Mobile communications

- Mobile communications is wireless – and almost all wireless communications is radio
- Wireless and radio basics: Electromagnetic waves, transmission, reception, bandwidth

EM waves and early radio

- Radio makes use of electromagnetic radiation propagating through free space
- James Clerk Maxwell showed that a changing electric field induces a changing magnetic field, and vice versa (circa 1861-73)
- Maxwell's Equations showed that electromagnetic waves, composed of changing electric and magnetic fields coupled with each other, can propagate in free space
- Heinrich Hertz devised the first experiment proving the existence of EM waves (1886) – a spark gap transmitter and receiver
- Hertz on the applications of EM waves: "It's of no use whatsoever"
- First uses of radio in communication: 1893-1895; various inventors (Tesla, Marconi, Popov)
- Basic radio system

(Fig. 1)

- transmitter induces alternating current in the transmitting antenna; antenna induces alternating magnetic field; EM wave propagates; EM wave induces alternating electric field in the receiving antenna; electric field induces alternating current in the receiver
- Alternation: Relatively high frequencies are needed

Signal transmission

(Fig. 2)

- f_c is the carrier frequency
- x(t) is a waveform to be transmitted (Baseband signal)
- Modulation: multiply by sinusoid at carrier frequency f_c
- i.e. boost frequency components

(Eq. 1)

Signal reception

(Fig. 3)

- y(t) is the current induced in the receiver antenna
- Amplification: $Ay(t) = x(t)\cos(2\pi f_c t)$
- Demodulation: multiplied by the same sinusoid

(Eq. 2)

- We now have x(t) and high-frequency components – lowpass filter to get rid of them
- And we have recovered x(t)
- **Noise:** The received signal y(t) might be extremely weak, so that the random thermal motions of electrons in the amplifier are significant by comparison
- These random motions introduce an additive, random noise term
- In fact the amplifier output is $x(t)\cos(2\pi f_c t) + n(t)$
- More about this later

Bandwidth

- Under the Fourier transform, x(t) can be represented as a collection of sinusoids: X(f) – can go back and forth between the two domains. (i.e., x(t) and X(f) are equivalent)
- The bandwidth of x(t) is the largest f such that X(f) is negligible
- Generally, larger bandwidth = more quickly changing signal = more information (Nyquist sampling)
- Telephone line: 8 kHz; CD-quality audio: 20 kHz; NTSC video: 6 MHz
- Let X(f) and X*(f) represent Fourier transforms of signals each with bandwidth B

(Fig. 4)

- modulate X(f) with carrier frequency f_c
- (example, for each f < B)
- modulate X*(f) with carrier frequency f_c+2B

(Fig. 5)

- Demodulate X(f) with frequency f_c – and lowpass filter
- Demodulate X*(f) with frequency f_c+2B – and lowpass filter

(Fig. 6)

- Thus two (or more) signals can share frequency space, as long as their bandwidth is finite

Wireless Spectrum

- The collection of all wireless devices in the world must work together to share all the available bandwidth (the wireless spectrum).
- The effectiveness of an antenna at a given frequency is highest if its length is proportional to the wavelength (usually $\lambda/4$ or $\lambda/2$), and $\lambda = c/f$ (c = speed of light = 3×10^8 m/s)
- E.g. f = 1 MHz, λ = 300 m; f = 100 MHz, λ = 3 m; f = 1 GHz, λ = 0.3 m
- This sets a practical limit on the lowest frequencies that are usable (also, less bandwidth is available)

- Atmospheric absorption limits the highest frequencies to around 300 GHz (contemporary applications do not go above 100 GHz) – beyond this is infrared

Name	Band	Application
ELF, SLF, VLF LF	Up to 30 kHz 30—300 kHz	Military, navigation
MF	300 KHz—3 MHz	AM radio
HF	3—30 MHz	Shortwave radio, amateur radio
VHF	30—300 MHz	FM radio, TV, point-to-point comm
UHF	300 MHz—3 GHz	TV, cellphones, ISM bands
SHF EHF	3—30 GHz 30—300 GHz	ISM, Satellites, high-capacity links

- Each slice of the wireless spectrum is licensed for a specific task, and by law cannot be used for any other purpose
- E.g. radio stations are licensed to the station using them by the government, and by law can only be used by that station for approved purposes
- An exception is the "ISM band" (ISM = Industrial, Scientific, Medical) – these are unlicensed bands which may be used for any purpose (subject to power restrictions)
- There are several in various parts of the wireless spectrum, but the most popular is from 2.4-2.5 GHz; used for WiFi, Bluetooth, Zigbee, etc.

Path Loss

- Wireless signals decay with distance from the transmitter

- In free space, the decay is proportional to the size of the wave front (conservation of energy)

(Fig. 7)

- in empty three-dimensional space, the surface area of a sphere is proportional to r^2 (more precisely it's $4\pi r^2$), so this is how fast signals decay
- If signal decay is proportional to r^d, then d is called the *path loss exponent*
- Real-world signals can bounce off the ground, buildings, etc., so in practice d can be between 2 and 4
- On the other hand, if the signal is guided (e.g. by a tunnel), d can be less than 2 (doesn't happen often).

(Fig. 8)

- example: 1 W signal, 100m from tx to rx, if rx power is 0.1 μW with d=2, assuming same constant of proportionality, what is rx power with d=3 and d=4?
- Answer: 1 μW, 0.01 μW

More on signal propagation

- Main effects on radio signals other than attenuation: scattering, shadowing, diffraction, reflection

(Fig. 9)

Antenna Design

- Antennas have an antenna pattern, the relative signal strength (either transmit or receive) from all directions
- Some examples: isotropic, dipole, sector

(Fig. 10)

- A directional antenna is designed by creating physical reflectors/blockers (e.g., dish) or by combining many antenna elements (e.g., array)
- Project signal where wanted, reject interference from unwanted sources, eliminate extra paths

The Mobile Radio Environment: Multipath Fading

- As mentioned, wireless signals can follow many paths from transmitter to receiver (reflection, scattering, etc.)
- These signals have different lengths and therefore different phases

(Fig. 11)

- For a sinusoid with frequency f_c, phase at the rx is

(Eq. 3)

- Say you have two paths, one with distance d_1, the other with distance d_2. They combine at the receiver to form

(Eq. 4)

- Trig identity: $\sin a + \sin b = 2 \sin((a+b)/2) \cos((a-b)/2)$

(Eq. 5)

- since sin is always between -1 and 1, the effect of the fading is to diminish the signal amplitude, possibly to nothing!

- Example. $d_1-d_2 = 1.5$m, f=100 MHz

Power vs. Amplitude

- Amplitude refers to the field strength of an E or M field
- Power is proportional to the square of the field strength
- So power decaying proportional to r^2 (path loss exponent) means field strength decays proportional to r
- Example. Received amplitude as given last class (Eq. 3) has power

(Eq. 6)

Plane Earth Path Loss

- As a practical example of two-ray multipath fading, consider a signal propagating over a flat plane

(Fig. 12)

- h_T is tx height, h_R is rx height, d is the distance from tx to rx, R is the reflection coefficient (i.e. coefficient multiplied by reflection)
- received signal is

(Eq. 7)

- If d is sufficiently large, R = -1 (from physics)
- Trig identity: sin a – sin b = 2 cos((a+b)/2) sin((a-b)/2)

(Eq. 8)

- $d_2 – d_1$ is … using the approximation $(1+u)^{1/2} \approx 1 + u/2$

(Fig. 9)

(Eq. 10)

- furthermore sin u ≈ u for small u

(Eq. 11)

- so the path loss exponent for plane earth propagation is 4.
- Limitations: Earth needs to be very flat and a very good conductor in order for this to hold (e.g., calm ocean)
- On dry ground, path loss exponent is close to 2 as long as line-of-sight path exists from tx to rx

Rayleigh fading

- What about lots and lots of sinusoids?
- Remember that sin(a+b) = sin a cos b + cos a sin b, so

(Eq. 12)

- The A cos θ and A sin θ terms are random (because path length is random, and phase is random)
- Mean zero, variance small (but the same for both)
- So in the end we get

(Eq. 13)

- Thanks to the central limit theorem, sums of large numbers of random variables approach the Gaussian distribution, so U and V are Gaussian random variables

- Another trig identity: $U \cos a + V \sin a = (U^2+V^2)^{1/2}\sin(a+b)$, where

(Eq. 14)

- The signal strength is then $(U^2+V^2)^{1/2}$, which has the Rayleigh distribution, which has CDF

(Eq. 15)

- The extra phase doesn't matter – can be tracked
- r^2 is the average power of the sinusoid (given)
- tells us the probability that the signal strength will be less than a given value in an environment with lots of paths
- Example: Average power is 1 W. What is the probability that the signal strength falls below 1?

Flat vs. frequency-selective fading

- How different is fading from frequency to frequency?
- Recall Eq. 3

(Eq. 16)

- if d_1-d_2 is large, a small change in frequency will have a huge effect (frequency-selective fading)
- on the other hand if d_1-d_2 is small, a small change in frequency will have a negligible effect (flat fading)

Rayleigh Fading

- Example: Average power is 1 W. What is the probability that the signal amplitude falls below 1?
- Recall Eq. 9 from last week's notes

(Eq. 17)

Flat vs. frequency-selective fading

- How different is fading from frequency to frequency?
- Recall Eq. 3 from last week's notes

(Eq. 18)

- if d_1-d_2 is large, a small change in frequency will have a huge effect (frequency-selective fading)
- on the other hand if d_1-d_2 is small, a small change in frequency will have a negligible effect (flat fading)
- note that $(d_1-d_2)/c$ is equal to the time delay of arrival between the two signals, so if $B(d_1-d_2)/c$ is small, then fading is flat
- $(d_1-d_2)/c$, or maximum delay between first and last significant paths, is called the delay spread

Effect of Motion on Wireless Communication

- Suppose the transmitter antenna is stationary, while the receiver antenna is mounted on a moving vehicle
- Rx antenna starts out at distance d and moves at a velocity of v with respect to tx

- Pure sinusoid is transmitted at carrier frequency f_c

(Fig. 13)

- Signal is now

(Eq. 19)

- collecting terms

(Eq. 20)

- The frequency $f_c v/c$ is called the Doppler frequency – arises from the doppler effect
- Causes periodic dropouts in signal strength if compensation is not used in the receiver

(Eq. 21)

Example. $f_c = 1$ GHz, v = 108 km/hr

- Causes periodic dropouts in signal strength in multipath
- Say two paths exist, one stationary and the other in motion

(Fig. 14)

- Received signal is now

(Eq. 22)

Example. f_c = 1 GHz; one path stationary: d_1 = 100m; second path moving: d_2 = 100m initially, v = 54 km/hr

Fast vs. Slow Fading

- Combining multipath with doppler, we have a signal that changes with time
- How quickly does the signal change?
- Best to consider this from the perspective of a packet

(Fig. 15)

- If the signal changes much more quickly than the packet length, this is called “fast fading”
- If the packet length is much longer than signal changes, this is called “slow fading”
- Generally, fast fading is better – getting stuck with a low signal strength is bad, but in fast fading you will probably get a good signal on average
- Unfortunately most of the fading in the world is slow

Link Budgeting

- How much power do you need to be reasonably assured of good communication? – make a link budget.
- Aside: Decibels

- Gains and losses in wireless communication are multiplicative, e.g. amplifier gain, fading
- Multiplicative gains are hard to deal with intuitively … but … if the gains are ABCD, they can be made additive by taking the log: log(ABCD) = log A + log B + log C + log D
- You can express a quantity x in decibels by taking $10\log_{10}x$
- Examples: x=1=0dB; x=10=10dB; x=100=20dB; x=1000=30dB; …
- Other fun stuff to know: x=2≈3dB
- Adding in dB is equivalent to multiplying in normal domain; subtracting in dB is equivalent to dividing in normal domain
- E.g. 20=2x10=3dB+10dB=13dB; 500=1000/2=30dB-3dB=27dB
- dBm = dB referenced to 1 mW : e.g., 30 dBm = 1000 mW = 1 W.
- Link budgets are usually expressed as POWER (not amplitude) and in terms of dB.
- Take the starting power (at the transmitter), add all the gains, and subtract all the losses

(Eq. 23)

- P_T: Transmitter power; G_T: Transmitter antenna gain; G_R: Receiver antenna gain; L_P: Path loss; L_F: Fading margin; L_O:Other losses; P_R:Receiver power (all in dB)

Link Budgeting

- How much power do you need to be reasonably assured of good communication? – make a link budget.
- Link budgets are usually expressed as POWER (not amplitude) and in terms of dB.

- Take the starting power (at the transmitter), add all the gains, and subtract all the losses

(Eq. 24) $P_R = P_T + G_T + G_R - L_P - L_F - L_O$

- P_T: Transmitter power; G_T: Transmitter antenna gain; G_R: Receiver antenna gain; L_P: Path loss; L_F: Fading margin; L_O:Other losses; P_R:Receiver power (all in dB)

Path Loss

- Path loss term is L_P
- Before we said that the path loss is proportional to d^a, where a is the path loss exponent
- In dB, we have

(Eq. 25) $L_P = 10 \log_{10} kd^a = 10 \log_{10} k + 10a \log_{10} d$

Fading margin

- this amount is allocated to ensure a high probability that fading will not disrupt the signal
- Can use the probability of various amplitudes in Rayleigh fading to obtain an adequate margin

Example. You need -10dBm of power to ensure reliable communication. Required range is 500 m, with a path loss exponent of 3 (ignore the constant of proportionality). The transmit antenna has G_T=3 dB, and the receive antenna is isotropic. Allow 10 dB for the fade margin, and 0 dB for other losses. What is the required power at the transmitter in dBm?

(Answer: $P_R = P_T + G_T + G_R - L_P - L_F - L_O$ … solve for P_T. The only tricky part is that $G_R = 0$ dB because the antenna is isotropic.)

Data Link Layer: Access Control and Multiple Access

- The role of the data link layer is to ensure reliable communication between two connected terminals, and to allocate access to a shared medium
- E.g., in wired communication, Ethernet is a data link protocol
- These protocols become very important in wireless communications, because everyone in the world is sharing the same medium (i.e., the air)
- Compared to wired systems, the random delays and frequency shifts inherent in wireless systems are problematic and require special attention in protocol design
- Also at this layer is error detection/correction

Fixed multiple access protocols

- There are two traditional multiple access protocols, which divide up the medium in terms of time or frequency – both ensure that the users do not interfere with each other
- Frequency division multiple access (FDMA): The entire frequency band is divided up (equally or not) among the users. In each user's allocated band, the user can do as s/he pleases.
- Guard bands between users are given to prevent doppler frequency shifts from causing interference
- Example. 1 MHz of bandwidth to be shared among 10 users, with a 10 kHz guard band between each user. So

each user gets 90 kHz (don't forget about half a guard band before the first user, and half a guard band after the last user).

- Time division multiple access (TDMA): Time is broken up into frames. Each frame is divided up (equally or not among the users. In each user's allocated time, the user can do as s/he pleases.
- Guard times between users are given to prevent random delays from causing interference
- Example. Frame duration of 10ms, guard time of 0.1ms, 10 users. So each user gets 0.9ms per frame.
- Problem with FDMA and TDMA: They require central control and make it difficult to quickly reuse resources … best for cellphone-like systems, where people need a full channel for a long period of time; bad for packet data systems, where data is sent less frequently

Exposed/hidden terminal problem

- Carrier sense multiple access (CSMA) solves this problem in Ethernet (i.e., sense whether the channel is free and then transmit) … this is decentralized and appropriate for packet data. Can we do the same thing here?
- Not exactly: there are two problems, the hidden terminal problem, and the exposed terminal problem

- Hidden terminal problem:
- There are three nodes: A, B, C
- Say B is in radio range of A and C. However, A and C are NOT in radio range of each other.
- A wants to send to B. A senses the medium and sees that it is clear, so A sends.
- While A is sending to B, C decides to send to B. C senses the medium – C can't tell that A is transmitting because A

and C are not in radio range. Thus, C decides that the medium is clear and transmits.
- Messages from A and C collide at B, corrupting each other.

- Exposed terminal problem:
- There are four nodes: A, B, C, and D.
- A is in range of B, B is in range of A and C, C is in range of B and D, and D is in range of C. (i.e., the nodes are arranged in a line as A B C D, and each node can only see its neighbors)
- B sends to A, and at the same time, C wants to send to D. C senses the medium and decides the medium is busy, so C does not send. However, B's transmission cannot reach D, so in fact it is safe for C to transmit, and that time/bandwidth is wasted.

(For the above, draw figures to help the students understand the relationships among the users)

- One solution: Multiple Access with Collision Avoidance (MACA)
- Clears radio neighborhoods in advance of transmission; avoids both the hidden terminal and exposed terminal problems
- Protocol design:
    - Two nodes: X and Y. X has a packet of data to send to Y.
    - X sends a control message to Y called Request-to-Send (RTS). This lets Y know that a packet is coming.
    - When Y is ready to receive the packet, Y sends a control message to X called Clear-to-Send (CTS).
    - X sends the data packet to Y.
    - Y confirms that it received the packet by sending a control message to X called Acknowledgment (ACK).

- Main rule of operation:
    - If a node observes a CTS message destined for a node other than itself, it remains silent until it observes the corresponding ACK message.
- Solves the hidden terminal problem:
    - Say Z is in radio range of Y, but not in radio range of X. Then Z is a "hidden terminal".
    - In MACA, Z does not see X's RTS, but Z does see Y's CTS. Thus, Z remains silent until Y transmits ACK to X.
- Solves the exposed terminal problem:
    - Say nodes are arranged W, X, Y, Z in a line, where neighbors on the line are in radio range of each other (i.e. X in range of W, W and Y in range of X, X and Z in range of Y, Y in range of Z.)
    - Then Z is an "exposed terminal" for transmissions from X to W.
    - Say X sends RTS to W. Y observes the RTS.
    - W sends CTS to X. Since Y is not in radio range of W, Y does not observe W's CTS.
    - Since Y observed X's RTS but not CTS, Y concludes that its transmissions will not interfere with X's.
- This protocol is used in IEEE 802.11 (WiFi)

(At this point please do a couple of simple examples.)

- A couple of problems with MACA

- First problem: Suppose round-trip times between nodes are much larger than the data packet length.
- Example: Wireless signals to and from a satellite in geostationary orbit:

    - Altitude of geostationary satellite: roughly 36,000 km; speed of light = 3 x 10^8 m/s; ground to satellite time = d/c = 0.12 seconds
    - So the shortest possible duration between CTS and ACK is: 0.12s + (t_{DATA}+0.12s) + (t_{ACK} + 0.12s), where t_{DATA} and t_{ACK} are the durations of the data and ACK packets, respectively.
    - At high data rates t_{DATA} and t_{ACK} are << 0.36s, so this this is inefficient.
- A good alternative in this case: ALOHA, where there is no access control and nodes transmit whenever they please. Packets are simply discarded and retransmitted in the event of collisions.

- Second problem: What if not everyone agrees to use MACA? Need a protocol that is robust to interference (i.e., if a collision occurs, the system must handle that situation)
- This situation occurs for point-to-point wireless links in the ISM band, e.g., Bluetooth, which connects consumer electronic devices to a central hub
- Frequency hopping spread spectrum (FHSS): divide time and the available bandwidth into equal-sized chunks.
- In each time segment, choose one of the frequency segments at random from all the possibilities, and only transmit in that segment
- The chance that two devices will choose to transmit in the same segment is very low, so interference is limited
- Also, the chances that anybody else is using a particular band is very low, so they will also cause little interference even if they are not using the same FHSS protocol
- Need a way to recover from the (relatively few) corrupted packets, e.g., retransmission or error-control coding
- FHSS is in fact used in Bluetooth

Mobile Network Layer

- The most important service at the network layer is routing
- IP (internet protocol) is a network layer protocol
- The most widely known feature of IP is the IP address: a unique number assigned to every computer attached to the internet
- E.g., red.cse.yorku.ca = 130.63.96.21
- In multi-hop routing, routers maintain routing tables to ensure that data is passed along the correct route

(Fig. 16)

- Problems with mobile routing:
    - An IP address is assigned on the assumption that the host will remain connected to the same subnet for the entire communication session
    - This allows routing to occur hierarchically, by subnet
    - E.g., 130.63.96.21 is in the 130.63.96 subnet, and routing is handled accordingly (i.e., in a simplistic sense, there is a router assigned to handle all 130.63.96.* traffic)
    - However, this may not be true in mobile networks

(Fig. 17)

- abandoning the subnet structure of the internet is not practical – every device connected to the internet would need a routing table for every other device, AND that table would need to be constantly updated

- mobile devices require a “topologically correct address” to ensure packets are sent to the right place

(Fig. 18)

- protocols like Dynamic Host Configuration Protocol (DHCP) do this already
- problem: what if the IP address changes in the middle of a session? Using DHCP, the connection is basically reset and remaining packets are dropped – not good if you’re downloading a large file or receiving a stream
- So we really need two addresses:
    - Care-of address (COA): This is the topologically correct address, representing the location of the mobile node from an IP perspective
    - Home network address (HNA): A permanent IP address from the perspective of the outside world, so that IP addresses don’t change in the middle of a session
- Two agents are needed, one for each address:
    - The foreign agent (FA), located in the foreign network, corresponds to the mobile node’s current location. For our purposes, the COA will be the IP address of the FA.
    - The home agent (HA), located in the home network, keeps track of the HNA. It also maintains a location registry: for each HNA, it keeps track of the current COA.
    - As the mobile node moves through different foreign networks, the foreign agent communicates the new COA to the HA. The HA then updates its location registry.

- Example. Internet-enabled smart phone. HA=cell provider's server; FA=server located on cell tower.

IP packet delivery

(Fig. 19)

- Suppose we set up a session with some "corresponding node" (CN) on the internet (e.g., YouTube)
- The process:
    - Packet starts out at CN, destined for HNA; home agent is on the path from CN to HNA.
    - Home agent sees the packet destined for HNA. Home agent looks up the HNA's entry in the location registry to find the COA.
    - Home agent repackages the packet as the data in a new IP packet, with the COA as the destination.
    - Packet arrives at the foreign agent (COA). Foreign agent examines the original header of the packet to determine the true destination in the foreign network.

Packet delivery – review

- Suppose we set up a session with some "corresponding node" (CN) on the internet (e.g., YouTube)
- The process:
    - Packet starts out at CN, destined for HNA; home agent is on the path from CN to HNA.
    - Home agent sees the packet destined for HNA. Home agent looks up the HNA's entry in the location registry to find the COA.

    - Home agent repackages the packet as the data in a new IP packet, with the COA as the destination.
    - Packet arrives at the foreign agent (COA). Foreign agent examines the original header of the packet to determine the true destination in the foreign network.
- In reverse, the mobile node returns packets to the source with its HNA as the source and the CN as the destination.
- These packets are sent directly through the internet and the intervention of the HA is not required.

(Fig. 20)

- Thus a virtual network tunnel is set up from CN to mobile node, and the mobility of the mobile node is hidden from the CN (which is what we needed).
- (Example)

Mechanics of Mobile IP

- How does the mobile node figure out what agents to use? Or how does it know that it has moved?

- Two methods:
    - Agent advertisement: Foreign agents and home agents will periodically advertise their presence by sending agent advertisement messages. This method uses the Internet Control Message Protocol (ICMP) and is similar to the way routers advertise themselves on the wired internet.
    - Agent solicitation: If a mobile node doesn't receive an agent advertisement, it can ask for one, again similarly to looking for routers on the wired internet.

    - (Example)
- Registration: needed to inform HA of the mobile node's current location – allows HA to update its location registry
- After discovering the FA, the mobile node sends a registration request to the FA – the FA then forwards it to the HA.
- On receipt, the FA updates its registry by setting up a mobility binding
- Mobility bindings contain the HNA, the COA, and a lifetime – the binding is deleted once the lifetime has expired
- (Example)

(Fig. 21)

Analysis

- What happens when the mobile node moves from one foreign network to another?
- Agent advertisement/solicitation messages and registration must be performed
- However, in the meanwhile, the mobile node has switched networks, and is in the "wrong" subnet as far as routing is concerned
- Message traffic during network change

(Fig. 22)

- Packets may be in transit during the switchover in networks. How many packets are lost?
- Notation:
  - $t_{MN \rightarrow FAi}$ = delay from mobile node to i-th FA
  - $t_{HA \rightarrow FAi}$ = delay from HA to i-th FA
  - t_{DATA} = duration of data packet
- Assumptions: control packets have negligible duration; no delay for computation; no errors.
- Based on the figure, the total delay between the link breaking and being re-established is

(Eq. 26)

- Two options: retransmit or buffer
  - Retransmit: all lost packets are dropped and retransmitted by the CN. Advantage: simple; disadvantages: wastes bandwidth, larger latency.
  - Buffer: a required number of packets are retained by the HA in a buffer in case the foreign network changes. Advantages: low latency; disadvantages: complexity at the HA.
- (Example.)
- Aside from packet loss, out-of-order packet delivery is a problem, whether you buffer or retransmit (though it is better with buffering)

- Soft handover: Maintain simultaneous links to two foreign networks during the handover process to ensure that packets are not lost – also helps with out-of-order delivery
- Disadvantage: Wasteful of bandwidth, though usually for only a short time

(Fig. 23)

Micro-mobility in IPv6

- Cellular IP: a hierarchy of "anchor points"
- MN attaches to a BS, which attaches to a CIP gw
- Similarly in IPv6 we have "hierarchical mobile IPv6" (HMIPv6)
- A mobile anchor point (MAP) appears between the HA and MN
- Attached to the MAP are several access routers
- The MAP acts like a local HA – within the MAP's domain, the MN acquires a new IP address called a link COA (LCOA)
- Packets can be forwarded by the access routers from an old LCOA to a new LCOA

(Fig. 1)

Mobile considerations at the transport layer

- Transport layer services:
    - Sockets by application (e.g., 80=http, 22=ssh)
    - Virtual circuit (in-order packet delivery)

    - Flow control and congestion avoidance
- Broadly speaking:
    - Network layer ensures packets get from one host to another
    - Transport layer maintains an end-to-end connection
- Two most common transport layer protocols:
    - User Datagram Protocol (UDP) – provides sockets but no virtual circuit or flow control – simplest transport layer protocol
    - Transmission Control Protocol (TCP) – provides all major transport layer services
- Sockets don't change in mobile, so no changes are made to UDP in mobile applications – we will focus on TCP

Review of TCP flow control

- from the mobile perspective, only the flow control aspects of TCP are important
- Flow control necessary because links on the internet have different bandwidths – must have a way of avoiding buffer overflow

(Fig. 2)

- Example. Downloading a 1-gigabyte file over a two-hop connection: first link 1 Gbps (e.g., backbone), second link 1 Mbps (e.g., DSL) – if the intermediate router's buffer is small, lots of packets get dropped and retransmitted (congestion)
- Idea of flow control is to use the network maximally without causing lots of congestion
- Key concepts:

    - After each packet is sent, the final destination returns an ACK to indicate successful delivery
    - Congestion window: Number of packets that source can transmit while waiting for ACK
    - Slow start (SS): First phase of TCP – size of congestion window starts at 1 and doubles with every successfully acknowledged packet
    - Congestion avoidance (CA): Second phase of TCP – size of congestion window increases by one with every successfully acknowledged packet
    - Congestion threshold: Size of congestion window at the transition between SS and CA
- Example: Congestion threshold=8

(Fig. 3)

- Eventually, the congestion window will grow so large that congestion occurs – when this happens, intermediate routers will drop packets and those packets will not be ACKed
- Source waits for ACKs until a timeout expires, then declares them lost
- If packets are lost, the congestion threshold is set to one half the current congestion window size, and goes back to slow start
- Example

(Fig. 4)

- The assumption in wired TCP is that all packet losses are caused by congestion, and none are caused by noise
- This is not a valid assumption in wireless networks and causes more slow starts than are necessary – thus lower throughput than necessary

Indirect TCP

- The simplest modification to TCP for wireless links
- The access point (AP) (e.g., foreign agent, base station), which is connected to the wired network, acts as the end point for the TCP connection

(Fig. 1)

- Packets that arrive at the AP are immediately acknowledged (TCP ACK) and are relayed to the mobile node
- Across the wireless link, the mobile node acknowledges packets as they arrive – these ACKs are only used by the AP
- The AP can quickly tell when packets are lost to noise – in this case the AP quickly retransmits without affecting the TCP connection (because all such packets have been acknowledged already by the AP)
- Whenever a packet is lost to congestion, the AP does not return ACK – this is how TCP is expected to work
- Thus, noise on the wireless link is completely isolated from TCP on the wired link
- Handover presents a big problem:
    - The AP holds several packets in its buffer at any given time

    - Say the mobile node switches to a new AP (e.g., new foreign network or foreign agent). The "state" of the old AP (buffered packets, sequence numbers, ports, etc.) must be forwarded to the new AP, since the buffered packets have already been acknowledged.

(Fig. 2)

- Disadvantages of indirect TCP:
    - AP crash is catastrophic: since packets in the AP's buffer have already been acknowledged, they will never be retransmitted – connection is effectively broken
    - Handover latency is a problem – all the buffered packets must be forwarded to the new AP before the TCP connection can resume.

Snooping TCP

- As we saw, segmentation causes disadvantages in indirect TCP – is it possible to maintain end-to-end TCP connection?
- Snooping TCP:
    - The AP/FA is on the path from source to mobile node, but does not segment the link into two; the mobile node is the endpoint of the TCP connection
    - The AP buffers packets as it forwards them, and watches (i.e. "snoops" the connection) for ACKs as the mobile node transmits them
    - If a packet is lost to noise, the AP does not see the ACK, and retransmits the packet from its buffer
    - if this happens fast enough, the mobile gets the packet and sends an ACK before the TCP link times out

    - The AP never acknowledges packets on its own! Thus, complicated handoffs are not needed, and AP crash is not catastrophic

(Fig. 3)

- Disadvantages of snooping TCP:
    - Isolation of wireless link is less effective
    - AP timeout must be much smaller than TCP timeout in order for this method to work properly

- A different problem: what about short disconnections (e.g., a dead zone between APs). This would lead to a lost connection or large buffers in the other two methods.

(Fig. 4)

Mobile TCP (note: not an official standard like Mobile IP)

- Mobile TCP keeps TCP connections alive during short disconnections
- Wireless signal strength is assumed to be good – any dropped packet (via noise or congestion) is forwarded directly from the source. (End-to-end connection is maintained, like in snooping TCP).
- AP monitors the TCP connection, like in snooping TCP – if several ACKs are missed, the AP assumes that the link is disconnected
- Once this happens, AP sends a control message setting source's window size to zero – TCP connection enters

"persistent mode", keeping TCP connection alive but preventing source from sending packets
- Once AP sees ACKs again (either the old AP or a new one), TCP window size is reset to its previous value (control message)

Indirect TCP Analysis – Material not covered Winter 2012

- Some assumptions:
  - Cumulative acknowledgment is used (consequence: an entire window is lost if an acknowledgment fails)
  - Round trip time >> segment length
  - TCP timeout ~= round trip time
  - Indirect TCP is used, so packets are acknowledged when they arrive at the router

(Fig. 1)

- Say the downstream link can send up to k packets at once
- So the router buffer will accumulate packets if more than k packets are sent in a round trip time
- (Note that this assumption is not realistic to analyze slow start – in reality the router would advertise a window size of k, and congestion would never occur)
- What does this look like?

(Fig. 2)

- Area above dotted line is the contents of the router queue
- Say the router queue size is n – if the area above the dotted line is greater than n, packets are dropped
- If ACKs are dropped, all the packets in that column are dropped (due to cumulative acknowledgment)
- For convenience, choose n small enough so that router queue has a chance to drain during slow start
- Example. k=5, n=8.

(Fig. 3)

- Total throughput in steady state: 33 segments per 8 round trip times, or 4.125 segments per round trip time
- Note that link capacity is 5 segments per round trip time
- Now let's suppose that, after each slow start, the 10th segment is lost to fading, and indirect TCP fails to return acknowledgments for these lost packets

(Fig. 4)

- Total throughput in steady state: 6 segments per 4 round trip times, or 1.5 segments per round trip time
- This is much less than the capacity of the link, even accounting for losses
- Now consider packet losses when indirect TCP is working. For each packet loss to fading, an extra transmission is needed from the router, so at worst we will add one to the size of the queue

(Fig. 5)

- steady state throughput is the same as for regular TCP
- so indirect TCP delivers a large improvement!

Cellular Radio Telephone Systems: The Cellular Concept

- Mobile handsets communicate with “base stations”
- Base stations are connected to the normal telephone exchange / internet (for data)
- Normally, a handset will communicate with the closest base station – thus, space is divided into regions that are closest to a given base, and all mobiles within that region communicate with the corresponding base station
- These regions are called “cells” hence “cellular radio”

Why Cellular?

- Main idea: Spatial reuse of bandwidth
- Say you have a bandwidth B. Say the bandwidth is shared using FDMA, and each user takes u. Then the number of (active) users is B/u.
- Obviously if two systems are far enough apart, they can use the same bandwidth. (No interference.)
- If all you have is B, then each region needs to be surrounded by a "dead zone".
- But what if regions in the "dead zone" used different frequencies? Each one is allocated a bandwidth B of different frequencies; again as long as they are far enough apart, that bandwidth can be reused
- Main idea: A repeating pattern of regions (i.e., cells) that use the same frequencies; as long as cells using the same frequencies are far enough apart, no interference
- Let B_T be the total bandwidth – the cellular reuse factor (the number of cells in a reuse pattern) is k – then the bandwidth per cell $B = B_T/k$
- Example

Cellular networks and mobility

- Problem: Crossing cell boundaries, need to reconnect to a new cell
- Handoff – disconnect from old base station, reconnect to new base station
    - Soft handoff – for a period of time, can connect to multiple cells at once (e.g. on the border)
    - Hard handoff – Old connection dropped, new connection established

History

- Concept of mobile telephony dates to the 1940s

    - First city with a mobile telephone system: St. Louis
- Cellular radio: (AMPS, 1G)
    - First implemented in Chicago, 1977
    - Strictly analog
- Digital telephony (2G)
    - Digital voice, GSM, SMS
    - First deployed in Finland, 1991
- High speed IP data to phones (3G)
    - Idea: Broadband speeds to mobile devices using IP
    - 3G "dongles" (connection points to traditional machines)
    - First city with 3G: Tokyo, 2001
- LTE, 4G
    - All-IP, all-packet-switched networks
    - Currently being rolled out around the world
    - First in Stockholm and Oslo, 2009.
    - Goal: Ultra-broadband access everywhere to any device

Cellular telephone systems: GSM

- GSM = Global System for Mobile Communication
- Most widely used cellular telephone service in the world
- Let's first consider the radio interface – uses a combination of FDMA and TDMA
- Frequencies
    - GSM 900: 890-915 MHz and 935-960 MHz
    - GSM 1800: 1710-1785 MHz and 1805-1880 MHz (different frequencies in the USA)
- Duplexing

    - A “full-duplex” link maintains a completely bi-directional channel at all times
    - All telephone connections are full-duplex (distinct from half-duplex, where the uplink and downlink can only exist one at a time, like a walkie-talkie)
    - GSM uses frequency division duplexing – this is why there are two frequency ranges given above
    - uplink is the lower set of frequencies, and downlink is the upper set – when a call is set up you are assigned both an uplink and a downlink channel
- Division of the channel
    - Take GSM 900 uplinks as an example (downlink and GSM 1800 are similar)
    - FDMA: Frequencies divided into 124 channels of 200 kHz each
    - TDMA: Within each channel, time is divided into frames of 4.615 ms each; each frame contains 8 slots of duration 577 μs (a slot is what is assigned to a user)

(Fig. 1)

    - Slots consist of 546.5 μs “burst” and 30.5 μs of guard time (why is guard time needed?)
    - Burst contains 148 bits (raw bit rate = 271 kbps) – contains 6 “tail bits” (3 on each end), 2 “S bits” (control), 26 training bits, and 114 data bits
    - Bit rate per user = 114/0.004615 = 24.7 kbps

(Fig. 2)

GSM Architecture

- Three major subsystems:
  - Radio subsystem – handles radio tasks, consists of mobiles and base stations
  - Network and switching subsystem – routes calls, performs handovers, localizes users worldwide
  - Operation subsystem – Security, authentication, billing.
- Radio subsystem:
  - Base stations are organized hierarchically into base station subsystems (BSS), containing several cells – tasks are divided into base transceiver station (BTS) and base station controller (BSC)
  - BTS: radios, antennas, signal processing, and amplifiers – one per cell
  - BSC: organizes radio frequencies, handles handovers within a given BSS – one per BSS
  - Recall Cellular IP

(Fig. 3)

- Mobile station (MS) is also part of the radio subsystem – most important task, aside from the obvious, is to hold the subscriber identity module (SIM)
- Network and switching subsystem:
  - Mobile services switching center (MSC): connect to several BSCs, each other, and possibly the outside world (gateway MSC) – handovers, connection to the PSTN, etc.
  - Home location register (HLR): Master database of all user data, e.g., identity, location

    - Visitor location register (VLR): One per MSC, contains copied information of all HLR information in that MSC's area
    - Compare HLR and VLR with home agent and foreign agent

(Fig. 4)

GPRS

- Global Packet Relay Service – a simple data standard associated with GSM
- Main idea: Allocate unused slots within a GSM frame to data
- Time-division duplex: some slots are used for data transmission, others for data reception
- Data rates from 9 kbps to 171 kbps are possible, though the instantaneous rate depends on the load in the cell (slots may be in use by other users).

Wireless LAN protocols

- Why wireless LAN?
    - Convenience and flexibility. No need to carry wires or retrofit buildings, (limited) mobility, no connectors (e.g. Macbook Air).
- Why not wireless LAN?
    - Not good if very high bandwidth is required
    - RF compatibility and other safety issues
    - Security (?)

IEEE 802.15.1 (Bluetooth)

- First let's consider Bluetooth, a very simple wireless LAN protocol – named for King Harald Bluetooth, a 10th century Danish king
- Bluetooth connects small devices over a very short range (~1-10m), depending on "class" – though longer ranges are possible – "personal area network" for connecting devices in your immediate vicinity

Bluetooth architecture

- Hierarchical "master-slave" system
- A collection of connected Bluetooth devices is a "piconet"
- Piconets consist of four kinds of device:
    - Master (one per piconet)
    - Slaves (up to seven per piconet, directly connected to the master, actively transmitting data)
    - Parked (known to the master but not transmitting data)
    - Standby (idle)

(Fig. 1)

- Any device can act as either master or slave; also, piconets can overlap (so it's maybe not appropriate to think of the master as a "base station")
- Active devices (master + slaves) are assigned a 3-bit active member address (AMA) (this is why there can be at most 7 slaves)

- Parked devices are assigned an 8-bit parked member address (can be upgraded to active members as needed or as a slot becomes available)
- It is possible for slaves to belong to more than one piconet; it is also possible for a master to be a slave in another piconet – however not simultaneously – the devices jump back and forth

(Fig. 2)

- it is not possible for a device to be a master in two piconets (in that case the piconets would merge into one)

Bluetooth Radio Interface

- Bluetooth operates in the 2.4GHz ISM band
- Multiple access via frequency hopping spread spectrum (FHSS)
    - 79 "hop carriers" each with 1 MHz of bandwidth
    - 1600 hops/s (625 μs per "slot")
    - so every 625 μs, the system occupies a different, pseudo-randomly selected 1 MHz frequency range

(Fig. 3)

- Why do this?
    - Devices in the ISM band must tolerate interference from other devices in the same band

    - By randomly jumping from frequency to frequency, it is unlikely (but not impossible!) that interference will be encountered
- When a piconet is formed, the master sets the hopping pattern
- All devices have a unique 48-bit device ID, which establishes the hopping pattern (when master)
- Time-division duplexing: transmissions proceed as master, slave 1, master, slave 2, master, etc.
- It is possible for transmissions to occupy 3 or 5 slots, as needed – in this case no hopping is performed during the block, and intermediate hops are skipped

(Fig. 4)

- data payload is up to 343 bits for single-slot packets
- What is the data rate? – given 1 slave, 2 slaves ... – for the master

IEEE 802.11 (WiFi)

- The world's most widely used wireless LAN standard
    - 802.11 (1997): 1-2 Mbps
    - 802.11a (1999): 27-54 Mbps (however, limited range – only 802.11 outside of 2.4 GHz)
    - 802.11b (1999): 11 Mbps
    - 802.11g (2003): 22-54 Mbps
    - 802.11n (2010?): 108-600 Mbps
- Compare with Ethernet – note – Cat 5 wire is most commonly used with 100 Mbps and is temperamental at

higher data rates (e.g., 1 Gbps Ethernet) – with 802.11n, all you need to do is upgrade the terminals

WiFi architecture

- WiFi architecture is simple and closely resembles Ethernet architecture
- "stations" are connected to "access points" within "basic service set" (BSS) areas
- multiple BSS can be connected together into an "extended service set (ESS) area (e.g. AirYork)

(Fig. 1)

- access points in WiFi support roaming!
- Protocol architecture:
    - Like Ethernet, WiFi only specifies the bottom two layers (physical and MAC)
    - Intentional: so that WiFi can fit with Ethernet routers

(Fig. 2)

WiFi Radio Interface

- most versions operate in the 2.4 GHz ISM band

- WiFi divides the ISM band into channels – number of channels varies with location, 11 in North America, 13 in Europe, 14 in Japan
- Channels are 22 MHz wide and spaced at intervals of 5 MHz, thus overlap
- Choose non-overlapping channels for a large-scale WiFi installation (similar to cells)

(Fig. 3)

- Information transmitted using frequency hopping spread spectrum (FHSS), as in Bluetooth, or direct sequence spread spectrum (DSSS), as in CDMA – DSSS in 802.11b

WiFi Medium access control

- Basic principle: Carrier-sense multiple access with collision avoidance (CSMA/CA)
- Similarly to Ethernet, nodes sense the medium and only transmit when it is idle
    - Once the medium is idle, everyone waits an additional period of time (Distributed Coordination Function Inter-Frame Spacing, DIFS) before transmitting – this allows high-priority messages to get through
    - After DIFS has expired, each node selects a random backoff time and continues to wait
    - If at any point up to the expiry of the backoff timer the medium is busy, the node must start over
    - Otherwise, once the backoff timer expires, the node may transmit

(Fig. 1)

- This is possibly unfair if a node repeatedly selects a large backoff time – that node won't be able to transmit
    - If a backoff timer is interrupted by a transmission, the node keeps the old value of the backoff timer and continues from where it left off
    - Maximum idle time is thus limited to the original value of the backoff timer
    - Example
- Shorter delays are possible for high priority messages
    - Short inter-frame spacing (SIFS) – for control messages, e.g., packet acknowledgments
    - Point coordination function inter-frame spacing (PIFS) – for time-bounded services.

(Fig. 2)

Wireless Sensor Networks

- By now, two-way digital radios are inexpensive, as is computing power
- In many applications, the more data collected (from as close as possible to the phenomenon), the better
    - Example. Wildlife tracking.
    - Conventional method: Radio transponder collars.

    - New method: Sensor networking collars that continuously monitor themselves and their neighbors

(Fig. 3)

- Sensor network features
    - Inexpensive – many devices can be used, okay to lose them, add more as needed
    - Robust – Tolerant of device failure/addition
    - Distributed – auto-configuration, no single master or point of failure
    - Ubiquitous – measurements taken from as many locations as possible, as close to the phenomenon as possible
- Sensor network challenges
    - Power management – how to optimize devices for extremely long life?
    - Low complexity – how to deal with devices that have reduced computing requirements?
    - Organization – how to perform network control in a distributed manner?
    - Routing – how to get data from one place to another in a dynamic network?

IEEE 802.15.4

- IEEE 802.15.4 is a broad wireless networking standard that addresses some of the challenges in sensor networking

- As in Bluetooth and WiFi, only the bottom two layers are specified – certain protocols, e.g. ZigBee, are extensions of 802.15.4 into higher layers
- 802.15.4 physical layer:
    - ISM bands at 2.4 GHz (worldwide); other channels in North America and Europe
    - Data rates ranging from 20-250 kbps
- 802.15.4 MAC layer:
    - Direct sequence spread spectrum; slots and/or CSMA/CA (same as WiFi).
- 802.15.4 architecture:
    - Unlike Bluetooth, distinguishes between "reduced function devices" (RFD) and "full function devices" (FFD)
    - Only FFDs can be "coordinators" (like masters in Bluetooth); RFDs only connect to FFDs

IEEE 802.15.4

- 802.15.4 architecture:
    - Unlike Bluetooth, distinguishes between "reduced function devices" (RFD) and "full function devices" (FFD)
    - Only FFDs can be "coordinators" (like masters in Bluetooth); RFDs only connect to FFDs
    - Star and peer-to-peer topologies are possible
    - Star topologies can be organized into a hierarchy
        - Network coordinator (global)
        - PAN coordinator (local star cluster)
        - Non-coordinator FFD and RFD (leaves)

(Fig. 1)

Beacons and Superframes

- IEEE 802.15.4 permits the organization of time into "superframes"
- Coordinators send "beacon" packets to organize the superframes (for synchronization and control)
- A superframe consists of:
    - Contention access period (CAP) – e.g., access using CSMA
    - Contention free period (CFP) – guaranteed slots for certain devices, assigned by coordinator
    - Inactive period
    - Beacon interval and superframe duration specified by coordinator in the beacon packet
    - Why have an inactive period? – to allow devices to sleep.

(Fig. 2)

Routing in sensor networks: AODV

- AODV = Ad hoc On demand Distance Vector routing
- One of the routing techniques allowed in ZigBee
- Features:
    - Strictly on-demand: routes don't keep routing tables unless in an active route; routes are only formed when needed
    - Avoid stale routes: network is dynamic so all routes must be maintained "fresh"
    - Local and distributed: global coordination and routing not needed

- Path Discovery
    - When a node needs a route to the destination, it transmits a Route Request (RREQ) to its neighbors
    - RREQ consists of: source address, source sequence #, broadcast ID, destination address, destination sequence #, hop count
    - Source address and broadcast ID identify the RREQ
    - Source sequence # maintains “freshness” of reverse route to the source
    - Destination sequence # is the last known sequence # used by the destination – source won’t accept any route with a more “stale” sequence number
    - If the neighbor has a route to the destination, it informs the source
    - If no route to the destination, neighbor broadcasts the RREQ to its neighbors & increments hop count

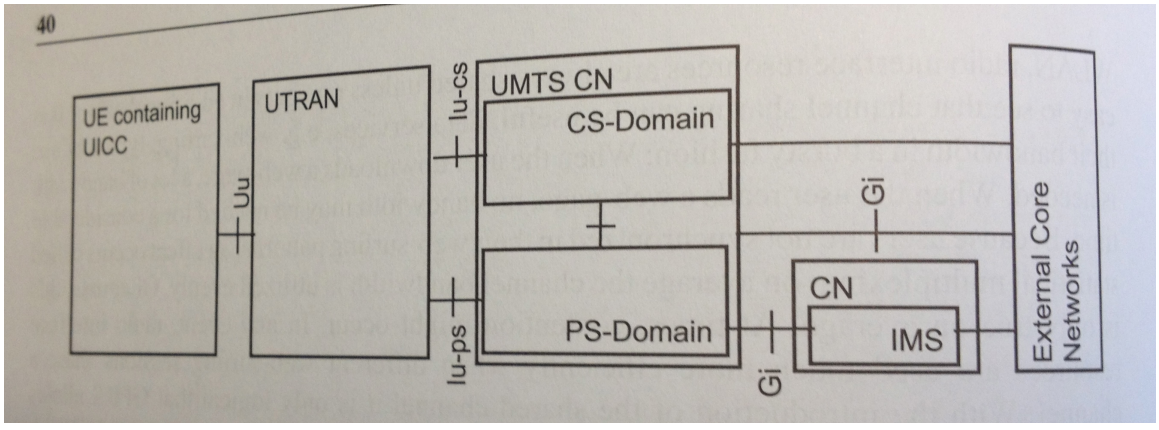
- Reverse Path
    - RREQ travels from the source to various possible intermediate routes – sets up an automatic reverse path
    - Reverse path entries are maintained by each node for at least enough time for the route information to propagate back to the source
- Forward path setup
    - Eventually a path is found – either the RREQ arrives at the destination, or a node with a route to the destination
    - Freshness: If not the destination, the destination sequence #s are compared – if the node’s dsn is smaller than the source’s dsn, this is a “stale” route and the RREQ is forwarded again
    - Otherwise the node returns a route reply (RREP) to the neighbor from which it received the RREQ

- RREP propagates back towards the source, and the route is set up
- Other nodes drop the routing information after a timeout has expired

(Fig. 3)

Introduction to 3G, 4G, and LTE wireless networks

- 3G: Two main standards
  - UMTS (Universal Mobile Telecommunication System)
  - CDMA2000
- UMTS evolved from GSM
  - Designed to work with GSM networks
  - Simplified UMTS architecture



(From C. Kappler, *UMTS Networks and Beyond*, Wiley, 2009.)

UE = user equipment (mobile)
UTRAN = UMTS Terrestrial Radio Access Network (RAN)
CS-domain = Circuit-switched (phone)

PS-domain = Packet-switched (internet)
CN = core network
IMS = IP multimedia subsystem

UMTS physical layer

- Based on CDMA (Code Division Multiple Access)

3GPP (3rd Generation Partnership Project)

- The global standards body for high speed wireless communication
- Originally to create a global standard for "3G" digital wireless mobile communication
- Currently has three missions:
    - Maintain GSM
    - Develop wireless standards beyond 3G
    - Develop IP Multimedia Subsystems (even beyond wireless – we won't talk about this much)

- Membership of 3GPP
    - Organizational Partners (determine policy and set standards)
        - Industry consortia from around the world (Asia, Europe, North America)
        - Membership includes hundreds to thousands of companies worldwide: providers, equipment vendors, government agencies, etc.
    - Market Representation Partners (advisory)
        - "Forums" representing specific technologies and industries

- 3GPP standards are organized into "releases"
    - Weirdly, the releases are numbered like this:

        - Phase 1, Phase 2
        - Release 96, Release 97, Release 98, Release 99
        - Release 4, Release 5, …, Release 12
    - Release 4 was the 4th release since the start of 3GPP (Phase 1, Phase 2, Release 96 preceded it)
    - Release 99 = 3G (UMTS)
    - Release 8 = LTE (considered 4G – see below)
    - Current deployed release: Release 10

- BUT – there are other, competing standards bodies in the world (IEEE, ITU etc.) that have some say in defining terms
- ITU describes what counts as 4G (also called IMT-Advanced):
    - Based on an all-Internet Protocol (IP) packet switched network[1]
    - Interoperability with existing wireless standards[2]
    - A nominal data rate of 100 Mbit/s while the client physically moves at high speeds relative to the station, and 1 Gbit/s while client and station are in relatively fixed positions.[3]
    - Dynamically share and use the network resources to support more simultaneous users per cell.
    - Scalable channel bandwidth 5–20 MHz, optionally up to 40 MHz[4][5]
    - Peak link spectral efficiency of 15 bit/s/Hz in the downlink, and 6.75 bit/s/Hz in the uplink (meaning that 1 Gbit/s in the downlink should be possible over less than 67 MHz bandwidth)
    - System spectral efficiency of up to 3 bit/s/Hz/cell in the downlink and 2.25 bit/s/Hz/cell for indoor usage[4]
    - Seamless connectivity and global roaming across multiple networks with smooth handovers[1][6]
    - Ability to offer high quality of service for multimedia support
- (Source: Wikipedia)
- LTE does not achieve this (e.g. 1 GBit/s is beyond its theoretical capabilities) – but ITU has otherwise recognized LTE as a 4G technology

Brief aside: spectral efficiency

- Multiple bits per signaling interval
- Effect on bandwidth

3GPP Technologies

HSPA: High Speed Packet Access

- HSDPA (Release 5)
    - Custom transport layer: High-speed downlink shared channel (HS-DSCH) – channel quality information is explicitly exchanged, data rate modified to suit
        - One slot: Announce data is coming
        - Next slot: ACK, with channel quality
        - Next slot: Data transmitted at an appropriate rate for channel conditions
    - Packet scheduling – adaptively routing more data to users that can handle it (e.g. high signal quality)
    - HARQ: error correction, with retransmission for severe errors
    - Up to 14 Mbit/s
- HSUPA (Release 6)
    - Similar improvements for HSDPA but for uplink
    - More challenging because UE is less capable than base station
    - Up to 5.76 Mbit/s
- HSPA+ (Release 7)
    - MIMO
    - Higher order modulation (up to 6 bits/symbol)
    - 168 Mbit/s downlink
    - 22 Mbit/s uplink
- Dual-cell HSDPA (Release 8)
    - Can assign carriers from more than one cell to UE
    - That is, can exceed the available bandwidth of a single cell to the UE
    - Also: Joint optimization of carriers and resource assignments across pairs of cells
- Release 9:
    - Dual-cell HSUPA, Dual-cell HSDPA with MIMO

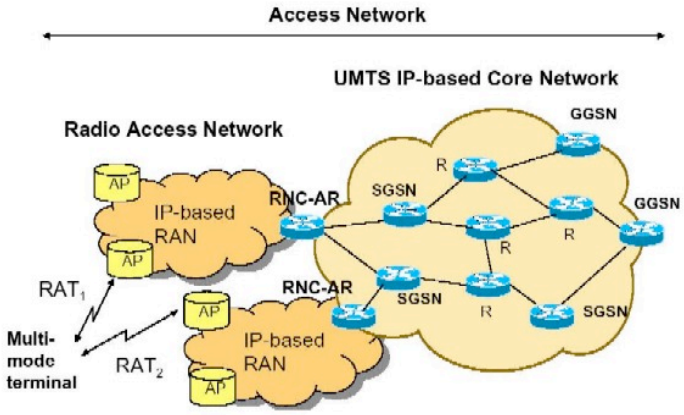
- Release 10:
    - 4-cell HSDPA

- Generic Access Networks (GAN) (Release 6)
    - Remember 3G/4G networks are all-IP, packet switched networks
    - GAN provides external IP access into their core networks (allowing them to integrate with regular internet services)
    - Two examples:
        - Handoff from provider's base station to home WiFi router
        - Femtocells [brief intro to femtocells]
    - Since this equipment is connected to the generic "internet" and not a provider's equipment, it needs access to the core network through IP

- Example: iPhone
    - iPhone 3G: UMTS/HSDPA (Release 5)
    - iPhone 3GS: UMTS/HSDPA (Release 5)
    - iPhone 4: UMTS/HSDPA/HSUPA (Release 6)
    - iPhone 4S: UMTS/HSPA+ (14.4 Mbit/s) (Release 7)

UTRAN and eUTRAN

- Radio Access Network: Connectivity from mobile device to core network (normally through base stations, base station controllers – remember GSM)
- UTRAN: The RAN for UMTS (UTRAN = "Universal Terrestrial Radio Access Network"
-



- source: http://www.aroma-ist.upc.edu/Imagenes/fig1.jpg

- Access points in UTRAN are called “Node B” – these are basically the base stations
- Although with multi-cell HSDPA and MIMO the “cellular” paradigm is not the best way to describe the system any more
- RNC = Radio network controller

- Future releases under 3GPP will move away from HSPA and towards an evolved UTRAN, called eUTRAN

- eUTRAN moves away from a CDMA-based network (UMTS) and towards an OFDMA-based network