Chapter 2
Cellular Wireless Communication

2.1 Introduction

Originally, the focus of mobile radio systems design was towards increasing the coverage of a single transceiver. A single powerful base station was employed to provide connectivity to all mobile devices in a service area. Since spectrum length allocated for private communication is limited, it led to spectral congestion in the service area when large number of mobile clients became concurrently active. Typically, in radio communication system, a user requires about 30 kHz for voice communication. Therefore, if a high power antenna is mounted on a large tower to cover an entire town, it can support just about 25 MHz/30 kHz = 833 users, assuming that 25 MHz spectral band is available for private communication. An obvious way to get around the technical limitations and increase both the capacity and coverage is to reuse allocated frequencies without interferences. The idea was driven by a simple counter thought to the use of high powered transceiver. When a limited range transceiver is used then the wireless connectivity can be provided only in a small finite area of few hundred square meters. However, the frequency of an already deployed transceivers can now be reused by deploying another similar transceiver at a distance where the new transceiver does not interfere with the transceivers which were deployed earlier. In other words, spectral congestion can be eliminated by developing an architecture that would allow spatial multiplexing.

The concept of cellular architecture [3, 5, 6, 9] became the turning point in wireless communication technologies based on frequency reuse. The success of spatial multiplexing depends not only on elimination of interferences but also to provide continuous uninterrupted coverage. To provide continuous coverage, the uncovered gaps in coverage area should be serviced by the transceivers operating with frequencies different from the previously deployed transceivers. The proposed deployment of transceivers is equivalent to partitioning of a large coverage area using certain small finite continuous area which may be appropriately called as a cell and serviced by a single transceiver.

R.K. Ghosh, *Wireless Networking and Mobile Data Management*,
DOI 10.1007/978-981-10-3941-6_2

The frequency reuse problem can be viewed in terms of map coloring. In a map, regions are typically demarcated by different colors. If two adjacent regions are colored by same color then it is difficult to distinguish one from the other. Since, multi-colored printing is expensive, as few colors as possible should be used for coloring of a map. It is well known that map coloring can be accomplished by use of four colors [11]. In a way, frequency reuse can be seen as similar to reuse of colors in coloring of a map. There are, however, many differences. Frequency reuse requires a minimum separation between cells which is dependent on the strength of signal interferences between the cells. However, one single color can be used for coloring two regions in a map provided a separation of one region exists in between the two. The problems like co-channel and adjacent channel interferences which are encountered in planning of frequency reuse have no parallels in the map coloring.

Under ideal scenario, wireless signals may be assumed to propagate equally in all directions. Therefore, a cell representing the coverage area of an antenna can be considered as a circle. To provide continuous coverage, we need to find a packing of the desired area using circles, each having an area equal to that of a cell. The packing has to be done in a way such that there are no uncovered gaps in the service area. This is possible in the way as illustrated by Fig. 2.1 where the circles overlap minimally. The common chords of a circle with adjacent overlapping circles define a hexagonal area within that circle as indicated by the hexagon in the figure. The collection of hexagons like the one shown in the figure, packs the service area without leaving gaps. From here on, we will consider a cell to be a regular hexagon as explained. In reality, however, this is not the case. Radio waves like light waves are affected by reflection, refraction, diffraction, absorption, polarization and scattering. In reality terrain structures could be very different from one coverage area to another. There may be tall buildings, hillocks and tall trees which come in the way of signal

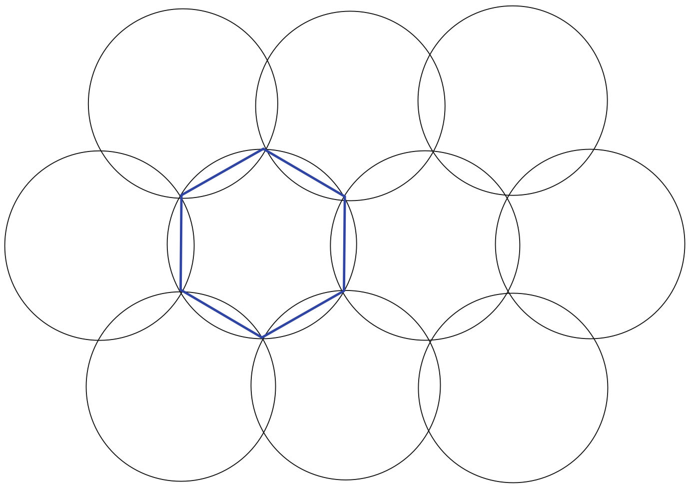


Fig. 2.1 Cell packing of a coverage area

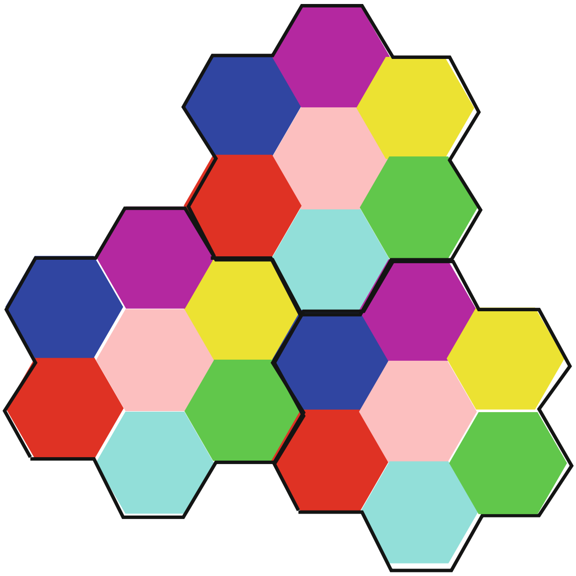
transmission. The actual shape of a cell is typically determined by field measurements radio signal propagation. But for convenience in presentation of basic ideas, assuming the cell to have a hexagonal boundary would suffice.

2.2 Frequency Planning

Figure 2.2 illustrates the frequency reuse concept. The set of cells in same shade, use the same frequency. Typically, an antenna is designed to work over the entire frequency spectrum. But it is possible to block most part of the spectrum and leave only a small group of selected channels on an antenna. So, a cell can be assigned a selected group of channels for communication inside it. The group of channels allocated to one cell will thus be different from the groups of channels assigned to the antennas of its geographically adjacent cells. The process of allocating channel groups to cells in a coverage area is called frequency reuse or frequency planning.

If a system has C duplex channels then each cell x can be allocated a set of C_x channels, where $C_x < C$. Assuming that C channels are equally divided among a group of N cells where each cell gets C_x number of channels, $C = N \times C_x$. The group of N cells which collectively share C channels is called a cluster. If there are R_c replicated clusters, the capacity K of the system is $K = R_c \times N \times C_x = R_c \times C$. This means that the capacity of a cellular system is directly proportional to the number of clusters. Since a cluster defines a full reuse of the set of allocated frequencies, the

Fig. 2.2 Frequency reuse concept



size of a cluster is the determining factor for capacity of a cellular based wireless communication system.

The cells which use same frequencies for communication are referred to as co-channel cells. Co-channel interferences can be minimized by increasing the reuse distance. So, if the cell radius is not changed, the cluster size N should be increased in order to increase the co-channel distance. But if N becomes large then a given service area can be covered by only a few number of clusters. So, the capacity of the service area, which depends on the replication of clusters, cannot be increased substantially. If the value of N is decreased then more clusters will be needed to cover the service area, so the capacity will increase. When cluster size becomes small, the distance between co-channel cells will become small. Though it allows reuse of the same frequency more frequently, the co-channel interferences increase. Thus, in order to maximize the capacity over a coverage area while it is desirable to keep N low, care should be taken so that co-channel interferences do not begin to affect the quality of communication.

The effect of frequency reuse can be best understood through a simple example. Suppose, initially a single high power transceiver was operating with a private spectrum of 25 MHz. This transceiver was replaced by 28 low power transceivers. Let the cluster size be 7. Then the effective spectrum allocation increases by four fold to 25 MHz $\times$ 4 = 100 MHz after replacing the high power transceiver. Each cell gets an allocation of 25 MHz/7 = 3.57 MHz of spectrum under the new system. Assuming voice communication to require 30 KHz, each cell can support up to 3.57 MHz/30 kHz = 119 users. The number of users in each cluster remains at 119 $\times$ 7 = 833. However, due to replications of clusters, the total number of users that can be supported under the new system increases by four fold to 833 $\times$ 4 = 3332.

With underlying assumption of a hexagonal layout of cells, it is possible to derive the formula for reuse distance. Let us first examine the geometry of a hexagonal cell. As shown in Fig. 2.3, the length of a side of a regular hexagon inscribed in a circle of radius R is also equal to R. Each side of this regular hexagon is at a distance $R\sqrt{3}/2$ units from the center of hexagon. The area of such a hexagon is $6(\sqrt{3}/4)R^2$ unit^2.

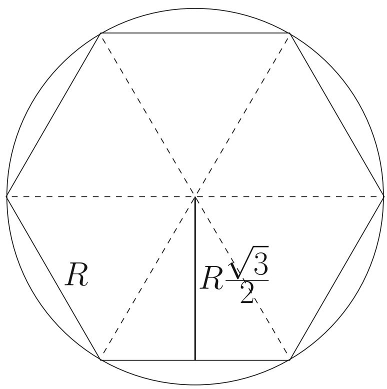


Fig. 2.3 A regular hexagon inscribed in circle of radius R

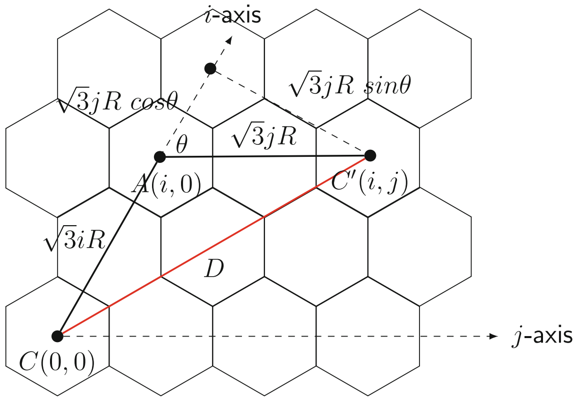


Fig. 2.4 Cell distance using hexagonal cell geometry

Now let us see how the reuse distance can be calculated. But, before that the unit for measurement of distance should be abstracted out from consideration. To realize it, we assume the centers of cells to be at integer coordinates. The arrangement of the cell centers according to hexagonal layout, thus, form a grid system whose axes are at an angle of $\pi/3$. Using the above fact and some simple properties of regular hexagon, the following lemma provides the relationship between the distance of a co-channel cell from a cell.

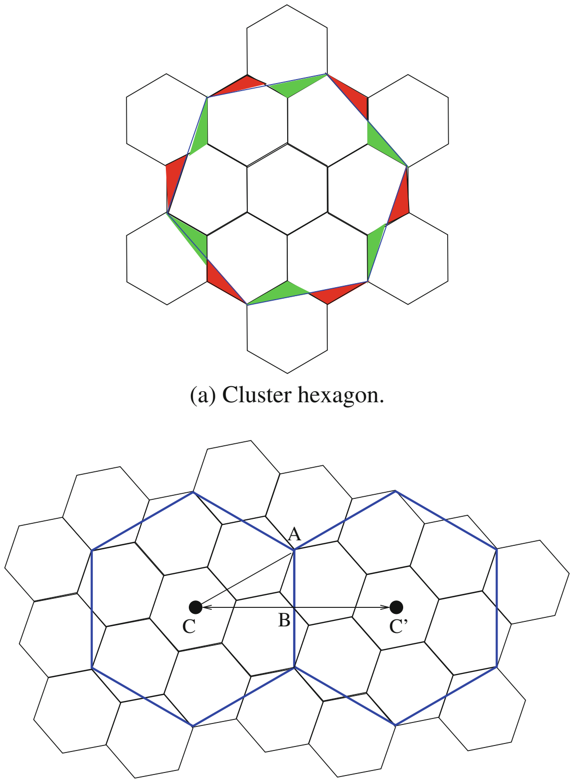
Lemma 2.1 *Let the coordinate of a cell C be (0, 0) and that of its co-channel cell C' be (i, j). Then co-channel distance $D = CC'$ is equal to $R\sqrt{3(i^2 + ij + j^2)}$, where R denotes the length of a side of the regular hexagon representing a cell.*

Proof Consider the Fig. 2.4. Since, the coordinates of the cell centers are given by a pair of integers, and the axes of the coordinate system between them form an angle of $\theta = \pi/3$, C' can be reached from C by traversing i cells in one axial direction then turning $\pi/3$ in clockwise and then hopping j cells in the other axial direction. From the above figure, distance $AB = \sqrt{3}jR\cos\pi/3$, $BC' = \sqrt{3}jR\sin\pi/3$. So

$$
\begin{aligned}
D^2 &= (\sqrt{3}iR + \sqrt{3}jR\cos\pi/3)^2 + (\sqrt{3}jR\sin\pi/3)^2 \\
&= R^2(3i^2 + 6ij\cos\pi/3 + 3j^2\cos^2\pi/3 + 3j^2\sin^2\pi/3) \\
&= R^2(3i^2 + 3ij + 3j^2) \\
&= 3R^2(i^2 + ij + j^2)
\end{aligned}
$$

Each cell in a cluster is assumed to have identical wireless range, and equal number of channels. So, the number of cells, N, per cluster can be obtained by determining the area of the cluster and then dividing it by the area of a cell. The area of a cluster can be found easily from the simple analysis of hexagonal geometry.

Fig. 2.5 Cluster area approximates to a regular hexagon



(a) Cluster hexagon.

(b) Area of cluster hexagon.

Lemma 2.2 *The cluster size* $N = i^2 + ij + j^2$.

Proof Consider the diagram in Fig. 2.5a. Let the large regular hexagon enclosing a cluster be referred to as cluster hexagon for brevity. Consider the shaded tiny triangles in the figure. A pair of adjacent shaded triangles one inside and the other outside the cluster hexagon are congruent. So the area of the cluster of cells is equal to the cluster hexagon indicted in Fig. 2.5a.

Now consider Fig. 2.5b. The distance between the centers C and C' of two adjacent cluster is D. The angle CAB is equal to $\pi/3$. So, the length $CA = (D/2)/\sin\pi/3 = D/\sqrt{3}$ which is equal to the radius of the cluster hexagon.

As explained earlier, the area of a single hexagonal cell of side R is $6R^2\sqrt{3}/4$. Since, a cluster is also a regular hexagon of side $D/\sqrt{3}$, its area is given by $Cluster_{area} = 6\frac{D^2}{4\sqrt{3}}$. Since the area of cluster divided by the area of one cell should be equal to N, we have

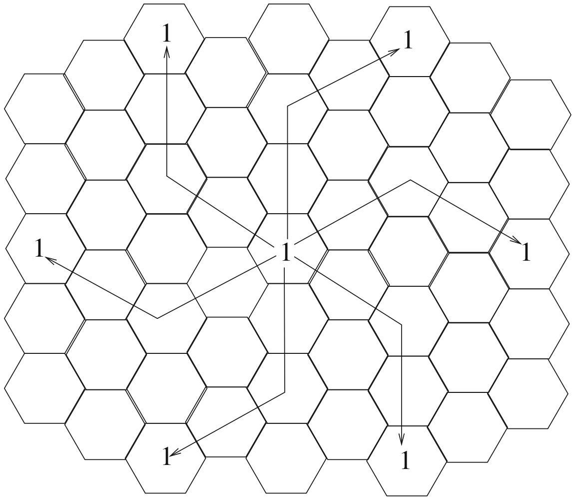


Fig. 2.6 Co-channel cells

$$N = \left(\frac{D^2}{4\sqrt{3}}\right) / \left(\frac{R^2\sqrt{3}}{4}\right) = \frac{D^2}{3R^2} = \frac{3R^2(i^2+ij+j^2)}{3R^2}$$
$$= i^2 + ij + j^2$$

Figure 2.6 depicts co-channel cells of a cell which use the same frequency. Here $i = 2$ and $j = 2$, and cluster size is $N = 2^2 + 2.2 + 2^2 = 12$. The ratio $D/R = \sqrt{3N}$ denoted by Q represents co-channel reuse ratio. A small value of Q means N is small. This leads to large capacity. However, when Q is large, it leads to low interference; hence better transmission quality. A trade off has to be made between capacity and transmission quality.

2.2.1 Co-channel Interference

Co-channel interference is one of the major problem faced by service providers in setting up wireless communication service. For a good understanding of how interference could affect communication, it is important to examine radio signal measurement and propagation.

DeciBel is the unit used for the measurement of relative strengths of radio signals from two different radios. Ten deciBels (dB) equals one Bel (B) which represents the power ratio of 1:10. A power ratio of 1:100 equals 2 B or 20 dB. Similarly, the power ratio 1:1000 is measured as 3 B or 30 dB. Use of log scale in measurement of relative power strengths simplifies the calculation. The log scale expression $\log_{10}(P_2/P_1)$, gives the measurement of relative power strength due to amplification in Bels. For example, if an amplifier outputs 100 W with an input of 100 mW, then the power gain due to amplification is $\log_{10}(100/0.1) = \log_{10} 1000 = 3$ B or 30 dB. In the case of radio signal measurements, power ratios are usually very small. So, deciBels is the preferred unit for expressing smaller power ratios. Accordingly, the formula

$$10 \log_{10} (P_2/P_1),$$

is normally used for measuring relative power strengths between the two radio sources.

When the measurement of power strength is made with reference to 1 W then the unit of measurement is denoted by dB or dBW. Sometimes power strength is measured with reference to 1 mW, and the unit of measurement is denoted by dBm. Unless stated otherwise, the unit of measurement here will be dBm. Zero dBm equals 1 mW. In general, a power of P watts equals x dBm, where

$$x = 10 \log_{10} \left(\frac{P}{10^{-3}} \right) = 30 + 10 \log_{10} P$$

A negative gain indicates a signal decay or attenuation, which is represented by a power ratio less than 1.

Let us consider an example to understand the significance of log scale in computation. Suppose, a micro-wave system uses a 10 W transmitter. This transmitter is connected by a cable with 0.7 dBm loss to a 13 dBm antenna. Let the atmospheric loss be 137 dB on transmission. The receiver antenna with 11 dBm gain connected by a cable with 1.3 dBm loss to a receiver. Then the power at the receiver can be calculated as follows:

Transmitter output: 10 W = 10000 mW.
Amplification gain at transmitter: $10\log_{10}(10000/1) = 40$ dBm.

Then the relative strength of power at the receiver equals $(40 - 0.7 + 13 - 137)$ dBm $= -84.7$ dBm including the amplification gain, atmospheric loss and loss due to cable connecting the amplifier to its antenna. On the receiver side, the antenna and the connecting cable together lead to (11–1.3) dBm gain. So the net power received by the receiver is $(-84.7 + 9.7)$ dBm $= -75$ dBm.

The co-channel interferences experienced by a cell is due to the use of same group of channels in the nearby cells. The quality of the signals received from the current base station is affected by co-channel interferences. As Fig. 2.7 depicts, the wireless signals from the base stations in co-channel cells around a cell ripple out much like waves in water when a stone is thrown into the center of a pond. The ratio of the signal strength from the current cell and the strength of co-channel interferences, provides a measure for the quality of communication. The Signal to Interference Ratio (SIR) should be monitored by individual mobile terminals. If the strength of interference increases compared to the strength of signal, then the quality of communication deteriorates. A Mobile Terminal (MT) experiencing low SIR should try to switch to a neighboring cell which may provide better SIR. The SIR is given

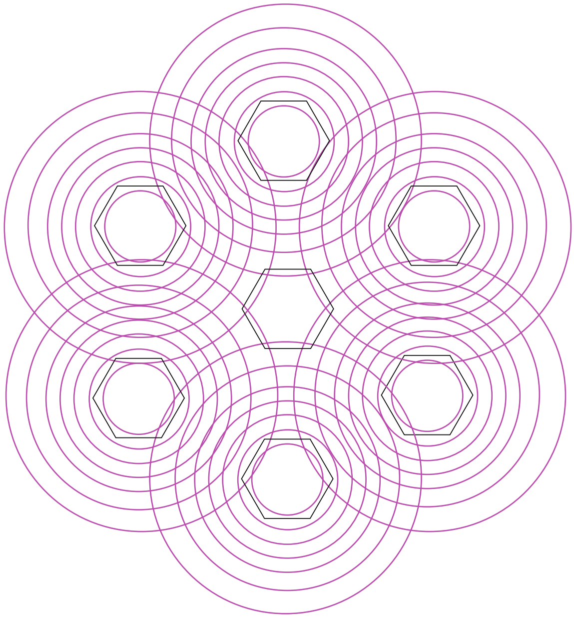


Fig. 2.7 Interferences due to propagation of co-channel signals

$$S/I = S/\left(\sum_{i=1}^{i_0} I_i\right),$$

where I_i is the interfering signal received from co-channel i, and i_0 is the number of adjacent co-channel cells.

In free space, the average signal strength known to decay according a power law involving the distance between the transmitter and the receiver. Let d be the distance between the transmitter and the receiver, and P_0 be the power received at a near-by reference point in the far-field region (with a distance of 2λ) of the transmitter antenna. Then the average received power P_r at receiver from the transmitting antenna is given by:

$$P_r = P_0 \left(\frac{d}{d_0}\right)^{-n}.$$

In reality P_r is proportional to the expression in the right hand side of the equation. Here, the constant of proportionality is assumed to be 1. However, its actual value will depend on antenna gain, path loss exponent and carrier wavelength. The above equation for power received, when expressed in terms of dBm, becomes

$$P_r(\text{dBm}) = P_0(\text{dBm}) - 10n \log_{10}\left(\frac{d}{d_0}\right).$$

Now let us examine the topological configuration of cells, and study how a MT may arrive at a better approximation of SIR. Suppose, the MT is at a distance D_i from i-th co-channel cell. The distance is measured from the cell center where the corresponding base station is located. So, the signal attenuation from ith co-channel cell is proportional to D_i^{-n}. The signal strength beamed from the current base station to the mobile terminal is proportional to R^{-n}, where R is the radio range of the base station. Assuming all interfering co-channel base stations to be at equal distance from the MT, the SIR (in dB) is equal to

$$\begin{aligned} S/I &= \left(R^{-n} / \left(\sum_{i=1}^{i_0} D_i^{-n}\right)\right) \\ &= \frac{(D/R)^n}{i_0} \\ &= \left(\sqrt{3N}\right)^n / i_0. \end{aligned} \tag{2.1}$$

Let the minimum value of SIR for good voice quality be 18 dBm. In order to check whether a cluster size of $N = 7$ in a cellular based wireless system with the path loss exponent $n = 4$ would meet the requirement of voice quality, let us compute SIR value received by a MT which is located at center of its current cell. Using Eq. 2.1:

$$\begin{aligned} 10\log_{10}(S/I) &= 10\log_{10}\left(\left(\sqrt{3N}\right)^4 / i_0\right) \\ &= 10\log_{10}\left(\left(\sqrt{21}\right)^4 / 6\right) \\ &= 10\log_{10} 73.5 \\ &= 10 \times 1.865 \\ &= 18.65 \end{aligned}$$

As the SIR value is above the required threshold of 18 dBm, a cluster size of 7 is alright when path loss exponent is 4.

When cell sectorization is used with 120° sectors, the number of co-channel cells which is reduced from 6 to 2 for $N = 7$. Therefore,

$$S/I = \frac{1}{2}\left(\sqrt{3N}\right)^n$$

Thus the increase in SIR with sectorization is 3 times more than that without sectorization.

The scenario that a MT will be located at the center of the current cell, occurs very rarely. So, topological configuration of the current cell, its co-channel cells and the position of MT should be taken into account for computing SIR. Figure 2.8 [10] depicts a realistic topological configuration. It is assumed that MT is located at the

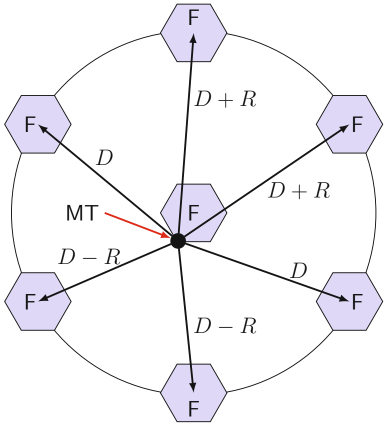


Fig. 2.8 Co-channel interference considering exact cell geometry

boundary of its current cell. The signal strength received by MT at this position from is at the lowest possible level. With cluster size of 7, there will be two co-channel cells at a distance $D - R$ from MT, two at a distance D and two others at a distance $D + R$ as indicated in Fig. 2.8. So the ratio of power strengths of the current base station and the other interfering base stations, with path loss factor $n = 4$, is

$$S/I = \frac{R^{-4}}{2(D-R)^{-4} + 2D^{-4} + 2(D+R)^{-4}}$$
$$= \frac{1}{2(\sqrt{21}-1)^{-4} + 2(\sqrt{21})^{-4} + 2(\sqrt{21}+1)^{-4}}$$

The above expression is equal to 49.56. In terms of deciBels the value of SIR will thus be $10 \log_{10} 49.56 = 17$ dBm. Since the value of SIR is lower than acceptable threshold the voice quality will not be good.

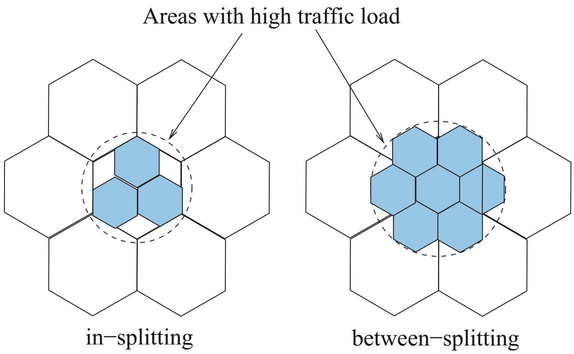
2.2.2 Cell Splitting and Sectoring

As observed earlier, frequency reuse can be increased using smaller cells. But smaller cells will lead to shorter co-channel distance; and therefore, an increase in co-channel interferences. The challenge is to keep interference low and increase the capacity. After a cellular service area has been planned and the infrastructure is in place, any incremental change like adding channels to a cell for handling the increase in traffic load is expensive and difficult to execute. However, two simple ideas, namely,

1. Cell splitting
2. Cell sectoring

are found to be quite effective for handling the increase in traffic load.

Fig. 2.9 Cell splitting



Cell splitting allows creation of smaller cells out of a standard cell. The advantage of cell splitting is that it uses the same idea of spatial multiplexing with smaller cells called microcells. The antennas can be placed on the top of buildings, hills and even on the lamp posts. Smaller cells are placed in or between large cells. Figure 2.9 illustrates the two possible splittings where radius of every microcell is half the radius of a standard cell. Splitting increases the number channels because the number of channels per unit area increases. But the trade off is in terms of increase in co-channel interference. As Eq. 2.1 in Sect. 2.2.1 indicates, co-channel interference can be controlled by keeping the ratio D/R unchanged. If D is decreased, R must also be decreased to keep the ratio at same value. Thus, the transmission power of newly introduced microcells must be reduced to avoid the co-channel interference.

Let P_s be the output power of the transceiver of a standard cell, and P_m be the output power of the transceiver of a microcell. Then the received power P_r at cell boundaries of the two cells are:

$$P_r[\text{original cell}] \propto P_s R^{-n}$$
$$P_r[\text{micro-cell}] \propto P_m (R/2)^{-n}$$

It means that the ratio of transmit powers of a microcell versus a normal cell is 1:16 when path loss exponent is 4. It is not necessary to split all the cells. Sometimes it becomes difficult to exactly identify the coverage area that would require cell splitting. So in practice different cell sizes may co-exist. This calls for careful fine-tuning of power outputs by transceivers so that a safe distance is maintained among the co-channel cells and the co-channel interference is kept at a minimum level. But it makes the channel assignment quite complicated. As we observed, two different transmit powers will be needed to support cell splitting. The channels in a normal cell needs to be divided into two groups: (i) the one corresponding to normal transmit power, and (ii) the other corresponding to a reduced transmit power. The splitting determines the sizes of two channel groups. Initially, only a few channels belong to the reduced transmit power group. But as traffic grows, more and more channels will be required, causing the group of channels with reduced transmit power to grow in size. The splitting continues until the channels used in an area are all of low transmit power. However, the presence of dual cell size calls for better channel switching with mobility of the users. The movements within microcells will lead to more frequent channel switching compared to the movements within normal cells.

Cell sectoring is also used to address the problem of capacity increase. With this approach transmit power of channel is concentrated into a finite sector of the cell. The omni-directional antenna of the cell is replaced by several directional antennas. The sectoring causes receipt of co-channel interference and transmission only within a specified region of the cell. So, it leads to greater reuse of frequencies. Normally a cell is partitioned into three sectors of 120° or six sectors of 60° each as in Fig. 2.10. When sectoring is used the channels of a cell are partitioned into sectored groups, and the channels of a group is used only in one sector. As may be observed from Fig. 2.11, with cluster size of 7, the cells labeled *F* are co-channel cells. Each cell can

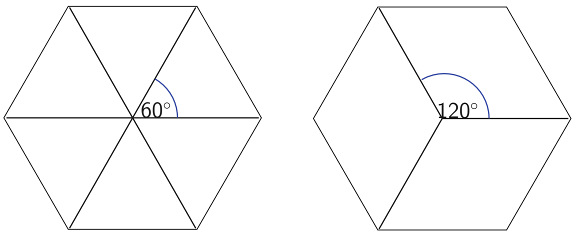


Fig. 2.10 Cell sectoring

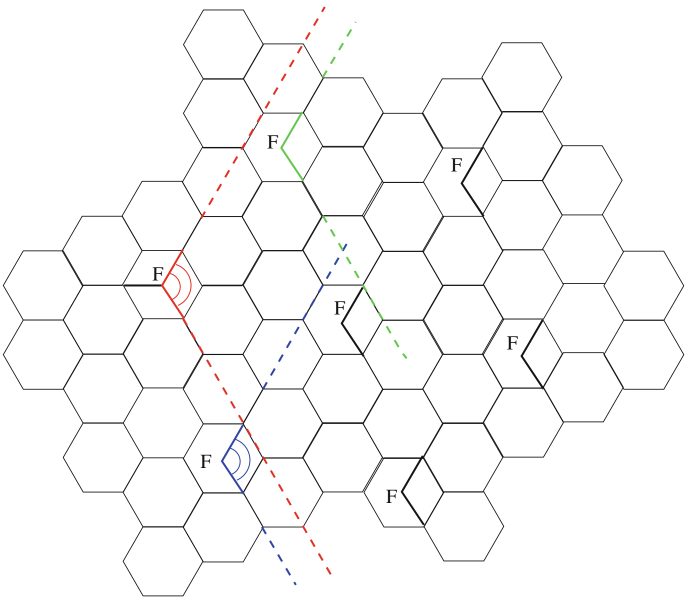


Fig. 2.11 Interference pattern with cell sectoring

receive signals only from two co-channel cells to its left. So, the cells at the center, which is within the signal propagation cones of its two co-channel cells on the left, may receive signals only from these two cells. The signal propagation cone of the co-channel cell vertically up to the left does not include the center cell, so the latter is not affected by co-channel interference due to the former cell.

2.3 Traffic Intensity

An important measurement related to capacity is traffic intensity. In a Telecom system, this measurement is an estimation of traffic pattern, pick load, and channel requirements. The traffic intensity varies over the day. The probability of securing a connection in busy hours is directly related to traffic intensity. The unit of measurement is *Erlang* [2]. One Erlang is equal to the total traffic generated by the voice calls amounting to 60 min. For example, if there are 40 calls in 1 h with each call having an average duration of 5 min, then the traffic in Erlang is:

$$\text{Traffic in hour } = (40 \times 5)/60 = 3.33 \text{ Erlangs.}$$

Formally, in a lossy system, Grade of Service (GoS) determines the probability of call blocking, which is computed by *Erlang B* traffic model. Erlang B traffic model is used by Telecom companies to determine the number of lines required to run a service based on anticipated traffic and future expansion.

Let λ be arrival rate and μ be service rate. Then $1/\lambda$ is the average time between arrival of two consecutive requests and $1/\mu$ is the average service time. For example, if the average duration of a connection is 3 min, then $1/\mu = 0.05$ (hour), equivalently, in an average $\mu = 20$ calls can be serviced per hour. To illustrate a concrete case, refer to Fig. 2.12 depicting the arrival of call requests and the servicing of these requests for 5 users. The intervals $I_i = a_{i+1} - a_i$, for $1 \le i \le 4$ represent the inter-arrival time. The duration of call services are represented by intervals $S_1 = d_1 - a_1, S_2 = d_2 - d_1$, $S_3 = d_3 - d_2, S_4 = d_4 - d_3, S_5 = d_4 - d_5$. The arrival and service rates are given by expressions $1/E(I_i)$ and $1/E(S_i)$ respectively, where $E(.)$ represents the expected values of the corresponding intervals.

The inter-arrival times for connection requests are typically modeled by Poisson distribution [13]. A Poisson process is a sequence of events which are randomly spaced in time. In a wireless network different users seek connections at different times independent of one another. Therefore, the call requests (representing events) in a cell can be represented by a Poisson process. The rate λ of a Poisson process is the average number of number events (arrival of call requests) per unit time over

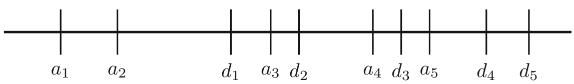


Fig. 2.12 Distribution of the arrival of requests and their servicing

a long period. The probability of n call requests arriving during an interval of time $[0, \delta)$ under Poisson process is,

$$Pr[n_{t+\delta} - n_t = n] = \frac{(\lambda\delta)^n}{n!}e^{-\lambda\delta}, \text{ for } n = 0, 1, \ldots,$$

where n_t denotes the number of arrivals since the time $t = 0$, and δ is the call inter-arrival time. For example, if we observe a system from some arbitrary time $t \geq 0$ during a small interval of time $\delta \geq 0$, the probabilities of arrival of number of call requests are:

$$\begin{aligned} Pr[n_{t+\delta} - n_t = 0] &= 1 - \lambda\delta + O(\delta^2) \\ Pr[n_{t+\delta} - n_t = 1] &= \lambda\delta + O(\delta^2) \\ Pr[n_{t+\delta} - n_t \geq 2] &= O(\delta^2), \end{aligned}$$

where $O(\delta^2)$ represents the probability of more than 1 call request arriving in time δ. Since, δ is small, no more than 1 call request can arrive at the system during this interval. Therefore, the event of two or more calls arriving in the system within an interval of δ can be considered as impossible. In other words, terms of $\mathrm{O}(\delta^2)$ can be safely ignored.

Assume that the number of channels is C. It means C connection requests can be serviced concurrently. Therefore, we have $M/M/C$ kind of queuing system [2] with following parameters:

- The arrival process is Poisson with arrival rate λ.
- The service time is exponential with servicing rate μ.
- The number of servers or the channels for serving the connection requests is C.
- The capacity or number clients which can be in the queue is also C.

The service scenario is best understood as a system with a limited number service lines connecting a large number of input lines (one for each call request) as shown in Fig. 2.13a. Figure 2.13b illustrates the same, but depicts how some of the service requests are met and others are dropped with probability of dropping a request being P_b.

Initially, 0 channels are used by system. Over a small interval the system may continue in state 0 with a probability of $1 - \lambda\delta$. The system will change to state 1 from state 0 with a probability $\lambda\delta$. But if one channel is already in use (i.e., system in state 1) then transition to state 0 will take place with a probability of $\mu\delta$. This implies that the systems continues in state 1 with a probability of $1 - \lambda\delta - \mu\delta$. So the system states and transitions is depicted by Fig. 2.14. Over a long period of time, the system reaches steady state. In a steady system if n channels remain occupied then, writing global balance equation for the steady state, we get

$$\begin{aligned} \lambda\delta P_{n-1} &= n\mu\delta P_n, n \leq C, \\ \lambda P_{n-1} &= n\mu P_n, \\ P_1 &= (\lambda P_0)/\mu. \end{aligned}$$

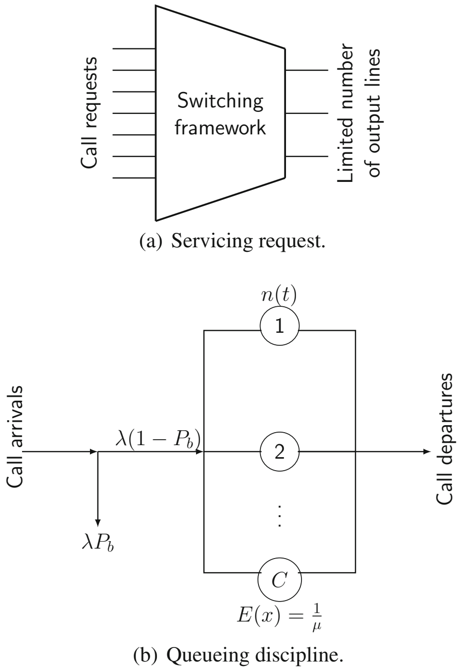


Fig. 2.13 Queuing discipline and service of call requests

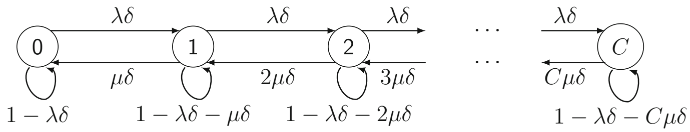


Fig. 2.14 Markov chain representing system state and transitions

The equation expresses the fact that transition from P_{n-1} to P_n is same as the transition from P_n to P_{n-1}. Solving the recurrence in balance equation,

$$P_n = P_0 \left(\frac{\lambda}{\mu}\right)^n \frac{1}{n!}.$$

Note that the sum of probabilities of the system being any of the states $n = 0, 1, 2, \ldots$, is 1. So, we have $\sum_0^C P_n = 1$. Now substituting for P_n in summation in terms of P_0,

$$\sum_0^C P_0 \left(\frac{\lambda}{\mu}\right)^n \frac{1}{n!} = 1$$

Therefore,

$$P_0 = \frac{1}{\sum_{n=0}^{C} \left(\frac{\lambda}{\mu}\right)^n \frac{1}{n!}}.$$

We already know that $P_C = P_0 \left(\frac{\lambda}{\mu}\right)^C \frac{1}{C!}$. This leads us to the expression for P_C as follows:

$$P_C = \frac{\left(\frac{\lambda}{\mu}\right)^C \frac{1}{C!}}{\sum_{n=0}^{C} \left(\frac{\lambda}{\mu}\right)^n \frac{1}{n!}}$$

The traffic intensity is determined by the ratio of arrival and departure rates. This ratio should not be allowed to exceed 1, otherwise request queue will build up. So, traffic intensity is actually a measure of congestion in the system. Let the traffic intensity be represented by $A = \lambda/\mu$. Now substituting A for traffic intensity Erlang B formula becomes:

$$P_C = \frac{A^C \frac{1}{C!}}{\sum_{n=0}^{C} A^n \frac{1}{n!}}.$$

The above formula is known as *blocked call cleared* formula as it determines the GoS for traffic system without a queue for blocked calls. Using Erlang B formula, the probability that a client's call request will not be serviced in a blocked call cleared system is:

$$Pr[\text{call blocking}] = \frac{A^C/C!}{\sum_{n=0}^{C} A^n/n!}.$$

To make computation of probability simple, the right hand side of the above equation can be rewritten as follows:

$$\frac{A^C/C!}{\sum_{k=0}^{C} A^n/n!} = \frac{1}{1 + \sum_{1}^{C} \left(\frac{C}{A}\right)\left(\frac{C-1}{A}\right)\ldots\left(\frac{C-n+1}{A}\right)}$$

The expression under summation in the numerator can be unrolled and recast using Horner's rule as follows:

$$\frac{C}{A} + \frac{C}{A}.\frac{C-1}{A} + \ldots + \frac{C}{A}.\frac{C-1}{A}\ldots\frac{1}{A} = \frac{C}{A}\left(\ldots\left(1 + \frac{2}{A}\left(1 + \frac{1}{A}\right)\right)\ldots\right).$$

The use of above expression restricts round-off and truncation errors in computation.

For an illustration of Erlang B formula, consider the following example. Suppose there are 240 connection requests per hour in peak time. So, the arrival rate is $\lambda = 240$. Let the average call duration be 3 min, or 0.05 h, then the service rate $\mu = 20$. It gives $A = 240/20 = 12$. Note that average number of request per hour $1/\lambda$, and the average call duration is $1/\mu$. Hence, the ratio $A = \lambda/\mu$ being equal to the product of the average number of requests and the average duration of the calls, is called the *busy hour traffic* (BHT). If there are 25 channels then the probability of call being blocked due to non-availability of a channel is given by

$$P_b = \frac{12^{25}\frac{1}{25!}}{\sum_{n=0}^{25} 12^n \frac{1}{n!}} = 3.78 \times 10^{-4}.$$

For a fixed value of A, we can indeed prove that the call blocking probability for busy hour traffic decreases with increase in C. When the number of channels increased by 1, the number of channels become $C + 1$. In steady state, balance equation tells us that $(C + 1)\mu P_{C+1} = \lambda P_C$. Applying a sequence of simplifications we get:

$$\begin{aligned} P_{C+1} &= \frac{\lambda}{\mu}\frac{P_C}{C+1} = \frac{A}{C+1}.\frac{A^C/C!}{1+\sum_1^{C+1} A^k/k!} \\ &< \frac{A^C/C!}{\frac{C+1}{A}\sum_1^{C+1} A^n/n!} \\ &= \frac{A^C/C!}{\frac{C+1}{A}\sum_0^{C} A^{n+1}/(n+1)!} \\ &= \frac{A^C/C!}{\sum_0^C ((C+1)/(n+1))(A^n/n!)} \\ &< \frac{A^C/C!}{\sum_0^C A^n/n!} = P_C. \end{aligned}$$

The boundary values for the above probability distribution function are $Pr[C = 0]$ = 1, and $Pr[C = \infty]$ = 0. Since the function is monotonic, with a given value of A, and a given value of call blocking probability p, there is a smallest integer C, such that the inequality $Pr[C] < p$ holds. This implies that the method of bisection can be applied to obtain the value of C.

2.4 Channel Assignment

Frequency planning consistent with the twin objective of increasing the capacity and guaranteeing the minimum co-channel interference is very important for efficient communication over wireless interfaces. The objective of frequency reuse can be met by employing different techniques such as partitioning the spectrum along frequency, time, or code. Frequency division scheme partitions spectrum allocating distinct frequency bands. A small guard band is placed between two adjacent bands to separate the frequencies. Time division achieves channel separation by disjoint time intervals called slots, while code division ensures channel separation by using different modulation codes. It is possible to combine different channel separation schemes. For example, time division and frequency division can be combined to divide each frequency band (obtained from frequency division) into time slots. The bottom line of the division principles is to maximize the separation of a channel with desired quality and also to maximize the channel utilization.

Broadly speaking channel assignments can be classified either as fixed or dynamic. The choice of specific channel assignment policy affects the performance. The impacts are felt in terms of quality of services as a connected mobile device moves from one cell to another. For example, an active communication may get terminated if mobile terminal moves from one cell to another which has no free channel. Thus channel assignment is one of the critical element of mobile communication.

Each cell is allocated a fixed number of channels in a fixed channel assignment scheme. A new connection in a cell can be activated only if there is a free channel. If all the channels are occupied then the request for a connection cannot be accommodated. It means an active communication will get terminated if a connected mobile terminal moves from a cell to another having no free channel. Many variations to fixed channel assignment scheme exists in order to eliminate the unpleasant interruptions during an ongoing communication. The process by which an active connection can be maintained as a mobile terminal moves from one cell to another is called *handoff*. Handoff provides the basic capability of mobility with an active connection in a cellular based wireless communication system, a detailed discussion on handoff can be found in Sect. 2.1.

In dynamic channel allocation schemes, no channel is permanently allocated to any cell. Each time a base station requires a channel to be allocated for a call, it requests the mobile switching center (MSC) for the allocation of a channel. The switch then allocates a channel using some sophisticated algorithms that can take care of the future call blockings, the volume of inter-cell and intra-cell handoffs, and co-channel interferences among other things. The effectiveness of dynamic channel allocation depends on MSC's ability to collect real-time data on channel occupancy, traffic distribution and received signal strength indication (RSSI) of all channels on a continuous basis. So, dynamic channel allocation increases both storage and computational load on the MSCs.

Channel assignment schemes can be implemented either in centralized or in decentralized fashion. In a centralized assignment, a central controller assigns the channels. In a decentralized assignment, many possibilities exists. The channels may be assigned by the local cell, or by the cell where the call originated, or even selected by mobile devices.

A general approach for the solution of the channel assignment problem is to use graph abstraction for representing cellular system, and transform the problem into a graph coloring problem. Then, the existing solutions for graph coloring can be applied for solving channel assignment problem.

Interference graph plays an important role in planning channel assignment. Such a graph is defined as follows:

- Every vertex represents a cell or a base station.
- Each edge (u, v) is associated with a weight $W(u, v)$ which is proportional to the maximum strength of signal interference between cells represented by u and v.
- There is also a system wide value W representing the minimum separation (maximum permissible signal interference) between any pair of channels allocated to the same cell.
- Each vertex v is also associated with a non-negative integer which indicates the channel requirements for the cell represented by v.

In theory, such a graph is a complete graph with lot of edges having 0 weights. In order to avoid interference, two channels a and b can be allocated to different cells u and v provided $|a - b| \leq W(u, v)$. This means frequency band representing channel a and that representing channel b cannot have mutual interference exceeding the prescribed weight $W(u, v)$. The value W is used to ensure minimum channel separation for channels allocated to a single cell. That is, if channels a and b are used in the same cell, then $|a - b| \leq W$.

The graph theoretic formulation given above is not quite concrete. From the point of view of a service provider, the important formulations are those that can handle both resource allocation and traffic constraints in different cells. Let us, therefore, discuss about such formulations. Suppose the number of cells be N_{cell}, and T_i is projected traffic requirement in cell C_i, $i = 1, 2, N_{cells}$. Then along with the constraints as stated earlier, new constraints concerning traffic requirements should be added. Therefore, the formulation of channel assignment is as follows:

Problem formulation *Given a number of cells C_i, $1 \leq i \leq N_{cells}$, the requested traffic T_i per cell C_i, the interference graph with weights W_{ij} and local interference constraint W, assign channels to cells so that*

1. *All frequency constraints are met.*
2. *Cell C_i gets T_i channels chn_{ij}, $1 \leq i \leq N_{cells}$, $1 \leq j \leq T_i$.*
3. *The quantity*
$$N_{channels} = \max\{chn_{ij}\}$$
is minimized.

The above formulation implies that each cell gets channel allocation according to its traffic requirements, and the overall system requirements for the channels is as small as possible.

In most general setting, the channel assignment can be posed as constraint satisfaction problem. An $n \times n$ symmetric matrix $C = \{c_{ij}\}$, known as compatibility matrix is defined, where c_{ij} represents the minimum frequency separation required between cells i and j, and n represents the number of cells. The minimum separation between the frequencies used by two cells ensures that the communication in each cell is free from the interference due to the communication in the other. Since the frequency bands are the evenly spaced, they can be identified by integers. The number of channels required for each cell is represented by a requirement vector $M = \{m_i\}$,

$i = 1, \ldots, n$. The frequency assignment vector $F = \{F_i\}$ is such that F_i is a finite subset of the positive integers which defines the frequencies assigned to cell i. F is admissible provided it satisfies the following constraints:

$$|F_i| = m_i, \text{ for } i = 1, \ldots n$$
$$|f - f'| \geq c_{ij}, \text{ where } f \in F_i \text{ and } f' \in F_j$$

The first constraint is needed to satisfy the channel requirement and the second constraint is needed to satisfy the interference free compatibility between each pair of cells. The largest integer belonging to F is known as the span of the frequency assignment. Note that the largest integer represents the minimum number of channels required for the frequency assignment. So, F with the minimum span constitutes the solution to the problem channel assignment. The problem is known to be NP hard [3].

2.4.1 Fixed Channel Assignment

In fixed channel assignment (FCA) scheme, a fixed number of channels is assigned to each cell. The set of available channels is partitioned into N disjoint sets, where N is the cluster size. As we already know $N = D^2/3R^2$ is related to reuse distance D, and cell radius R.

In a FCA scheme, each of the N cells in a cluster are assigned the same number of channels. If the distribution of traffic load is uniform, then the uniform channel distribution works fine. The overall average call blocking (connection not materializing due to non-availability of a channel) probability will be same as call blocking probability in a cell. But, hardly such an ideal situation ever occurs. There may be both temporal and spatial fluctuations in traffic across the cellular service area. Due to short term variations in traffic, most of the times a FCA scheme is not able to maintain the quality of service and network capacity that may be attainable even with static traffic demand.

2.4.1.1 Borrowing

To remedy this, when a request for a connection is placed in a cell that has no nominal free channels, a common sense driven approach is to borrow a free channel from one of its neighboring cells. The cell which borrows is referred to as the *acceptor* while cell which lends is known as the *donor*. The selection of a free channel from a donor cell should be such that it does not adversely affect the donor cell's ability to satisfy a subsequent request for a connection, and at the same time the borrowed channel does not introduce added interferences on the existing connections in the acceptor cell. In other words, the selection of both the donor cell and the channel to be borrowed should minimize:

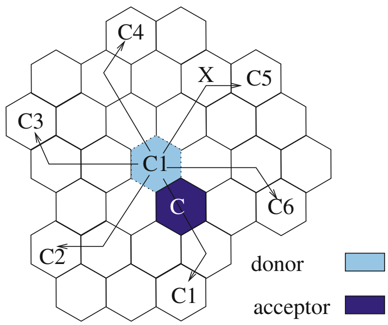
1. The probability of donor cell not being able to service a connection request due to non availability of a nominal channel, and
2. The interference due to communication over the borrowed channel on the active communications in the acceptor cell.

The first condition can be met by selecting the donor from one of the neighboring cells (of the acceptor cell) that has the largest number of free channels. The strategy of Borrowing From the Richest (BFR) will also most likely meet the second condition. Because with the availability of more channels, the probability of the borrowed channel interfering with the other active channels in the acceptor's cell will be expectedly low. Once the borrowed channel has been identified, it is locked in the co-channel cells of the donor cell which are at a distance smaller than channel reuse distance from the acceptor cell. Furthermore, the channel that is allocated should be the one that locked in most cells.

To understand the above constraints let us consider an example as shown in Fig. 2.15. Let $C1$ be richest neighboring of C which requires a channel. C will then borrow a channel from $C1$. The borrowed channel should be locked in cells C_1, C_2 and C_6 as the distances of these cells from C are smaller than the prescribed reuse distance. Therefore, borrowing of a channel affects several channels. However, if borrowing is planned carefully then the same channel may concurrently serve as a borrowed channel in different acceptor cells. For example, if cell X now needs to borrow a channel from its neighboring cell $C5$, then $C5$ can loan out the channel that C has borrowed from $C1$. This is because, C and X are at a distance more than the required reuse distance from each other. Therefore, using the same borrowed channel at C and X will not create any additional interference in communication.

Any sophisticated borrowing method will incur penalties for complex searchings. So, a simpler option could be to Borrow from the First Available (BFA) channel from a neighboring cell. However, for implementing BFA, the initial channels assignment is bit different than direct assignment of channels to cells as is done in other FCA schemes. The set of channels is first divided into groups and each group is assigned to cells at a reuse distance D. The channels in a group are numbered in sequence. A group of channels assigned to a cell is subdivided further into two subgroups A and B. The channels belonging to A are used exclusively for calls originating from

Fig. 2.15 Borrowing affects co-channels cells of donor cell



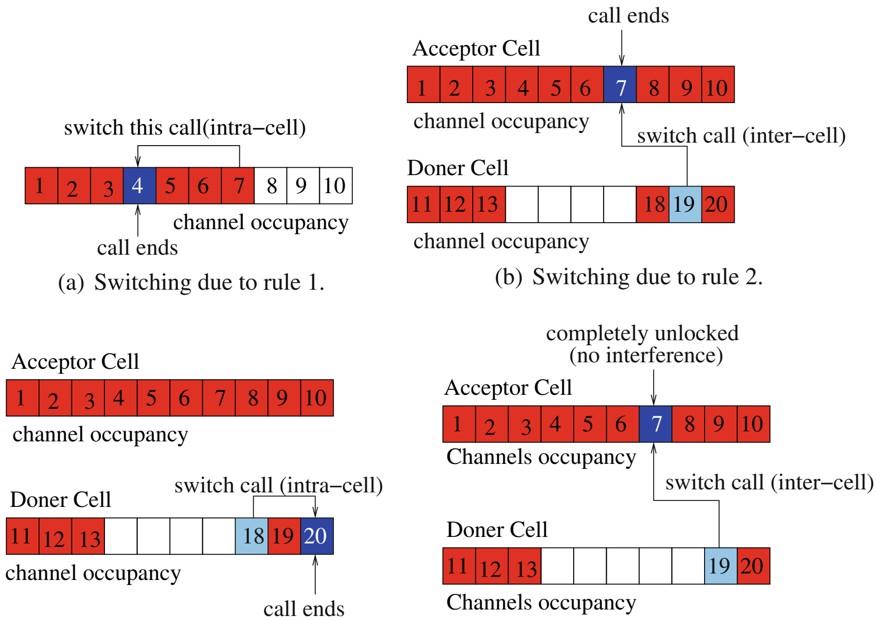
a cell. A cell C can lend channels to another neighboring cells C' only when one is available in subgroup B of C. On arrival of a call request, the nominal channels in the cell are scanned in order and the first available channel is assigned for the call. If no nominal channel is found, then (acceptor) cell C' searches for a free channel in an adjacent cell C having the largest number of channels in group B. As explained earlier, a channel is considered available for borrowing if it is also free in two other co-channel cells (using directional locking strategy) of that chosen neighboring cell. The first free channel encountered in this order is borrowed. This method is known as Borrowing with Channel Order (BCO). Interestingly, BCO compares favorably with the system that performs exhaustive complex searches to identify the channel to be borrowed from a neighboring cell. Furthermore, it has an advantage over other methods by being computationally less expensive.

After the communication over the borrowed channel is complete, that the borrowed channel should be returned to the donor cell. Since borrowing is done due to non-availability of a nominal free channel, the question which naturally arises is whether the borrowed channel should be returned as soon as one of the nominal channels of the acceptor cell becomes free. Interestingly, the decision whether or not to free a borrowed channel could influence systems performance.

Channel reallocation concept is used in conjunction with directional lockings to improve the performance of the system and minimize connections on the borrowed channels. The rules for the channel reallocation are as follows.

1. When a nominal channel becomes free and there is an active connection on a higher order nominal channel in the same cell then this connection is transferred to the lower order nominal channel.
2. When a nominal channel in local cell becomes free and there is an active connection on a borrowed channel from a neighboring cell, then the borrowed channel is returned to neighboring by transferring the connection to a newly freed nominal channel.
3. When borrowed channel becomes free due to termination of a connection in neighboring cell, and there is another active connection on a lower order borrowed channel in the same cell, then communication on lower order channel is transferred to the higher order channel. This requires a channel switching neighboring cell.
4. A nominal channel in a cell may be blocked due to lending by a co-channel cell to a different cell. If such a channel becomes is completely unlocked (in all directions), then any active connection on either on a borrowed channel or on a higher order channel is immediately switched to the unlocked channel. It leads to a channel switching in local cell.

The reallocation rules are illustrated by Fig. 2.16. Figure 2.16a shows a channel switching within the same cell. It happens because there is ongoing communication on a lower order nominal channel, i.e., channel no. 7 when a higher order nominal channel, i.e., channel number 4 becomes free. Figure 2.16b shows a channel switching that takes place between two cell. The acceptor cell on the top has borrowed channel no. 19 from the donor cell at the bottom. Once nominal channel 7 becomes



(a) Switching due to rule 1. (b) Switching due to rule 2. (c) Switching due to rule 3. (d) Switching due to rule 4.

Fig. 2.16 Channel switching with borrowing

free in the acceptor cell then channel 19 is returned back to the donor cell. An inter-cell handoff is executed to realize such a channel switching. Figure 2.16c illustrates the switching, where the borrowed channel of the donor is switched to a lower order channel if such a channel becomes free in the donor cell. Finally, Fig. 2.16d illustrates the switching when a locked nominal cell become unlocked in the acceptor cell. Note this case refers to situation where borrowing is implemented with directional locking, and channel 7 is also be locked in other directions due to borrowing by another adjacent cell.

The channel borrowing strategies, outlined above, result in lower blocking than fixed allocation under light and moderate traffic conditions [15]. Under heavy traffic condition, channel borrowing could create domino effect due to blocking of borrowed channels in respective co-channel cells. These cells being deprived of their nominal channels would in turn have to resort to borrowing and, thus, creating the domino effect that would require a comprehensive channel reallocation strategy as in [15]. So fixed channel allocation sometimes may provide better performance than schemes relying on channel borrowing.

2.4.2 Dynamic Channel Assignment Policies

Due to temporal bursts in traffic, FCA schemes are found to be inadequate for handling traffic and lead to inefficient channel utilizations. DCA schemes have been proposed as remedies, where no channel is allocated permanently to any cell. Channels are allocated to cells on need basis. After a connection terminates, the channel is returned back to the central pool. So, the key idea behind DCA scheme is to evolve a procedure for the evaluation of cost in using each candidate channel for a connection request. The cost function should take into account: (i) radio signal strength measurements at the end devices, (ii) the acceptable average call blocking, (iii) the number of future call blocking, (iv) the distribution of channel occupancy under the current traffic conditions, (v) co-channel/adjacent channel interferences, and (vi) the QoS related requirements by the clients. So, DCA strategies can be designed to adaptively adjust to the traffic conditions as well as assurance of desired GoS and QoS. The design of a perfect cost function is understandably difficult as it needs to balance complex trade offs in optimization involving several attributes as indicated above. These difficulties in design of cost function made DCA an area of intensive research [5].

DCA algorithms are of two types based on the type of control, viz., centralized and distributed. In centralized DCA, a central controller determines the channel allocation for each connection request. There is no fundamental difference between different centralized DCA schemes except for the difference in cost function used in the selection of a channel. Some of the known approaches under this category are:

- First available (FA): it assigns the first available channel by ensuring that the channel reuse constraints are not violated. The search for the candidate channel is fast, and the scheme is simple. FA is found to perform better than most of the FCA schemes by 20% in low to moderate traffic conditions.
- Locally optimized dynamic assignment (LODA): it assigns the channel in response to a request from a cell by minimizing the future call blocking possibilities in the nearby cells.
- Channel reuse optimization: it tries to optimize reuse distance while allocating a channel for a new request. The fundamental idea behind this approach is to maximize spatial multiplexing. The shorter is the reuse distance, the greater is the channel reuse over the service area. Therefore, with this scheme network capacity also increases.
- Maximum use in reuse ring: it selects the channel for allocation by finding the one that is used in most cells in co-channel set. Note that co-channel cell form a reuse ring as explained earlier in Figs. 2.6 and 2.7.
- Mean square: it selects the channel that minimizes the mean square of the distances among cells using same channel.
- Nearest neighbor: it selects the available channel occupied in the nearest cell in distance $\geq D$.

Most of the channel reuse optimization schemes [5] as described above try to employ local optimizations. The 1-clique [7] scheme, however, employs a global

Fig. 2.17 ACO matrix of cell i

Cell	Channel Number										Free
Number	1	2	3	4	5	6	7	8	...	M	Channels
i		x						x	...		0
i_1	x			x			x		...		0
i_2			x						...		2
i_3	x				x		x		...		0
i_4			x			x			...	x	4
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
i_{k_i}			x		x				...		3

optimization scheme. It builds a graph for every channel, where each vertex represents a cell, and two vertices in this graph are connected by an edge if and only if the cells corresponding to the end vertices do not have co-channel interference. So, each graph reflects channel allocations possibilities. An actual channel assignment is done from several possibilities so that as many vertices as possible, still remain available for allocation. The scheme works fine for small number of channels and a small service area (with few cells). But for a coverage area with large number of cells and large number of channels the computation time becomes prohibitive.

Distributed DCA (DDCA) schemes depend either on local information about availability of free channels in the neighborhood of the requesting cell or rely on signal measurements [12]. The focus of the most of the work in DDCA schemes has been on channel reuse [1] sometimes even at the cost of interference.

The cell based DDCA schemes rely on local information. Each cell keeps track of free channels by storing the information in an augmented channel occupancy (ACO) matrix along with sufficient additional information that enables a base station to determine if a channel can be assigned. ACO matrix at cell i is an $(M + 1) \times (k_i + 1)$ matrix, where M is the number of channels in the system and k_i is the number of neighboring cells within the co-channel interference distance from cell i. Figure 2.17 shows ACO matrix for some cell i. When the column corresponding to a channel has no entry, it means that channel is free and can be assigned to cell i. Each entry in the last column gives the number of assignable channels available in the cell represented by the entry. Note that this number would be equal to the number of empty columns in ACO matrix of the corresponding cell. For example, $ACO[i, M + 1] = 0$ represents the fact that the cell i does not have any assignable free channel. Therefore, the ACO matrix for cell i has no empty column. Similarly, $ACO[i_2, M + 1] = 2$ indicates that there are 2 free channels in cell i_2. Equivalently, ACO matrix of cell i_2 has 2 empty columns.

Since no free channels are available in a cell as in i, when cell i needs an additional channel, it searches for channel that corresponds to an empty column in the first row of the ACO matrix. However, assigning any such channel may lead to a number of reassignments of other channels. The reassignment cost should also be minimized. Therefore, the candidate channel should be one that has no entry in the first row and also has minimum number of total entries. For example, column 6 of the ACO matrix has no entry in first row, and has only one entry for row i_4. It indicates that channel 6 is currently occupied by only cell i_4. Since, $ACO[i_4, M + 1] = 5$, cell i_4 has 5 assignable free channels. Therefore, the ongoing call in cell i_4 on channel 6 can be shifted to one of the free channels. If the shifting is successful then channel 6 becomes assignable in cell i. The contents of ACO matrix is updated by collecting channel occupancy information from the interfering cells. The cell corresponding to each entry of the chosen column is requested to shift its ongoing call from the current channel to a free channel. After each of the requested cells reports completion of call shifting, the chosen channel is declared free and assignable in the requester cell. However, requests to multiple cells for shifting could cause unpredictable delays and lead to higher probabilities of failures. So, from implementation prospectives, processing only one channel shifting request appears practical.

In addition to co-channel interference, adjacent channel interference (ACI) caused by use of close frequency channels results in premature handoffs, call drops, and cross connectivity. The effects of ACI can be minimized by use of sophisticated channel filters, but cannot be eliminated altogether. DCA methods discussed so far have focused only on co-channel interference. But it is possible to address the issue of ACI by adding extra restriction on the channel selection from the ACO matrix in DDCA algorithm described above. Assume ACI effects to be negligible with channel separation of N_{adj}. Then for every column which have an entry in the first row of the ACO matrix for a cell, $N_{adj} - 1$ adjacent columns either to the left or the right cannot have an entry in the first row. It means if a channel k has been allocated to a cell i then all channels corresponding to adjacent columns $k - j$, and also those corresponding to adjacent columns $k + j$, for $j = 1, \ldots, N_{adj} - 1$, cannot be allocated by the cell i.

At the time of assigning a new channel c to cell i, the algorithm should ensure that the channels corresponding to $N_{adj} - 1$ adjacent columns to the left or to the right of column c for the row i in ACO matrix do not have a entry in the first row of ACO matrix for cell i. It means those group of channels should not have been allocated earlier to cell i.

The channel allocation algorithm based on the above strategy works as follows. When cell i receives a connection request, it searches the first row of the ACO matrix for a consecutive group of $2N_{adj} - 1$ empty entries such that column corresponding to the column at the center of the group is empty. If the search is successful, the cell i assigns the channel represented by the central column. If no group of $2N_{adj} - 1$ empty columns can be found, then cell i looks for a group of consecutive $2N_{adj} - 1$ empty columns in the first row where center column c has a single entry and the related cell (let it be j) has an assignable free channel as indicated by a non-zero entry in ACO$[j, M + 1]$. After the column j has been identified, cell i requests cell j to shift the ongoing call on the channel c to one of its free channels. Once the shifting is over, and j reports the completion, cell i can assign channel c to new call request.

As an example, suppose $N_{adj} = 2$. To satisfy a request for assigning a new channel, a search is initiated for a group of 3 consecutive columns having no entries in first row of the ACO matrix provided in Table 2.17. Any such group can only be identified between columns 3–7. However, no group of 3 empty columns can be found between columns 3–7. So, we try to find a group of 3 columns such that center column has just one entry and that the cell related to the entry has assignable free channels.

The columns 5–7 provide such a group, with only cell i_4 occupying the channel 6. Since i_4 has 4 assignable free channels it can be requested to shift the ongoing communication from channel 6 to some free channel satisfying the adjacent channel interference restrictions. After cell i_4 vacates channel 6, cell i can allocate the same to the new request.

Some DDCA techniques considers the channel allocation as a resource sharing problems [4]. The underlying idea is that channel is a resource which must be shared among neighboring cells in a mutual exclusive manner. There are two important differences in classical mutual exclusion (ME) and channel sharing among neighboring cells.

- Firstly, in classical ME two competing processes are not allowed to share a resource, but a channel sharing among cells are permitted as long as they satisfy the constraint of reuse distance.
- Secondly, in channel allocation, not a single but a collection of resources should be handled.

Many DCA schemes exists for the channel assignment. An interested reader may look at some excellent references [5] for more details on the topic.

2.5 Handoff

Mobility is the key feature of a cellular based wireless communication system. However, in a bid to increase the capacity, increasingly smaller cell size are used. Maintaining continuity of service poses a challenge with use of reduced cell size. Normally on cell crossings, a mobile user experiences deterioration in the quality of ongoing communication. The deterioration of service is attributed to one of the following two reasons:

1. **Signal quality deterioration**. The quality of a signal is determined by a set of parameters such as, RSS, Signal to Noise Ratio (SNR), and bit error rate (BER). RSS attenuates with distance from the serving base station, slow fading (shadow fading or lognormal fading) and fast fading (Rayleigh fading) [10].
2. **Traffic load**. When the capacity of the current cell has either reached or is about to reach its maximum capacity.

With smaller cells, the frequency of cell crossing by a mobile user is expected increase. Under this circumstance, the continuity of an active connection for a mobile user can only be maintained by handoffs (or handovers) through a sequence of intermediaries (intermediate cells).

The RSS from the base station of the cell from which a mobile user is moving away decays gradually. As opposed to this, the RSS from the cell, into which the mobile is moving, increases gradually. If the link to the new base station is formed either before or almost immediately around the time when the link to old base station goes down, then it is possible to keep connection active. Thus, a handoff is the transition of signal transmission from one base station to another geographically adjacent base station. By executing a handoff *in-session* mobility to a user can be supported. As explained in Sect. 2.4, a frequency switching may sometimes be needed when a mobile terminal is moving within a cell. Such a switching is called an intra-cell handoff. Our attentions here, however, is limited to only inter cell handoffs.

Cells overlap, so a mobile terminal or the device could be within the range of multiple base stations around the boundary of a cell. It facilitates maintenance of an active connection while the user migrates from one cell to a geographically adjacent cell. The network decides—with or without the assistance of user's mobile handset (MH)—which base station will handle the transmission of signals from and to the MH. Handoff should be transparent to the user. In other words, it should provide assurance that the communication will neither be lost nor be terminated unless user's hardware is out of the range of all base stations.

The important first step of a handoff process is the early detection of the handoff condition. The handoff can then be initiated before the current connection is broken. Once an active connection is completely severed there is no way to restore it. A particular threshold of signal is needed for acceptable level of ongoing communication. In the case of voice communication it could be between −90 and −100 dBm. A slightly stronger signal level is used as the threshold for initiation of handoff process. The margin between the threshold signal for handoff and the minimum usable signal level is known as the handoff *hysteresis* Δ. Formally, hysteresis can be defined as

$$\Delta = S_{handoff} - S_{min},$$

where $S_{handoff}$ represents the threshold for performing handoff, and S_{min} denotes the minimum level of signal strength needed for the voice communication. The value of Δ can neither be too large nor be too small for the handoff to work smoothly.

Figure 2.18 depicts the different handoff situations when a mobile terminal travels from station BS_1 to BS_2 along a straight line. At time t_0 the mobile host receives signal only from BS_1. At time t_1 received signal strengths from both base stations become

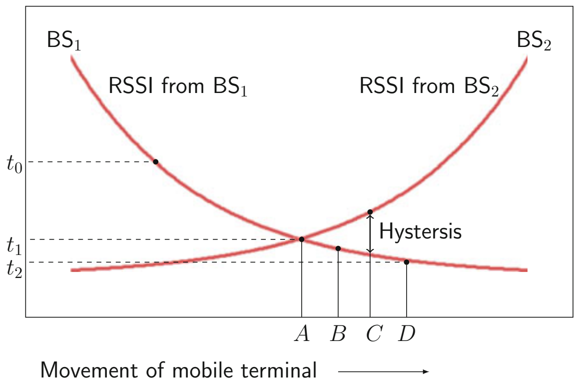


Fig. 2.18 Handoff scenarios at cell boundaries

comparable. But as the mobile user reaches point B, the signal strength received from BS_2 dominates, even though the signal from BS_1 is still above the acceptable level. The handoff must begin after mobile terminal reaches point A and completed before it reaches point C. If handoff latency is δt then, the handoff must be performed δt time before reaching the threshold C. Therefore, handoff should start when strength of signal from BS_1 is little higher and should be completed on or before reaching C. If handoff is not complete before the mobile reaches point C then the signal from BS_1 deteriorates fast. So, when the mobile reaches point D, the call gets terminated.

It has already been indicated that handoff should be performed when the received signal strength is below a certain threshold. However, the signal measurement mechanism for handoff should not be considered as a simple one-off instantaneous measurement. The reason is due to multipath fading, instantaneous measurement of signal may not provide a correct estimation of distance between the mobile and its current base station. So, the signal strength should be measured over a certain time interval, and the average of such measurements should be the basis for the decision to perform a handoff.

The window of time for performing handoff is provided by length of hysteresis. To reduce the overhead of handoffs, the attempt should to optimize Δ. The value of Δ depends on several factors such as:

- Environment.
- Speed of direction of mobile's movement.
- Time required to perform handoff.

The use of hystersis is aimed at reducing the ping-pong effect. However, the delay is also not without cost. It increases interference, reduces the quality of service. The delay could as well lead to a call dropping. A velocity adaptive handoff algorithm could enforce more frequent measurement of signal strengths by adjusting length of averaging window in which RSSs from neighboring BSes are averaged. The direction parameter assigns more weightage to execution of handoff in favor of BSes towards which the mobile is moving [8].

2.5.1 Handoff Policies

A simple approach would be to prioritize the channel assignments for handoff requests ahead of the requests for new calls. It results in a tolerable increase in call blocking probability while reducing the probability of dropped calls.

Another approach is to pre-allocate a certain number of handoff channels called guard channels. Only these channels should be used for handoffs. If the guard channels are not available then the handoff will be serviced by other channels.However, if all guard channels are occupied then, a handoff call should compete with new call for allocation of the channels outside the guard band of channels. This strategy keeps the probability of a call blocking under an acceptable level, but increases the probability of a dropped call. There is also a possibility that guard channels are under utilized, while new call requests cannot be met due to non-availability of free channels. Such a situation leads to an inefficient spectrum utilization. The concept of reserved channels would be appropriate for dynamic channel assignment scheme. Because the guard channels can be allocated from a central pool. So the non-availability of a free channel for new call requests does not become a local problem.

Typically, when no channel is available, new call requests are made to wait in a queue with the hope that handoff calls or some of the ongoing connections may release their respective channels. Once a channel becomes free, one of the queued new call request can be serviced. The strategy works well because, a user any way has to wait for an interval of time before expecting establishment of a connection against his/her request. In the case of a handoff too, there is a certain finite time interval during which the existing call connection is retained, and a new channel is assigned for the handoff. Therefore, it is also possible to enqueue a handoff request albeit for a small interval of time. The position of handoff request in the queue depends on how close the mobile is to the boundary of the cell. Higher priorities can be assigned to the mobiles that are close to a cell boundary or moving very fast while lower priorities can be assigned to the mobiles still inside a cell boundary, or are moving slowly.

Whichever entity (the network or the mobile terminal) that makes the handoff decision uses some metrics. It applies the relevant algorithms on the basis of those metrics, and assures required performance guarantees. The most critical metric for a handoff decision is the measurements of signal strengths received at the mobile from the current base station and the neighboring base stations which are probable candidates to handle the frequency switching. The measurement process should be able to avoid unnecessary handoffs. For example effecting handoff during a temporary fading may result in *ping pong* effect and put pressure on network due to unnecessary handoffs. On the other hand, if a distance dependent fading (caused by mobile user moving away from current base station) is not properly detected it could result in termination of the connection before handoff can be initiated.

2.5.2 Handoff Protocols

Three types of entities are involved in a handoff:

1. User's mobile handset (MH),
2. Base station (BS) to which MH is currently connected and BSes in the neighborhood of MH's movements, and
3. MSCs controlling the above group of BSes.

So, both network entities (BSes and MSCs) as well as user's MH may initiate and control a handoff process. Based on the nature of controlling entity or the entities, the handoff protocols can be classified into three basic types.

- Network controlled.
- Mobile assisted.
- Mobile controlled.

In network controlled protocol, the decision for handoff is based on measurements of RSSs of mobile terminal at a number of adjoining base stations. The entire handoff process which includes measurements, channel switching and network switching, etc., takes approximately around 100–200 ms. As opposed to this, in mobile assisted handoff process MH measures the RSSs it receives from BSes in its neighborhood and the decision for handoff is made by the network entities. Mobile assisted handoff may take upto 1 s. In mobile controlled handoff user's MH takes decision to execute handoff. This type of handoff requires a short reaction time, just about the order of 0.1 s. MH measures RSS of neighboring BSes and interference levels of all channels. Then it initiates handoff if the signal strength from the serving BS is lower than another neighboring BS by a pre-determined threshold. Mobile controlled handoff being a completely decentralized process, relieves the network from a high overhead specially in a high density micro-cellular system.

Several different mechanisms exist for realizing handoff. However, the common goals which every handoff mechanism should try to achieve are:

1. Handoffs should be performed quickly.
2. Interruption in connection due to a handoff should be imperceptible to users.
3. Handoffs should be performed infrequently.
4. Handoffs should be performed successfully.

Figure 2.19 illustrates the generic procedure for execution of handoff. The mobile terminal reports about the signal measurements to serving base station. If the serving base station decides about the handoff, it informs the Mobile Switching Center (MSC) that a handoff is required. The MSC then send a handoff request to the new base station. The new base station allocates the required resources for the handoff and sends handoff request accept to the MSC. Then the MSC issues handoff command to the old base station which informs the same to the mobile terminal. The mobile terminal requests for a link establishment with the new base station. After link has been established, handoff completion command is issued by the new base station to the MSC. Following this, the MSC instructs the old base station to flush the resources for the mobile terminal being used for communication. The old base station flushes the resources and the handoff is complete. The basic handoff procedure described above may be augmented by other participating components such as one or more base station controllers (BSC) and one or more MSCs depending on the structure of the network. There are many resources to be managed in wireless cellular communication system. These include channel assignment, signal to interference ratio (SIR), transmit power control, etc. All these may influence handoff decisions in some way or the

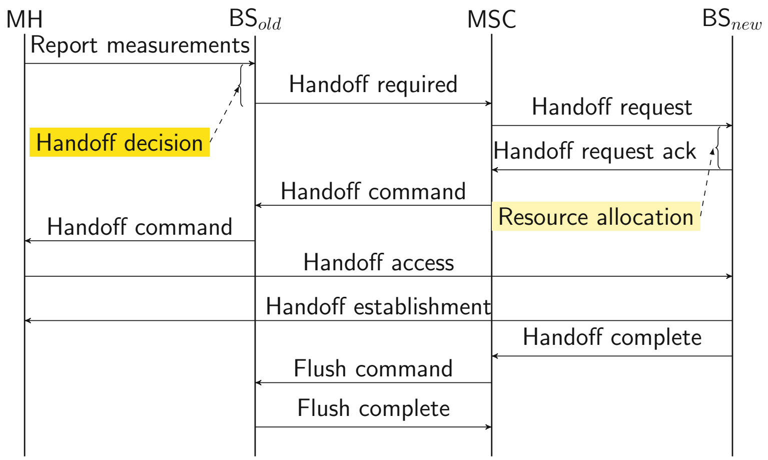


Fig. 2.19 The generic procedure of handoff

other. So, for a better overall performance, adaptive handoff protocols have been designed by integrating such resource optimization with handoff decisions [14].

References

1. F.A. Cruz-Pérez, D. Lara-Rodríguez, Distributed dynamic channel assignment with violation to the reuse pattern for microcellular networks. IEEE Trans. Veh. Technol. **51**(6), 1375–1385 (2002)
2. D. Gross, C.M. Harris, *Fundamentals of Queueing Theory*. Wiley Series in Probability and Statistics, 4th edn. (Wiley, New York, USA, 2008)
3. W.K. Hale, Frequency assignment: theory and applications. Proc. IEEE **68**, 1497–1514 (1980)
4. J. Jiang, T.H. Lai, N. Soundarajan, On distributed dynamic channel allocation in mobile cellular networks. IEEE Trans. Parallel Distrib. Syst. **13**(10), 1024–1037 (2002)
5. I. Katzela, M. Naghshineh, Channel assignment schemes for cellular mobile telecommunication systems: a comprehensive survey. IEEE Pers. Commun. **3**, 10–31 (1996)
6. V.H. MacDonald, AMPS: the cellular concept. Bell Syst. Tech. J. **58**, 15–41 (1979)
7. K. Okada, F. Kubota, On dynamic channel assignment strategies in cellular mobile radio systems. IEICE Trans. Fundam. Electron Commun. Comput. Sci. E75-A:1634–1641 (1992)
8. T. Onel, C. Ersoy, E. Cayirci, G. Parr, A multicriteria handoff decision scheme for the next generation tactical communications systems. Comput. Netw. **46**, 695–708 (2004)
9. G.P. Pollini, Trends in handover design. IEEE Commun. Magaz. 82–90 (1996)
10. T.S. Rappaport, *Wireless Communications: Principles and Practice*, 2nd edn. (Prentice Hall PTR, 2002)
11. N. Robertson, D. Sanders, P. Seymour, R. Thomas, The four-colour theorem. J. Comb. Theor. **70**, 2–44 (1997)
12. M. Serizawa, D.J. Goodman, Instability and deadlock of distributed dynamic channel allocation, *43rd IEEE VTC*, 1993, pp. 528–531
13. D. Stirzaker, G. Grimmett, *Probability and Random Processes* (Oxford Science Publication, Oxford University Press, 1992)

14. N.D. Tripathy, J.H. Reed, H.F. VanLandingham, Handoff in cellular systems. IEEE Pers. Commun. **5**, 26–37 (1998)
15. M. Zhang, T.P. Yum, Comparisons of channel-assignment strategies in cellular mobile telephone systems. IEEE Trans. Veh. Technol. **38**(4), 211–215 (1989)

Springer

http://www.springer.com/978-981-10-3940-9

Wireless Networking and Mobile Data Management
Ghosh, R.K.
2017, XLIV, 546 p. 249 illus., Hardcover
ISBN: 978-981-10-3940-9