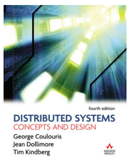
Slides for Chapter 3: Networking and Internetworking



From **Coulouris, Dollimore and Kindberg**
Distributed Systems: Concepts and Design
Edition 4, © Pearson Education 2005

Networking Issues (1)

- Performance:
  - *Latency* (time between send and start to receive)
  - *Data transfer rate* (bits per second) [max]
  - Transmission time = latency + length / transfer rate
  - System bandwidth, throughput [actual]: total volume of traffic in a given amount of time
  - Using different channels concurrently can make bandwidth > data transfer rate
  - traffic load can make bandwidth < data transfer rate
  - network speed < memory speed (about 1000 times)
  - Access to local disk is usually faster than remote disk
  - Fast (expensive) remote disk + fast network
    - can beat slow (cheap) local disks

Networking Issues (2)

- scalability
- reliability
  - corruption is rare
  - mechanisms in higher-layers to recover errors
  - errors are usually timing failures, the receiver doesn't have resources to handle the messages
- security
  - firewall on gateways (entry point to org's intranet)
  - encryption is usually in higher-layers
- mobility--communication is more challenging: locating, routing,...
- quality of service--real-time services
- multicasting--one-to-many communication

Instructor's Guide for Coulouris, Dollimore and Kindberg Distributed Systems: Concepts and Design Edn. 4
© Pearson Education 2005

Types of Networks (1)

- Local Area Networks (LAN)
  - floor/building-wide
  - single communication medium
  - no routing, broadcast
  - segments connected by switches or hubs
  - high bandwidth, low latency
  - Ethernet - 10Mbps, 100Mbps, 1Gbps
  - no latency guarantees (what could be the consequences?)
  - Personal area networks (PAN) [ad-hoc networks]: blue tooth, infra-red for PDAs, cell phones, …

Instructor's Guide for Coulouris, Dollimore and Kindberg Distributed Systems: Concepts and Design Edn. 4
© Pearson Education 2005

Types of Networks (2)

- Metropolitan Area Networks (MAN)
  - city-wide, up to 50 km
  - Digital Subscriber Line (DSL): .25 - 8 Mbps, 5.5km from switch
    - BellSouth: .8 to 6 Mbps
  - Cable modem: 1.5 Mbps, longer range than DSL
    - Bright house w/ Road Runner: .5 to 10Mbps

Instructor's Guide for Coulouris, Dollimore and Kindberg Distributed Systems: Concepts and Design Edn. 4
© Pearson Education 2005

Types of Networks (3)

- Wide Area Networks (WAN)
  - world-wide
  - Different organizations
  - Large distances
  - routed, latency .1 - .5 seconds
  - 1-10 Mbps (upto 600 Mbps)

Instructor's Guide for Coulouris, Dollimore and Kindberg Distributed Systems: Concepts and Design Edn. 4
© Pearson Education 2005

Types of Networks (4)

- Wireless local area networks (WLAN)
  - IEEE 802.11 (WiFi)
  - 10-100 Mbps, 1.5km
    - 802.11 (1997): upto 2 Mbps, 2.4 GHz
    - 802.11a (1999): upto 54 Mbps, 5 GHz, ~75 feet outdoor
    - 802.11b (1999): upto 11 Mbps, 2.4 GHz, ~150 feet [most popular]
    - 802.11g (2003): upto 54 Mbps, 2.4 GHz, ~150 feet [backward compatible with 802.11b, becoming more popular]
- Wireless metropolitan area networks (WMAN)
  - IEEE 802.16 (WiMax)
  - 1.5-20 Mbps, 5-50km

Types of Networks (5)

- Wireless wide area networks (WWAN)
  - worldwide
  - GSM (Global System for Mobile communications)
  - 9.6 – 33 kbps
  - 3G ("third generation"): 128-384 kbps to 2Mbps

Types of Networks (6)

- Internetworks
  - connecting different kinds of networks
  - routers, gateways

Network performance

	Example	*Range*	*Bandwidth (Mbps)*	*Latency (ms)*
Wired:				
LAN	Ethernet	1-2 km	10-1000	1-10
MAN	ATM	250 km	1-150	10
WAN	IP routing	worldwide	.01-600	100-500
Internetwork	Internet	worldwide	0.5-600	100-500
Wireless:				
WPAN	Bluetooth (802.15.1)	10 - 30m	0.5-2	5-20
WLAN	WiFi (IEEE 802.11)	0.15-1.5 km	2-54	5-20
WMAN	WiMAX (802.16)	550 km	1.5-20	5-20
WWAN	GSM, 3G phone nets	worldwide	0.01-2	100-500

Network principles (1)

- Packet transmission
  - message: logical unit of informatio
  - packet: transmission unit
  - restricted length: sufficient buffer storage, reduce hogging

Network principles (2)

- Data Streaming
  - audio/video
  - Need 120 Mbps (1.5 Mbps compressed)
  - play time: the time when a frame need to be displayed
  - for example, 24 frames per second, frame 48 must be display after two seconds
  - IP protocol provides no guaranteesIPv6 (new) includes features for real-time streams, stream data are treated separately
  - Resource Reservation Protocol (RSVP), Real-time Transport Protocol (RTP)

Network principles (3)

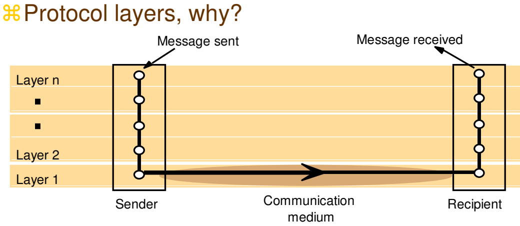
- Switching schemes (transmission between aribitrary nodes)
  - Broadcast: ethernet, token ring, wireless
  - Circuit switching: wires are connected
  - Packet switching:
    - store-and-forward
    - different routes
    - “store-and-forward” needs to buffer the entire packet before forwarding
  - Frame relay
    - Small packets
    - Looks only at the first few bits
    - Don’t buffer/store the entire frame

Network principles (4)

- Protocols
  - Key components
    - Sequence of messages
    - Format of messages

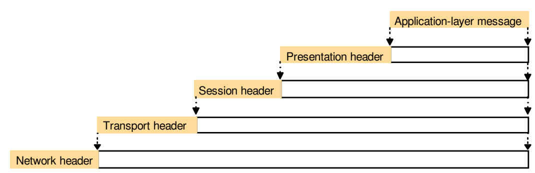
Network principles (5)

- Protocol layers, why?



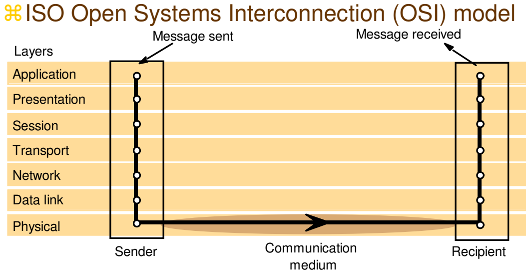
Network principles (6)

- Encapsulation in layered protocols



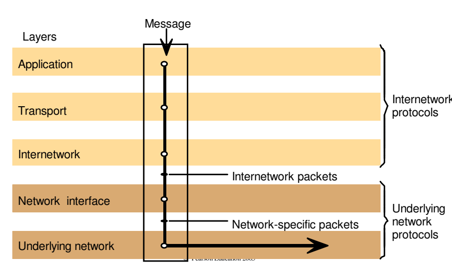
Network principles (7)

- ISO Open Systems Interconnection (OSI) model



Network principles (8)

- Internet layers
  - Application = application + presentation
  - Transport = transport + session



Network principles (9)

- Packet assembly
  - header and data
  - maximum transfer unit (MTU): 1500 for Ethernet
  - 64K for IP (8K is common because of node storage)
- ports: destination abstraction (application/service protocol)
- addressing: transport address = network address + port
  - Well-known ports (below 1023)
  - Registered ports (1024 - 49151)
  - Private (up to 65535)

Network principles (10)

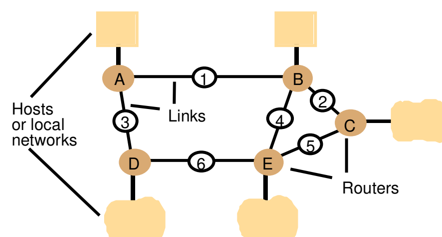
- Packet delivery (at the network layer)
  - Datagram packet
    - one-shot, no initial set up
    - different routes, out of order
    - Ethernet, IP
  - Virtual circuit packet
    - initial set up for resources
    - virtual circuit # for addressing
    - ATM
- Similar but different pairs of protocols at the transport layer (connection-oriented and connectionless)

Network principles (11)

- Routing
  - LAN?
  - Routing Algorithm
    - decide which out-going link to forward the packet
      - for circuit switching, the route is determined during the circuit setup time
      - for packet switching, each packet is routed independently
    - update state of the out-going links
  - Routing Table
    - a record for each destination
    - fields: outgoing link, cost (e.g. hop count)

Network principles (12)

- Router example



Network principles (13): Routing tables

Routings from A

To	*Link*	*Cost*
A	local	0
B	1	1
C	1	2
D	3	1
E	1	2

Routings from B

To	*Link*	*Cost*
A	1	1
B	local	0
C	2	1
D	1	2
E	4	1

Routings from C

To	*Link*	*Cost*
A	2	2
B	2	1
C	local	0
D	5	2
E	5	1

Routings from D

To	*Link*	*Cost*
A	3	1
B	3	2
C	6	2
D	local	0
E	6	1

Routings from E

To	*Link*	*Cost*
A	4	2
B	4	1
C	5	1
D	6	1
E	local	0

Network principles (14)

- Router information protocol (RIP)
  - "Bellman-Ford distance vector" algorithm
  - Sender: send table summary periodically (30s) or changes to neighbors
  - Receiver: Consider *A* receives a table from *B*, *A* updates
    1. *A* -> *B* -> … -> *X*: *A* updates--*B* has more up-to-date (authoritative) info
    2. *A* -> not *B* -> … -> *X*: Does routing via *B* have a lower cost?
    3. *B* -> … -> *X*: *A* does not know *X*
    4. *[B* -> *A* -> … -> *X]*: *A* doesn't update--*A* has more up-to-date info
    5. Faulty link, cost is infinity
  - RIP-1 (RFC 1058)
  - More recent algorithms
    - more information, not just neighbors
    - link-state algorithms, each node responsible for finding the optimum routes

Network principles (15): Pseudocode for RIP routing algorithm

- *Tl* is the table local table; *Tr* is the received remote table

```
Send: Each t seconds or when Tl changes, send Tl on each non-faulty outgoing link.
Receive: Whenever a routing table Tr is received on link n:
    for all rows Rr in Tr {
        if (Rr.link != n) {  // destination not routed via the receiver
            Rr.cost = Rr.cost + 1;
            Rr.link = n;
            if (Rr.destination is not in Tl) add Rr to Tl;
             // add new destination to Tl
            else for all rows Rl in Tl {
                if (Rr.destination = Rl.destination and
                        (Rr.cost < Rl.cost or Rl.link = n)) Rl = Rr;
                // Rr.cost < Rl.cost : remote node has better route
                // Rl.link = n : remote node is more authoritative
            }
        }
    }
```


Network principles (16)

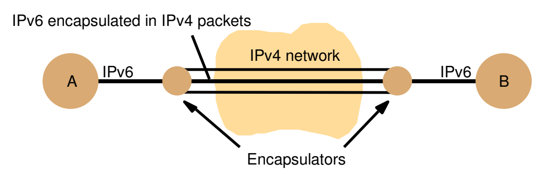
- Congestion control
  - high traffic load, packets dropped due to limited resources
  - reducing transmission rate: "choke packets" from sender to receiver

Networking principles (17)

- Network connecting devices
  - Hubs: extending a segment of LAN (broadcast)
  - Switches: switching traffic at data-link level (different segments of a LAN), making temporary hardware connections between two ports (or store and forward) [switches do not exchange info with each other]
  - Routers: routing traffic at IP level
  - Bridges: linking networks of different types, could be routers as well

Networking principles (18)

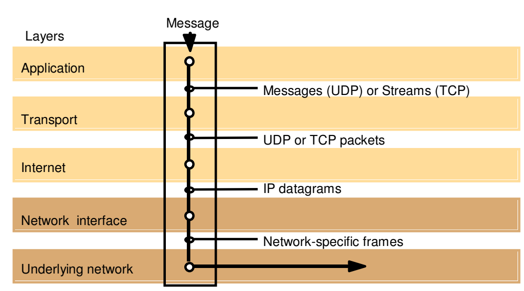
- Tunneling
  - communicate through an "alien" protocol
  - "Hide" in the payload
  - IPv6 traffic using IPv4 protocols



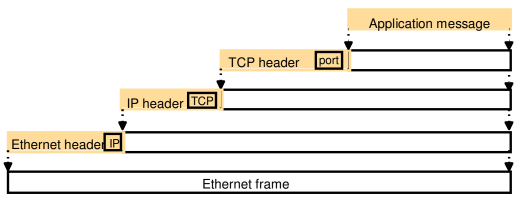
Internet protocols (1)

- IP (Internet Protocol)
  - "network" layer protocol
  - IP addresses
- TCP (Transmission Control Protocol)
  - transport layer
  - connection-oriented
- UDP (User Datagram Protocol)
  - transport layer
  - connection-less

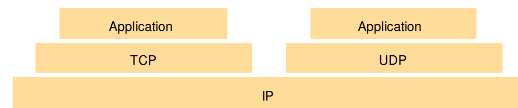
Internet protocols (2): TCP/IP layers



Internet protocols (3): layer encapsulation

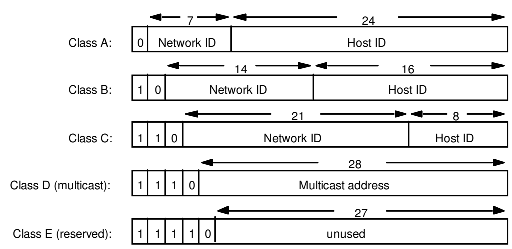


Internet protocols (4): Programmer's view



Internet protocols (5): Internet address structure

- 32-bit



Internet protocols (6): Decimal representation

- 163.118.131.9 (www.fit.edu)

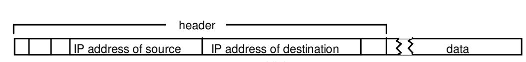
	octet 1	octet 2	octet 3		Range of addresses
Class A:	1 to 127	0 to 255	0 to 255	0 to 255	1.0.0.0 to 127.255.255.255
Class B:	128 to 191	0 to 255	0 to 255	0 to 255	128.0.0.0 to 191.255.255.255
Class C:	192 to 223	0 to 255	0 to 255	1 to 254	192.0.0.0 to 223.255.255.255
Class D (multicast):	224 to 239	0 to 255	0 to 255	1 to 254	224.0.0.0 to 239.255.255.255
Class E (reserved):	240 to 255	0 to 255	0 to 255	1 to 254	240.0.0.0 to 255.255.255.255

(Class A: Network ID = octet 1; Host ID = octets 2–4. Class B: Network ID = octets 1–2; Host ID = octets 3–4. Class C: Network ID = octets 1–3; Host ID = last octet. Class D: Multicast address.)

Internet protocols (7)

- Classless interdomain routing (CIDR)
  - shortage of Class B networks
  - add a mask field to indicate bits for network portion
  - 138.73.59.32/22 [subnet: first 22 bits; host: 10 bits]

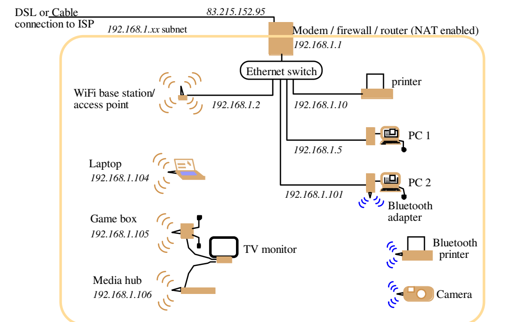
Internet protocols (8)



Internet protocols (9): Network Address Translation

- Sharing one “global” IP address at home
- Routers with NAT
  - Router has a “global” IP address from ISP
  - Each machine has a “local” IP address via DHCP
  - Machine -> router
    - Router stores the local IP addr and source port #
    - Table entry indexed by a virtual port #
  - Router -> outside
    - put the router IP addr and virtual port # in the packet
  - Outside -> router
    - Reply to the router IP addr and virtual port #
  - Router -> machine
    - Use the virtual port # to find table entry
    - Forward to the local IP address and port #
- What happens if we want the device to be a server, not a client?

Internet protocols (10)



Internet protocols (11)

- Server with NAT
  - Fixed internal addr and port #
  - Fixed entry in the table
  - All packets to the port on the router are forwarded to the internal addr and port # in the entry
- What if more than one internal machines want to offer the same service (port)?

Internet protocols (12)

- IP Protocol
  - unreliable or best-effort
  - lost, duplicated, delayed, out of order
  - header checksum, no data checksum
  - IP packet longer than MTU of the underlying network, break into fragments
  - before sending and reassemble after receiving
  - Address resolution (on LANs)
    - mapping IP address to lower level address
    - ARP: address resolution protocol
    - ethernet: cache; not in cache, broadcast IP addr, receive Ethernet addr
  - IP spoofing: address can be stolen (not authenticated)

Internet protocols (13)

- RIP-1: discussed previously
- RIP-2: CIDR, better multicast routing, authentication of RIP packets
- link-state algorithms: e.g., open shortest path first (OSPF)
- Observed: average latency of IP packets peaks at 30-seconds intervals [RIP updates are processed before IP]
  - because 30-second RIP update intervals, locked steps
  - random interval between 15-45 seconds for RIP update
- large table size
  - all destinations!!
  - map ip to geographical location
  - default route: store a subset, default to a single link for unlisted destinations

Internet Protocols (14): IPv6

- IP addresses:128 bits (16 bytes)
  - 3×10^{38} addresses (7×10^{23} addresses per square meter!)
- routing speed
  - no data checksum as before
  - no fragmentation – need to know the smallest MTU in data-link layer
- real-time and special services
  - traffic class: priority, time-dependent (expired data are useless)
  - flow label: timing requirements for streams (reserving resources in advance)
- “next” header field
  - extension header types for IPv6
  - routing information, authentication, encryption ...
- Anycast: at least one nodes gets it
- security
  - currently handled above the IP layer
  - extension header types
- Migration from IPv4
  - backward compatibility: IPv6 addresses include IPv4 addresses
  - Islands of IPv6 networks, traffic tunnels though other IPv4 networks

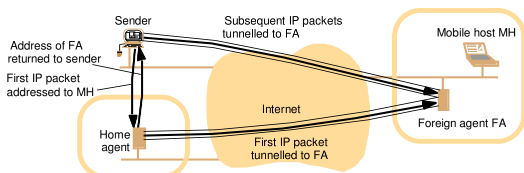
Internet protocols (15):

Version (4 bits)	Traffic class (8 bits)	Flow label (20 bits)
Payload length (16 bits)	Next header (8 bits)	Hop limit (8 bits)
Source address (128 bits)		
Destination address (128 bits)		

Internet Protocols (10): Mobile IP

- Dynamic Host Configuration Protocol (DHCP)
  - assign temporary IP address
  - provide addresses of local resources like DNS
- Routing to maintain continuous access
  - IP routing is subnet-based, fixed relative locations
  - Home agent (HA) and Foreign agent (FA)
  - HA - current location (IP addr) of the mobile host
    - is informed by the mobile host when it moves
    - proxy for the host after it moves
    - inform local routers to remove cached records of the host
    - responds to ARP requests
  - FA - informed by the host when it arrives
    - new temp IP addr
    - contacts HA what the new IP address is
  - HA - receives the new IP address and may tell the sender the new IP addr

Internet protocols (11): MobileIP routing mechanism



Internet protocols (12)

- Transport protocols: TCP and UDP
  - network protocol: host to host
  - transport protocol: process to process
  - Port #'s to indicate processes
- UDP
  - no guarantee of delivery
  - checksum is optional
  - max of 64 bytes, same as IP
  - no setup costs, no segments

Internet protocols (13)

- TCP
  - arbitrarily long sequence
  - connection-oriented
  - sequencing of segments
  - flow control: acknowledgement includes "window size" (amount of data) for sender to send before next ack
  - interactive service: higher frequency of buffer flush, send when deadline reached or buffer reaches MTU
  - retransmission of lost packets
  - buffering of incoming packets to preserve order and flow
  - checksum on header and data

Internet protocols (14)

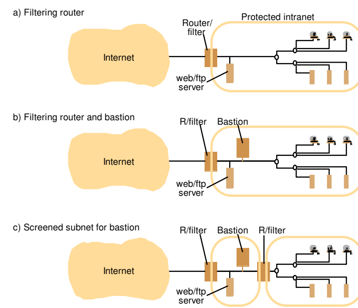
- Domain names
- DNS
  - distributed data
  - each DNS server keeps track of part of the hierarchy
  - unresolved requests are sent to servers higher in the hierarchy

Internet protocols (15)

- Firewalls
  - monitor and filter communication
  - controlling what services are available to the outside
  - controlling the use of services
  - controlling internal users access to the outside
- Filtering at different protocol levels
  - IP packet filtering: addresses, ports..
  - TCP gateway: check for correctness in TCP connections
    - e.g., are they partially opened and never used (why?)
  - Application-level gateway: proxy for applications
    - no direct communication between the inside and outside
    - e.g., smtp proxy can check addresses, content...

Internet protocols (16)

- Bastion (tcp/ application filter)
- C): two router filters
  - Access to web/ftp server, but not LAN
  - Hide internal IP addresses
    - Bastion has the mapping
    - Second router is the second IP filter (invisible to the outside)



Internet protocols (17)

- Virtual Private Network (VPN)
  - extending a secured internal network to an external unsecured host
  - e.g. IPSec tunneling through IP

Network Case Studies (1): Ethernet and WiFi

IEEE No.	*Name*	*Title*	*Reference*
802.3	Ethernet	CSMA/CD Networks (Ethernet)	[IEEE 1985a]
802.4		Token Bus Networks	[IEEE 1985b]
802.5		Token Ring Networks	[IEEE 1985c]
802.6		Metropolitan Area Networks	[IEEE 1994]
802.11	WiFi	Wireless Local Area Networks	[IEEE 1999]
802.15.1	Bluetooth	Wireless Personal Area Networks	[IEEE 2002]
802.15.4	ZigBee	Wireless Sensor Networks	[IEEE 2003]
802.16	WiMAX	Wireless Metropolitan Area Networks	[IEEE 2004a]

Network Case Studies (2): Ethernet

- Ethernet, CSMA/CD, IEEE 802.3
  - Xerox Palo Alto Research Center (PARC), 1973, 3Mbps
  - 10,100,1000 Mbps
  - extending a segment: hubs and repeaters
  - connecting segments: switches and bridges
  - Contention bus
  - Packet/frame format
    - preamble (7 bytes): hardware timing
    - start frame delimiter (1)
    - dest addr (6)
    - src addr (6)
    - length (2)
    - data (46 - 1500): min total becomes 64 bytes, max total is 1518
    - checksum (4): dropped if incorrect

Network Case Studies (3)

- Carrier Sensing Multiple Access / Collision Detection (CSMA/CD)
  - CS: listen before transmitting, transmit only when no traffic
  - MA: more than one can transmit
  - CD: collision detected when *signals transmitted are not the same as those received* (listen to its own transmission)
    - After detection of a collision
      - send jamming signal
      - wait for a random period before retransmitting
- T (Tau): time to reach the farthest station
- When is the collision detected?
  - A and B send at the same time
  - A sends, B sends within T seconds
  - A sends, B sends between T and 2T seconds
  - A sends, B sends after 2T seconds
- Minimum length of packet for collision detection:
  - packet length > 2T, between T and 2T, and < T ?

Network Case Studies (4)

- Physical implementation:
  - <R><B><L>
  - R: data rate in Mbps
  - B: medium signaling type: baseband [one channel] or broadband [multiple channels]
  - L: max segment length in 100meters or T (twisted pair cable, hierarchy of hubs)

Network Case Studies (5): Ranges and speeds

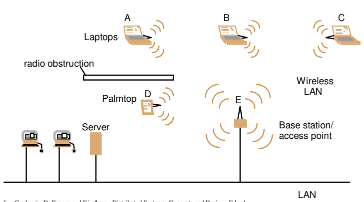
	10Base5	*10BaseT*	*100BaseT*	*1000BaseT*
Data rate	10 Mbps	10 Mbps	100 Mbps	1000 Mbps
Max. segment lengths:				
Twisted wire (UTP)	100 m	100 m	100 m	25 m
Coaxial cable (STP)	500 m	500 m	500 m	25 m
Multi-mode fibre	2000 m	2000 m	500 m	500 m
Mono-mode fibre	25000 m	25000 m	20000 m	2000 m

Network Case Studies (6): WiFi

- IEEE 802.11 wireless LAN
  - up to 150m and 54Mbps
  - access point (base station) to land wires
  - Ad hoc network--no specific access points, "on the fly" network among machines in the neighborhood
  - Radio Frequency (2.4, 5GHz band) or infra-red

Network Case Studies (7): Problems with wireless CSMA/CD

- Hidden station: not able to detect another station is transmitting
  - A can't see D, or vice versa
- Fading: signals weaken, out of range
  - A and C are out of range from each other
- Collision masking: stronger signals could hide others
  - A and C are out of range from each other, both transmits, collide, can't detect collision, Access point gets garbage



Network Case Studies (8)

- Carrier sensing multiple access with collision *avoidance* (CSMA/CA)
  - reserving slots to transmit
  - if no carrier signal
    - medium is available,
    - out-of-range station requesting a slot, or
    - out-of-range station using a slot

Network Case Studies (9)

- Steps
  1. Request to send (RTS) from sender to receiver, specify duration
  2. Clear to send (CTS) in reply
  3. in-range stations see the RTS and/or CTS and its duration
  4. in-range stations stop transmitting
  5. acknowledgement from the receiver
- Hidden station & Fading: CTS, need permission to transmit
- RTS and CTS are short, don't usually collide; random back off if collision detected
- Should have no collisions, send only when a slot is reserved