UNIT II
CELLULAR ARCHITECTURE

Different multiple access schemes:

Frequency Division Multiple Access

FDMA is the oldest, and conceptually most simple, multi-access method. Each user is assigned a frequency (sub) band – i.e., a (usually contiguous) part of the available spectrum. The assignment of frequency bands is usually done during call setup, and retained during the whole call. FDMA is usually combined with the Frequency Domain Duplexing (FDD), so that two frequency bands (with a fixed duplex distance) are assigned to each user: one for downlink (BS-to-MS) and one for uplink (MS-to-BS) communication. Pure FDMA is conceptually very simple, and has some advantages for implementation:

• The transmitter (TX) and receiver (RX) require little *digital signal processing*. However, this is not so important in practice anymore, as the costs for digital processing are continuously decreasing.

• (*Temporal*) *synchronization is simple*. Once synchronization has been established during the call setup, it is easy to maintain it by means of a simple tracking algorithm, as transmission occurs continuously. However, pure FDMA also has significant disadvantages, especially when used for speech communications. These problems arise from spectral efficiency considerations, as well as from sensitivity to multipath effects:

• *Frequency synchronization and stability are difficult*: for speech communications, each frequency subband is quite narrow (typically between 5 and 30 kHz). Local oscillators thus must be very accurate and stable; jitters in the carrier frequency result in adjacent channel interference. High spectral efficiency also requires the use of very steep filters to extract the desired signal. Both accurate oscillators and steep filters are expensive, and thus undesirable. If they are not admissible, guard bands can be used to mitigate filter requirements. This, however, reduces the spectral efficiency of the system.

• *Sensitivity to fading*: since each user is assigned a distinct frequency band, these bands are narrower than for other multiaccess methods (compare TDMA, CDMA) – i.e., 5–30 kHz. For such narrow subbands, fading is flat in practically all environments. This has the advantage that no equalization is required; the drawback is that there is no frequency diversity. Remember that frequency diversity is mainly provided by signal components that are more than one channel coherence bandwidth apart.

• *Sensitivity to random Frequency Modulation (FM)*: due to the narrow bandwidth, the system is sensitive to random FM: the Bit Error Rate (BER) due to random FM is proportional to $(v\max T_S)2$ Thus, it is inversely proportional to the square of the bandwidth. On the positive side, appropriate signal-processing schemes can not only mitigate these effects but even exploit them to obtain time diversity. Note that the situation here is dual to wideband systems, where delay dispersion can be a drawback, but equalizers can turn them into an asset by exploiting frequency diversity.

• *Intermodulation*: the BS needs to transmit multiple speech channels, each of which is active the whole time. Typically, a BS uses 20–100 frequency channels. If these signals are amplified by the same power amplifier, third-order modulation products can be created, which lie at undesirable frequencies – i.e., within the transmit band. We thus need either a separate amplifier for each speech channel, or a highly linear amplifier for the composite signal – each of these solutions makes a BS more expensive.

Time Division Multiple Access

For TDMA, different users transmit not at different frequencies but rather at different times. A time unit is subdivided into N timeslots of fixed duration, and each user is assigned one such timeslot.During the assigned timeslot, the user can transmit with a high data rate (as it can use the whole system bandwidth); subsequently, it remains silent for the next $N - 1$ timeslots, when other users take their turn. This process is then repeated periodically. At first glance, this approach has the same performance as FDMA: a user transmits only during $1/N$ of the available time, but then occupies N times the bandwidth. However, there are some important practical differences:

• Users occupy a larger bandwidth. This allows them to exploit the frequency diversity available within the bandwidth allocated to the system; furthermore, the sensitivity to random FM is reduced. On the flipside, equalizers are required to combat InterSymbol Interference (ISI) for most operating environments; this increases the effort needed for digital signal processing.

• Temporal guard intervals are required. A TX needs a finite amount of time to ramp up from 0-W output power to “full power” (typically between 100mW and 100 W). Furthermore, there has to be sufficient guard time to compensate for the runtime of the signal between the MS and BS. It is possible that one MS is far away from the BS, while the one that transmits in the subsequent timeslot is very close to the BS and thus has negligible runtime. As the signals from the two users must not overlap at the BS, the second MS must not transmit during the time it takes the first signal to propagate to the BS.3 Note, however, that there is no need for frequency guard bands, as each user completely fills up the assigned band.

• Each timeslot might require a new synchronization and channel estimation, as transmission is not continuous. Optimization of timeslot duration is a challenging task. If it is too short, then a large percentage of the time is used for synchronization and channel estimates (in GSM, 17% of a timeslot are used for this purpose). If the timeslot is too long, transmission delays become too long (which users find annoying especially for speech communications), and the channel starts to change during one timeslot. In that case, the equalizer has to track the channel during transmission of a timeslot, which increases implementation effort (this was required, e.g., in the – now defunct – Interim Standard (IS)-136 cellular standard). If the time between two timeslots assigned to one user is larger than coherence time, the channel has changed between these two timeslots, and a new channel estimate is required.

• For interference-limited systems, TDMA has a major advantage: during its period of inactivity, the MS can "listen" to transmission on other timeslots.4 This is especially useful for the preparation of handovers from one BS to another, when the MS has to find out whether a neighboring BS would offer better quality, and has communications channels available.

Carrier Sense Multiple Access

A TX can determine (*sense*) whether the channel is currently occupied by another user (*carrier*). This knowledge can be used to increase the efficiency of a packet-switched system: if one user is transmitting, no other user is allowed to send a signal. Such a method is called *CSMA*. It is more efficient than ALOHA, because a TX does not disturb other users that are already on the air. The most important parameters of a CSMA system are detection delay and propagation delay.

Detection delay is a relative measure for how long it takes a TX to determine whether the channel is currently occupied. It depends essentially on the hardware of the system, but also on the desired false alarm probability and the SNR. *Propagation delay* is the measure of how long a data packet takes to get from the MS to the BS. It can happen that at time $t1$, TX 1 determines that the channel is free, and thus sends off a packet. At time $t2$ another TX senses the channel. If $t2 - t1$ is shorter than the time it takes data packet A to get from TX 1 to TX 2, then TX 2 determines that the channel is free, and sends off data packet B. In such a case a collision occurs. This description makes it clear that detection delay and propagation delay should be much smaller than packet duration.

CSMA:

A TX can determine (*sense*) whether the channel is currently occupied by another user (*carrier*). This knowledge can be used to increase the efficiency of a packet-switched system: if one user is transmitting, no other user is allowed to send a signal. Such a method is called *CSMA*. It is more efficient than ALOHA, because a TX does not disturb other users that are already on the air. The most important parameters of a CSMA system are detection delay and propagation delay.

Detection delay is a relative measure for how long it takes a TX to determine whether the channel is currently occupied. It depends essentially on the hardware of the system, but also on the desired false alarm probability and the SNR. *Propagation delay* is the measure of how long a data packet takes to get from the MS to the BS. It can happen that at time $t1$, TX 1 determines that the channel is free, and thus sends off a packet. At time $t2$ another TX senses the channel. If $t2 - t1$ is shorter than the time it takes data packet A to get from TX 1 to TX 2, then TX 2 determines that the channel is free, and sends off data packet B. In such a case a collision occurs. This description makes it clear that detection delay and propagation delay should be much smaller than packet duration.

Implementation of Carrier Sense Multiple Access

There are different methods of implementing CSMA.

• *Nonpersistent CSMA*: the TX senses the channel. If the channel is busy, the TX waits a random time duration until retransmission.

• *p-Persistent CSMA*: this method is applied in slotted channels. When a TX determines that a channel is available, it transmits with probability p in the subsequent frame; otherwise, it transmits one timeslot later.

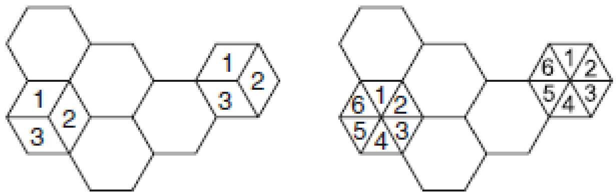
• *1-Persistent CSMA*: the TX constantly senses the channel, until it realizes that the channel is free; then it immediately sends off the packet. This is obviously a special case of p-persistent transmission, with $p = 1$.

• *CSMA with collision detection*: in this method, a node observes whether two TXs start to transmit simultaneously. If that is the case, transmission is immediately terminated. This approach is not commonly used for wireless packet radio.

• *Data Sense Multiple Access* (DSMA): in this approach, the downlink includes a control channel, which transmits at periodic intervals a “busy/available” signal that indicates the state of the channel. If a user finds the channel to be free, it can immediately send off a data packet. Note that for peer-to-peer networks, implementation of the control channel is more difficult than in a scenario with a central node (BS).

Methods that increase the system capacity:

System capacity is the most important measure for a cellular network. Methods for increasing capacity are thus an essential area of research.
1. *Increasing the amount of spectrum used*: this is the “brute force” method. It turns out to be very expensive, as spectrum is a scarce resource, and usually auctioned off by governments at very high prices. Furthermore, the total amount of spectrum assigned to wireless systems can change only very slowly; changes in spectrum assignments have to be approved by worldwide regulatory conferences, which often takes ten years or more.
2. *More efficient modulation formats and coding*: using modulation formats that require less bandwidth (higher order modulation) and/or are more resistant to interference. The former allows an increase in data rate for each user (or an increase in the number of users in a cell while keeping the data rate per user constant). However, the possible benefits of higher order modulation are limited: they are more sensitive to noise and interference , so that the reuse distance might have to be increased. The use of interference-resistant modulation allows a reduction in reuse distance. The introduction of near-capacity-achieving codes – turbo codes and low-density parity check codes) – is another way of achieving better immunity to interference, and thus increases system capacity.
3. *Better source coding*: depending on required speech quality, current speech coders need data rates between 32 kbit/s and 4 kbit/s. Better models for the properties of speech allow the data rate to be decreased without decreasing quality . Compression of data files and music/video compression also allows more users to be served.
4. *Discontinuous Voice Transmission* DTX: exploits the fact that during a phone conversation each participant talks only 50% of the time. A TDMA system can thus set up more calls than there are available timeslots. During the call, the users that are actively talking at the moment are multiplexed onto the available timeslots, while quiet users do not get assigned any radio resources.
5. *Multiuser detection*: this greatly reduces the effect of interference, and thus allows more users per cell for CDMA systems or smaller reuse distances for FDMA systems
6. *Adaptive modulation and coding*: employs the knowledge at the TX of the transmission channel, and chooses the modulation format and coding rate that are “just right” for the current link situation. This approach makes better use of available power, and, among other effects, reduces interference
7. *Reduction of cell radius*: this is an effective, but very expensive, way of increasing capacity, as a new BS has to be built for each additional cell. For FDMA systems, it also means that the frequency planning for a large area has to be redone.8 Furthermore, smaller cells also require more handovers for moving users, which is complicated, and reduces spectral efficiency due to the large amount of signaling information that has to be sent during a handover.
8. *Use of sector cells*: a hexagonal (or similarly shaped) cell can be divided into several (typically three) sectors. Each sector is served by one sector antenna. Thus, the number of cells has tripled, as has the number of BS *antennas*. However, the number of BS *locations* has remained the same, because the three antennas are at the same location .
9. *Use of an overlay structure*: an overlay structure combines cells with different size and different traffic density. Therefore, some locations may be served by several BSs simultaneously. An *umbrella cell* provides basic coverage for a large area. Within that coverage area, multiple microcells are placed in areas of high traffic density. Within the coverage area of the microcells, most users are served by the microcell BS, but fast-moving users are assigned to the umbrella cell, in order to reduce the number of handovers between cells. Around 2010, so-called *femtocells* have been introduced, which are intended to be installed in apartments and offices. The BSs of those femtocells are connected to the cellular network via the internet connection (cable, Digital Subscriber Line, (DSL)) of the customer on whose premises the femtocell is located. The main purpose of such a femtocell is to provide good coverage/datarate for the apartment/office in which it is installed. The main challenge is the integration of femtocells into the overall network; since it can be anticipated that a large number of such cells will exist in the future, Self Organizing Networks (SONs) seem the best way for such integration.



(a) 120°-sectorization (b) 60°-sectorization

10. *Multiple antennas*: these can be used to enhance capacity via different scenarios:
(a) diversity (Chapter 13) increases the quality of the received signal, which can be exploited to increase capacity – e.g., by use of higher order modulation formats, or reduction of the reuse distance;
(b) multiple-input multiple-output systems (Section 20.2) increase the capacity of each link;
(c) space division multiple access (Section 20.1) allows several users in the same frequency channel in the same cell to be served.
11. *Fractional loading*: this system uses a small reuse distance, but uses only a small percentage of the available timeslots in each cell. This leads to approximately the same average capacity as the "conventional" scheme with large reuse distance and full loading of each cell. However, it has higher flexibility, as throughput can be made higher in some cells when throughput in other cells is low.
12. *Partial frequency reuse*: in this scheme, the available spectrum is divided into $N + 1$ subbands.One subband is used in *all* the cell centers, while the other subbands are used at the cell edges, employing a conventional frequency reuse (with cluster size N). The "cell edges" must be large enough so that interference from one cell center to another is sufficiently weak. We also note that the subbands need not all have the same bandwidth. Depending on the size of the cell center, the subband used in the center might be larger than the bands used at the edges.

Differences between TDMA, FDMA and CDMA:

Frequency division multiple access (FDMA):

It is a technology by which the total bandwidth available to the system is divided into frequencies. This division is done between non overlapping frequencies that are then assigned to each communicating pair (2 phones).FDMA is used mainly for analog transmission. Its not that this technology is not capable of carrying digital information, but just that it is not considered to be an efficient method for digital transmission. In FDMA all users share the satellite simultaneously but each user transmits at single frequency.

Code division multiple access (CDMA):

Unlike FDMA, CDMA separates calls by code. Every bit of a conversation is been tagged with a specific and unique code. The system gets a call, it allocates a unique code to that particular conversation, now the data is split into small parts and is tagged with the unique code given to the conversation of which they are part of. Now, this data in small pieces is sent over a number of the discrete frequencies available for use at any time in the specified range. The system then at the end reassembles the conversation from the coded bits and deliver it

Time division multiple access (TDMA):

Unlike FDMA and CDMA, In TDMA the division of calls happens on time basis. The system first digitizes the calls, and then combines those conversations into a unified digital stream on a single radio channel. Now it divides each cellular channel into three time slots that means three calls get put on a single frequency and then, a time slot is assigned to each call during the conversation, a regular space in a digital stream. The users transmit in rapid succession, one after the other, each using its own time slot. This allows multiple stations to share the same transmission medium (e.g. radio frequency channel) while using only a part of its channel capacity.
This technology enables three different users to use one frequency at the same time.
Here there is no need for three separate frequencies like in FDMA. As in FDMA, instead of monopolizing a single radio channel for a single call, TDMA efficiently carries three calls at the same time. BTW, this technology is the one used in our GSM system

FDMA (frequency division multiple access) is the division of the frequency band allocated for wireless cellular telephone communication into 30 channels, each of which can carry a voice conversation or, with digital service, carry digital data. FDMA is a basic technology in the analog Advanced Mobile Phone Service (AMPS), the most widely-installed cellular phone system installed in North America. With FDMA, each channel can be assigned to only one user at a time. FDMA is also used in the Total Access Communication System (TACS).
The Digital-Advanced Mobile Phone Service (D-AMPS) also uses FDMA but adds time division

multiple access (TDMA) to get three channels for each FDMA channel, tripling the number of calls that can be handled on a channel.

TDMA (time division multiple access) is a technology used in digital cellular telephonecommunication that divides each cellular channel into three time slots in order to increase the amount of data that can be carried.TDMA is used by Digital-American Mobile Phone Service (D-AMPS), Global System for Mobile communications (GSM), and Personal Digital Cellular (PDC). Each of these systems implements TDMA in somewhat different and potentially incompatible ways. An alternative multiplexing scheme to FDMA with TDMA is CDMA (code division multiple access), which takes the entire allocated frequency range for a given service and multiplexes information for all users across the spectrum range at the same time.TDMA was first specified as a standard in EIA/TIA Interim Standard 54 (IS-54). IS-136, an evolved version of IS-54, is the United States standard for TDMA for both the cellular (850 MHz) and personal communications services (1.9 GHz) spectrums. TDMA is also used for Digital Enhanced Cordless Telecommunications (DECT).

CDMA (Code-Division Multiple Access) refers to any of several protocols used in so-called second-generation (2G) and third-generation (3G) wirelesscommunications. As the term implies, CDMA is a form of multiplexing, which allows numerous signals to occupy a single transmission channel, optimizing the use of available bandwidth. The technology is used in ultra-high-frequency (UHF) cellular telephone systems in the 800-MHz and 1.9-GHz bands.CDMA employs analog-to-digital conversion (ADC) in combination withspread spectrum technology. Audio input is first digitized into binary elements. The frequency of the transmitted signal is then made to vary according to a defined pattern (code), so it can be intercepted only by a receiver whose frequency response is programmed with the same code, so it follows exactly along with the transmitter frequency. There are trillions of possible frequency-sequencing codes, which enhances privacy and makes cloning difficult. The CDMA channel is nominally 1.23 MHz wide. CDMA networks use a scheme called soft handoff, which minimizes signal breakup as a handset passes from one cell to another. The combination of digital and spread-spectrum modes supports several times as many signals per unit bandwidth as analog modes. CDMA is compatible with other cellular technologies; this allows for nationwide roaming.

The original CDMA standard, also known as CDMA One and still common in cellular telephones in the U.S., offers a transmission speed of only up to 14.4Kbps in its single channel form and up to 115 Kbps in an eight-channel form.CDMA2000 and Wideband CDMA deliver data many times faster.

Block diagram of a basic cellular system:

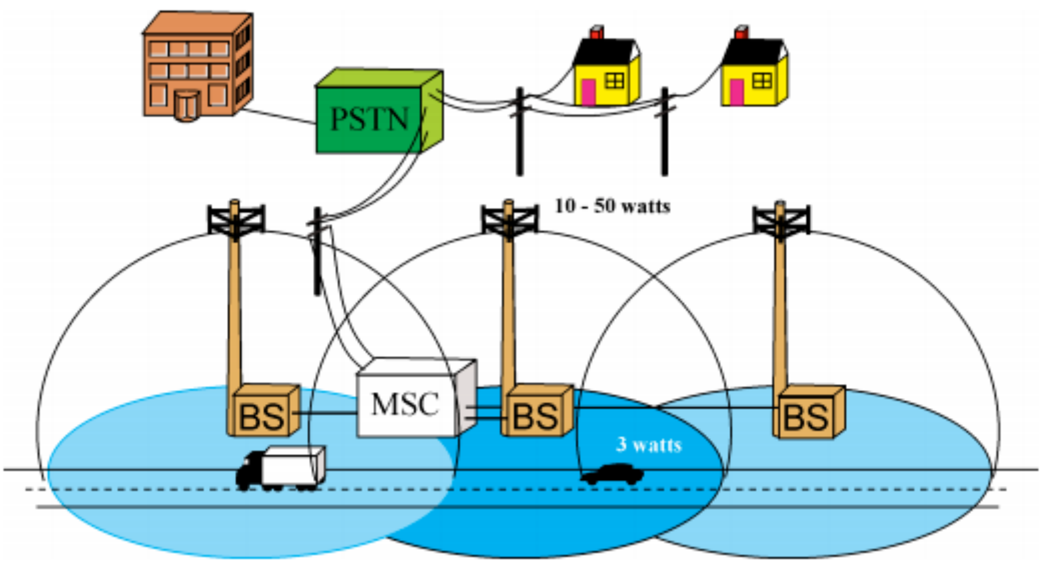


Figure 1.1, Basic Cellular System

1.Introduction

There are various cellular systems in the world, such as the GSM and CDMA. The design of these cellular systems are complicated but the architecture of most cellular systems can be brokendown into six basic components.

In this article, I will illustrate the six basic components that can be found in most cellular systems.

2. Six basic components of Cellular Systems

The architecture of most cellular systems can be broken down into the following six components:

a) Mobile Station (MS)
A mobile station is basically a mobile/wireless device that contains a control unit, a transceiver and an antenna system for data and voice transmission. For example, in GSM networks, the mobile station will consist of the mobile equipment (ME) and the SIM card.
b) Air Interface Standard
There are three main air interface protocols or standards: frequency division multiple access (FDMA), time division multiple access (TDMA) and code division multiple access (CDMA). These standards are basically the medium access control (MAC) protocols that define the rules for entities to access the communication medium.
These air interface standards allow many mobile user to share simultaneously the finite amount of radio channels.
c) Base Station (BS)
A base station is a fixed station in a mobile cellular system used for radio communications with mobile units. They consist of radio channels and transmitter and receiver antenna mounted on a tower.
d) Databases
Another integral component of a cellular system is the databases. Databases are used to keep track of information like billing, caller location, subscriber data, etc. There are two main databases called the Home Location Register (HLR) and Visitor Location Register (VLR). The HLR contains the information of each subscriber who resides in the same city as the mobile switching center (MSC). The VLR temporarily stores the information for each visiting subscriber in the coverage area of a MSC. Thus, the VLR is the database that supports roaming capability.
e) Security Mechanism
The security mechanism is to confirm that a particular subscriber is allowed to access the network and also to authenticate the billing.
There are two databases used for security mechanism: Equipment Identify Register (EIR) and Authentication Center (AuC). The EIR identifies stolen or fraudulently altered phones that transmit identity data that does not match with information contained in either the HLR or VLR. The AuC, on the other hand, manages the actual encryption and verification of each subscriber.
f) Gateway
The final basic component of a cellular system is the Gateway. The gateway is the communication links between two wireless systems or between wireless and wired systems. There are two logical components inside the Gateway: mobile switching center (MSC) and interworking function (IWF).
The MSC connects the cellular base stations and the mobile stations to the public switched telephone network (PSTN) or other MSC. It contains the EIR database.
The IWF connects the cellular base stations and the mobile stations to Internet and perform protocol translation if needed.

Call routing

Each time a call is placed for routing, the destination number (also known as the called party) is entered by the calling party into their terminal. The destination number generally has two parts, a prefix which generally identifies the geographical location of the destination telephone, and a number unique within that prefix that determines the specific destination terminal. Sometimes if the call is between two terminals in the same local area (that is, both terminals are on the same telephone exchange), then the prefix may be omitted.
When a call is received by an exchange, there are two treatments that may be applied:
Either the destination terminal is directly connected to that exchange, in which case the call is placed down that connection and the destination terminal rings.
Or the call must be placed to one of the neighboring exchanges through a connecting trunk for onward routing.
Each exchange in the chain uses pre-computed routing tables to determine which connected exchange the onward call should be routed to. There may be several alternative routes to any given destination, and the exchange can select dynamically between these in the event of link failure or congestion.
The routing tables are generated centrally based on the known topology of the network, the numbering plan, and analysis of traffic data. These are then downloaded to each exchange in the telephone operators network. Because of the hierarchical nature of the numbering plan, and its geographical basis, most calls can be routed based only on their prefix using these routing tables.
Some calls however cannot be routed on the basis of prefix alone, for example non-geographic numbers, such as toll-free or freephone calling. In these cases the Intelligent Network is used to route the call instead of using the pre-computed routing tables.
In determining routing plans, special attention is paid for example to ensure that two routes do not mutually overflow to each other, otherwise congestion will cause a destination to be completely

blocked.

Handoff scenario at cell boundary:

Handover or Handoff is a process where a mobile radio operating on a particular channel is reassigned to a new channel. It is often initiated either by crossing a cell boundary or by a deterioration in quality of the signal in the current channel. The process is often used to allow subscribers to travel throughout the large radio system coverage area by switching the calls (handover) from cell-to-cell (and different channels) with better coverage for that particular area when poor quality conversation is detected.

Handover is necessary for two reasons. First, where the mobile unit moves out of range of one cell site and is within range of another cell site. Second, a handover may be required when the mobile has requested the services of a type of cellular channel that different capabilities (e.g. packet data). This might mean assignment from a digital channel to an analog channel or assignment from a wide digital channel to a packet data channel.

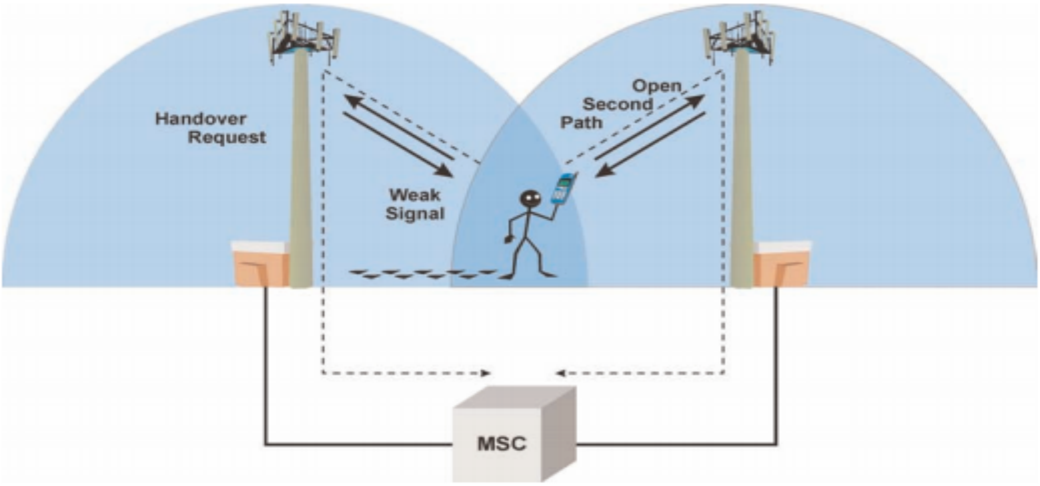


Figure 1.3, Handover Operation

Handoff is divided into two broad categories— hard and soft handoffs. They are also characterized by “break before make” and “make before break.” In hard handoffs, current resources are released before new resources are used; in soft handoffs, both existing and new resources are used during the handoff process.