2

WIRELESS WIDE AREA NETWORK (CELLULAR NETWORK)

Since long, people have sought to communicate on a worldwide basis with each other. It is no more a privileged service; rather it is a bare need. Worldwide communication has become possible due to technological advancements in diverse fields of telephony and the Internet. The dream of communication, even in geographically remote areas, has come true through the introduction of wireless networks.

The most popular way to categorize communication networks is based on their scale—that is, the magnitude of area coverage. The Wide Area Network (WAN) covers a large geographical area. The Internet, spanning the earth, is the most prominent example of WAN in the wired network domain. In the wireless domain, the major example of WAN is the cellular network. This chapter deals with the cellular network and investigates the internals of a wireless WAN.

2.1 THE CELLULAR CONCEPT

Traditionally, the WAN is a collection of Local Area Networks (LANs) dispersed geographically. The relatively new cellular network is the collection of small networks, operated by the service provider(s) local to the networks. This concept was introduced in the early 1970s at Bell Laboratories. One of the most successful initial implementations of cellular concept was the Advanced Mobile Phone System (AMPS). It has been popular in the United States since 1983.

The system capacity of a cellular mobile network can be interpreted as the maximum number of users that can be supported at a particular point of time. Support to a large number of users leads to a large area coverage. If a single transmitter is used to cover a large geographical area, a very powerful transmitter/antenna has to be installed. However, in reality, the system capacity offered by a single high power transmitter with full set of radio frequencies, allocated to the system, can't extend beyond a limit. Therefore, the alternative is to use a set of radio frequencies to serve a comparatively smaller geographical area and then reuse it for serving the other small areas. It avoids installation of the high power transmitter while keeping the desired system capacity. Care must be taken to ensure that the same set of frequencies, used to serve more than one geographical area, does not introduce interference among signals from the users in two different areas.

Partitioning a large coverage area with the target to provide small contiguous areas, supported by a low power transmitter with low antenna height in each of the areas, is the basis of a cellular communication network. The implementation of such a system follows the concept of cellular radios.

The next subsection defines the components of a cellular mobile network.

2.1.1 The Components

The basic components of a cellular mobile network developed around the cellular concept are as follows.

Transceiver: The device capable of simultaneously transmitting and receiving radio signals is called a transceiver.

Mobile Station (MS): Mobile station is an electronic device used either as a portable hand-held unit or mounted in a vehicle for communicating voice and data. It contains a transceiver, antenna and control circuitry.

Base Station (BS): A base station is a fixed station in the mobile radio system and is used for radio communication with mobile station. It is normally located at the centre (Figure 2.1) of a network cell. Alternatively, it may also be placed on the edge of a coverage region. A base station contains transmitter, receiver, control unit, and antenna; and it has radio channels for allocation.

Mobile Switching Centre (MSC): An MSC coordinates routing of calls in a large service area. It connects both the base station and mobile stations to a telephone network called Public Switch Telephone Network (PSTN). That is, the MSC is an interface between the radio system and PSTN. The MSC is also referred to as the Mobile Telephone Switching Office (MTSO).

Forward Voice Channel (FVC) or downlink channel: The channels used for voice transmission from a base station to the mobile station are called FVC or downlink channel.

Reverse Voice Channel (RVC) or uplink channel: The RVCs or uplink channels are used for voice transmission from a mobile station to the base station.

Forward Control Channel (FCC)/Reverse Control Channel (RCC): These two channels (FCC and RCC) are responsible for initiation of mobile calls and are often called set-up channels. The FCC continuously broadcasts all traffic requests targeted to the MSs and serves as beacon. In general, on an average 5% of the total available channels within a system are assigned as the control or set-up channels.

2.1.2 Cellular Architecture

Figure 2.1 illustrates cellular network architecture with its components, the MS, BS and MSC. The whole geographical area covered by the network is partitioned into a number of hexagonal cells. Each such cell is serviced by a BS and a set of BSs are connected to an MSC. In addition to its function as a switching exchange, an MSC handles/controls mobility management of MSs in coordination with the BSs. The MSCs of the network are connected to the PSTN through a wired connection. The only wireless connection in a cellular network is in between an MS and the BSs.

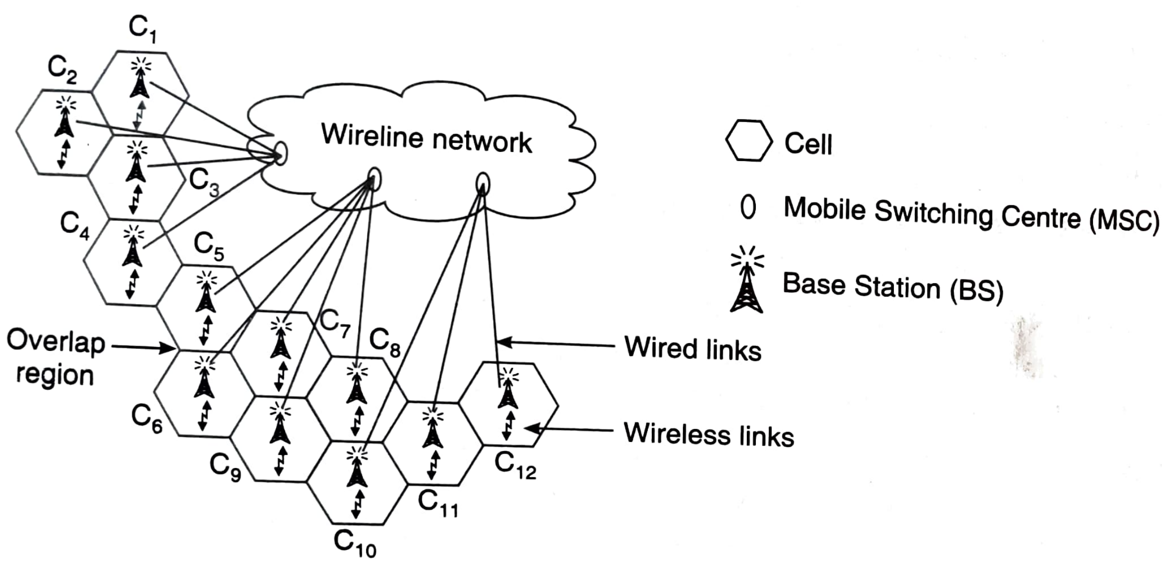


Figure 2.1 A cellular mobile network architecture.

The hexagonal cell shape considered in Figure 2.1 is purely imaginary. In reality, no such cell shape exists in the cellular network implementation. However, for proper planning of the system architecture, the hexagonal shape is universally adopted for several reasons.

Cell shape: In a cellular network, the geographical coverage region is considered to be partitioned into a number of small areas called cells. These cells are hexagonal in shape. The actual radio coverage of an antenna installed within a network cell is called the footprint. Although the real footprint is amorphous, the assumption of regular cell shape, in cellular network, is essential for a systematic system design. The probable candidate regular shapes can be the square, circular, hexagonal, etc. In a square cell based design, all the adjacent cell centres cannot be at the same distance. The normal practice of installing BS antenna at the centre of a cell, therefore, cannot ensure the equidistance among the antennas. The antennas placed at equidistant simplify the task of determining when to switch a user (MS) to an adjacent antenna and as well as to choose the right one.

Considering circular radiation pattern and free-space propagation of the omni-directional BS antenna, the circular cell shape is mostly desirable. However, the circular cell can't fully cover a geographical region without the overlapping areas. On the contrary, a hexagonal cell approximates the circular radiation pattern. The adjacent BS antennas placed at the centres of such hexagonal cells are equidistant. Moreover, for a given distance between the centre of a polygon (circle/square/hexagon) and the farthest points on its perimeter, the hexagon has the largest area coverage among the three shapes.

The above discussions point to the fact that a hexagonal cell shape is most desirable in cellular network to cover a given network service area (coverage area). However, the exact hexagonal shape of cells cannot be implemented due to topographical and some other practical limitations of installing base station at a desired site. In reality, the designers practice variations in cell shapes. The reasons for such variations are:

- Due to practical limitations, e.g. land dispute, location unsuitability, some of the BSs can't be placed at the desired point (centre of a hexagon).
- The antenna at BS has some directional bias, in spite of the fact that it is desired to be omni-directional.
- The signal propagation characteristics may vary at places. In some cases, the transmission path between a transmitter and the receiver can be the simple direct line of sight. In a number of situations, the transmission path may be obstructed severely by high-rise buildings, foliage and terrain.

2.2 CALL SET-UP

When a Mobile Station (MS) is turned on within the coverage of a network, the following tasks are executed:

(i) The MS scans the group of FCCs (Forward Control Channels) in search of the strongest BS signal. The BS from which the strongest signal is received works as serving BS of the call initiator MS.

(ii) It checks continuously the selected FCC to ensure that the signal is not below the usable level.

In a cellular network, the call may be initiated by an MS or by a land phone. The call initiation process followed by an MS is different from that of a land phone. The steps to be executed during a call initiation are summarized in the next two subsections.

2.2.1 Call Initiation by an MS

Despite variations in the characteristics of MSs, the following sequential steps are commonly performed while a call is initiated by an MS:

(i) A call request is sent from the MS to the serving BS on RCC (Reverse Control Channel). The request message contains the MIN (Mobile Identification Number), ESN (Electronic Serial Number) and phone number of the called party (MS/land phone).

(ii) On receiving the request message, the serving BS transfers the request to the concerned MSC (Mobile Switching Centre).

(iii) The MSC, on receiving the message, validates the request/MS.

(iv) After successful validation, the MSC instructs the BS to assign a pair of unused FVC and RVC to the caller MS.

(v) MSC connects the MS with the called party. If the called party is also an MS,

   (a) MSC dispatches the request to all its BSs.

   (b) The MIN of the called MS is paged by all the BSs and instructs the called MS to use the unassigned FVC and RVC pair.

   (c) The called MS responds over RCC.

   (d) The serving BS sends the acknowledgement to MSC.

   (e) The serving BS sends an alert message to the called MS to ring.

(vi) Both the caller and called MSs begin voice transmission and reception.

2.2.2 Call Initiation by a Land Phone

The call initiation from a land phone leads to the following steps of activities while connecting an MS.

(i) The MSC receives the call request from a land phone through PSTN.

(ii) The MSC connects the land phone with the called party (MS)

   (a) MSC dispatches the call request to all its BSs.

   (b) On receiving the call request message the MIN of called MS is broadcast by all the BSs.

   (c) The called MS responds over RCC.

   (d) The serving BS sends an acknowledgement to the MSC.

   (e) The serving BS sends an alert message to the called MS to ring.

(iii) On receiving the acknowledgement, MSC instructs the BS to assign a pair of unused FVC and RVC to the called MS.

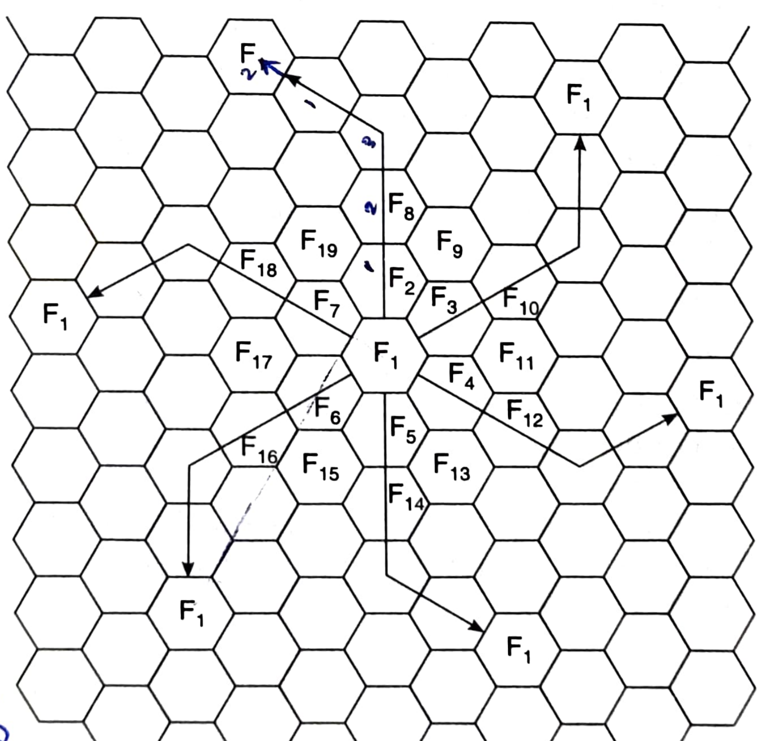
(iv) The MS responds to the caller party (land phone) over RVC for voice transmission.

2.3 FREQUENCY REUSE AND CO-CHANNEL CELL

In a cellular mobile network, the system may have to support a large number of simultaneous calls from and to the MSs. It requires allocation of a large number of forward and reverse channel pairs, thereby requiring a wide frequency bandwidth, which is a scarce resource. The objective of frequency reuse in a cellular system is to employ the same frequency in more than one network cell. This reuse reduces the demand of such scarce resource. However, use of the same frequency in nearby cells may interfere preventing the desired information flow. Therefore, the design challenge is to determine the distance between two cells that can use the same set of frequencies within the acceptable limit of interference.

The frequency reuse ensures proper distribution of total available frequencies allocated to a network. The total allocated frequencies are partitioned into a number of distinct sets for distribution in different network cells. A group of adjacent cells, called the cluster or compact pattern, is then determined. Each cell of a cluster is then allocated a distinct set of frequencies. It implies that in a cluster/compact pattern there is no scope of frequency reuse. A frequency f is allocated to two different cells X_i and X_j ensures that X_i and X_j belong to two different compact patterns.

Co-channel cell: The cells that can use the same set of frequencies as assigned to a cell X are called the co-channel cells of X. In Figure 2.2, the set of cells $S = \{F_1, F_2, \ldots, F_{19}\}$ form the cluster/compact pattern. Each cell of S uses a unique set of frequencies from the total allocated frequencies



i=3
j=2

# of cells
= i^2+ij+j^2

Figure 2.2 The co-channel cells.

of the network. The cells labeled with F_1 and are located outside the compact pattern are the nearest co-channel cells of the cell that is labeled with F_1 and is located at the centre of the compact pattern. The co-channel cells use the same set of frequencies for allocation to the MSs under their control. The minimum distance between the centres of any two co-channel cells that are nearest to each other is D. The distance D defines the effective measure of frequency reuse while avoiding any possibility of interference beyond the acceptable limit.

The repetition of compact pattern in the network coverage area places the co-channel cells in tiers. In general, a cell is surrounded by $6T$ co-channel cells in tier T. Figure 2.2 indicates 6 co-channel cells F_1 only at tier $1(T = 1)$—that is, the nearest co-channel cells. Similarly, tier 2 contains 12 co-channels. The co-channel cells of a network cell in each tier form a hexagon. The minimum distance between a cell (centre cell labeled with F_1) and its co-channel cell at tier T is TD.

The number of cells in a cluster $N = i^2 + ij + j^2$, where i and j are two positive integers. The size of cluster depends on the parameters i and j. In the figure $i = 3$, $j = 2$ and, therefore $N = 19$. The N is also considered as the measure of frequency reuse factor. In TDMA and FDMA, the frequency reuse factor is greater than 1 whereas in CDMA, this reuse factor is 1. In some literature $1/N$ is considered as reuse factor.

Example 2.1 Let us consider for a cellular system, 33 MHz bandwidth is allocated. Each of the full-duplex voice and control channels uses 50 kHz bandwidth. The number of channels can be allocated per cluster is, therefore, $33000/50 = 660$. Now, if the cluster size chosen is $N = 12$ ($i = 2$, $j = 2$), for a cluster of size 12, then each network cell gets $660/12 = 55$ channels.

To identify the nearest co-channel cells for a given cell X, and for particular values of i and j, the following steps are to be performed:

- Starting from X, move i cells along any chain of hexagons
- Turn 60° counterclockwise
- Move j cells

In Figure 2.2, $i = 3$ and $j = 2$.

Co-channel interference: A signal other than the signal required for desired communication within a network cell is called interference. Interference creates disturbance in communication and affects adversely the system performance. The interference caused by co-channel cell is called co-channel interference. Co-channel interference plays an important role in determining the quality of service (QoS)[1].

[1] The QoS is measured by several parameters such as coverage, call blocking (unable to initiate a call), call dropping (call is stopped during conversation), audio quality (cross-talk), etc. The audio quality is mainly affected by co-channel interference.

2.4 CELL DESIGN

The allocated frequency spectrum in a cellular network is very limited. The maximization of system capacity, keeping a reasonable quality of service, under the constraint of limited spectrum is the measure of performance of a system. However, the system capacity and quality of service are the two conflicting requirements. Thus while selecting the cell and cluster size, these two constraints are to be considered.

For example, let us assume that a cellular system has S duplex channels and a group of K channels are allocated to each cell of the network. If N cells are in a cluster, then $S = KN$. If a cluster is replicated M times, the total number of allocated duplex channels is $C = MKN = MS$, out of which S channels are unique. Here C denotes the capacity of the network. It signifies that the capacity is directly proportional to the number of times a cluster is replicated. Figure 2.3 shows the two parameters D and R (R is radius of a cell) for $N = 7$.

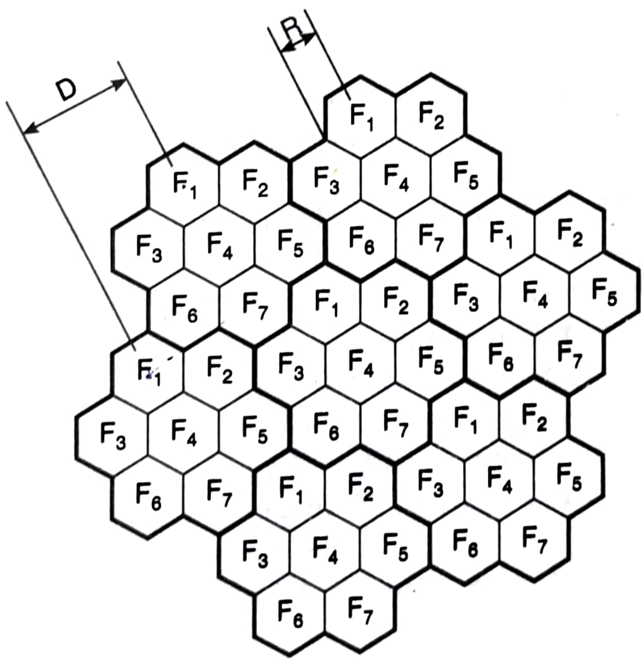


Figure 2.3 7-cell reuse pattern.

The ratio $Q = D/R$ is referred to as the frequency reuse ratio. From hexagonal cell geometry, the relation between frequency reuse ratio (Q) and the number of network cells per cluster (N) is

$$\boxed{Q = \frac{D}{R} = \sqrt{3N}} \tag{2.1}$$

The number of cells in a cluster (cluster size) is $N \propto D/R$. Now for a fixed cell size (R), if the cluster size (N) is increased, then D is to be increased. Alternatively, a large cluster size indicates that D/R is large. On the other hand, a network with small cluster size indicates its co-channel cells are located closer to each other and hence co-channel interference is significant.

The above discussions set the design criteria of a cellular network for the given coverage. These are (i) to maximize capacity ($C = M \times K \times N$), the smallest possible value of N is to be chosen and (ii) to reduce interference, the value of N—that is, the cluster size, is to be large. Thus to meet these two conflicting requirements (i) and (ii), a trade-off between the capacity and interference is desirable.

Example 2.2 Consider a cellular system having 1596 duplex channels to cover 2310 sq. km. The area of each cell is 6 sq. km. Assume that we need to compute the following:

(i) System capacity (C) for cluster size 7, 12 and 19.
(ii) M refers to number of times the cluster is replicated to cover the entire network area.

The total allocated channels in the system is $S = 1596$.

For the cluster size $N = 7$, 12 and 19, the number of channels per cell (K) = 1596/7, 1596/12 and 1596/19 = 228, 133 and 84 respectively.

The coverage area of a cluster of size 7/12/19 is $7 \times 6/12 \times 6/19 \times 6$ = 42/72/114 sq. km.

That is, the number of times a cluster is to be replicated is $M = 2310/42$, 2310/72 and 2310/114 = 55, 32 (approx) and 20 (approx) for the cluster size 7, 12 and 19 respectively.

Therefore, the system capacity for cluster size 7 is $C = M \times K \times N = 55 \times 228 \times 7 = 87780$.

For cluster size 12 and 19, $C = 32 \times 133 \times 12$ and $20 \times 84 \times 19$, that is, 51072 and 31920 respectively.

The system capacity $C = 87780$ signifies that at any point of time maximum 87780 simultaneous calls are allowed in the network with cluster size 7. Similarly, if the cluster sizes are 12 and 19, the number of simultaneous calls permitted is 51072 and 31920 respectively.

The results of the above example indicate that the system capacity of a cellular system decreases with the increase in cluster size.

2.5 INTERFERENCE IN CELLULAR SYSTEM

Interference creates disturbance in communication and affects the performance of a cellular network. It causes impairments ranging from the decrease in system capacity to dropping of calls. Signal received by the base station (BS) from a communicating mobile station (MS) may encounter interference due to:

- transmission from other mobiles in the same cell
- transmission from other mobiles in the neighbouring cells
- background noise
- other neighbouring base stations operating in the same frequency band
- noncellular system leaks energy into the cellular frequency band

Some of the reasons for interference are the improper channel assignment strategies, co-channel interference and adjacent channel interference out of imperfect cell design, etc. The co-channel interference is described in Section 2.3.

Adjacent channel interference: Signal impairment to a frequency due to the presence of another signal of very close frequency is called 'adjacent channel interference'. It mainly occurs due to the imperfect receiver filters that allow close/adjacent frequencies to leak into the passband[2].

The adjacent channel interference can have a serious effect if two MSs (adjacent channel users), one is located very close to the BS and the other is at a distance from the BS, simultaneously transmit equal power. The BS receives more power from the MS close to it. The SNR (signal-to-noise ratio) for this communication is much higher than that for the MS at a distance. Such a case is referred to as the 'near-far' problem. Further, if the MS close to the BS transmits a signal with order of magnitude much higher than the MS at a distance, the SNR for the second MS may not be detectable. This effectively creates jams in the communication channel of a network.

2.5.1 Signal-to-Interference Ratio

The quality of a signal communicated is also measured by a parameter called S/I (signal-to-interference) ratio. To compute S/I, the following interference sources are ignored as a matter of convention:

- co-channel interference from the second and other higher tiers of the network
- adjacent channel interference (negligible in comparison to the co-channel interference)
- background noise

The inter cell interference is mainly dominated by the co-channel interference.

The average channel quality is a function of distance dependent path loss[3]. Effects of other issues of communication are normally ignored. If c be the number of co-channel interfering cells and I_i be the interference

---

[2] In digital communication, the frequency band is split up into two main parts—the baseband and the passband. The passband contains all frequencies above a limiting frequency whereas the baseband refers to the frequencies below the limiting frequency. In radio communication, a baseband signal must be converted to the passband signal for its successful transmission.

[3] Path loss is the reduction in power density of the radio signal as it propagates. It is one of the dominating elements that causes impairment in the propagation channel as well as distorts the information-carrying signal as it propagates over medium. As an MS moves away from its serving BS, the received signal at BS becomes weaker because of the reduction of power density of the signal due to the growing distance between the MS and the BS.

power caused by transmissions from ith interfering co-channel cell to the BS, the S/I at the desired mobile receiver is:

$$\frac{S}{I} = \frac{S}{\sum_{i=1}^{c} I_i} \tag{2.2}$$

Now, if r be the distance between the mobile station and serving BS, the desired received signal power S is proportional to r^{-k}, where k is the path loss component (in general, $2 \le k \le 5$). Further, if D_i is the distance between ith co-channel cell and the mobile station, the I_i is proportional to $(D_i)^{-k}$. Thus the received S/I at the mobile station can be expressed as:

$$\frac{S}{I} = \frac{r^{-k}}{\sum_{i=1}^{c} (D_i)^{-k}} \tag{2.3}$$

If only first-tier co-channel cells are considered, then $C = 6$, and when an MS is in cell boundary ($r = R$), then $D_i \approx D$, for $i = 1, \ldots, 6$. Therefore, from Eq. (2.3), it can be deduced that:

$$\frac{S}{I} = \frac{\left(\frac{D}{R}\right)^k}{6} = \frac{Q^k}{6} = \frac{(\sqrt{3N})^k}{6}$$

or,

$$Q = \left(6 \times \frac{S}{I}\right)^{1/k} \tag{2.4}$$

It implies, from Eq. (2.4), for a given acceptable signal-to-interference (S/I) ratio, the frequency reuse ratio Q can be computed. From a known Q, the cluster size/compact pattern can also be settled (Eq. (2.1)) for a network.

Example 2.3 Let us assume that for a GSM TDMA system, the acceptable S/I is 15 dB or more. If path loss exponent $k = 3$, the frequency reuse ratio $Q = (6 \times 10^{1.5})^{1/3} \approx 3.13$. Then the compact pattern size $N = Q^2/3 = 3.26 \approx 3$. That is, a 3-cell frequency reuse system is to be used with 15 dB S/I.

2.5.2 Interference Reduction

It is reported in the earlier section that the co-channel interference is reduced if the distance between co-channel cells is increased. However, the large distance among the co-channel cells increases the cluster size, thereby reduces the system capacity.

On the other hand, the adjacent channel interference can be reduced through careful filtering and the proper channel assignments. Since each cell of a network is given only a fraction of the total available channels, a cell need not be assigned channels that are adjacent in frequency. This enables provision for the greatest possible frequency separation between the channels to ensure reduced adjacent channel interference within a cell

Controlling transmission power of mobile stations (MSs) is another effective option to reduce the overall signal-to-noise ratio caused due to reverse channel transmission. In such mechanism, the power levels transmitted by each mobile station are under constant control of the serving base station (BS) to ensure that each unit transmits the minimum power necessary to maintain a good quality link on the reverse channel. The transmission power control also helps to extend battery life and combats the 'near-far' problem. The MS closer to BS transmits less power compared to that of the MS at a distance while both the MSs are adjacent channel users. At the serving BS's receiver, it maintains almost the same SNR for all the MSs' transmitters.

2.6 CHANNEL ASSIGNMENT

At the time of installation of a cellular network, each cell is assigned a set of channels to provide services to the individual calls of the cell. The task of assigning frequency channels satisfying the frequency separation constraints is known as the channel assignment problem. As the use of mobile communication systems grows, more efficient channel assignment/allocation techniques are needed. The channels on a network are the scarce resources and its proper management is very much desirable. An efficient assignment is not only important for new call initiation but it has also a great impact on call management while a mobile station moves from a cell to another—that is, handoff.

Channel assignment strategies are classified as the fixed and dynamic assignment. In fixed assignment, each cell of the network is to be allocated a predetermined set of voice channels. To provide call service, the cell can only utilize channels from that set. Any new call attempt within the cell is served while considering the channels remain unused at that moment. If all the channels of that cell are occupied, the attempted call is blocked.

One of the variations of fixed channel assignment is the channel borrowing. In this strategy, a cell is allowed to borrow channels from a neighbouring cell when its channels are already occupied and a handoff/new call is to be serviced. The MSC supervises such borrowing processes ensuring that this does not disrupt or interfere with the calls in progress. That is, the channels currently not being used by the donor cell or by its co-channel cells can be borrowed.

In dynamic assignment, the voice channels allocation to cells are not predetermined. Each time a call request is made by an MS, the serving BS seeks a channel from the MSC. The MSC in turn allocates a channel to the cell, requesting for the channel. While assigning channels, an MSC should consider the following:

- Likelihood of future blocking within the cell
- Reuse distance
- Cost functions

A number of works are so far been reported on channel assignment strategies. The main objective of such works is to ensure maximum utilization of radio resource, keeping the call blocking probability as low as possible. Such a region/area is defined as the hotspot. Therefore, an effective channel assignment strategy has to take care of the hot-spot areas.

2.7 HANDOFF

During its move an MS (mobile station) may occasionally switch to a new cell of the network. Switching from a cell to its neighbouring cell implies change of BS (base station) as well as switching to new voice and control channels. The process of transferring an MS from the control of current serving BS to a new one is called handoff. That is, the handoff process includes identification of a new BS for the MS and also allocation of the voice and control channels associated with the new BS. Handoff must be performed successfully and as infrequently as possible to reduce overhead due to switching.

The early prediction of handoff is the requirement for an efficient handoff-handling scheme. The simplest implementation is that whenever the power level from an MS received by a BS crosses a threshold ($P_{rthreshold}$), the BS initiates the handoff process. However, the whole process should be transparent to the users. To realize it, the system designer must specify an optimal signal level ($P_{rthreshold}$) to initiate a handoff. This threshold signal level is defined as $P_{rthreshold} = P_{rminimum} + \Delta$, where $P_{rminimum}$ is the minimum usable signal for acceptable voice quality at BS receiver and Δ is a small value.

If $P_{rthreshold}$ is too large, there may be unnecessary handoffs. The unnecessary handoff is considered to be costly to an MSC. On the contrary, if $P_{rthreshold}$ is too small, the system may not get sufficient time to complete the handoff process. While the BS/MSC handling the handoff, the MS may move farther from the serving BS and the call may be lost due to weak signal. As the handoff is costly, prior to any handoff, the system has to ensure that the drop in the measured signal level at serving BS is not due to any other reason except the movement of the MS away from the serving BS. In reality, a serving BS monitors the signal level received from an MS for a certain period of time. If the MS moves very fast, the slope of the short-term average of received signal becomes steep indicating that the handoff is to be processed instantly.

2.7.1 Handoff Strategies

The handoff in a cellular network is a regular feature. A cellular network without handoff is of no practical use. Therefore, efficient strategies need to be devised to address this issue. The following handoff tackling strategies are considered to be effective in a cellular mobile network:

- Network-controlled handoff (NCHO)
- Mobile-assisted handoff (MAHO)

NCHO: In this technique, the BSs monitor the signal quality received from an MS and report it to the MSC. On receiving this information from the BSs, the MSC selects the new serving BS and initiates handoff. The new BS takes control of the MS.

MAHO: In mobile-assisted handoff, implemented in GSM, the process of handoff is jointly taken care of by the mobile station (MS) and the network. The MS measures the signal levels received by it from the various BSs using a periodic beacon, generated by the BSs, to keep track of the MSs. The MS then feeds the measured signal levels back to the MSC via the serving BS. The MSC finally takes the decision of handoff.

Based on the procedures followed to take care of the handoff issues in a cellular network, the handoff is classified as Hard and Soft handoff.

FDMA, TDMA

Hard handoff: In Time Division Multiple Access (TDMA) and Frequency Division Multiple Access (FDMA), the frequency reuse factor (N) is greater than one. The neighbouring cells within a cluster use different frequency bands of the available frequency spectrum in such an implementation. An MS in a cell X', receives signals from different cells (BSs) surrounding X'. When the signal strength received by an MS from a neighbouring cell (X) exceeds a predefined threshold, the MS is instructed to switch to it (X). A new frequency band is then allocated to the cell X to support the ongoing communication of the MS. Such a termination of the existing connection (with X') to establish a new connection (to X) is referred to as the Hard Handoff. In hard handoff, an MS can have the radio link only with a single base station at any point of time and, therefore, it is also known as the 'break before make'.

CDMA

Soft handoff: In soft handoff, the new connection between an MS and the new BS is established prior to the termination of existing link between the MS and the currently serving BS. Unlike hard handoff, during soft handoff, an MS can simultaneously communicate to more than one BS. The Code Division Multiple Access(CDMA) systems implement soft handoff techniques. As in CDMA systems for the cluster size $N = 1$ (Section 2.3), the spatial separation of frequency spectrum for allocating separate frequency band in cells is not necessary. Further, the interference from neighbouring cells is not too serious in such systems. Therefore, it can allow an MS to be connected to more than one BS simultaneously. If the MS enters a region in which the transmissions from two BSs are comparable, the MS then gets connected to the two BSs. The MS remains connected to the two BSs until one BS clearly predominates and the MS is assigned to that cell.

1st generation vs 2nd generation handoff

In 1st generation cellular systems, the voice transmission is analog. It uses FDMA technique to share the available spectrum. As cellular mobile systems have grown from analog to digital, the techniques of handling handoff get the maturity. In 1st generation analog cellular systems, the notable features of handoff processing techniques are:

- Handoff decisions are taken by MSC with assistance of BS (NCHO)
    - BS measures signal strength (SS)
    - MSC supervises
- Each BS constantly monitors SS of RVCs (reverse voice channels) and determines the relative location of each MS

On the other hand, in 2nd generation digital cellular systems, which use TDMA technology, the following are the important features for implementing handoff:

- Handoff decisions are mobile assisted (MAHO)
- MS measures the power levels received from the BSs of neighbouring cells
- The measured values are continually sent to the serving BS

In MAHO, handover of calls between the BSs is much faster than that realized in 1st generation systems. An MS continuously measures the parameter values required for the handoff. The MSC no longer remains constantly busy for monitoring the signal strength at the MS. MAHO is particularly suitable for micro cellular (to be discussed in Section 2.7.2) environments where handoffs are more frequent.

Prioritizing handoffs

From the user's point of view, abrupt termination of a connection/call during running conversation is more annoying than a call is blocked occasionally on a new call attempt. Hence prioritizing handoff over the new call attempt is desirable. That is, at an instant of time if the system has to support a handoff and a new call attempt simultaneously, then a strategy is to be adopted so that service to handoff is considered first. There are two standard handoff strategies—the guard channel concept and the queuing of handoff requests.

In the guard channel, a fraction of the total available channels within a network cell is reserved for providing service to the handoff requests from ongoing calls. It means a cell gives priority to the handoff calls. On the other hand, in the queuing based handoff handling technique, there is no such reserved channel for the handoff calls. At the time of handoff if there is no available channel to serve, the call request is queued. Therefore, it decreases probability of forced termination of a call but delays the handoff.

The handoff strategies, guard channel concept and queuing of handoff, limit the total carried traffic in a network. In guard channel based technique a comparatively fewer number of channels can be allocated for the new call attempts by the MSs as a set of channels is kept aside for handoff calls. This reduces the total carried traffic within the system at any instant of time. On the other hand, in queuing the probability of a successful handoff improves at the cost of increased blocking probability of originating calls.

2.7.2 Constraints

The practical constraints in performing handoff are the accommodating users (MSs) with different speeds and the cell dragging.

Users with different speeds

In the cellular network, the mobile users can be inside a high-speed vehicle or they can be simply pedestrians. For pedestrians carrying MSs, the handoffs do not take place very frequently. So, the cell size in the network can be kept smaller (micro cell approach) without any severe problem in processing handoff. On the other hand, for fast moving users (sitting in a high speed vehicle) micro cell approach is not desirable. Micro cell based design for such cases may result in frequent events of handoff, thereby, increases computational overhead at the system components e.g. BS, MSC, etc. For high-speed mobile stations, the cell size is to be large (macro cell approach) to ensure efficient handling of handoff avoiding the excess loads in system components.

Theoretically, the cellular concept provides additional system capacity through addition of new cell sites. However, in reality, it is difficult for the service providers to identify new cell sites in urban areas even for some non-technical reasons. The practical solution is to use different antenna heights as well as the different power levels in the same location, to enable the design of 'micro' and 'macro' cells. This technique is called the 'umbrella cell' approach (Figure 2.4). The umbrella cell configuration consists of one large BS (macro cell) with high transmission power and antenna height. This set-up serves as an 'umbrella' for a number of small BSs (cells) with low transmission power and small diameters. While a high-speed user

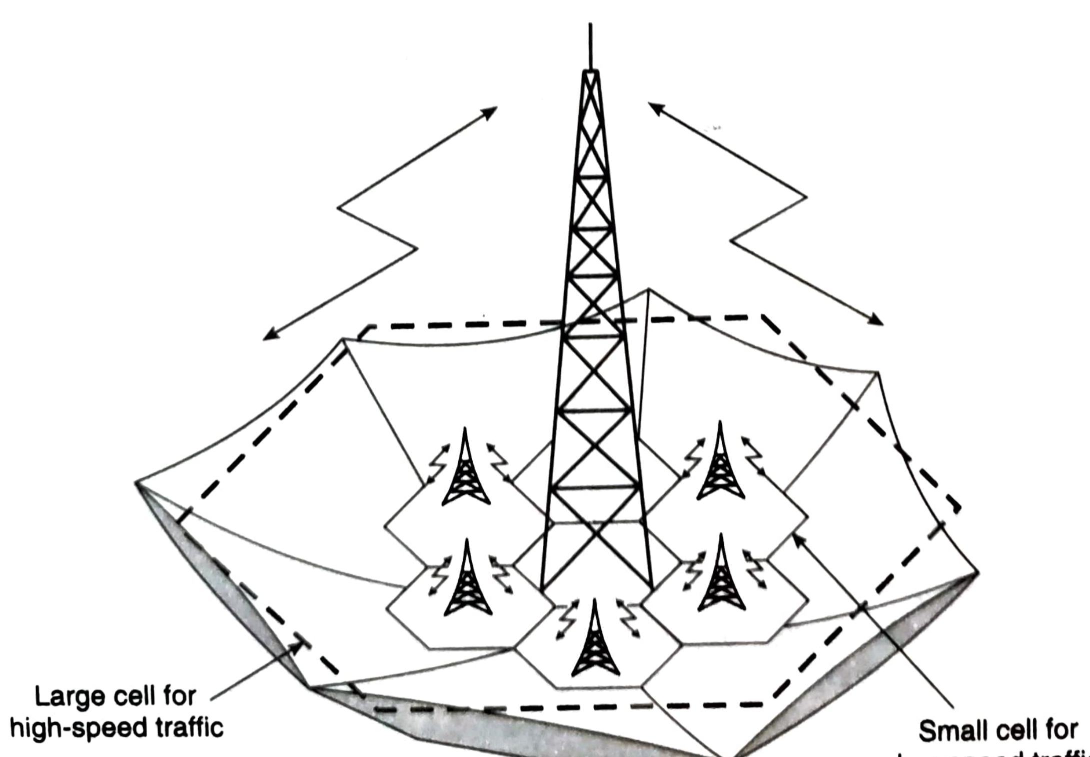


Figure 2.4 The umbrella cell.

(MS) within the large umbrella cell approaches the large BS, with rapidly decreasing speed, the BS decides to handover the MS into one of the co-located small BSs (micro cells) without any intervention of the MSC.

In GSM, the speed of a user can be determined almost accurately by the change of value of the parameter, timing advance (TA)[4]. Its value is updated in the base station controller (BSC) located between BS and the MSC every 480 ms.

The umbrella cell based system (service area is covered by co-located micro and macro cells) has the following advantages over a microcell (service area is covered by microcells only) based system:

- Since it is not necessary to cover the whole service area with microcells, infrastructure cost is saved.
- The number of handoffs is much less than in microcell based system. In such a system whenever an MS, irrespective of its speed, crosses a cell boundary, a handoff occurs. On the contrary, in an umbrella cell environment, MSs with higher speed are under control of the BS of macrocell and the MSs with lower speed are under the control of the BS of microcell. So the MSs under macrocell are subject to less number of handoffs in comparison to the MSs under a microcell. Total number of handoffs is reduced in this way.

Frequent handoffs result in a substantial increase of the signalling load for a network. It degrades the quality of signal at the end user. The umbrella cell based system eliminates the possibility of frequent handoff for high speed MSs and very much effective in urban environments that feature city highways.

Cell dragging

In microcell environment, a serving base station may sense very strong signal from a slow-speed MS (pedestrian) even the user carrying the MS is well beyond the designated range of the cell. The signal received at the BS may be higher than the handoff threshold and, therefore, no handoff takes place. This phenomenon typically occurs in an urban environment when there is a line-of-sight (LOS) radio path between the user (MS) and the serving BS. It also causes interference when the MS penetrates deep into an adjacent cell area. Such a state is called the cell dragging. The parameter $P_{rthreshold}$ for taking handoff decision is to be chosen carefully to get rid of this problem.

2.7.3 Roaming

The roaming is a mechanism by means of which the intersystem handoff is taken care of. While the signal strength received by a mobile station (MS) becomes weak and the mobile switching centre (MSC) cannot find

[4] In GSM, timing advance (TA) corresponds to the time a signal from the MS takes to reach to the base station.

a cell within its system to handover the MS, it is handed off to a cell under the control of a different MSC. The intersystem handoff is called roaming (Figure 2.5).

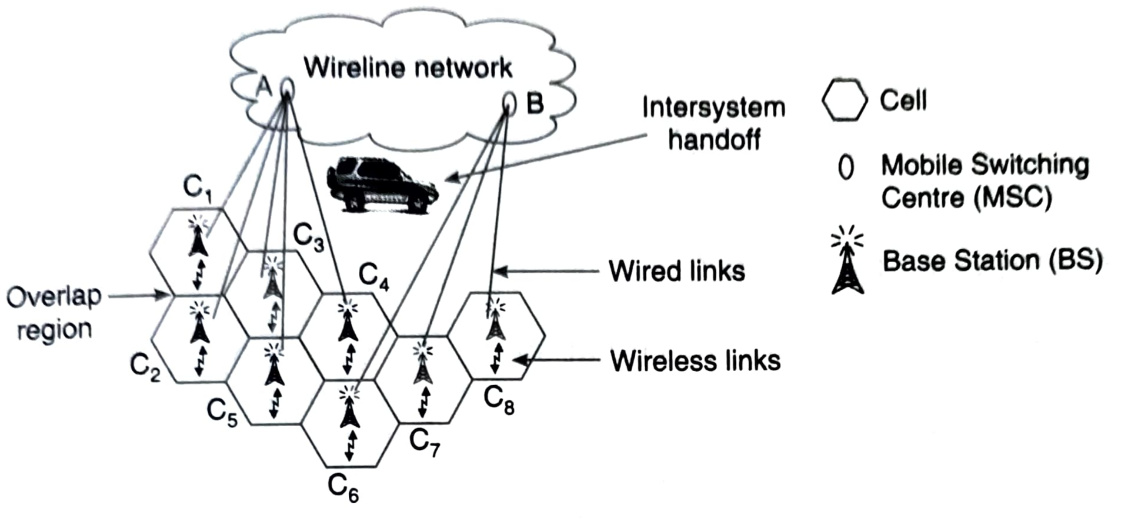


Figure 2.5 Intersystem handoff.

The important features of roaming are

- The MS moves out of its home network (to be described in Section 2.8).
- Local call becomes long distance call.
- Compatibility between two MSCs must be determined before implementing an intersystem handoff.

2.8 MOBILITY MANAGEMENT

The handoff management schemes are described in the earlier section. It is a part of mobility management in a cellular mobile network. The other important issue in mobility management is the location management. In this section, we concentrate on location management of a mobile station (MS) to keep track of its location for establishing seamless connection.

To start with an MS is assigned to a network called the home network of the MS. The association between the MS and the home network is made through a registration process. The home network is always aware of the current location of MS. While roaming from its home to a different network, an MS must register at the visiting network. Simultaneously, this update must reach its home network through the visiting network establishing an association between the home and the visiting networks.

A cellular network maintains two databases—Home Location Register (HLR) and Visiting Location Register (VLR). The HLR keeps the profile of mobile station and authenticates the subscriber before any updates. The VLR, on the other hand, authenticates the visiting MS in consultation with the HLR. The whole process which keeps track of a mobile subscriber's current location is called location management.

When a call/message destined to an MS is generated, the whole network is paged by the MSC to track the MS. In an alternative implementation, every time a mobile station crosses a location area (a group of cells), it sends its current location to the network (location update). That is, instead of paging the whole network, only cells in the location area are paged to track the desired mobile station. This method reduces the paging overhead at the cost of location update. There are many other schemes that target cost optimization for tracking an MS.

2.9 GRADE OF SERVICE

The concept of trunking in circuit-switching telephone is applied for estimating the system load of cellular network. The average amount of network traffic during a busy hour is estimated to fix the size of a network. According to ITU-T (International Telecommunication Union–Standardization Sector) recommendation, the busy hour is determined by considering the average of the busy hour traffic on the 30 busiest days of a year. Statistically, not all subscribers make calls at the same time and, therefore, it is reasonable to fix the size of network that can be able to handle expected level of load in the network. As a result of which whenever traffic load crosses the expected load, the probability of failure of an attempted call becomes non-zero and this may prohibit a user to access the trunked system. The ensured user access to a trunked system during the busiest hour is measured by the Grade of Service (GOS).

Cellular radio systems rely on trunking to accommodate a large number of users in a limited radio spectrum. There is a trade-off between the number of available pair of channels and the likelihood of a particular user finding that no channels are available during the peak time. C is the number of trunked channels offered (channel capacity) by a trunked radio system and A is the total offered traffic (offered load).

The trunking theory was developed by Danish mathematician Erlang. After his name a unit (Erlang) to measure the traffic intensity in a network is introduced. One Erlang represents the amount of traffic intensity carried by a channel when completely occupied. The traffic intensity A Erlang is expressed as $A = \gamma \times h$, where γ is the mean rate of connection requests attempted per unit time and h is the mean call holding time per successful connection. Thus A is the average number of calls arriving during the average holding period.

The traffic intensity A Erlang can also be expressed as per user traffic intensity (A_{pu}) and the number of users (n), i.e. $A = n \times A_{pu}$. If the average rate of connection request in a system is 25 calls/minute, and the average call holding time is 3 minutes, then the traffic intensity $A = 75$ (as $A = \gamma \times h$). This indicates if the channel capacity is exactly 75, it can meet the average demand. To meet the worst-case traffic load, the chosen capacity should be greater than 75. On the other hand, the channel capacity 150 indicates the channels in the system are remaining half-utilised.

In the trunked cellular system, when a user tries to communicate and all the channels are busy, then either the user is blocked, i.e. denied access to the system, or the user is queued until a channel is available. In the first case, the trunking is called blocked call cleared or lost call cleared (LCC); whereas the queuing is referred to as blocked call delayed or lost calls delayed (LCD). In general the cellular systems implement LCC. This model assumes the following:

- fixed arrival rate that follows poisson distribution
- memory-less arrivals of call requests, i.e. all users including blocked users may request a channel
- probability of a user occupying a channel is exponentially distributed so that long-duration calls are less likely to occur
- finite number of channels
- infinite number of users

An LCC model based system follows Erlang *B* formula to determine the blocking probability (*P*) of a new attempted call, where

$$P = \frac{\dfrac{A^C}{C!}}{\displaystyle\sum_{x=0}^{C} \dfrac{A^x}{x!}}$$

is the measure of GOS. For example, GOS = 0.001 implies that during a busy hour, on an average, one out of 1000 call requests may be blocked. Table 2.1 presents sample values of offered load to achieve the desired GOS for different number of channels.

Table 2.1 Capacity of an Erlang *B* System

Number of channels (C)	*Offered load*				
	GOS = 0.001	GOS = 0.002	GOS = 0.005	GOS = 0.01	GOS = 0.02
10	3.09	3.43	3.96	4.46	5.08
20	9.41	10.07	11.10	12.03	13.18
30	16.68	17.61	19.03	20.34	21.93
40	24.44	25.60	27.38	29.01	30.99
50	32.51	33.88	35.98	37.90	40.25
60	40.79	42.35	44.76	46.95	49.64
70	49.24	50.98	53.66	56.11	59.13
80	57.81	59.72	62.67	65.36	68.69
90	66.48	68.56	71.75	74.68	78.30
100	75.24	77.47	80.91	84.06	87.97

The table enables computation of—

- the GOS (*P*) for a given offered load (*A*) and the number of channels (*C*).

- the traffic load (A) that can be handled for a given number of channels (C) to achieve a desired GOS (P).

Example 2.4 Let us consider the network with C = 20 (number of channels/cell) and on an average each user makes 3 calls/hr. The average duration of a call is 2 minutes. The number of users supported in a cell with 1% blocking (GOS = 0.01) can be estimated from Table 2.1.

Now, as $C = 20$ and $P = 0.01$, then from Table 2.1, $A = 12.03$. Moreover, $A_{pu} = \lambda \times h = 3 \times (2/60) = 0.1$ Erlang. So, the number of users admissible to a cell = $12.03/0.1 \approx 120$.

Example 2.5 In a system with C = 120 channels, if the utilization of channel is 2/3, then the traffic intensity (load) in this network is $A = 2/3 \times 120 = 80$.

2.10 CAPACITY IMPROVING METHODS

The overall system capacity in a cellular network increases with the introduction of frequency reuse (Section 2.3). However, the frequency reuse cannot always avoid the cases of hotspot in a network. The channels assigned to the network cells may not be sufficient to accommodate all the users seeking connection. A set of cells may get over-committed resulting the hotspots (cells) within the network. This problem is taken care of by the technique called 'load balancing'.

To address the issue of hotspot, the capacity of a cellular mobile network can be increased simply by allocating/adding new channels. Alternatively, cells in the hotspot zone may borrow channels from the relatively cold cells (if any). However, a borrowed channel may be in use within the co-channel cell(s) of the lender cell. This causes channel interference and creates additional disturbances in information exchange. Therefore, during channel borrowing the necessary condition to be checked is that the lending cell has at least one free channel and that is not being used by any of its co-channel cell. Once borrowed, the channel in the lender as well as in all its co-channel cells are locked. Many variations of channel borrowing techniques are reported in the literature. However, it is still an active area of research. Detailed discussion on that is beyond the scope of this book.

Besides the manipulation of radio resources (such as channel), the other useful measures to increase the capacity of a system are 'cell splitting' and 'cell sectoring'. In 'cell splitting', the network cells are partitioned into smaller cells, whereas in 'cell sectoring', a cell is divided into several sectors resulting in increase in the system capacity.

2.10.1 Cell Splitting

The cell splitting is a mechanism of partitioning a network cell into a desired number of 'microcells', each having its own base station. When a cell becomes congested, it is partitioned into smaller subcells (microcells), as illustrated in Figure 2.6, to increase the system capacity. The heavy-traffic regions (e.g. hotspots) are normally split into such smaller areas to ensure the acceptable grade of service.

In a microcell, antenna height as well as the transmitter power is lowered. While a cell diameter ($2R$) ranges from 2 to 20 kilometres, the microcell diameter ranges from hundred metres to a kilometre. The cell splitting technique decreases R (Figure 2.6) while leaving Q (frequency reuse ratio) relatively unchanged. This mechanism allows a system to grow by replacing large cells with smaller cells without upsetting the channel allocation scheme required to maintain the minimum frequency reuse ratio (Q or D/R).

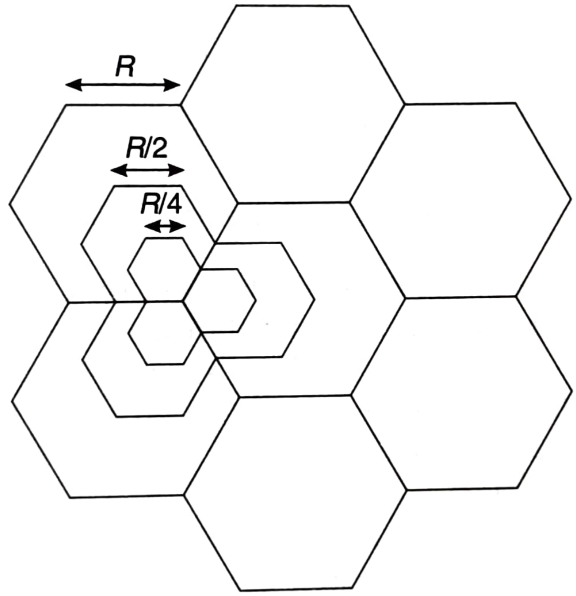


Figure 2.6 Cell splitting.

As an effect of cell splitting, the capacity of a network increases due to the additional number of channels per unit area. But the increased number of cells in a network adds more cell-boundaries. This may lead to more frequent handoffs. Therefore, the application of cell splitting is limited only to the cells with heavy traffic overloads. However, splitting only in a part of the system may result in serious channel assignment problems.

The designers favour cell splitting because of its capacity improving capability. The powers used by an MS and the BS after splitting is also comparatively low. The problem of unwanted handoff, however, is addressed with introduction of the concept of umbrella cells.

2.10.2 Cell Sectoring

Sectorization is the division of an omnidirectional (360°) view into non-overlapping slices called 'sectors' (Figure 2.7). Each sector has its own set of channels and considers directional antenna. The directional antennas at the BS are used to focus on the sectors. Usually this method divides a cell into three or six sectors. When a cell is sectored, the R (cell radius) is remained unchanged. Further, D is reduced, and the amount of frequency reuse is increased. Hence the capacity is increased. Therefore, bandwidth efficiency of the system is enhanced as the frequency can be reused more often.

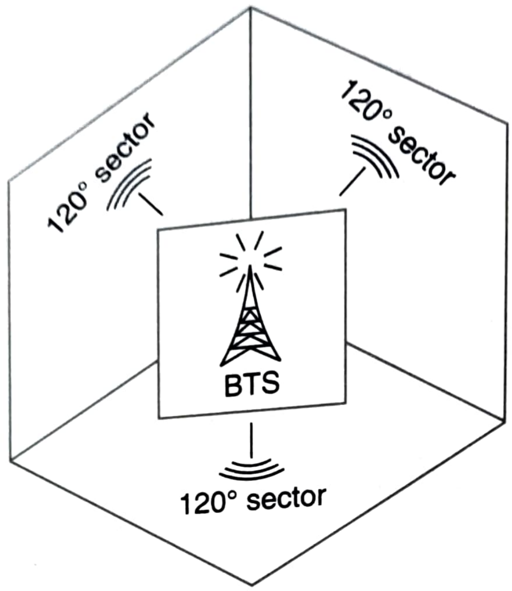


Figure 2.7 Cell sectoring.

The merit of cell sectoring lies on the fact that instead of interference received from all directions, it comes only from one direction causing the increase in signal to interference ratio. This enables to reduce the cluster size, and in effect allocates more channels per cell. However, within a cell an MS may need to handoff while crossing a sector boundary. The handoff is managed by the BS at the cell centre.

The cell sectoring adds complexity at the BSs due to introduction of additional antennas. Moreover, this mechanism decreases the trunking efficiency (queuing efficiency) while a large number of customers receive service from a set of servers. The proportional assignment of customers to different servers is a better solution.

2.11 USER VALIDATION IN CELLULAR COMMUNICATION

Just like any other network, user validation is an essential task in cellular network. In this section, the subscriber validation process, implemented in AMPS, is described.

2.11.1 Sources of Piracy

A common practice in unauthorized use of terminals is the cloning of stolen terminals. In AMPS, a terminal stores two ID numbers—the electronic serial number (ESN) and mobile identification number (MIN). The ESN is a 32-bit hardware-based serial number (unique 8-bit manufacturer's code + 24-bit ID for the mobile node for the given manufacturer's code). The MIN corresponds to a user telephone number (area code + phone ID) assigned when a subscriber account is opened. Both the ESN and MIN are stored in the mobile terminal (MS).

The intruder can extract the ESN of a subscriber from the EEPROM (Electrically Erasable Programmable ROM) of the mobile device. The conventional measure for preventing unlawful access to such IDs is the encryption. The ESN is encrypted before writing them into the mobile terminal's EEPROM. Whenever the terminal tries to send ESN, the terminal decrypts the ESN first and then transmits over the air. That is, an ESN is easily accessible to the intruder who taps the network transmission. Therefore, it requires further elaborated security measure that prevents access to the network resources by an intruder. One such popular scheme is described in the next section.

2.11.2 Validation

The validation procedure in AMPS relies on the two IDs of a terminal (MS)—the ESN and MIN. The MS sends both the (ESN, MIN) to the network for:

- Registration
- Initiating a call
- Roaming

The subscriber validation process in the early AMPS system implemented verification of the transmitted tuple (ESN, MIN) with that already registered in the network. It also maintained the list of stolen terminal IDs to protect the system from piracy (Section 2.11.1). However, as both the IDs were transmitted straight over the air, an intruder had the scope to clone the ESN and MIN. Therefore, protection for unlawful access to the network was not guaranteed.

The advanced AMPS system uses a key-based authentication procedure to validate a subscriber. The serving network provides an A-key (64-bit primary secret key) value for each MS. During authentication a shared secret data (SSD, 128-bit shared secret key) is generated from the A-key and the ESN. The mobile terminal then runs an authentication algorithm and generates a terminal authentication result (MS-AUTHR). The MS-AUTHR along with the ESN and MIN is then transmitted from the MS to the network. Upon receipt of the information, the network registers the terminal. The serving network also computes SSD separately with the stored A-key and based on this SSD a network generated NT-AUTHR is produced. Finally, the NT-AUTHR and the MS-AUTHR are compared. If comparing a match indicates the MS is a valid user, the network allows the call initiated from the MS to proceed.

An unauthorised MS is not supposed to know A-key and hence the SSD is not known to it, the authentication data MS-AUTHR sent by the unauthorised MS will not match the NT-AUTHR. Therefore, the system should be able to recognise the unauthorised MS and denies its service.The algorithmic steps of this authentication process (Figure 2.8) are noted below:

- An MS has a unique A-key supplied by the serving network.
- A SSD (shared secret key) is derived at the MS from A-key and ESN.
- MS runs an authentication algorithm and generates Authentication Result (MS-AUTHR) based on the SSD.
- The network produces a network generated AUTHR (NT-AUTHR) with the help of A-key stored in it.
- The NT-AUTHR and MS-AUTHR are compared at the network station. If there is a match, the network allows the call to proceed.

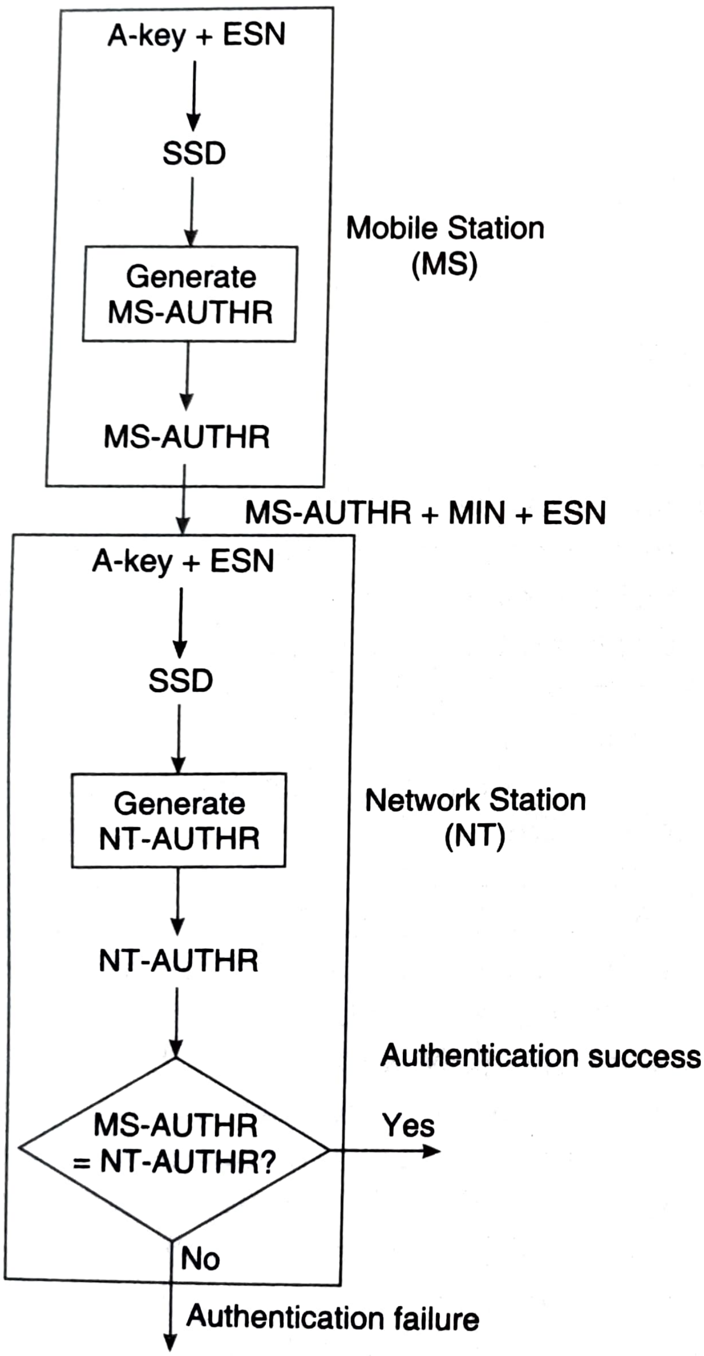


Figure 2.8 User validation process.