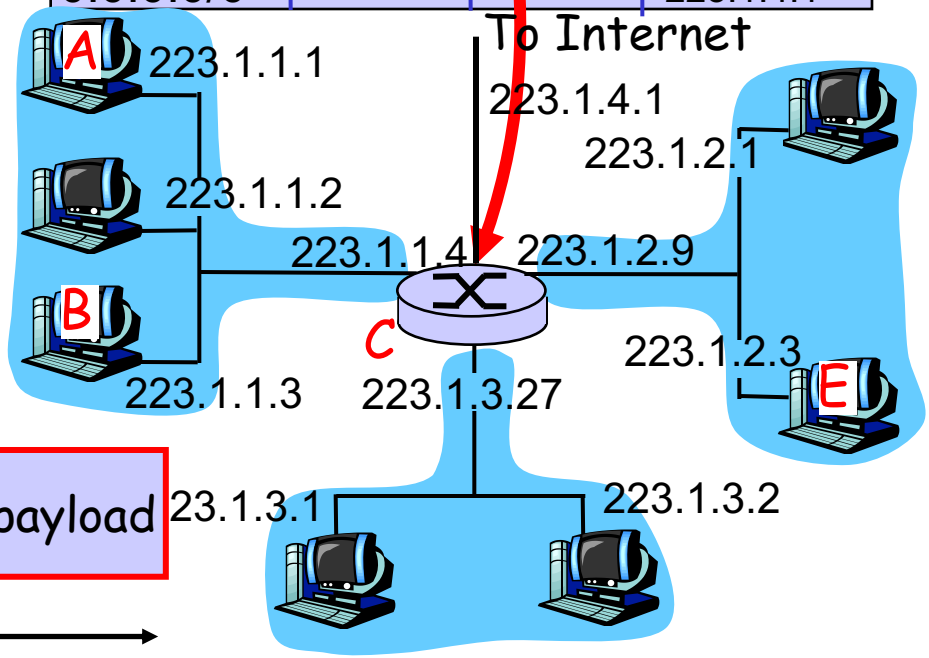
Putting it Together: Example 2 (Different Networks): A-> E

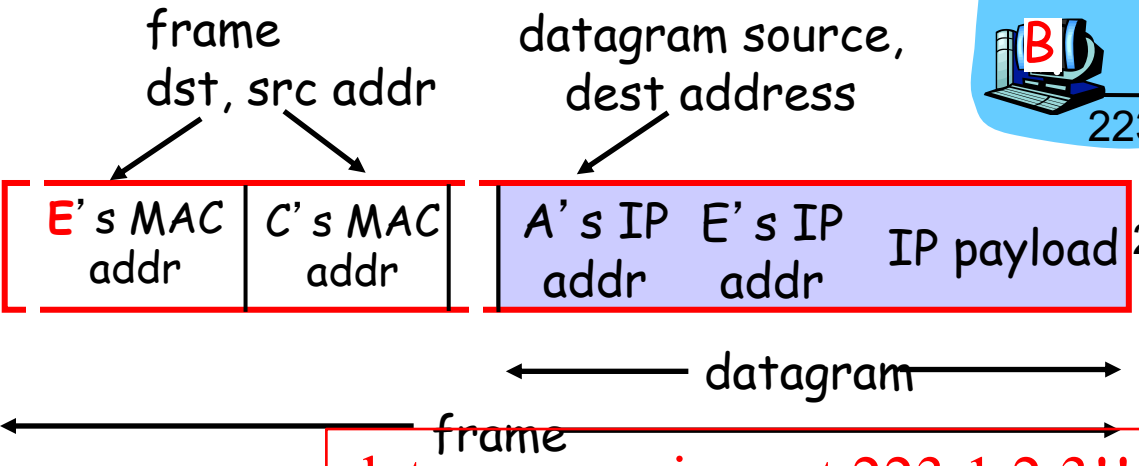
misc fields	223.1.1.1	223.1.2.3	data

- Setting: Packet above arrives at Router C's network layer.
- Action:
  - Router C conducts standard router actions
    - Assume packet correct, find next hop should be 223.1.2.9
  - Hand datagram to link layer to send inside a link-layer frame

forwarding table in router

Dest. Net	router	Nhops	interface
223.1.1/24	-	1	223.1.1.4
223.1.2/24	-	1	223.1.2.9
223.1.3/24	-	1	223.1.3.27
0.0.0.0/0	-	-	223.1.4.1





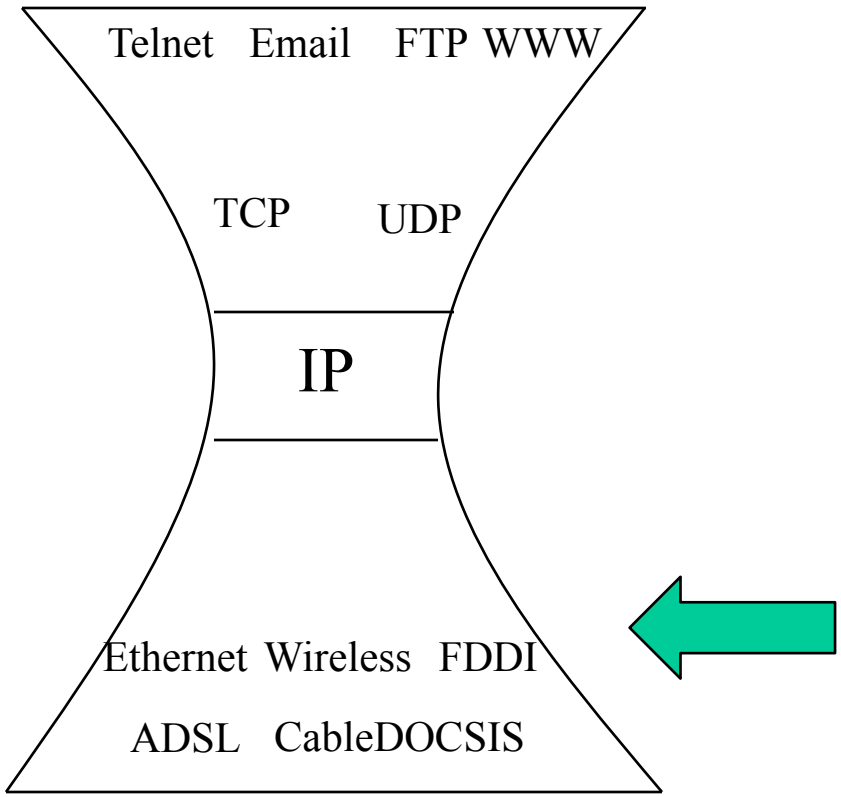
datagram arrives at 223.1.2.3!! (hooray!)

Summary of Network Layer

- We have covered the very basics of the network layer
  - routing and basic forwarding

- There are multiple other topics that we did not cover
  - Multicast/anycast
  - QoS
  - slides as backup
    just in case you need
    reading in the winter

Roadmap: The Hourglass Architecture of the Internet



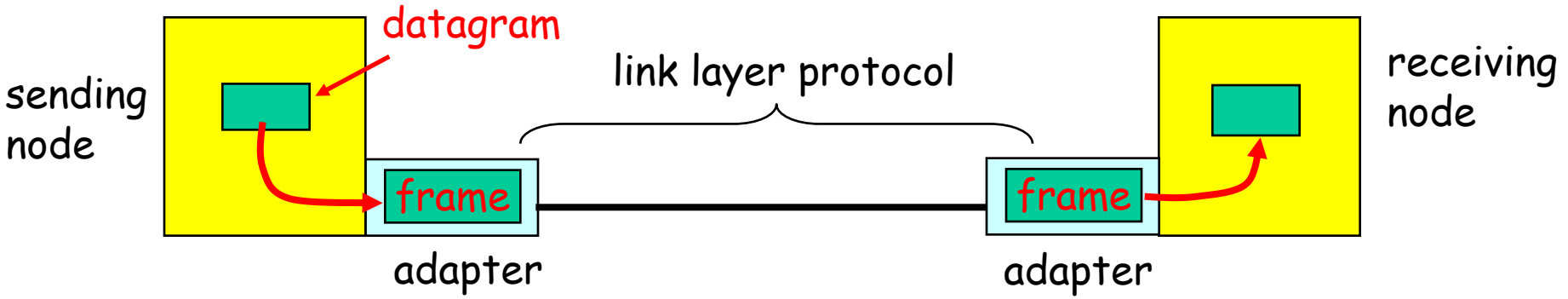
Outline

- Admin and recap
- Network layer
- Link layer

Link Layer Services

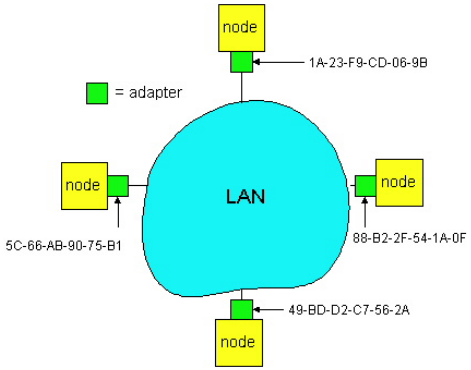
- Framing
  - encapsulate datagram into frame, adding header, trailer and error detection/correction
- Multiplexing/demultiplexing
  - frame headers to identify src, dest
- Reliable delivery between adjacent nodes
  - we learned how to do this already !
  - seldom used on low bit error link (fiber, some twisted pair)
  - common for wireless links: high error rates
- Media access control
- Forwarding/switching with a link-layer (Layer 2) domain

Adaptors Communicating



- link layer typically implemented in "adaptor" (aka NIC)
  - Ethernet card, modem, 802.11 card, cloud virtual switch
- adapter is semi-autonomous, implementing link & physical layers

- in most link-layer, each adapter has a unique link layer address (also called MAC address)

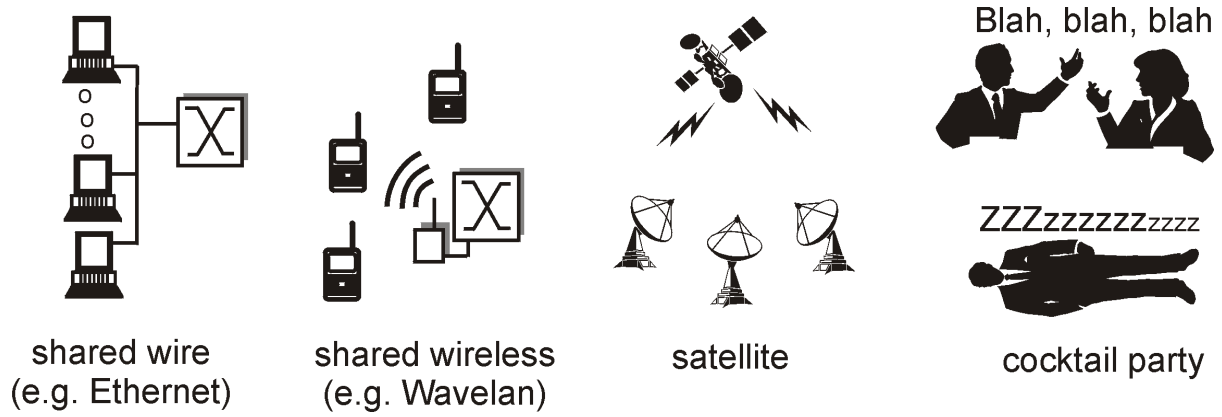


Outline

- Admin and recap
- Network layer
- Link layer
  - Overview
  - Media access
  - Link layer forwarding

Multiple Access Links and Protocols

- Many link layers use broadcast (shared wire or medium)
  - traditional Ethernet; Cable networks
  - 802.11 wireless LAN; cellular networks
  - satellite



- Problem: if two or more simultaneous transmissions, due to interference, only one node can send successfully at a time (see CDMA later for an exception)

Multiple Access Protocol

- Protocol that determines how nodes share channel, i.e., determines when nodes can transmit
- Communication about channel sharing must use channel itself !

- Discussion: properties of an ideal multiple access protocol.

Ideal Mulitple Access Protocol

Broadcast channel of rate R bps

- Efficiency: when only one node wants to transmit, it can send at full rate R
- Rate allocation:
  - simple fairness: when N nodes want to transmit, each can send at average rate R/N
  - we may need more complex rate control
- Decentralized:
  - no special node to coordinate transmissions
  - no synchronization of clocks
- Simple

MAC Protocols

Goals

- efficient, fair, decentralized, simple

Three broad classes:

- non-partitioning
  - random access
    - allow collisions
  - "taking-turns"
    - a token coordinates shared access to avoid collisions
- channel partitioning
  - divide channel into smaller "pieces" (time slot, frequency, code)

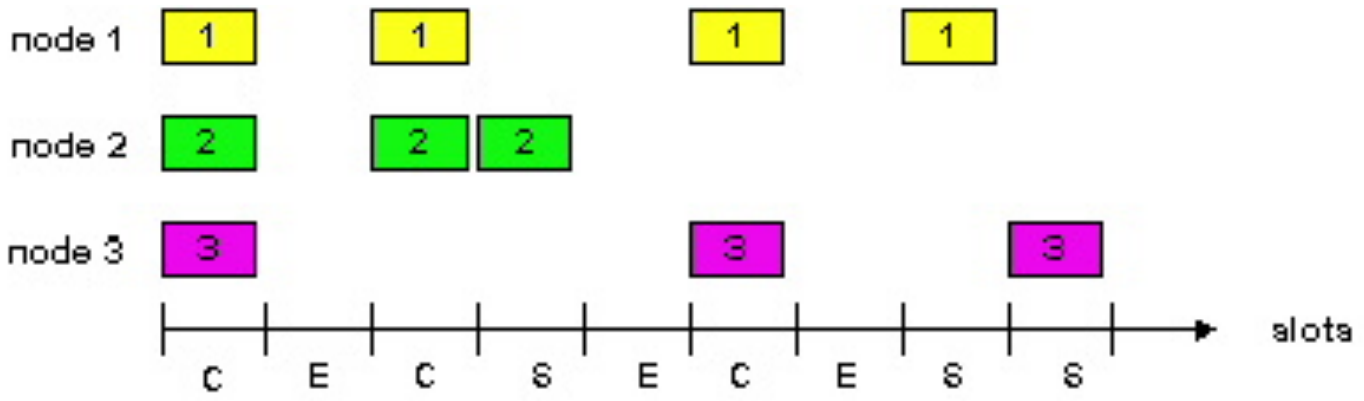
Focus: Random Access Protocols

- Examples of random access MAC protocols:
  - slotted ALOHA and pure ALOHA
  - CSMA and CSMA/CD, CSMA/CA
  - Ethernet, WiFi 802.11

- Key design points:
  - when to access channel?
  - how to detect collisions?
  - how to recover from collisions?

Slotted Aloha [Norm Abramson]

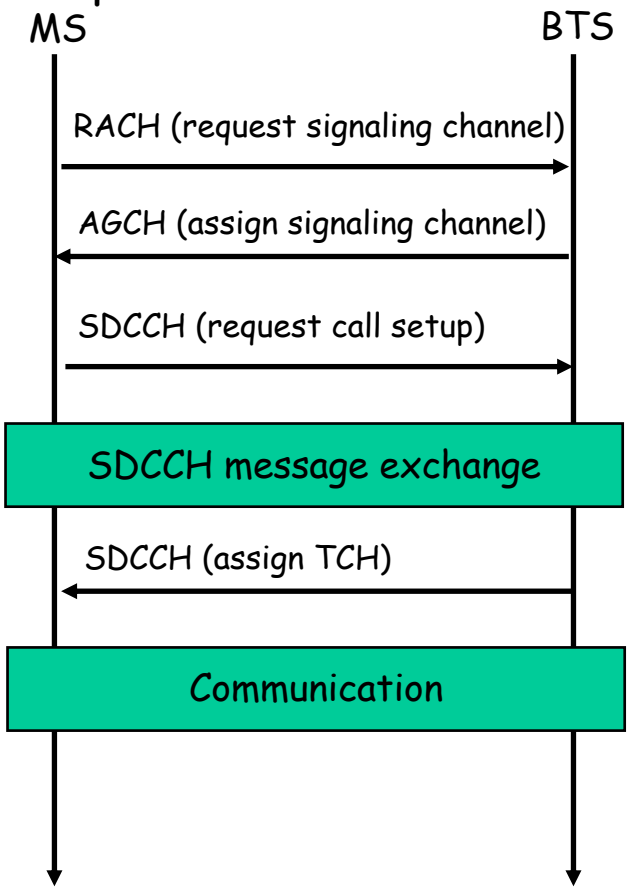
- Time is divided into equal size slots (= pkt trans. time)
- Node with new arriving pkt: transmit at beginning of next slot
- If collision: retransmit pkt in future slots with probability p, until successful.



Success (S), Collision (C), Empty (E) slots

Slotted Aloha in Real Life

- call setup in GSM



- Notations:
  - Broadcast control channel (BCCH): from base station, announces cell identifier, synchronization
  - Random access channel (RACH): MSs for initial access, **slotted Aloha**
  - access grant channel (AGCH): BTS informs an MS its allocation
  - standalone dedicated control channel (SDCCH): signaling and short message between MS and an MS
  - Traffic channels (TCH)

Slotted Aloha Efficiency

Q: What is the fraction of successful slots?

suppose n stations have packets to send

suppose each transmits in a slot with probability *p*

- prob. of succ. by a specific node: $p\,(1-p)^{(n-1)}$

- prob. of succ. by any one of the N nodes

  S(p) = n * Prob (only one transmits)

  = $n\,p\,(1-p)^{(n-1)}$