Master Course Computer Networks IN2097

Prof. Dr.-Ing. Georg Carle
Christian Grothoff, Ph.D.
Stephan Günther

Chair for Network Architectures and Services
Department of Computer Science
Technische Universität München
http://www.net.in.tum.de

Roadmap

Link Layer

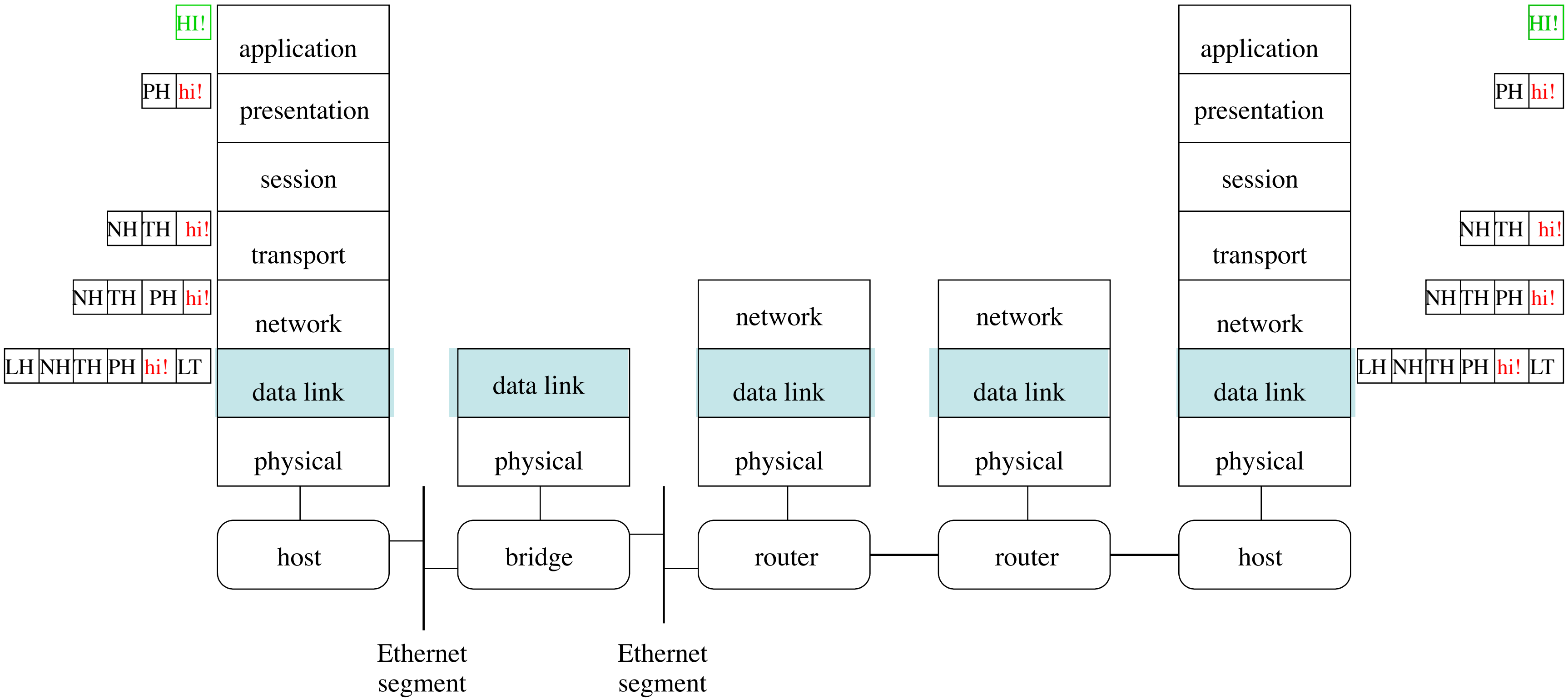
Internet Protocol

The Internet

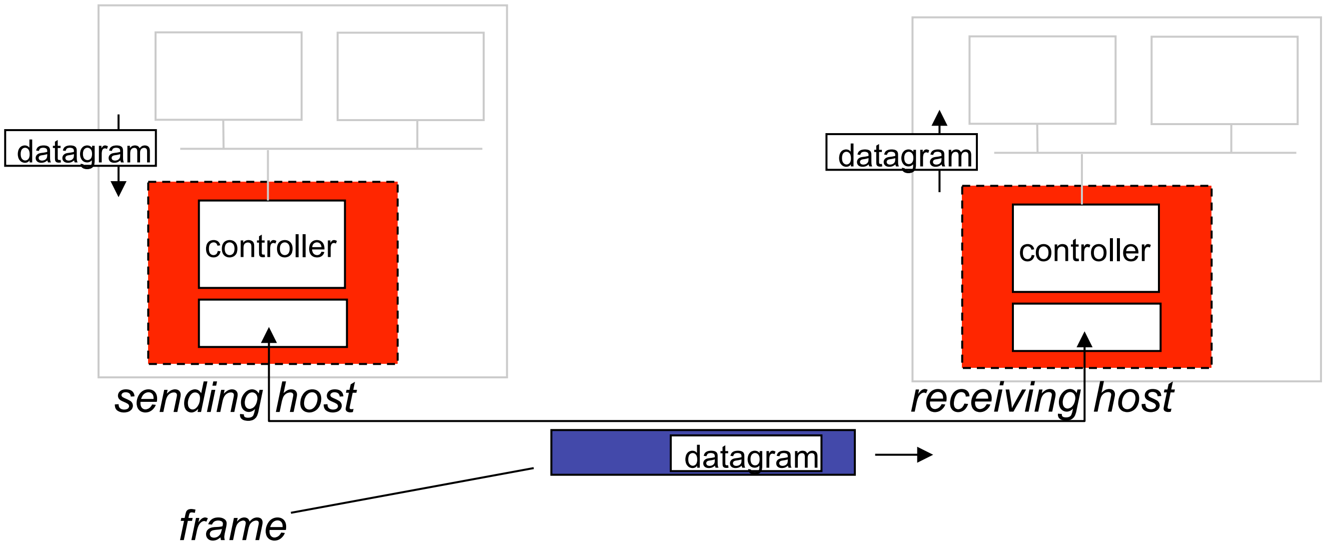
Delay, loss and throughput in packet-switched networks

Link Layer

Protocol Layering



Adaptors Communicating



- sending side:
  - encapsulates datagram in frame
  - adds error checking bits, flow control, etc.
- receiving side
  - looks for errors, flow control, etc.
  - extracts datagram, passes to upper layer at receiving side

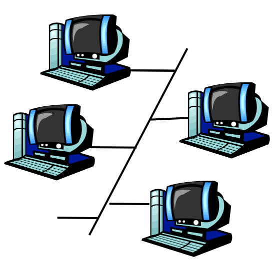
Link Layer

- Introduction and services
- Multiple access protocols
- Link-layer Addressing
- Ethernet
- Link-layer switches

Multiple Access Links and Protocols

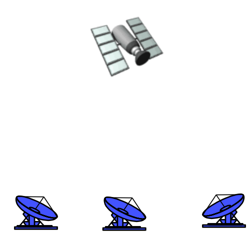
Two types of "links":

- point-to-point
  - PPP for dial-up access
  - point-to-point link between Ethernet switch and host
- broadcast (shared wire or medium)
  - old-fashioned Ethernet
  - upstream HFC
  - 802.11 wireless LAN



shared wire (e.g., cabled Ethernet)

shared RF (e.g., 802.11 WiFi)



shared RF (satellite)

humans at a cocktail party (shared air, acoustical)

Multiple Access protocols

- single shared broadcast channel
- two or more simultaneous transmissions by nodes: interference
  - *collision* if node receives two or more signals at the same time

Multiple access protocol

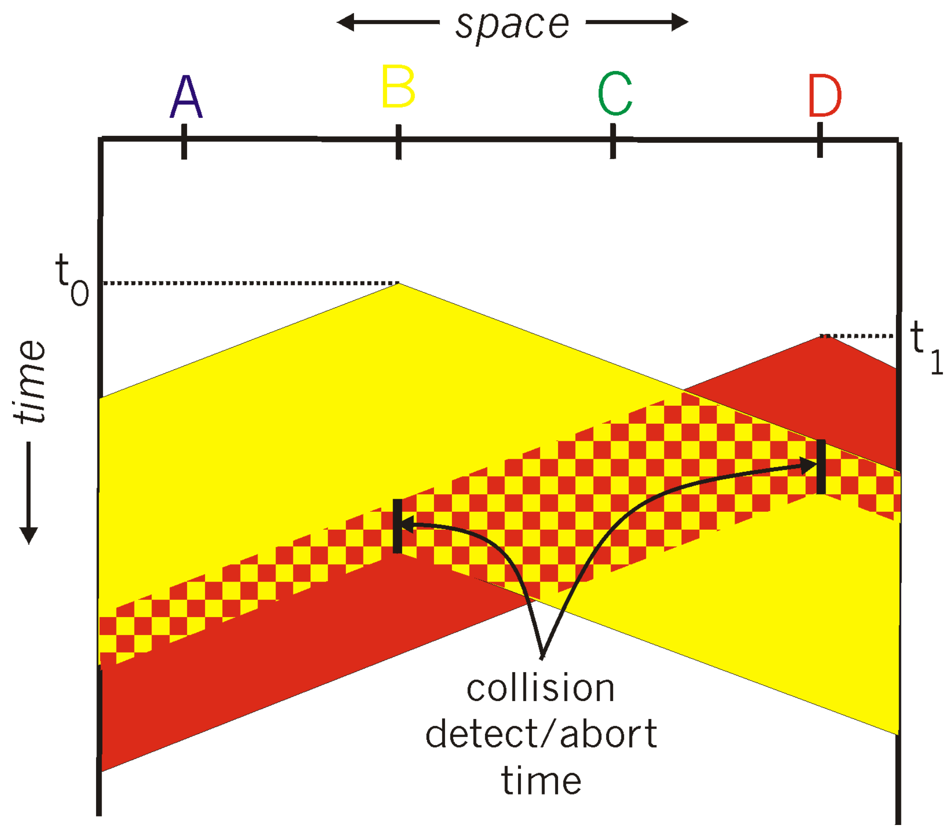
- distributed algorithm that determines how nodes share channel, i.e., determine when node can transmit
- communication about channel sharing uses channel itself, i.e. no out-of-band channel for coordination

MAC Protocols: a taxonomy

Three broad classes:

- Channel Partitioning
  - divide channel into smaller "pieces" (time slots, frequency, code)
  - allocate piece to node for exclusive use
- Random Access
  - channel not divided, allow collisions, "recover" from collisions
  - Examples of random access MAC protocols:
    - ALOHA, slotted ALOHA
    - CSMA, CSMA/CD, CSMA/CA
- "Taking turns"
  - nodes take turns, nodes with more to send can take longer turns
  - polling from central site, token passing
  - Bluetooth, FDDI, IBM Token Ring

CSMA/CD collision detection



Link Layer

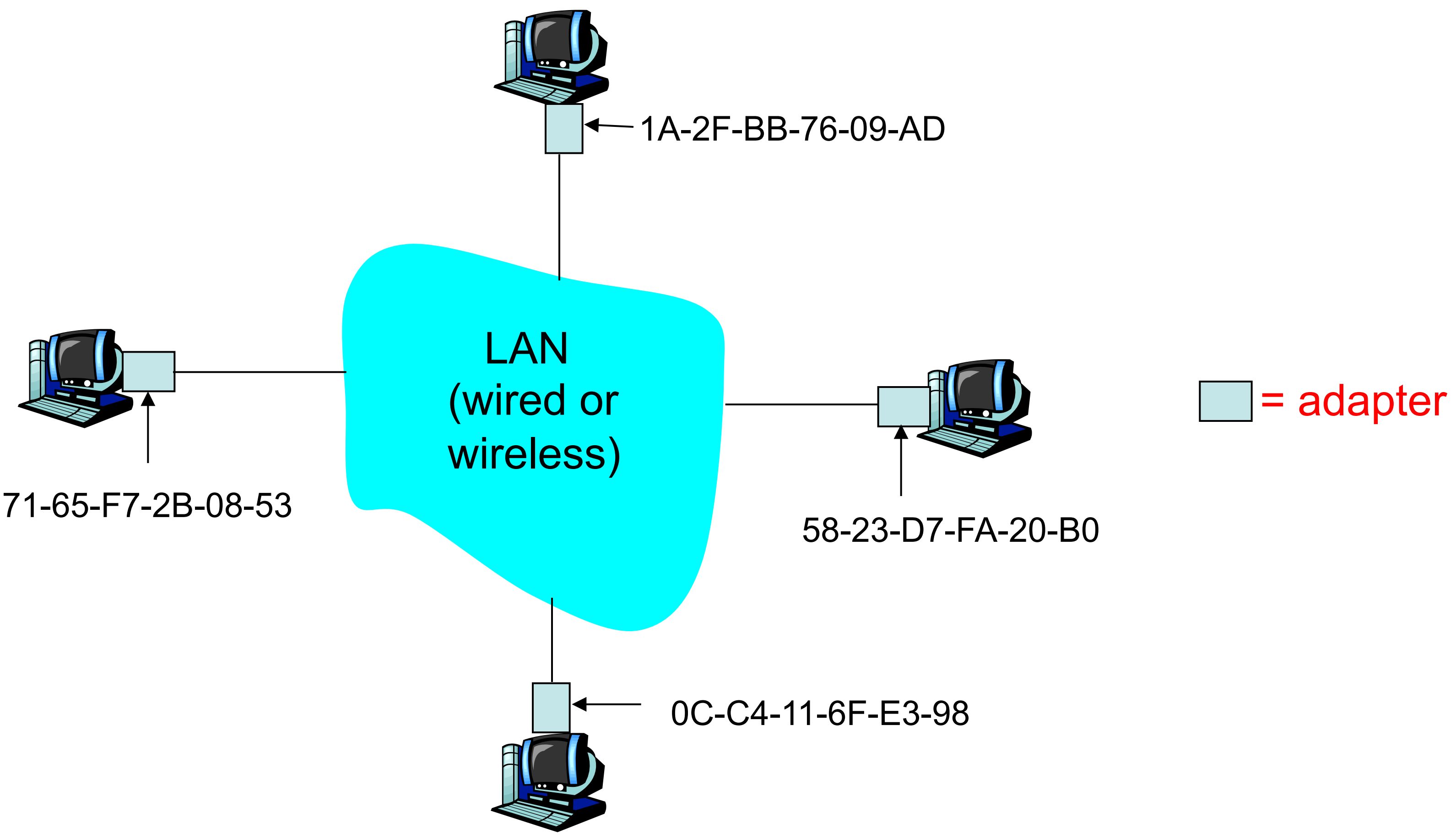
- Introduction and services
- Multiple access protocols
- Link-layer Addressing
- Ethernet
- Link-layer switches

MAC Addresses

- 32-bit IP address:
  - *network-layer* address
  - used to get datagram to destination IP subnet
- MAC (or LAN or physical or Ethernet) address:
  - function: *transmit frame from one interface to another physically-connected interface (same network)*
  - 48 bit MAC address (for most LANs)
    - burned in network adapter ROM, or software settable

LAN Addresses and ARP

- Each adapter on LAN has unique LAN address



LAN Addresses

- Transmission of addresses over the wire
  - Canonical form (also known as "LSB format" and "Ethernet format")
    First bit of each byte on the wire maps to
    least significant (i.e., right-most) bit of each byte in memory
    (c.f. RFC 2469)
  - Token Ring (IEEE 802.5) and FDDI (IEEE 802.6)
    do not use canonical form, but instead: most-significant bit first

- Human-friendly notation for MAC addresses: six groups of two hex digits, separated by “-”, in transmission order,
  e.g. 0C-C4-11-6F-E3-98

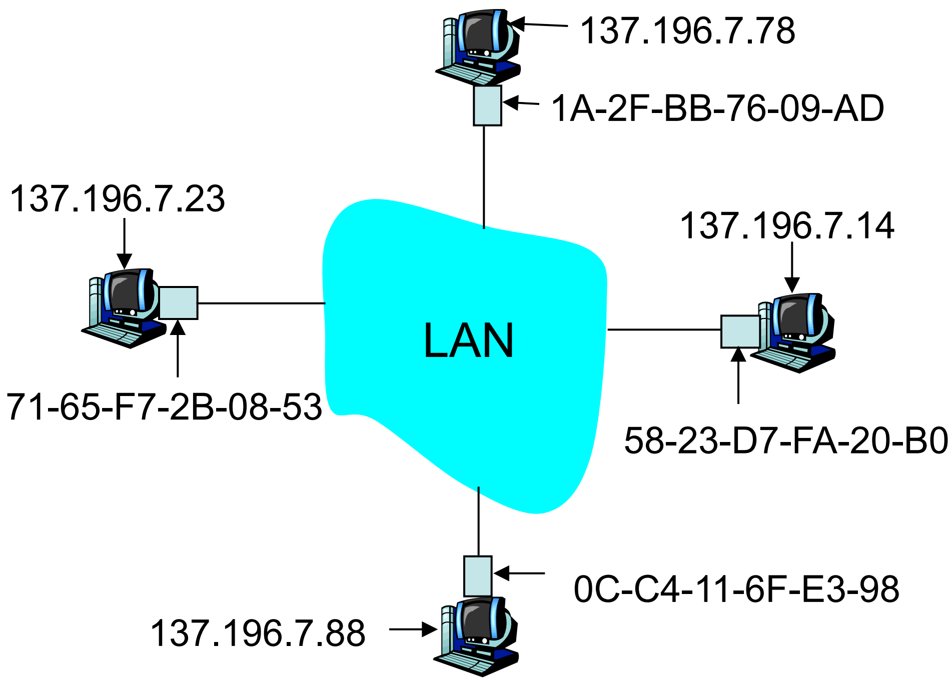
LAN Address

- MAC address allocation administered by IEEE
- Multicast and broadcast
  - Broadcast address: FF-FF-FF-FF-FF-FF
  - Multicast address: least-significant bit of first octet has value "1"
- Organisation Unique Identifier (OUI)
  - manufacturer buys portion of MAC address space (assuring uniqueness)
  - First 3 byte in transmission order
  - OUI enforced: 2nd least significant bit has value "0"

- MAC flat address ➜ portability
  - can move LAN card from one LAN to another
- IP hierarchical address NOT portable
  - address depends on IP subnet to which node is attached

ARP: Address Resolution Protocol

<u>Question:</u> how to determine MAC address of an I/F, knowing an IP address?



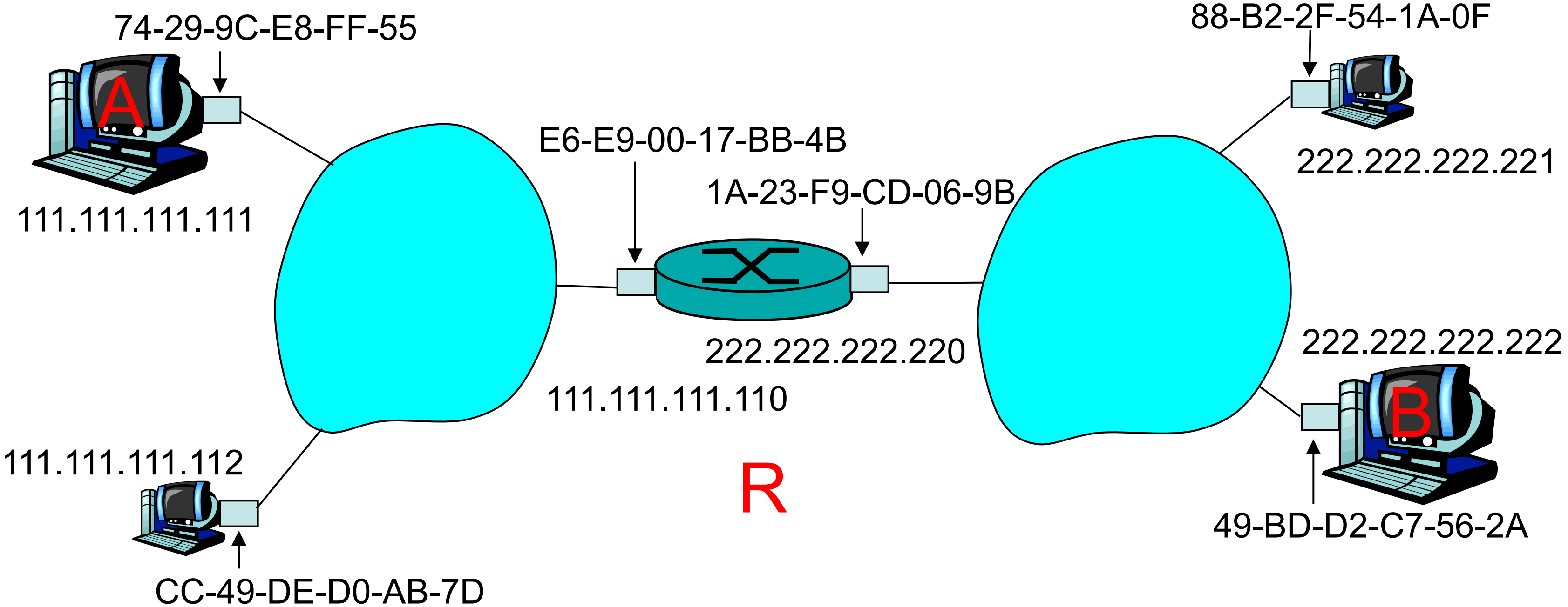
- Each IP node (host, router) on LAN has ARP (Address Resolution Protocol) table
- ARP table: IP/MAC address mappings for some LAN nodes
- <IP addr; MAC addr; TTL>
  - TTL (Time To Live): time after which address mapping will be forgotten (typically 20 min)

ARP protocol: Same LAN (network)

- A wants to send datagram to B, and B's MAC address not in A's ARP table.
- A broadcasts ARP query packet, containing B's IP address
  - dest MAC address = FF-FF-FF-FF-FF-FF
  - all machines on LAN receive ARP query
- B receives ARP packet, replies to A with its (B's) MAC address
  - frame sent to A's MAC address (unicast)
- A caches IP-to-MAC address pair in its ARP table until information times out
  - ***soft state***: information that times out (goes away) unless refreshed
- ARP is "plug-and-play":
  - nodes create their ARP tables without intervention from network administrator

Addressing: routing to another LAN

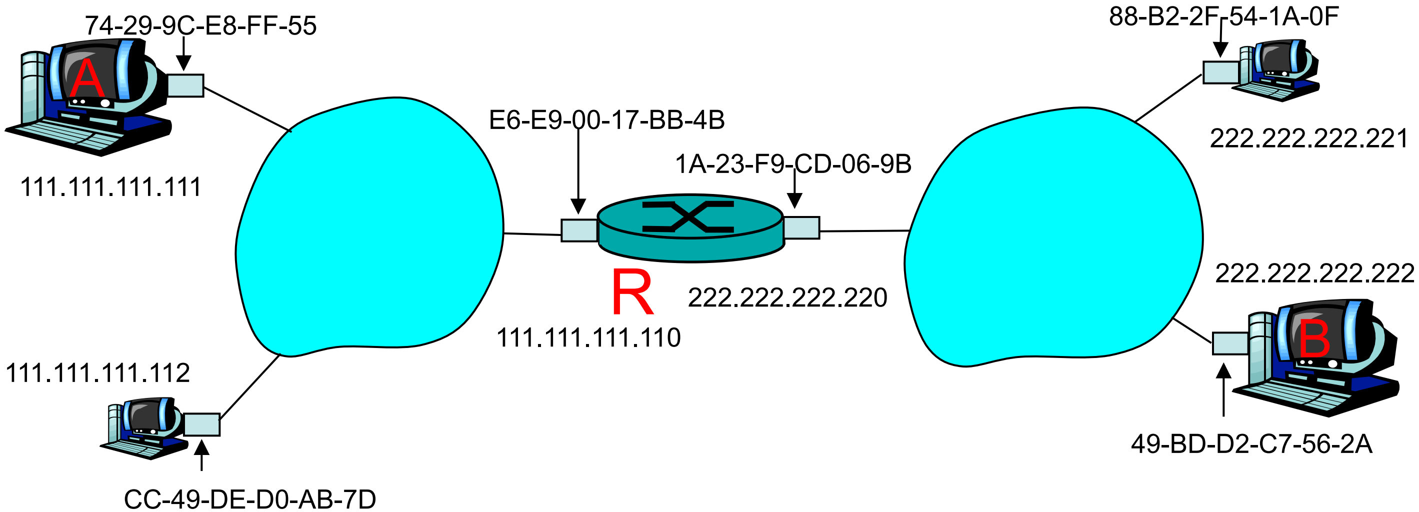
- example: send datagram from A to B via R
  (assumption: A knows B's IP address)



- two ARP tables in router R, one for each IP network (LAN)

Addressing: routing to another LAN (2)

- A creates IP datagram with source A, destination B
- A uses ARP to get R’s MAC address for 111.111.111.110
- A creates link-layer frame with R's MAC address as dest, frame contains A-to-B IP datagram
- A’s NIC sends frame
- R’s NIC receives frame
- R removes IP datagram from Ethernet frame, sees its destined to B
- R uses ARP to get B’s MAC address
- R creates frame containing A-to-B IP datagram sends to B

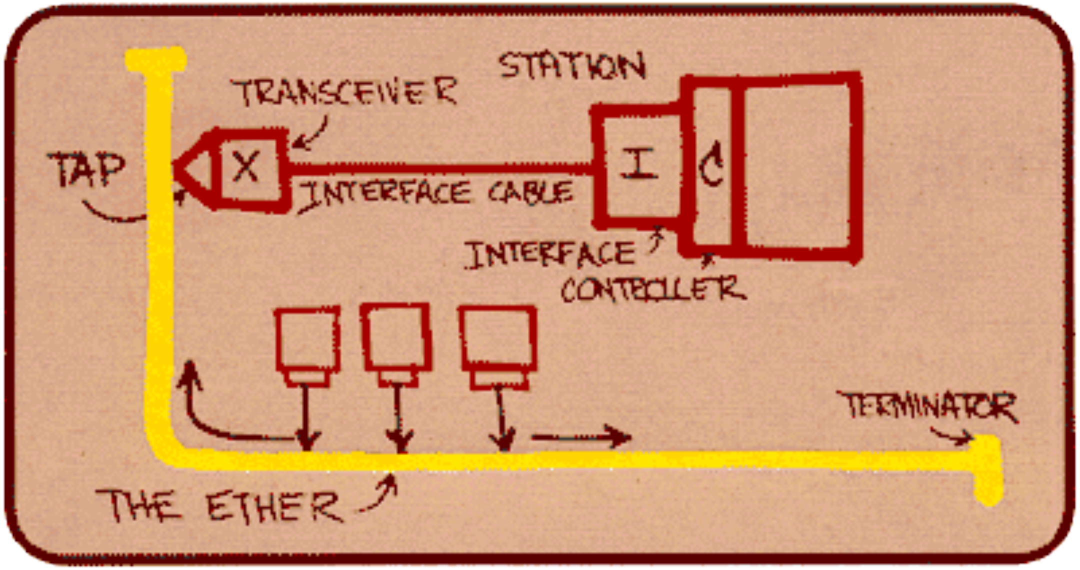


Link Layer

- Introduction and services
- Multiple access protocols
- Link-layer Addressing
- Ethernet
- Link-layer switches

Ethernet

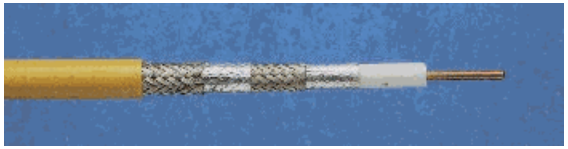
- "dominant" wired LAN technology:
- cheap $20 for NIC
- first widely used LAN technology
- simpler, cheaper than token LANs and ATM
- kept up with speed race: 10 Mbps – 10 Gbps

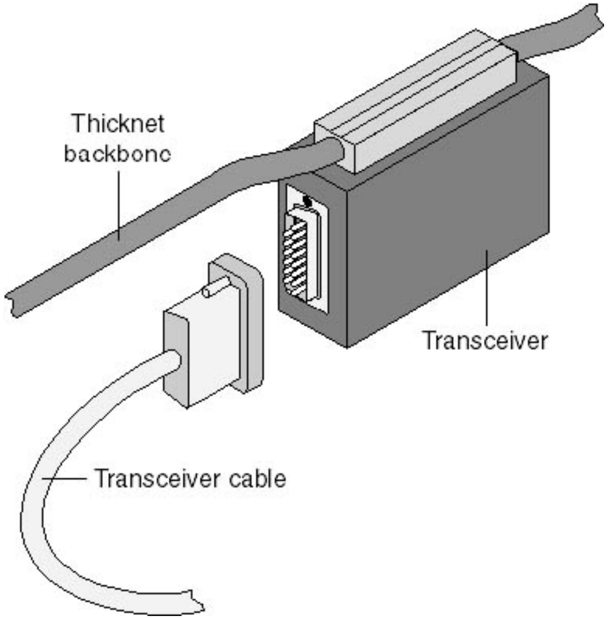


Metcalfe's Ethernet sketch 1976

10Base5 - Thick Ethernet

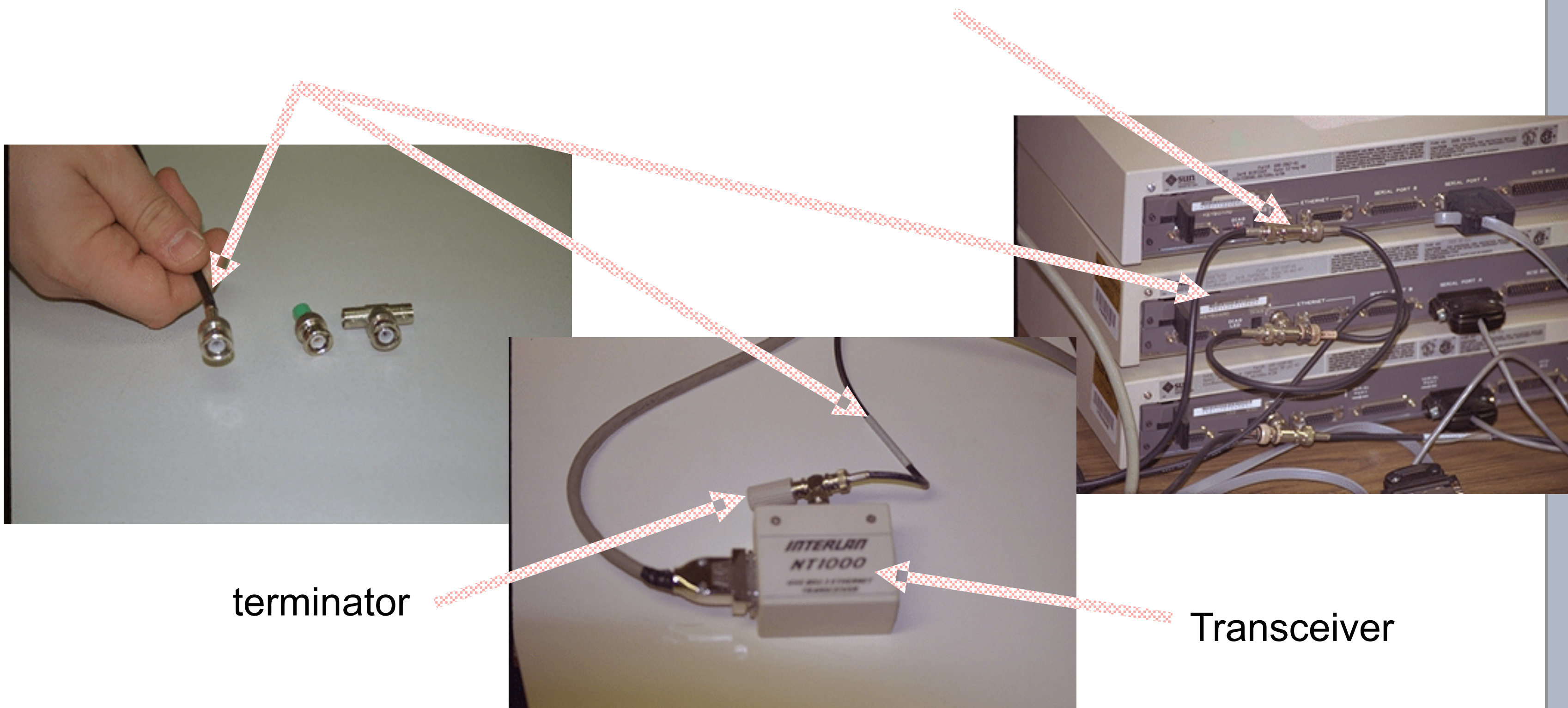
- thick coax cable (yellow)
- 10Base5: 10 Mbit/s,
- Segments of 500 m, can be coupled with repeaters (max. 5 segments)
- Transceiver (Transmitter & Receiver) MAU (Medium Attachement Unit) with „Carrier Sensing“ function
- Transceiver cable max. 50 m





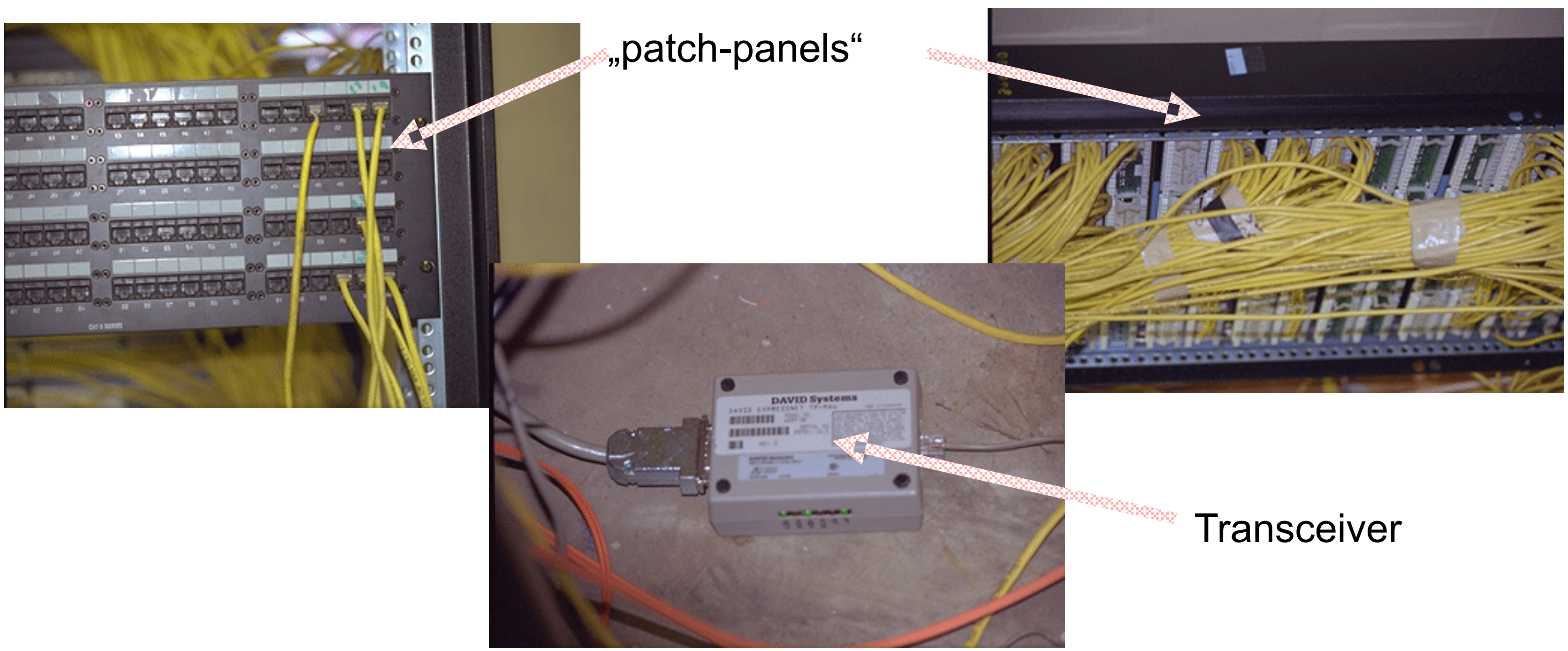
10Base2 - Thin Wire Ethernet

- 10 MBit/s, segments of max. 185 m
- Transceiver can be part of ethernet adapter



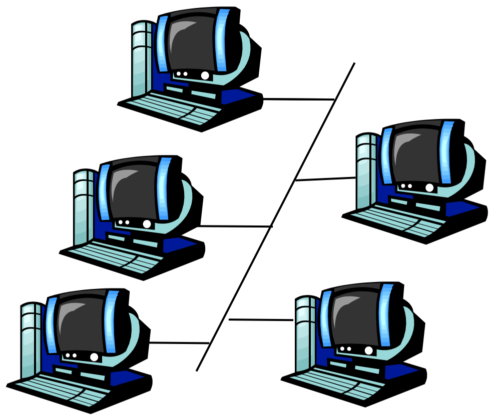
10Base-T - Twisted Pair

- 10 MBit/s
- max. 100 m point-to-point connection to multiport repeater

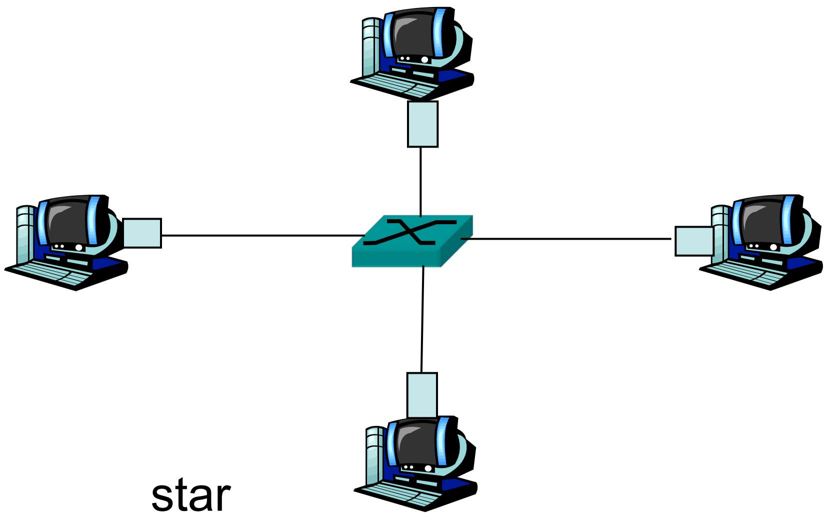


Star topology

- Originally: logical bus topology
  - physically: bus or star
  - nodes in collision domain
- today: star topology prevails
  - active **switch** in center
  - each "spoke" runs a (separate) Ethernet protocol (nodes do not collide with each other)



bus



star

Ethernet CSMA/CD algorithm

1. NIC receives datagram from network layer, creates frame
2. If NIC senses channel idle, starts frame transmission If NIC senses channel busy, waits until channel idle, then transmits
3. If NIC transmits entire frame without detecting another transmission, NIC is done with frame !
4. If NIC detects another transmission while transmitting, aborts and sends jam signal
5. After aborting, NIC enters **exponential backoff**: after *m*th collision, NIC chooses *K* at random from $\{0,1,2,\dots,2^m-1\}$. NIC waits K·512 bit times, returns to Step 2

Ethernet's CSMA/CD (more)

Jam Signal: make sure all other transmitters are aware of collision; 48 bits

Bit time: 0.1 microsec for 10 Mbps Ethernet ;
for K=1023, wait time is about 50 msec

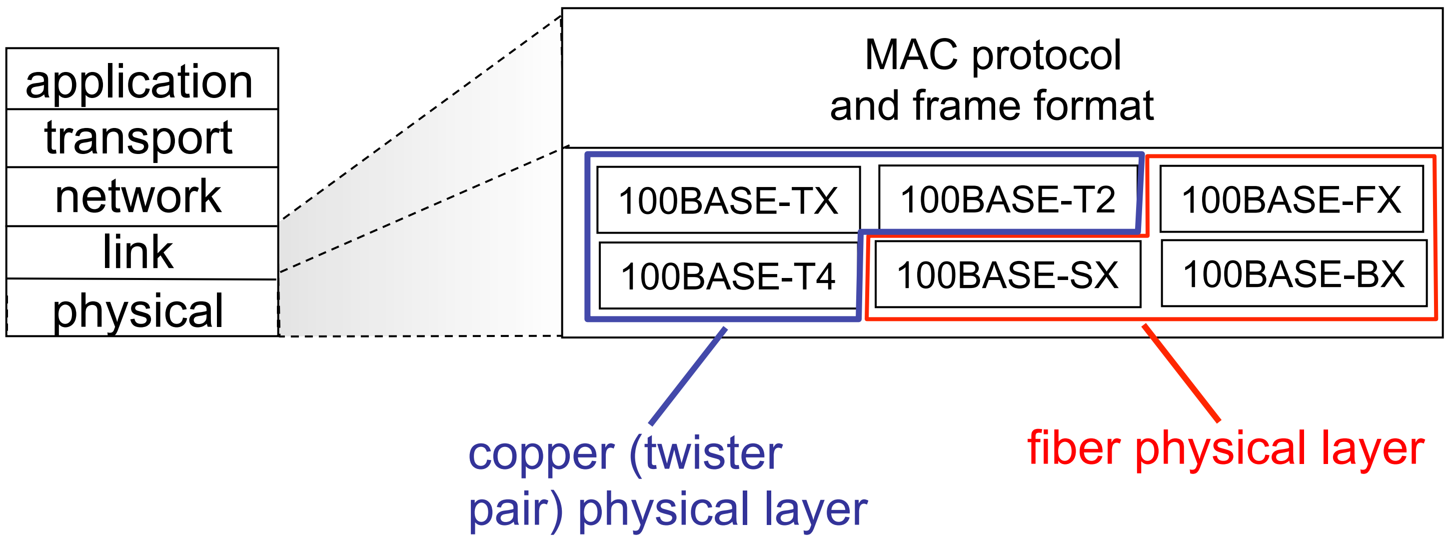
Exponential Backoff:

- *Goal*: adapt retransmission attempts to estimated current load
  - heavy load: random wait will be longer
- first collision: choose K from {0,1}; delay is K· 512 bit transmission times
- after second collision: choose K from {0,1,2,3}…
- after ten collisions, choose K from {0,1,2,3,4,…,1023}

See/interact with Java applet on AW Web site:
http://wps.aw.com/aw_kurose_network_5/
⇨student resources - recommended !

802.3 Ethernet Standards: Link & Physical Layers

- many different Ethernet standards
  - common MAC protocol and frame format
  - different speeds: 10 Mbps, 100 Mbps, 1 Gbps, 10 Gbps
  - different physical layer media: fiber, cable

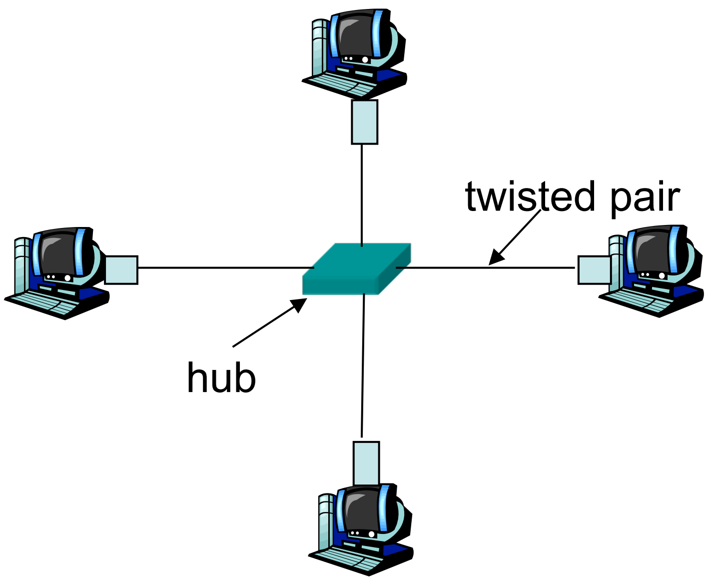


Link Layer

- Introduction and services
- Multiple access protocols
- Link-layer Addressing
- Ethernet
- Link-layer switches

Hubs

- … physical-layer ("dumb") repeaters:
  - bits coming in one link go out all other links at same rate
  - all nodes connected to hub can collide with one another
  - no frame buffering
  - no CSMA/CD at hub: host NICs detect collisions

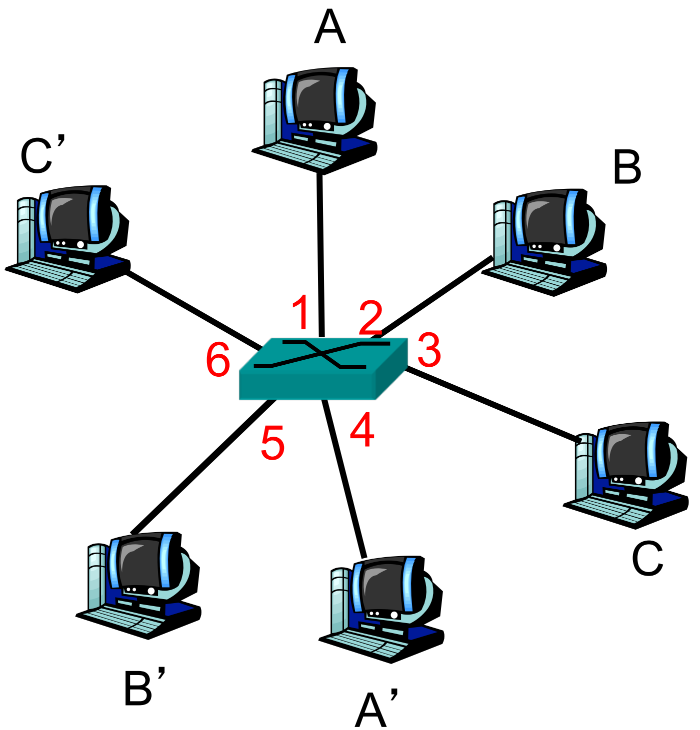


Switch

- link-layer device: smarter than hubs, take *active* role
  - store, forward Ethernet frames
  - examine incoming frame's MAC address, selectively forward frame to one-or-more outgoing links when frame is to be forwarded on segment, uses CSMA/CD to access segment
- *transparent*
  - hosts are unaware of presence of switches
- *plug-and-play, self-learning*
  - switches do not need to be configured

Switch: allows multiple simultaneous transmissions

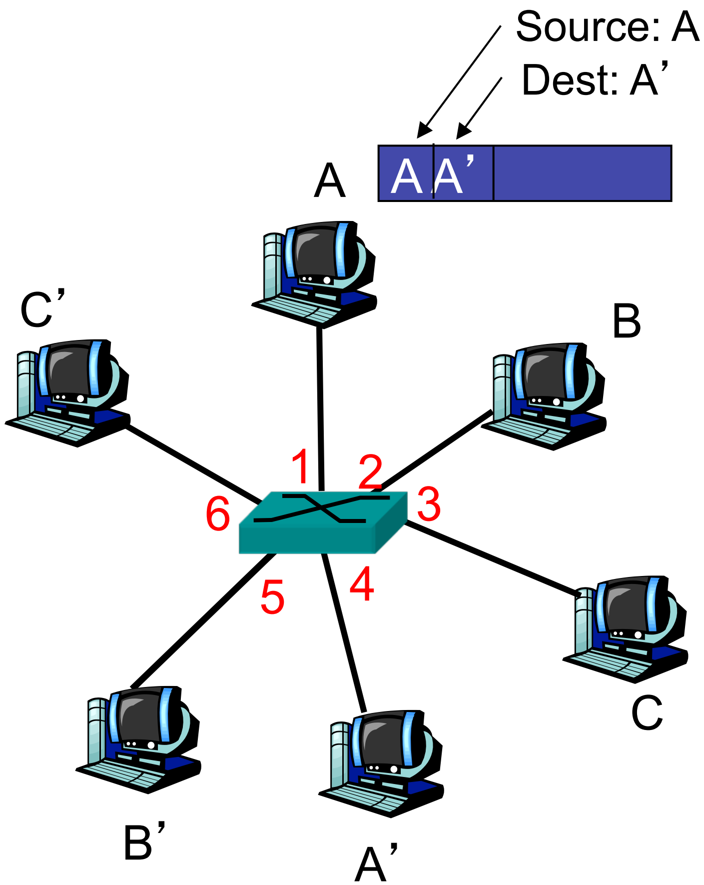
- hosts have dedicated, direct connection to switch
- switches buffer packets
- Ethernet protocol used on *each* incoming link, but no collisions; full duplex
  - each link is its own collision domain
- *switching:* A-to-A' and B-to-B' simultaneously, without collisions
  - not possible with dumb hub



switch with six interfaces (1,2,3,4,5,6)

Switch: self-learning

- switch *learns* which hosts can be reached through which interfaces
  - when frame received, switch "learns" location of sender: incoming LAN segment
  - records sender/location pair in switch table



MAC addr	interface	TTL
A	*1*	*60*

Switch table (initially empty)

Switch: frame filtering/forwarding

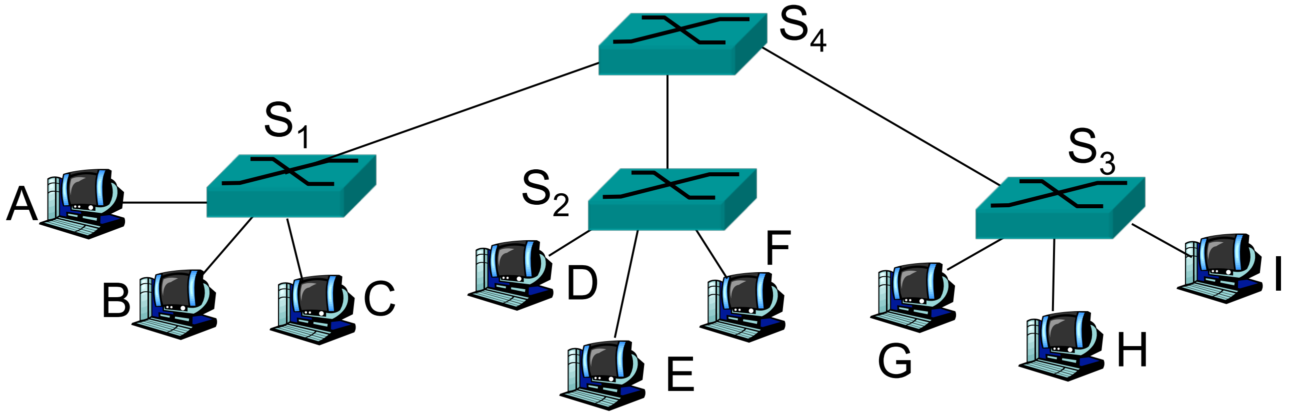
When frame received:

1. record link associated with sending host
2. index switch table using MAC dest address
3. **if** entry found for destination
   then {
   if dest on segment from which frame arrived
   then drop the frame
   else forward the frame on interface indicated
   }
   else flood

forward on all but the interface on which the frame arrived

Interconnecting switches

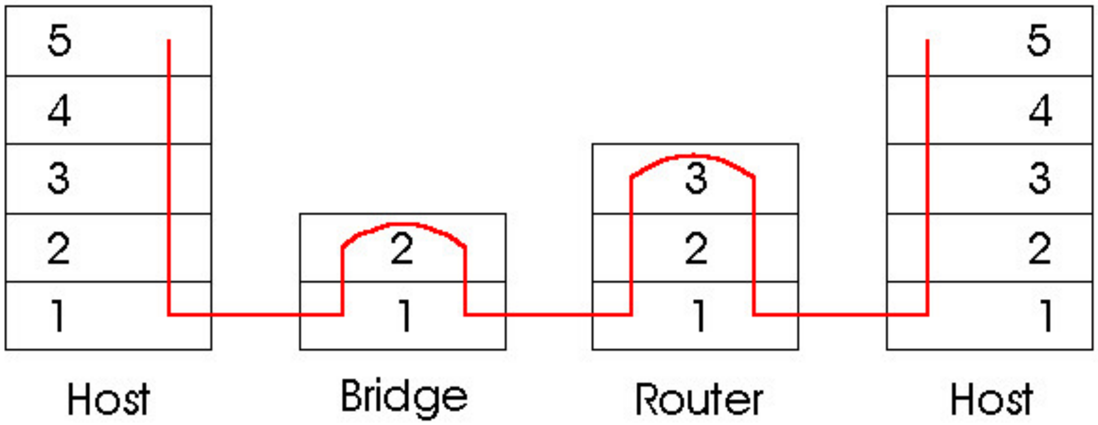
- switches can be connected together



- *Q:* sending from A to G - how does S_1 know to forward frame destined to F via S_4 and S_3?
- *A:* self learning! (works exactly the same as in single-switch case!)

Switches vs. Routers

- both store-and-forward devices
  - routers: network layer devices (examine network layer headers)
  - switches are link layer devices
- routers maintain routing tables, implement routing algorithms
- switches maintain switch tables, implement filtering, learning algorithms



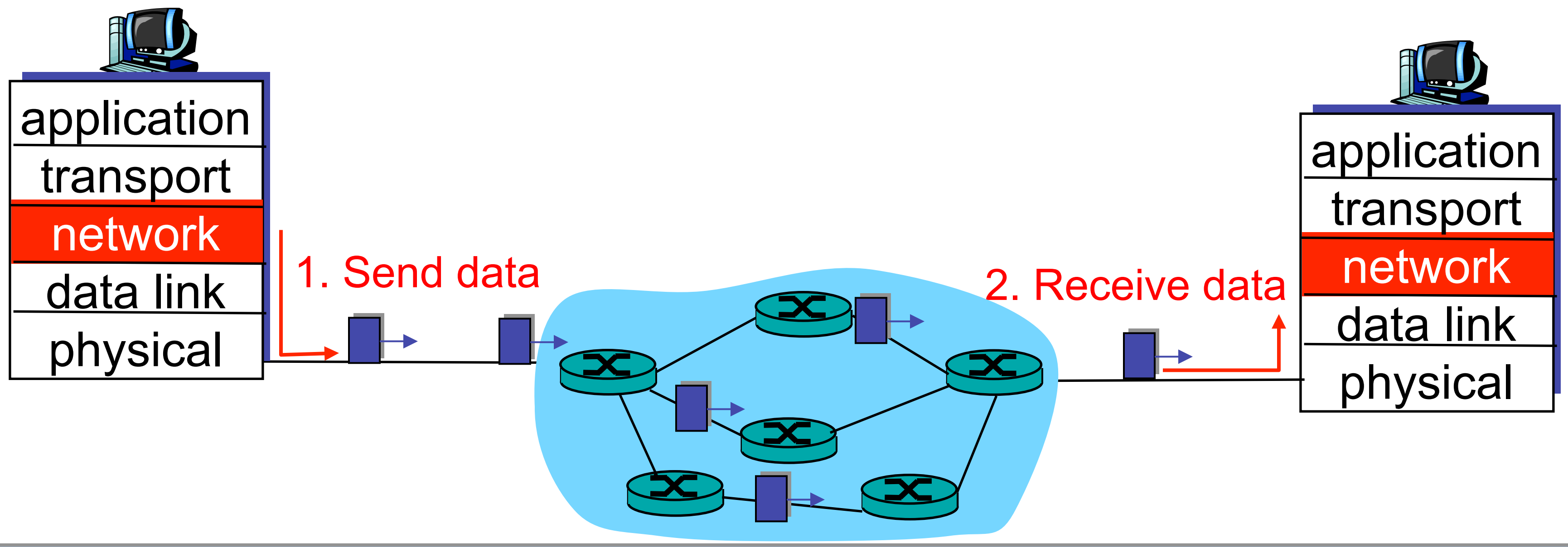
Link Layer

- Introduction and services
- Multiple access protocols
- Link-layer Addressing
- Ethernet
- Link-layer switches

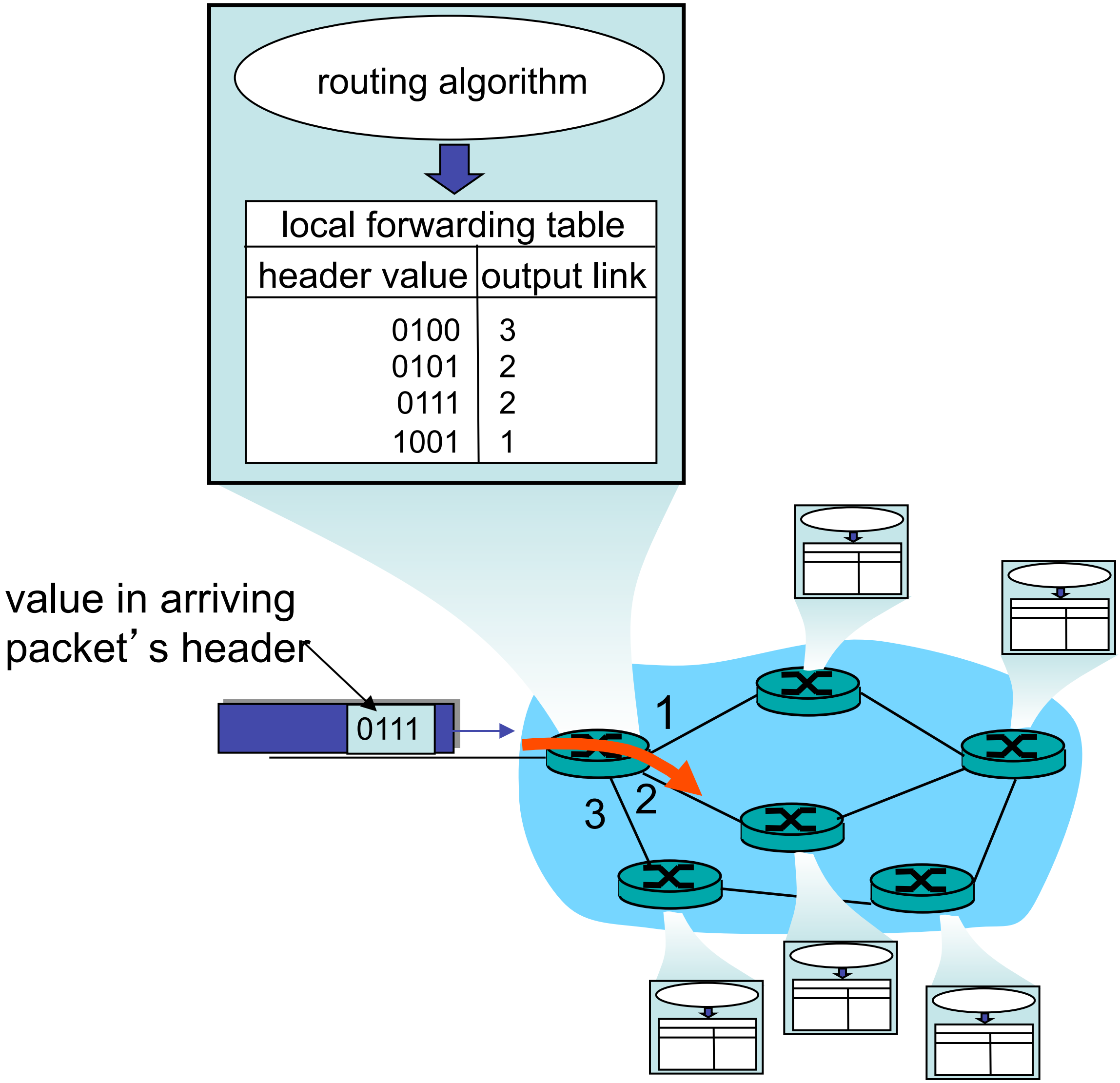
Internet Protocol

Datagram networks

- no call setup at network layer
- routers: no state about end-to-end connections
  - no network-level concept of "connection"
- packets forwarded using destination host address
  - packets between same source-dest pair may take different paths



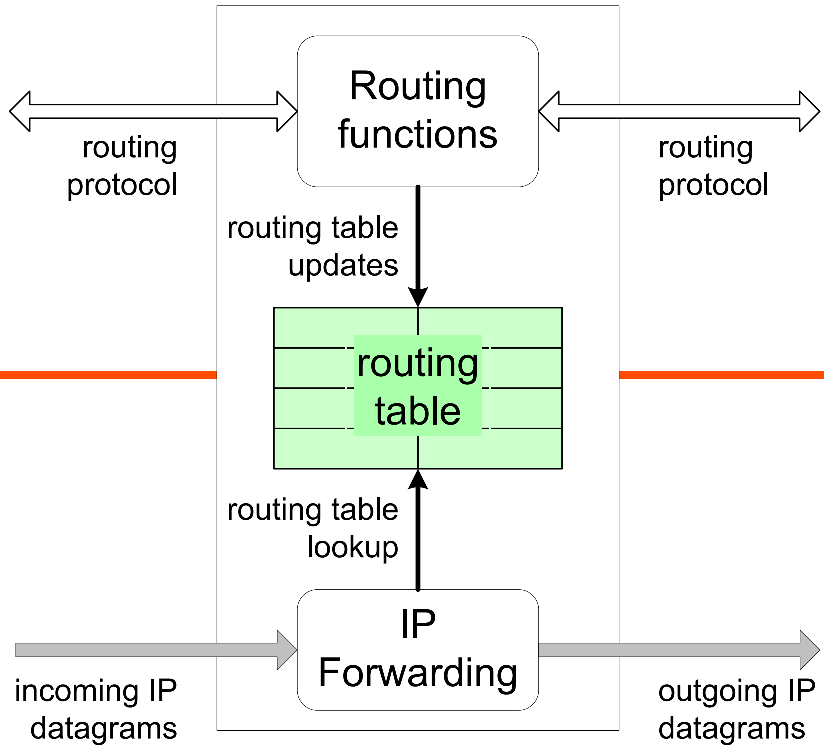
Interplay between routing and forwarding



Routers: Forwarding and Routing

- Forwarding: data plane
  - Directing a data packet to an outgoing link
  - Individual router using a forwarding table
- Routing: control plane
  - Computing the paths the packets will follow
  - Routers talking amongst themselves
  - Individual router creating a forwarding table

Functional Components



Control

Datapath: per-packet processing

Routing and Forwarding

Routing functions include:

- route calculation
- maintenance of the routing table
- execution of routing protocols

- On commercial routers handled by a single general purpose processor, called *route processor*

IP forwarding is per-packet processing

- On high-end commercial routers, IP forwarding is distributed (Most work is done on the interface cards)

Forwarding Table

2^{32} (~4 billion) possible entries

Destination Address Range	Link Interface
11001000 00010111 00010000 00000000 through 11001000 00010111 00010111 11111111	0
11001000 00010111 00011000 00000000 through 11001000 00010111 00011000 11111111	1
11001000 00010111 00011001 00000000 through 11001000 00010111 00011111 11111111	2
otherwise	3

Longest prefix matching

Prefix Match	Link Interface
11001000 00010111 00010	0
11001000 00010111 00011000	1
11001000 00010111 00011	2
otherwise	3

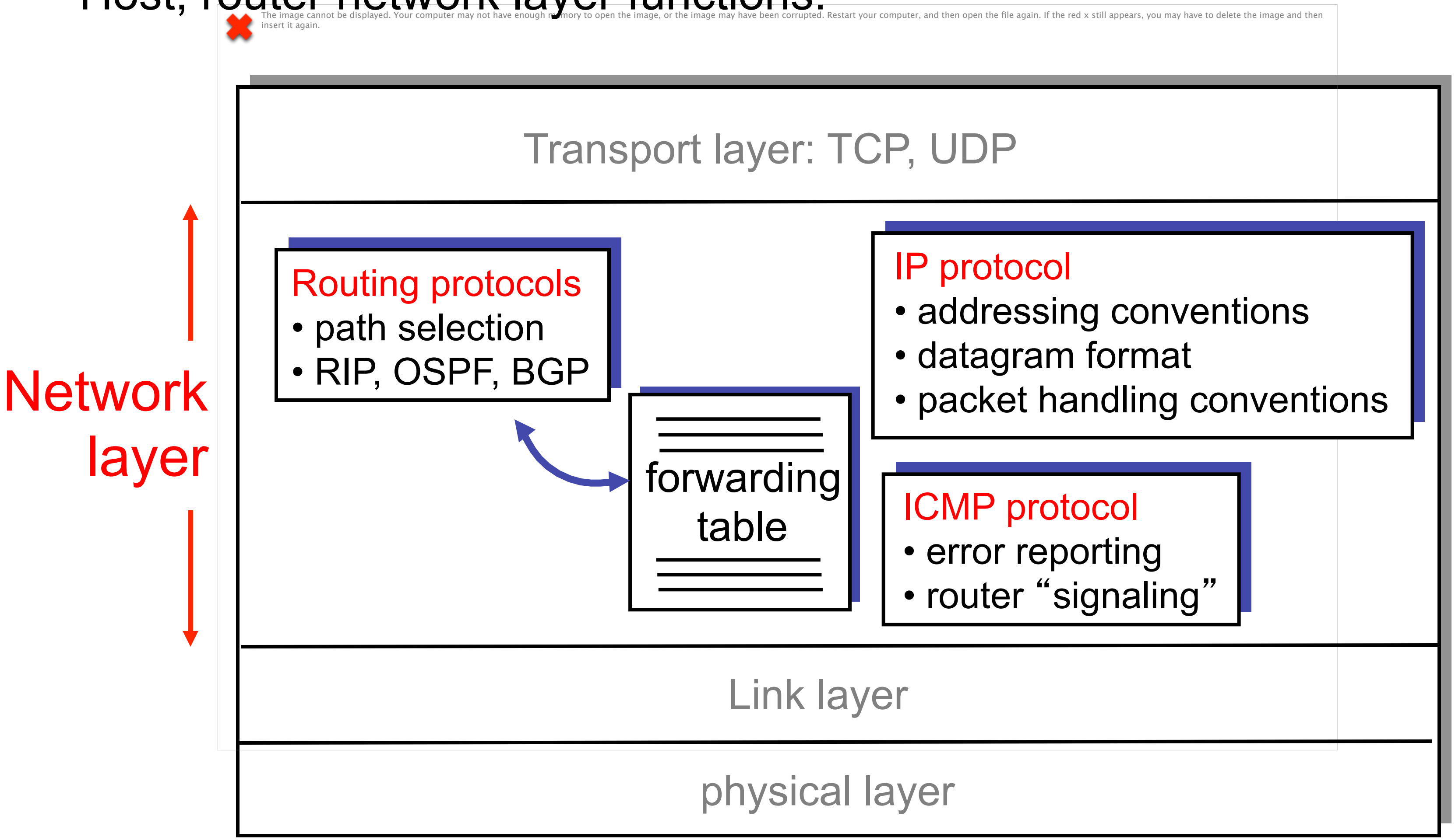
Examples

DA: 11001000 00010111 00010110 10100001 Which interface?

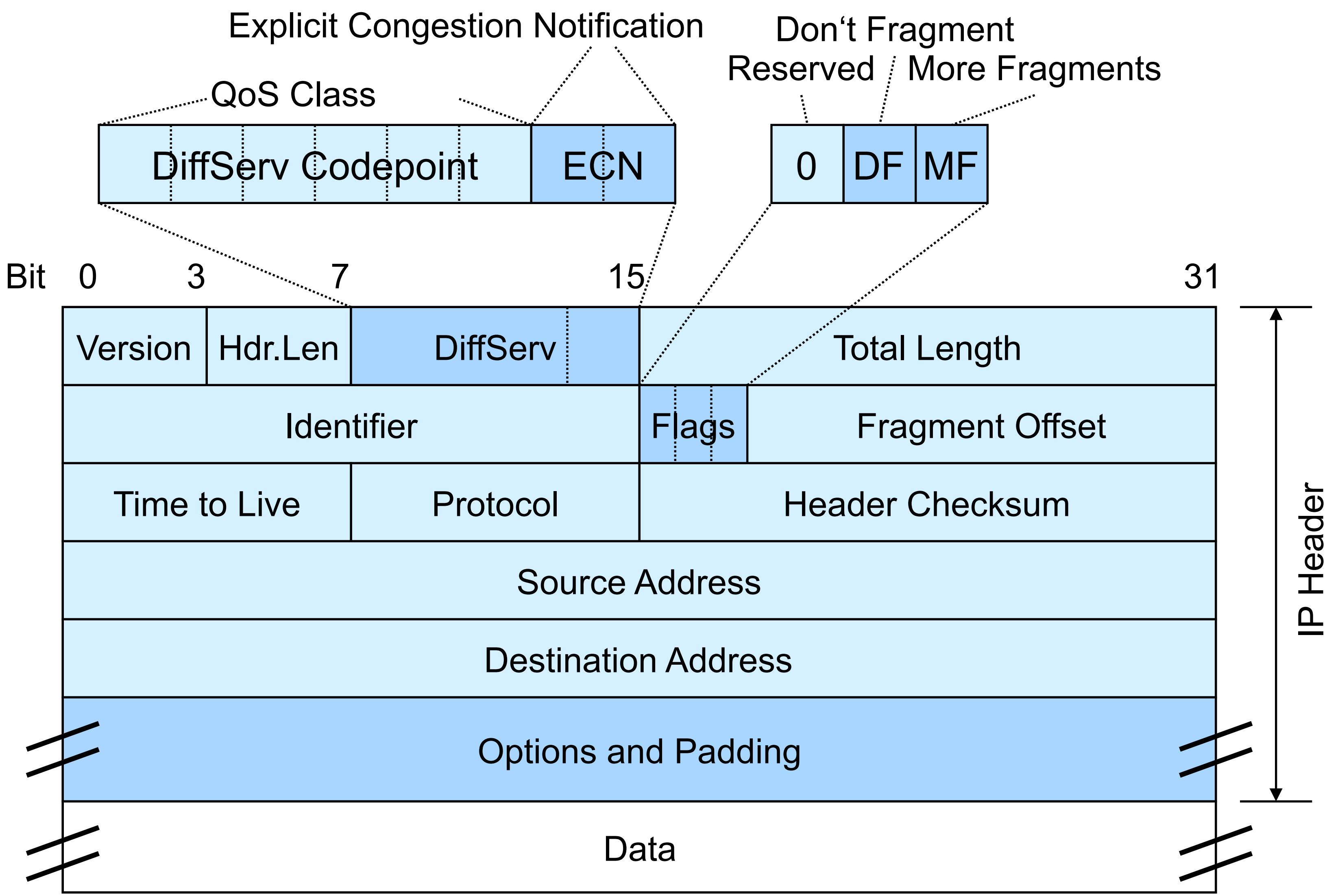
DA: 11001000 00010111 00011000 10101010 Which interface?

The Internet Network layer

Host, router network layer functions:



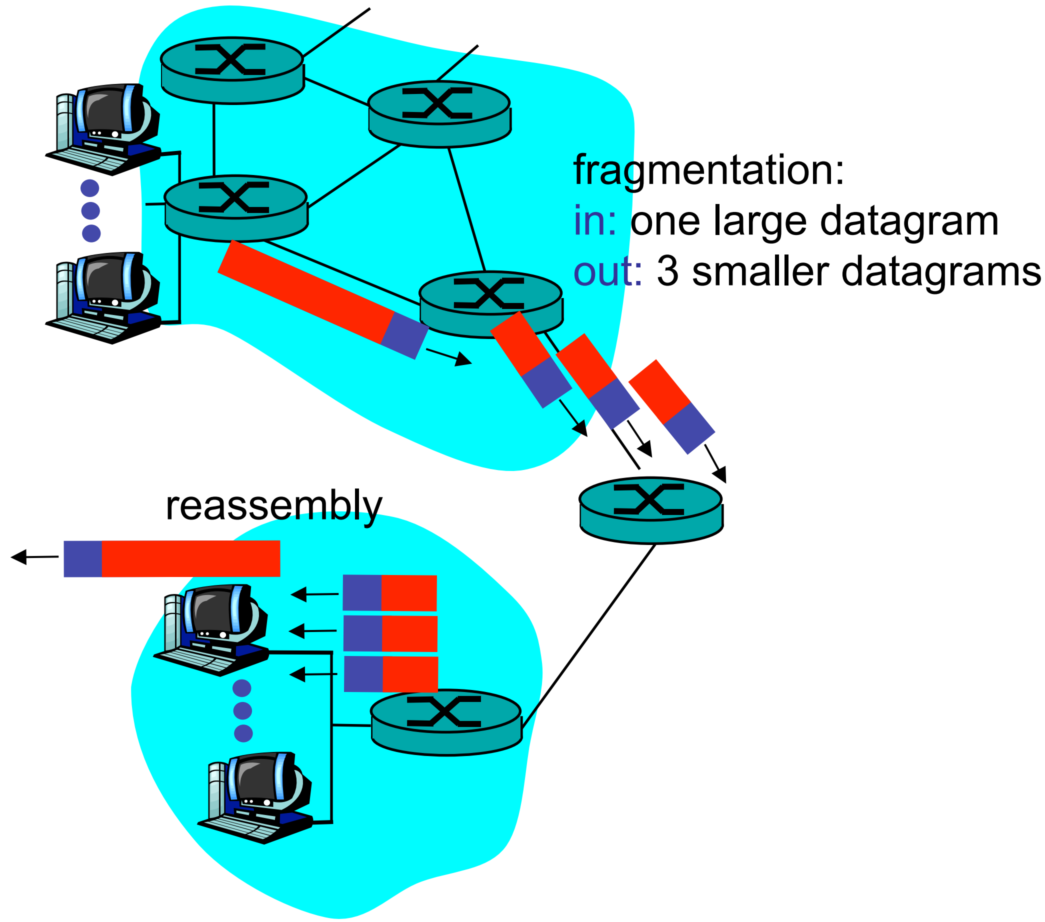
IP Datagram



Explicit Congestion Notification
QoS Class
Don‘t Fragment
Reserved
More Fragments
DiffServ Codepoint
ECN
0
DF
MF
Bit
0
3
7
15
31
Version
Hdr.Len
DiffServ
Total Length
Identifier
Flags
Fragment Offset
Time to Live
Protocol
Header Checksum
Source Address
Destination Address
Options and Padding
Data
IP Header

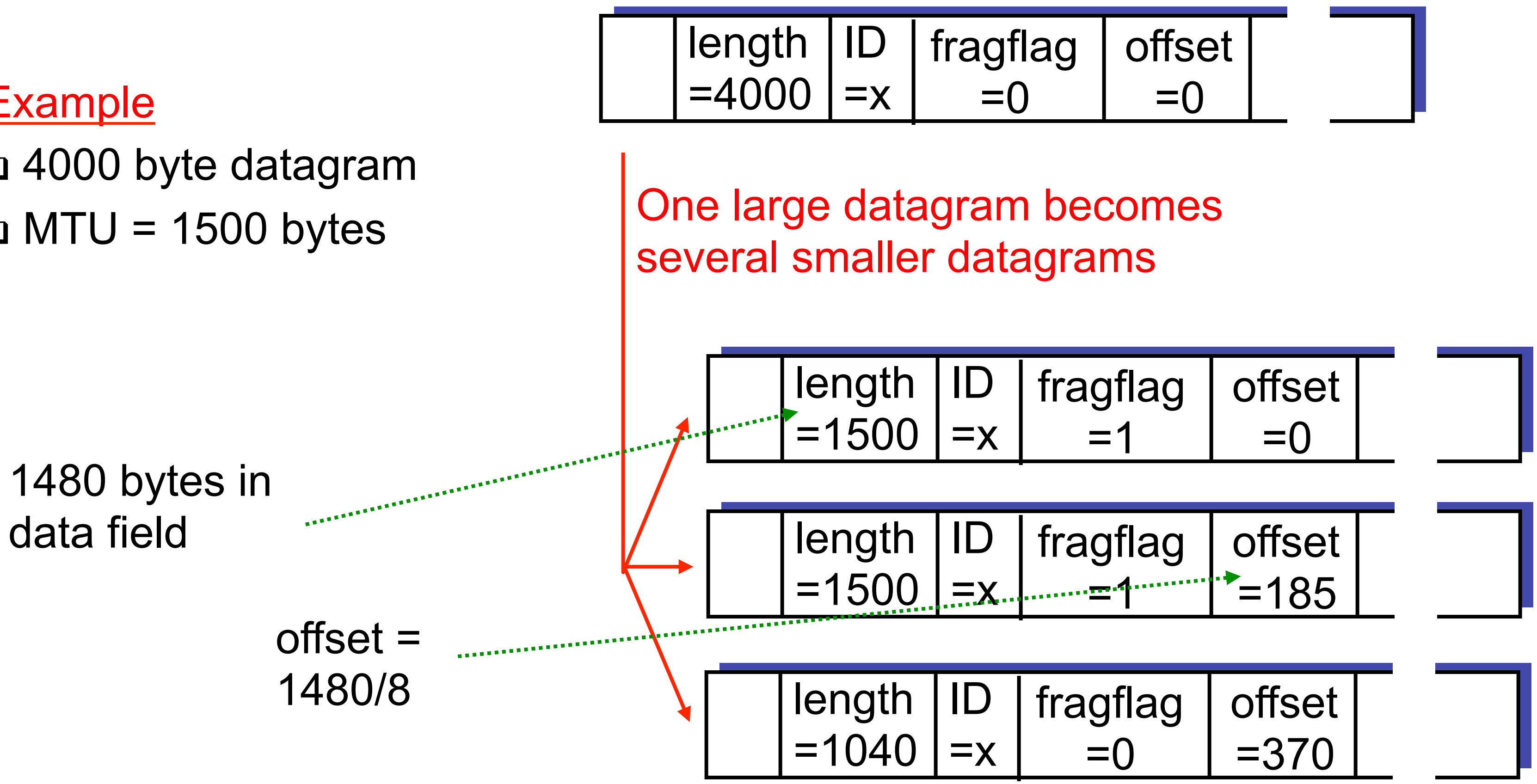
IP Fragmentation & Reassembly

- network links have MTU (max.transfer size) - largest possible link-level frame.
  - different link types, different MTUs
- large IP datagram divided ("fragmented") within net
  - one datagram becomes several datagrams
  - "reassembled" only at final destination
  - IP header bits used to identify, order related fragments



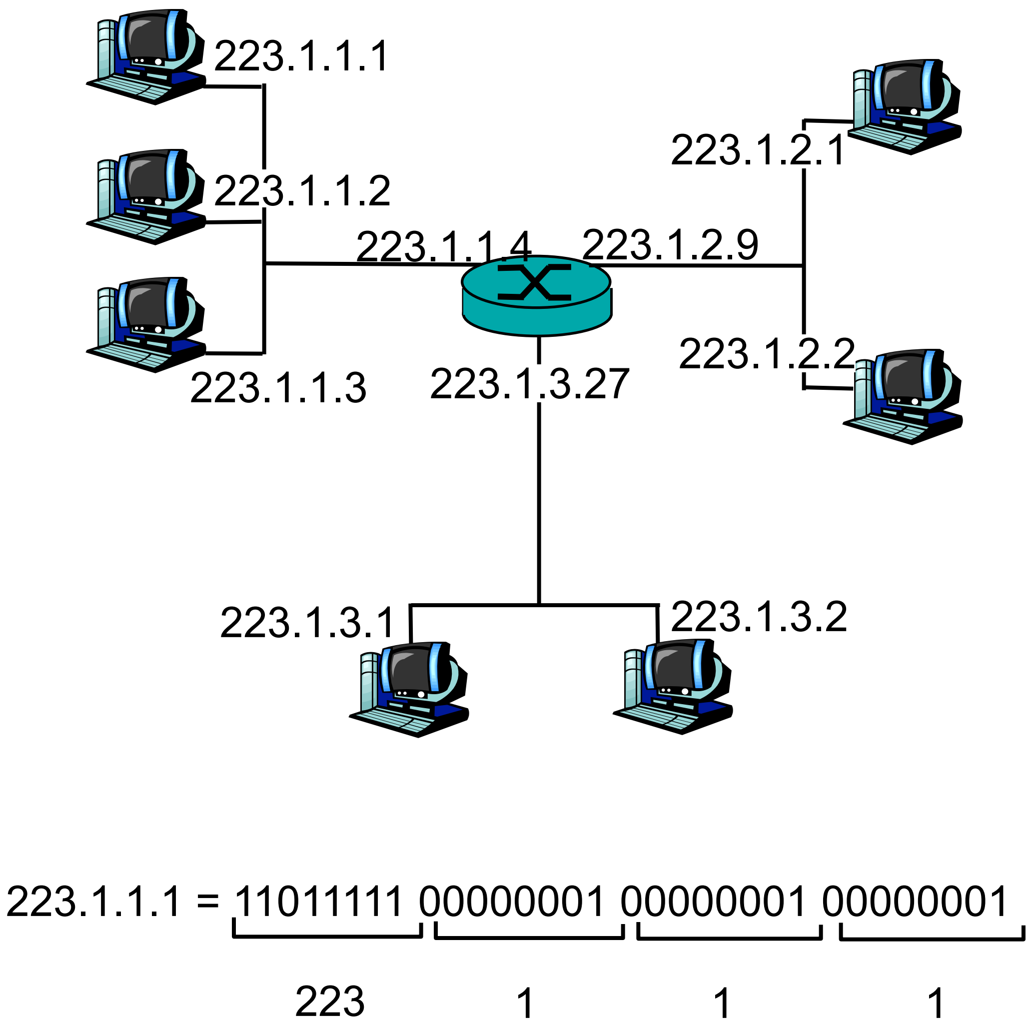
IP Fragmentation and Reassembly

Example
- 4000 byte datagram
- MTU = 1500 bytes



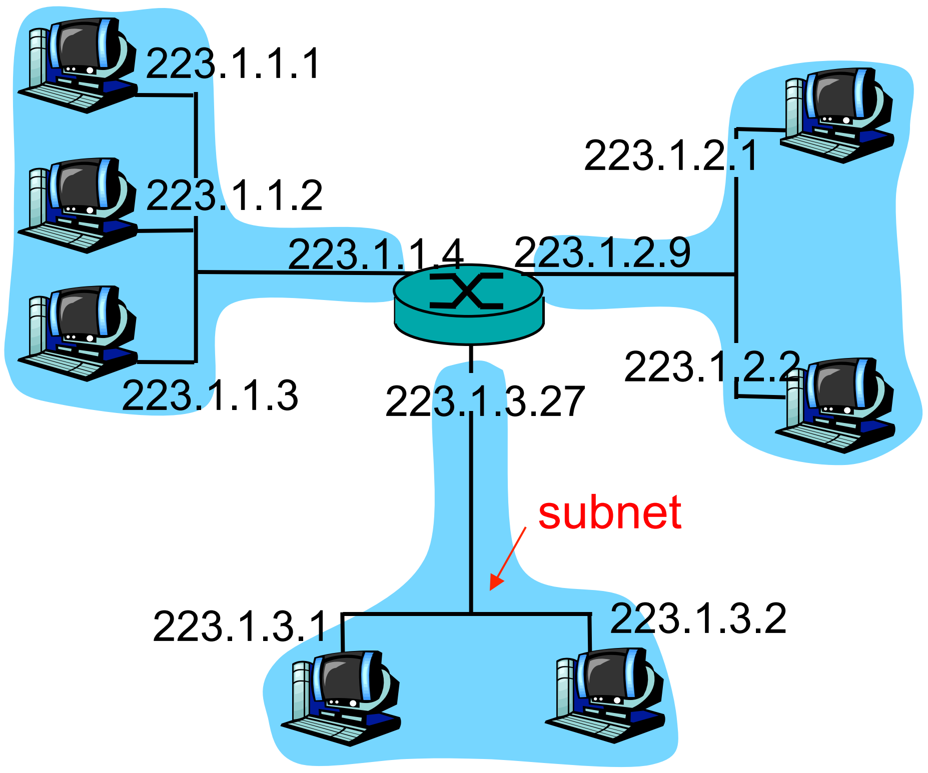
IP Addressing

- IP address: 32-bit identifier for host, router *interface*
- *interface:* connection between host/router and physical link
  - IP addresses associated with each interface



Subnets

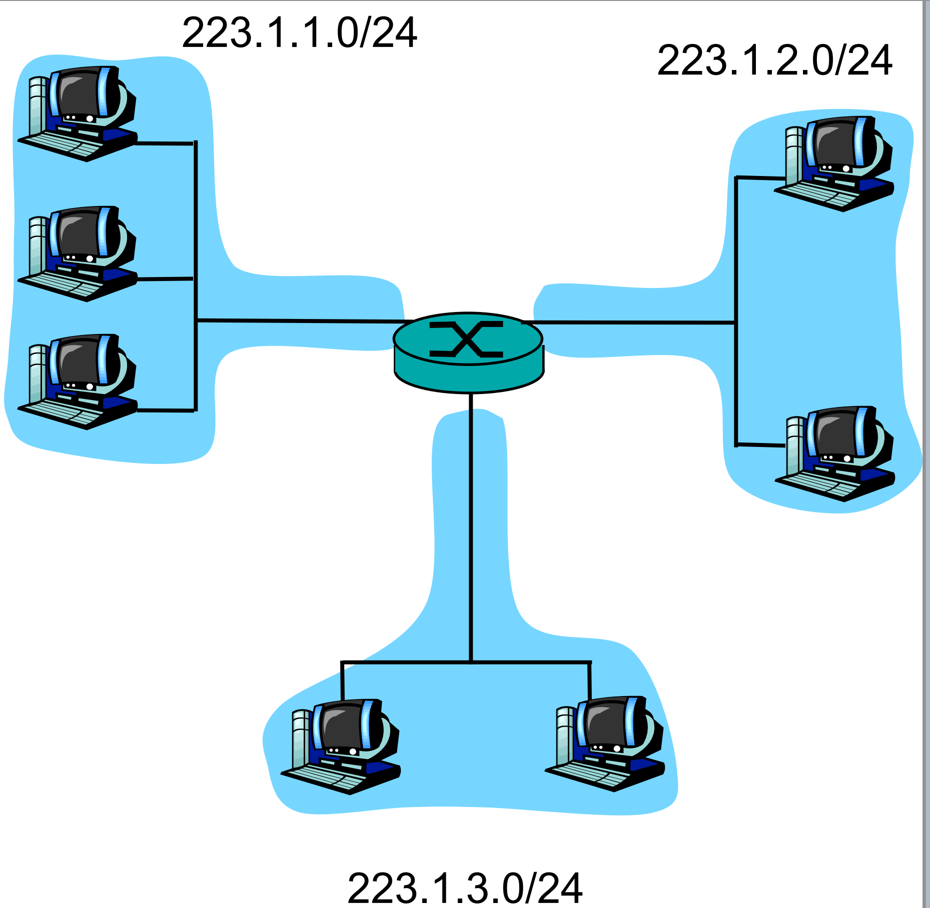
- IP address:
  - subnet part (high order bits)
  - host part (low order bits)
- *What's a subnet ?*
  - device interfaces with same subnet part of IP address
  - can physically reach each other without intervening router



network with 3 subnets

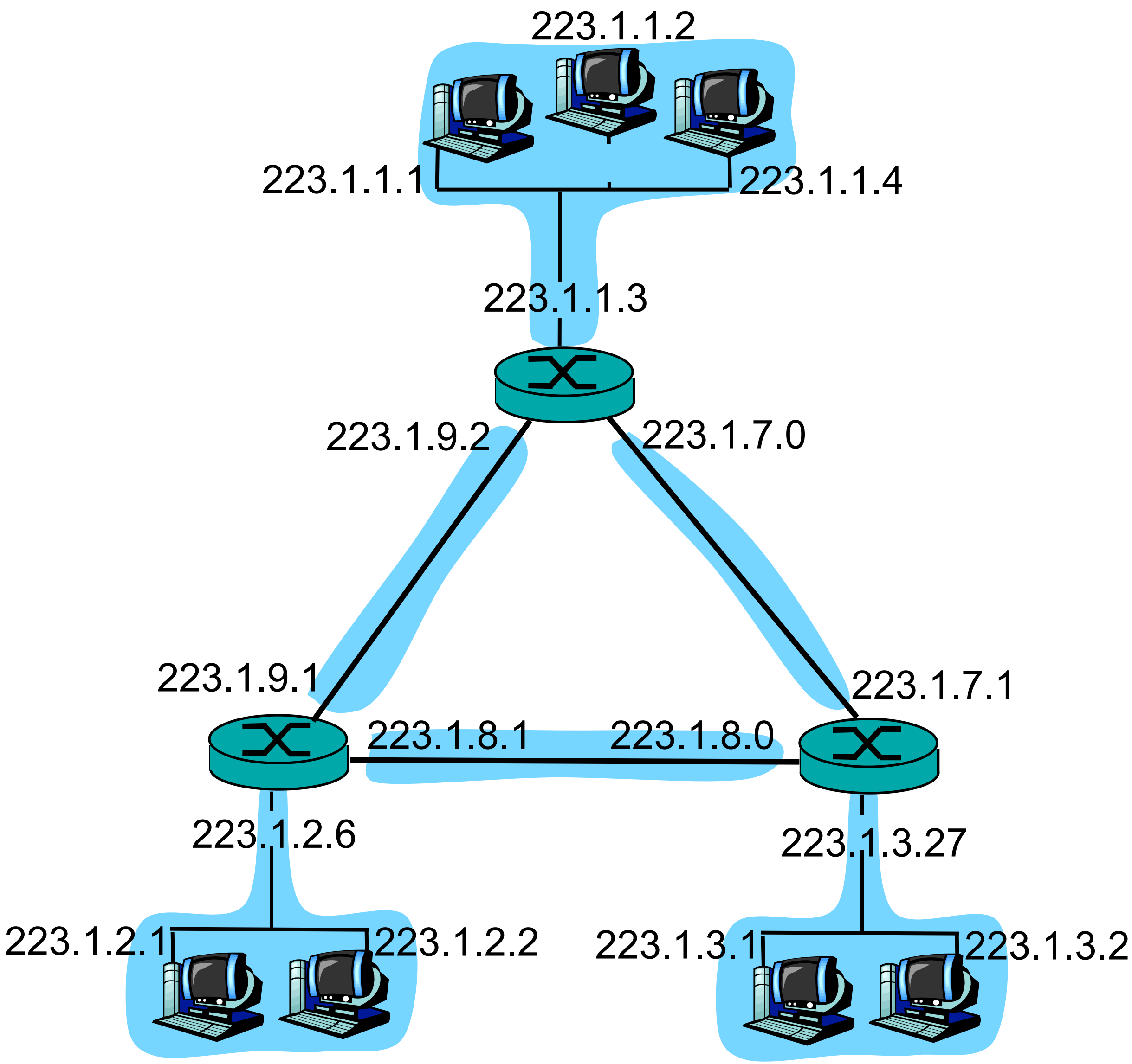
Subnets

- To determine subnets, detach interfaces from host or router



Subnet mask: /24

Subnets



223.1.1.2
223.1.1.1
223.1.1.4
223.1.1.3
223.1.9.2
223.1.7.0
223.1.9.1
223.1.7.1
223.1.8.1
223.1.8.0
223.1.2.6
223.1.3.27
223.1.2.1
223.1.2.2
223.1.3.1
223.1.3.2

IP addressing: CIDR

CIDR: Classless InterDomain Routing

- subnet portion of address of arbitrary length
- address format: a.b.c.d/x, where x is # bits in subnet portion of address

subnet part | host part

11001000 00010111 00010000 00000000

200.23.16.0/23

IP addresses: how to get one?

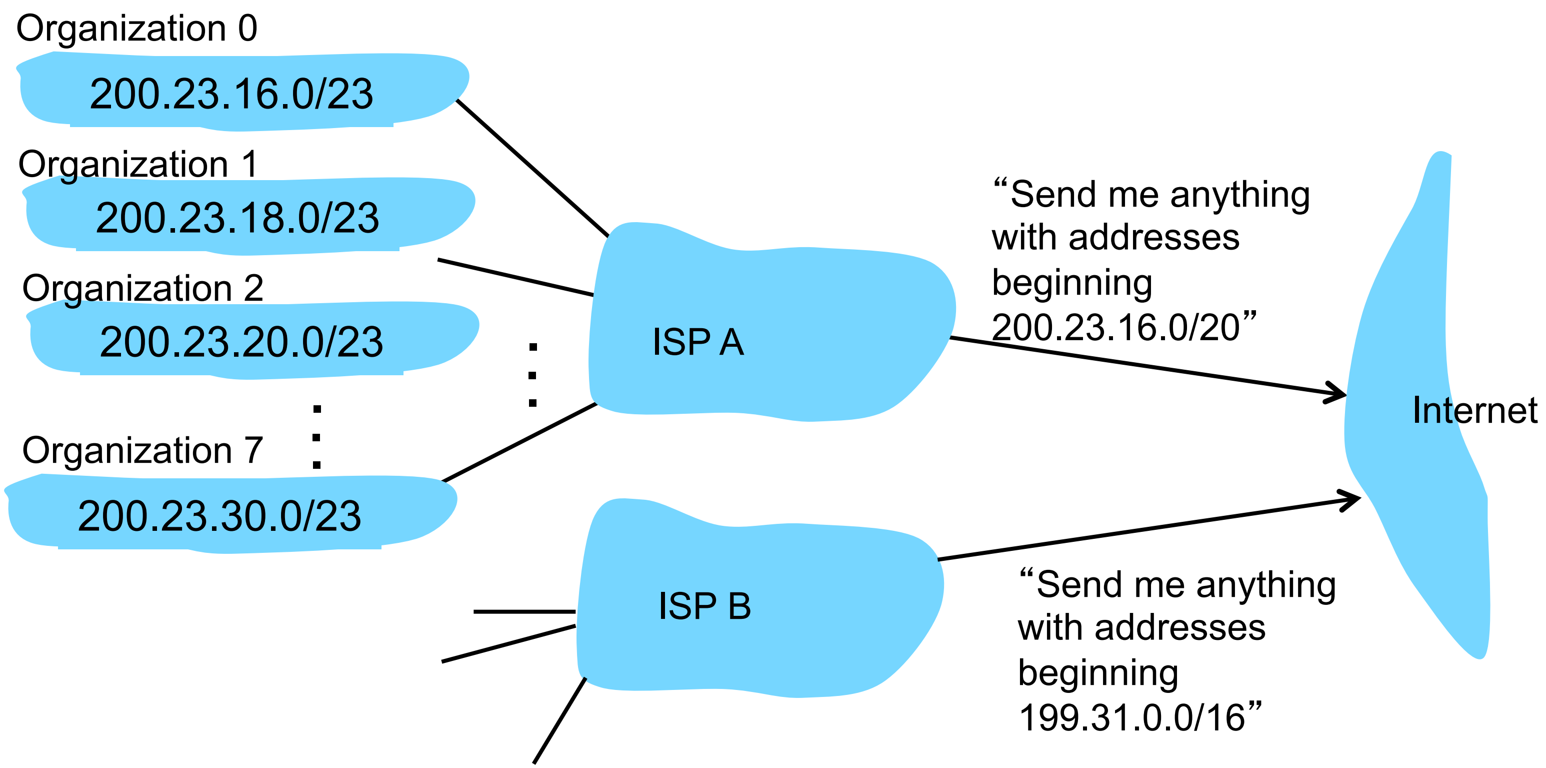
Q: How does *network* get subnet part of IP addr?

A: gets allocated portion of its provider ISP's address space

ISP's block	11001000 00010111 00010000 00000000	200.23.16.0/20
Organization 0	11001000 00010111 00010000 00000000	200.23.16.0/23
Organization 1	11001000 00010111 00010010 00000000	200.23.18.0/23
Organization 2	11001000 00010111 00010100 00000000	200.23.20.0/23
...	….. ….	….
Organization 7	11001000 00010111 00011110 00000000	200.23.30.0/23

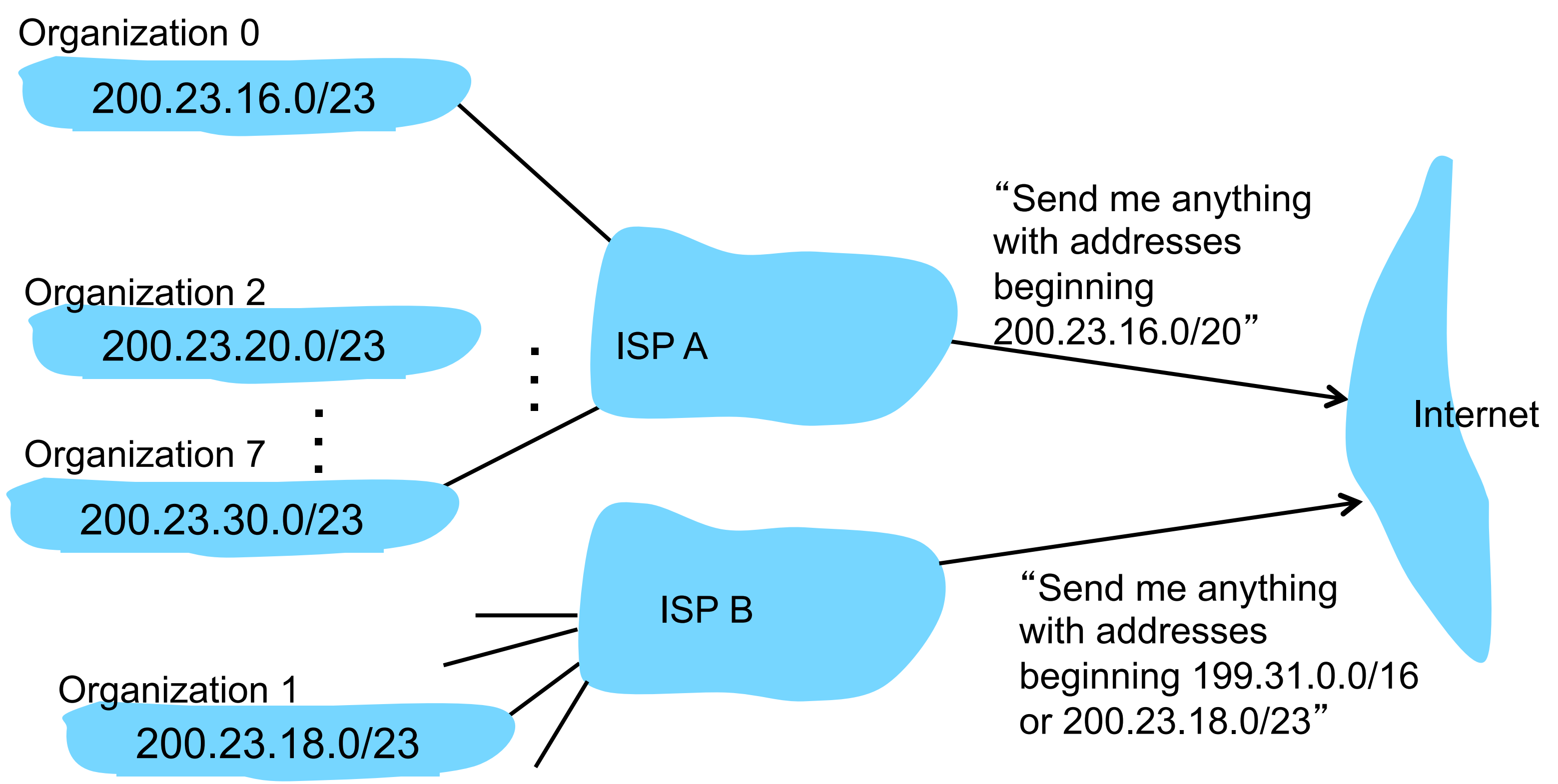
Hierarchical addressing: route aggregation

Hierarchical addressing allows efficient advertisement of routing information:



Hierarchical addressing: more specific routes

ISP B has a more specific route to Organization 1



IP addressing

Q: How does an ISP get block of addresses?

A: ICANN: Internet Corporation for Assigned Names and Numbers

- allocates addresses
- manages DNS
- assigns domain names, resolves disputes

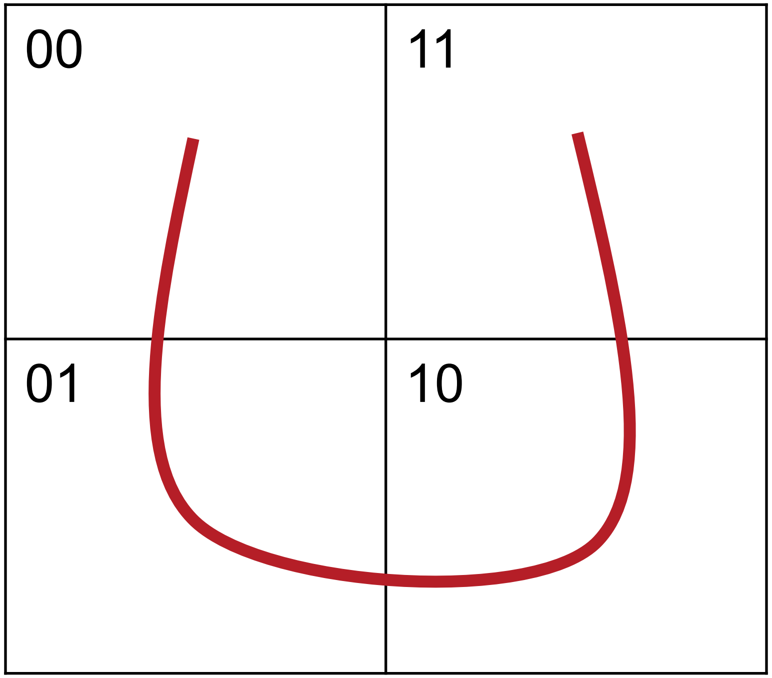
Visualisation of IP Addresses

- **Problem**: how to visualize 4 billion IP adresses?
  - Number line: length 2^{32} pixels not feasible (> 300 km with 300 DPI)
  - Bitmap: 2^{16} x 2^{16} pixels (25 m^2 with 300 DPI)
  - Visualisation of /24 networks (2^8 IP adresses per pixel)
    ⇨ bitmap with 2^{12} x 2^{12} Pixel (16 MPixel, A4 with 300 DPI)

- Requirement: meaningful neighbourhood properties of addresses in bitmap
  - Number line: neighbourhood properties correct
  - Bitmap: neighbourhood properties depend on 2D mapping
  - Approach: room-filling curves

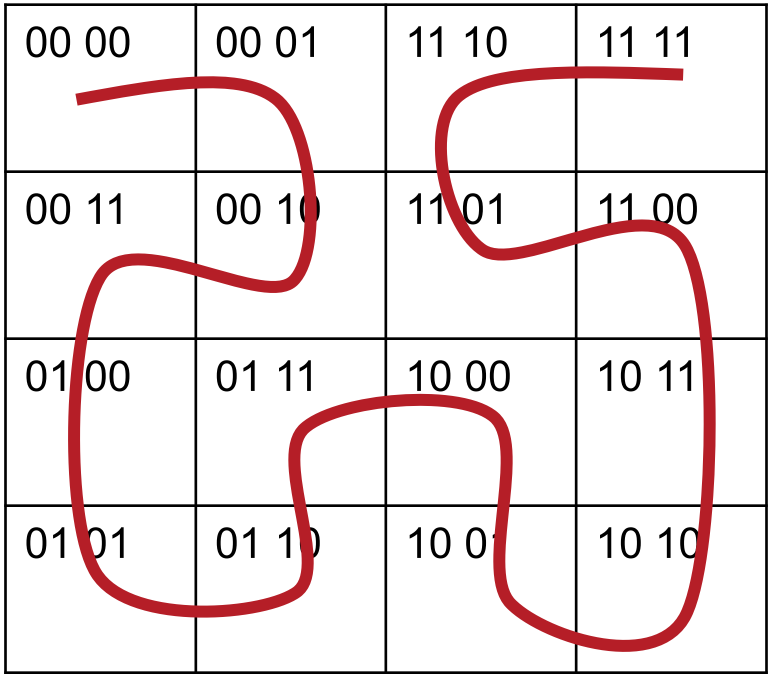
Room-Filling Curves

- **Approach**
  - Map curve to n-dimensional space
  - Requirement: complete filling of space with steady function
  - Recursion



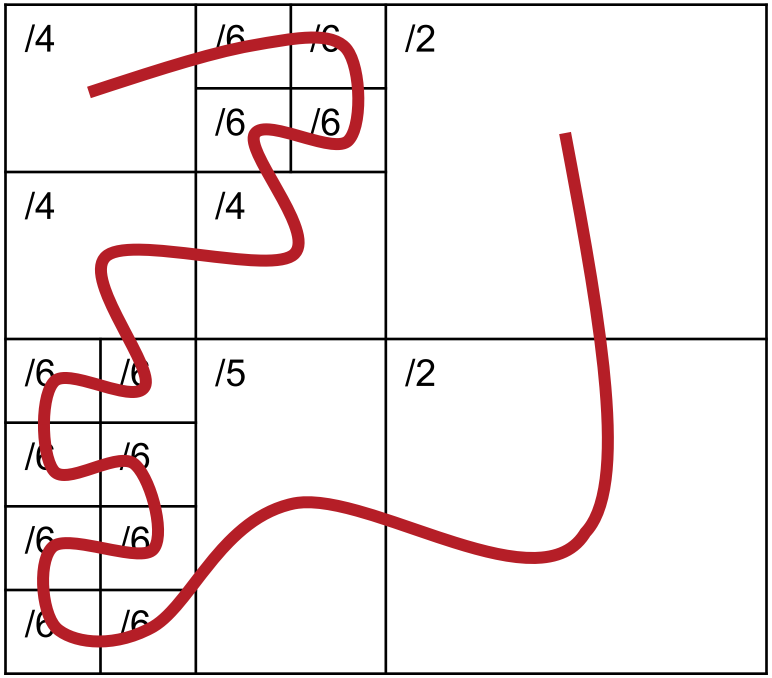
Room-Filling Curves

- **Approach**
  - Map curve to n-dimensional space
  - Requirement: complete filling of space with steady function
  - Recursion by *continuous fractal space-filling curve* using **Hilbert space-filling curve**



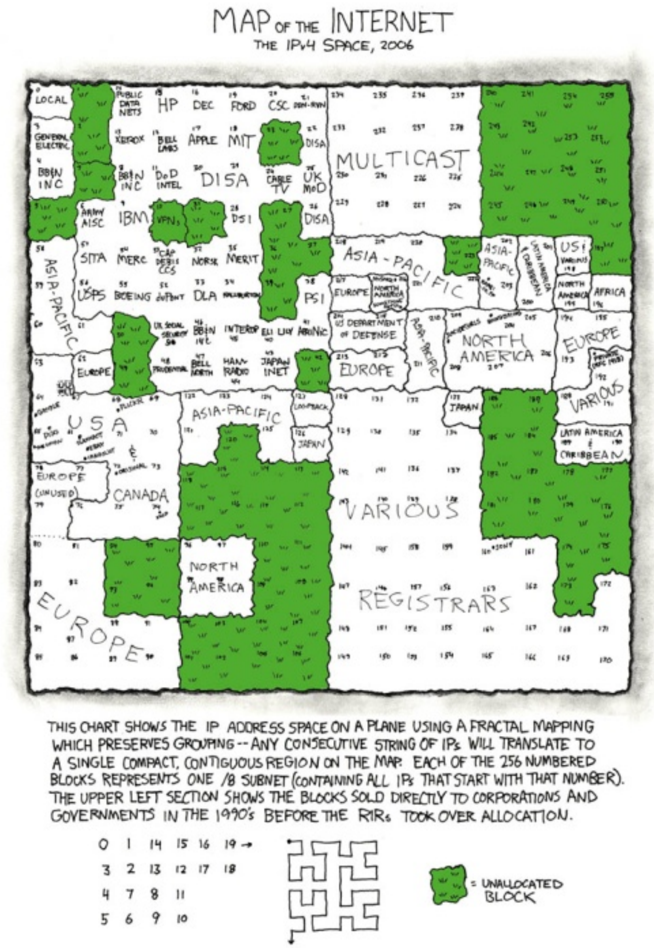
Room-Filling Curves

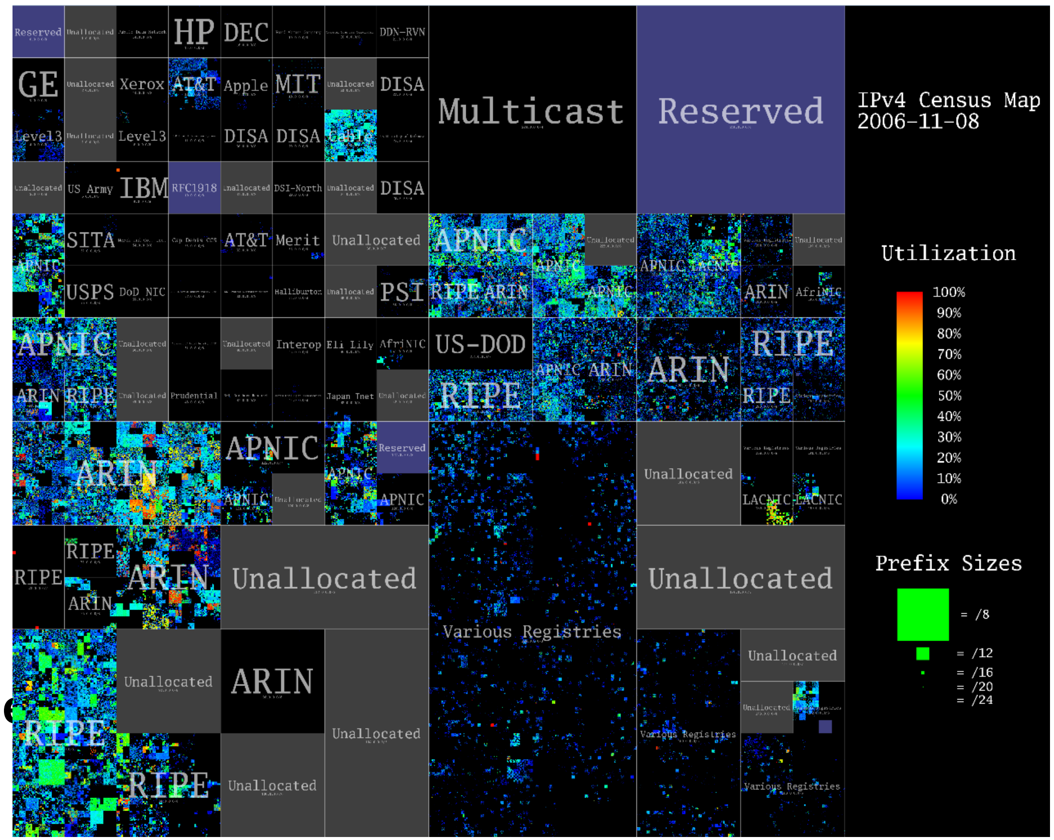
- **Approach**
  - Map curve to n-dimensional space
  - Requirement: complete filling of space with steady function
  - Recursion
    - base curve partitions room into 4 areas
    - rotation of base curve
    - continue up to needed depth



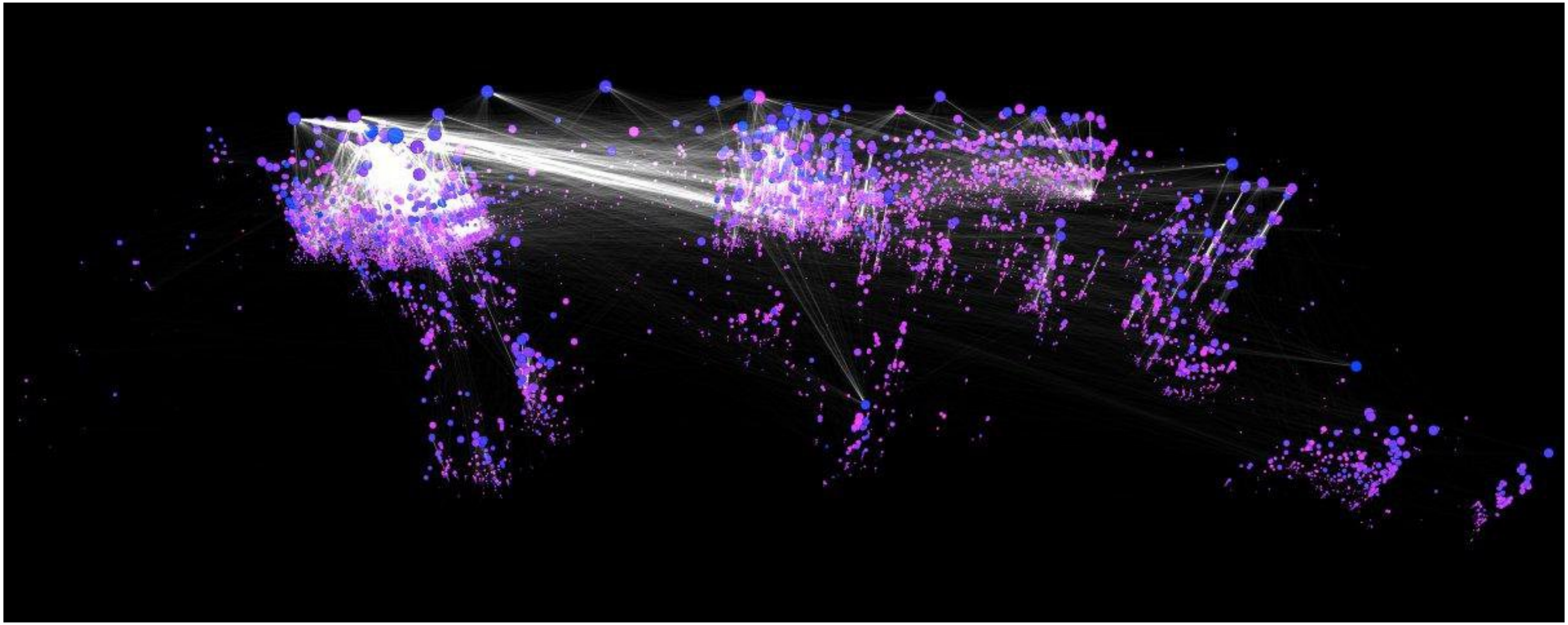
Room-Filling Curves

- **Hilbert curve** for 2D representation of IPv4 address space

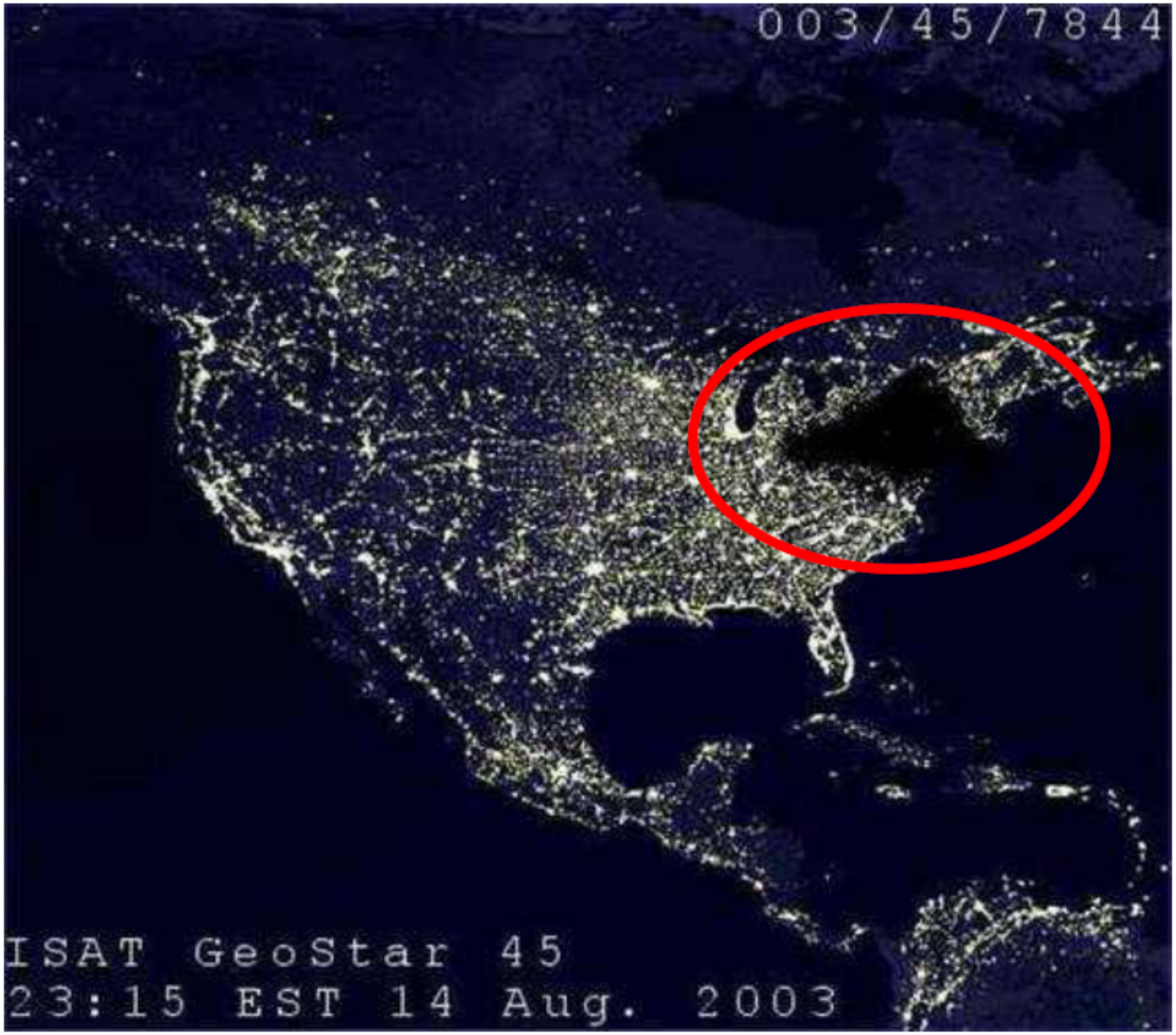




Internet Structure



Incident

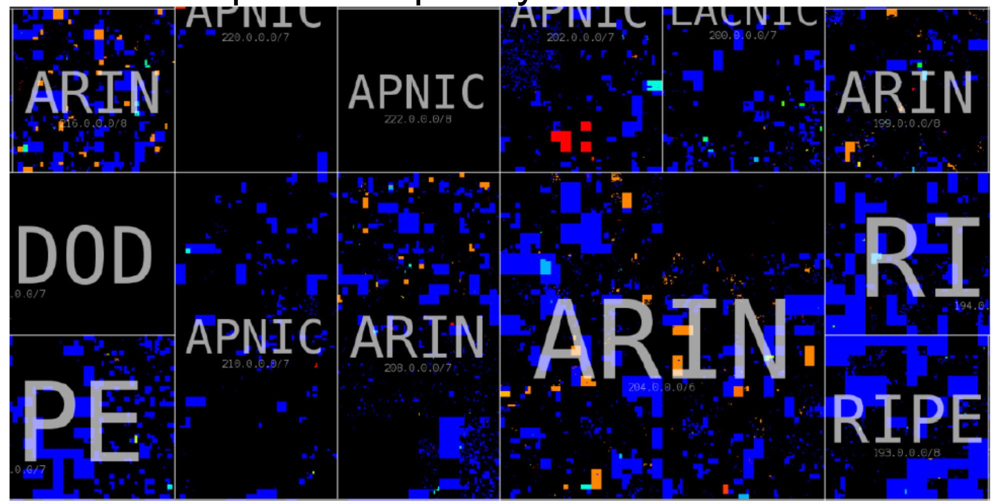


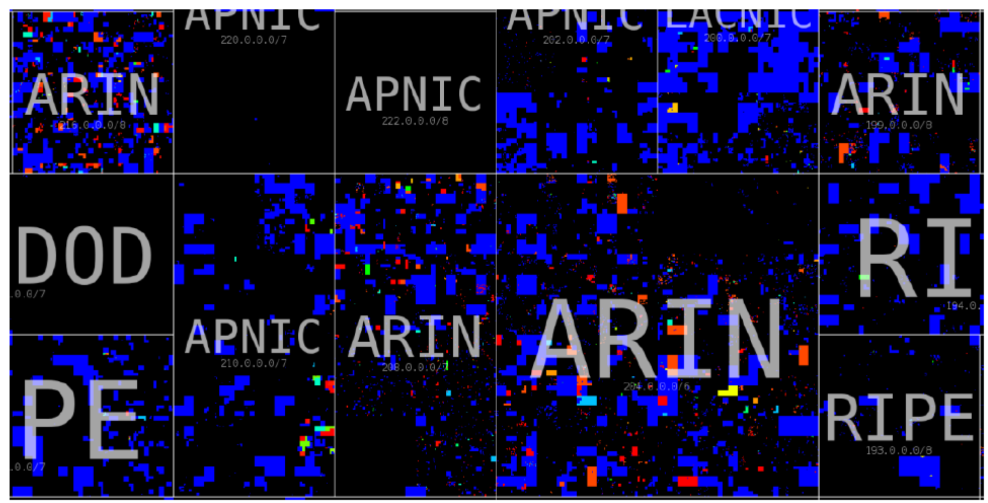
Blackout in North East USA, August 2003

Blackout USA, 2003

- > 100 power stations affected
- > 3.000 networks in > 1.700 organisations affected

Route update frequency 2 h before blackout





Route update frequency 2 h after blackout