7

Mobility Management in Wireless Networks

One of the salient features of wireless communications is the flexibility to support user roaming. With the desire to increase system capacity by using small radio cells in an array structure, users tend to move in and out of cells frequently. The frequency of cell boundary crossing increases with the speed of the user. To maintain continuity for a connection, when the mobile crosses the cell boundary, it is necessary to hand the connection off from the old base station to the new base station. Also, the new location of the mobile must be made known to its home location register where the mobile's permanent address resides in order to deliver messages destined for the mobile. Mobility management is thus a two-step process: handoff management and location management. This chapter is concerned with issues in, and methods for, handoff management and location management, and calculation of handoff traffic.

7.1 INTRODUCTION

A wireless network has the flexibility to support user roaming. However, the geographical coverage of a wireless network is somewhat limited. For wide area coverage, a backbone network, such as a wireline network or a satellite network, is needed to extend the coverage to provide personal communications globally. Such a network with wide area coverage is called a personal communications network (PCN). If the backbone network is a wireline network, then in general a PCN comprises both wireline and wireless links interconnecting traffic handling devices referred to as routers. The user terminals can be mobile (e.g., handsets) or fixed. A base station (BS) is employed as a hub to handle the information transfer between a source and a destination user. In this regard, the base station provides the resource for a mobile station to access the network. Hence the base station is also known as the access point (AP).

As stated in Chapter 1, the geographical area served by a single BS is called a radio cell. Depending on the cell size, a radio cell may be a picocell, a microcell, or a macrocell. With small cell sizes (e.g., microcells), a mobile can migrate out of the current cell and into a neighboring cell quite frequently. In this case, handoff can be too frequent. A cluster of neighboring cells can be grouped together and overseen by a common controller. The controller can collect and accumulate information from all cells in the cluster. In this way, user migration from one cell to another cell within the cluster does not involve the transfer of connection information. A mobile switching center (MSC), which has much more computing power and functionality than any individual base station, is the appropriate common controller to handle the handoff functions.

At a higher level of the network hierarchy, MSC's can also be interconnected by wirelines. A pictorial view of a mobile communications network is shown in Figure 7.1.

To facilitate user roaming, there must be an effective and efficient *handoff* mechanism so that the mobile's connection with its current serving base station is handed off to its target base station to maintain service continuity without the need to disconnect and reconnect. The handoff completion time must be within a prescribed tolerance. The need to perform handoff is induced by user mobility so that handoff management represents one component of mobility management.

Users subscribe to a regional subnetwork for communications services. This regional subnetwork is called the subscriber's home network. For identification purposes, the subscriber must register with its home network. The subscriber's identity is a permanent address that resides in the database of a location register, called the subscriber's home location register (HLR). When the subscriber moves away from its home network, the new network it enters is a foreign or visitor network. It must update its registration with the home location register through its visitor location register (VLR) in order to facilitate message delivery to the mobile in its new location. The procedure of maintaining an association between the mobile and its HLR when it is away from its home network is referred to as location management.

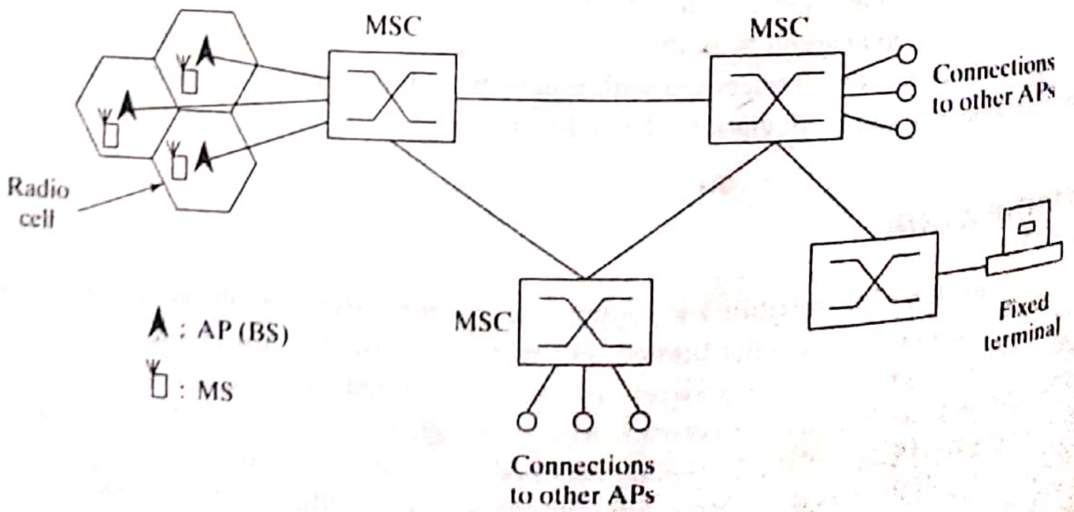

Figure 7.1 Wireless/wireline network.

Mobility management thus consists of two components:

a. *handoff management*, in which a connection is transferred from one access point to another access point, and
b. *location management*, in which the association between the mobile and its HLR is maintained as the mobile moves from one network to another.

As discussed in Chapter 6, the capacity of a cell is a hard number in FDMA and TDMA. Although the capacity of a CDMA system is soft, it is nevertheless limited by the E_b/I_0 specification that defines the quality of service (QoS) required by the participating users. For discussion purposes, suppose the capacity of the cell-site (base station) is defined as the number of basic channels, each of which is capable of handling the information transmission in one connection. Let N_c be the capacity of the cell under consideration. This total capacity is used to handle ongoing connections and to admit new and handoff requests. The mechanism that oversees the admission of new and handoff calls should protect the integrity of the network and satisfy user QoS requirements. This mechanism is referred to as call admission control.

The remainder of this chapter is organized as follows. Section 7.2 treats the issues that govern call admission control. The rationale behind handoff management and techniques for the development of handoff strategies are discussed in Section 7.3. The complexity of location management is a function of the population size. With higher transmission rates in third generation wireless networks, location management for third generation systems is expected to be relatively more involved than that for second generation systems. Location management issues are treated in Section 7.4 and Section 7.5. Traffic and handoff rate calculations are discussed in Section 7.6.

7.2 CALL ADMISSION CONTROL (CAC)

The performance requirements of users are measured in terms of quality of service (QoS) and grade of service (GoS). QoS is a packet-level factor which includes packet loss rate, packet delay, packet delay variation, and throughput rate. GoS is a call-level factor, which includes new call blocking probability (NCBP), handoff call dropping probability (HCDP), and connection forced termination probability (CFTP). CFTP is a measure of a call (connection) being forced to terminate at some point during the lifetime of the connection. In terms of CAC at the call-level, we are more concerned with NCBP and HCDP as GoS measures.

CAC ensures network integrity by restricting access to the network so as to avoid overload and congestion, and to ensure that QoS requirements of all ongoing connections are satisfied. The CAC problem can be phrased as follows. Suppose there are $(N-1)$ ongoing calls. When the Nth request arrives, the network calculates the amount of available resources. If there are enough resources to admit the Nth call, such that its QoS requirement and those of the ongoing calls are satisfied, then the new request is admitted. Otherwise, the call request is denied. This is the case whether the new request is initiated by a new call or a handoff call.

Before making admission decisions, the CAC algorithm needs to determine the amount of unallocated capacity (i.e., the number of basic channels available for accepting new and handoff requests). Also, a handoff request tries to maintain service continuity of an ongoing session. If the available capacity for accepting connection requests is limited, handoff call requests are admitted in preference to new call requests. For example, if only one basic channel is available and there are two competing requests, one new and one handoff, the decision should be to accept the handoff request and reject the new request. Thus, handoff requests should be offered a higher admission priority than new requests.

Let P_n be the new call blocking probability and P_h be the handoff call dropping probability. The GoS can be defined by jointly taking both P_n and P_h into account. That is

$$\text{GoS} = P_n + \alpha P_h, \tag{7.2.1}$$

where $\alpha > 1$ is a balancing factor between new call and handoff call GoS. A CAC algorithm may be designed subject to satisfaction of the GoS defined by Eq. (7.2.1).

7.2.1 Prioritized Call Admission

Handoff calls can be admitted at a higher priority than new calls. To manage the admission of requests based on priority, it is necessary to reserve capacity for admitting handoff requests. Let N_g be the number of basic channels reserved for admitting handoff requests. A common technique to reserve capacity for handling handoff requests is the *guard channel* method. This is the reason that we use the subscript g to denote the amount of reserved capacity for accepting handoff requests. Let N_{ua} denote the number of unallocated channels. With the guard channel method, the admission rule is the following:

a. If $N_{ua} > N_g$, admit a new or handoff request;
b. If $N_{ua} \le N_g$, admit a handoff request only.

The above guard channel method is a fixed reservation strategy. One can also introduce dynamic reservation methods to reserve the right amount of capacity to satisfy the demand-supply problem. The dynamic approach would offer a higher resource utilization efficiency, but may entail fairly complex parameter estimation issues. For example, if the controller knows exactly the number (or the rate) of handoff requests during any epoch, it can determine the value of N_g exactly. This information is never available; at best, the number or the rate of handoff may be estimated.

The focus of this chapter is more on handoff management and location management issues. The remainder of this chapter will be devoted to these two aspects of mobility management, rather than dwelling further on call admission control.

7.3 HANDOFF MANAGEMENT

When a mobile moves outside the footprint of its current serving base station, its connection must be transferred from the current serving base station to the target new base station. The handoff procedure consists of an *initiation phase* and an *execution phase*.

The initiation phase may employ a decision making strategy based on the measured received signal level, with or without hysteresis. Without hysteresis, a handoff is initiated as soon as the average signal level from the new base station exceeds that from the current base station. With hysteresis, a handoff is initiated when the average signal level from the new base station exceeds that from the current base station by a threshold amount specified by the hysteresis level. Also, the decision at any instant should be a function of previous decisions.

The execution phase will include the allocation of new radio resources (e.g., channel assignment) and the exchange of control messages.

7.3.1 Handoff Strategies

As mentioned earlier, an MSC is the appropriate device to oversee the handoff operation. An MSC is connected to a cluster of base stations through wirelines and obtains status information for all the base stations in the cluster on an ongoing basis.[1] In this way, the MSC has information concerning the channel occupancy status of all the BSs in the cluster. When the mobile moves between cells in the cluster overseen by the same MSC, handoff initiation and execution are handled by the MSC, and there is no need to copy connection identification states. This mode of operation is referred to as *intraswitch handoff*. When the mobile moves from one cluster of cells into another cluster of cells, overseen by another MSC, connection status information has to be transferred from the current serving MSC to the new MSC. The copying of state information from one MSC to another MSC takes time. This mode of operation is referred to as *interswitch handoff*. Because of the need to transfer state information, interswitch handoff tends to take longer than intraswitch handoff.

Depending on the information used and the action taken to initiate the handoff, the methods for handoff can be

a. mobile controlled handoff (MCHO),
b. network controlled handoff (NCHO), or
c. mobile assisted handoff (MAHO).

These handoff methods have the following significance:

a. MCHO is a desirable method because it reduces the burden on the network. However, this will increase the complexity of the mobile terminal.
b. In NCHO, the BSs or APs monitor the signal quality from the mobile and report the measurements to the MSC. The MSC is responsible for choosing the candidate AP and initiating the handoff. The mobile, on the other hand, plays a passive role in the handoff process.
c. MAHO is a variant of NCHO and is employed by GSM. In MAHO, the mobile measures the signal levels from the various APs using a periodic beacon generated by the APs (to keep track of the locations of the mobiles). The mobile collects a set of power levels from different APs and feeds it back to the MSC, via the serving AP, for handoff decision making.

7.3.2 Types of Handoff

The handoff procedure can be hard, soft, backward, or forward, as described below.

Hard Handoff This handoff mode is characterized by a mobile having a radio link with only one AP at any time. Thus, the old connection is terminated before a new connection is activated. This mode of operation is referred to as *break before make*.

[1] The size of a cell cluster in handoff management can be much larger than the size of a cell cluster in frequency reuse.

Soft Handoff In soft handoff, the mobile can simultaneously communicate with more than one AP during the handoff. That is, a new connection is made before breaking the old connection, and is referred to as *make before break*. CDMA systems use soft handoff techniques.

Backward Handoff In backward handoff, the handoff is predicted ahead of time and initiated via the existing radio link. A sudden loss or rapid deterioration of the radio link poses a major problem in backward handoff.

Forward Handoff In forward handoff, the handoff is initiated via the new radio link associated with the candidate AP. This type of handoff alleviates the problem of signaling on deteriorating old links. However, it can result in large delays.

7.3.3 Design Issues

The simplest scenario for investigating the handoff issue is a configuration involving a mobile station handing off from one access point, say AP_0, to another access point, AP_1, where the actual handoff decisions are made by the MSC. This scenario is illustrated in Figure 7.2.

Communication is a two way process since the mobile sends as well as receives information. During the dwelling time of the mobile within a cell, the mobile sends and receives information through the same access point. In the scenario shown in Figure 7.2, the serving access point is AP_0. When the mobile approaches the boundary of its current cell and a handoff is initiated, all conditions associated with sending and receiving information by the mobile must be taken care of during the *execution* phase. For example, during the transition time required to terminate transmission via AP_0 and start transmission via AP_1, it is possible that AP_0 has some leftover packets destined for the mobile after the mobile has moved into the footprint of AP_1. Therefore, the handoff strategy must ensure that packet delivery to the mobile after its connection has been handed off from AP_0 to AP_1 (see Figure 7.2) does not suffer undue loss. Also, the order of packets delivered to the mobile after handoff should be preserved.

An effective handoff scheme should have the following features:

a. fast and lossless
b. minimal number of control signal exchanges

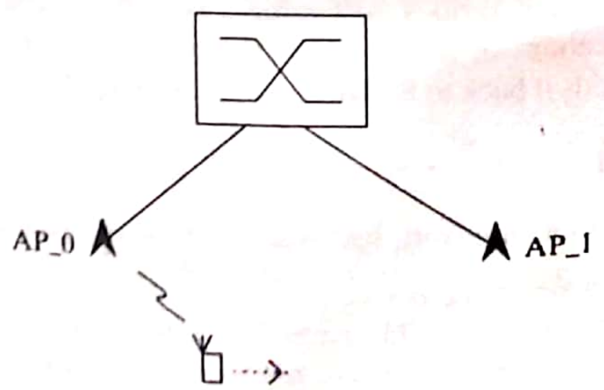

Figure 7.2 Scenario of MS moving from AP_0 toward AP_1.

c. scalable with network size
d. capable of recovering from link failures, such as abrupt loss of radio link (e.g., due to shadowing).

7.3.4 Feedback-Based MAHO Strategy

Rather than considering handoff management issues in a very general way, we shall focus attention on the movement of a tagged mobile, which is currently located in its serving cell and moving toward an adjacent cell. On the downlink from the BS to the MS, the APs broadcast pilot (beacon) signals to the mobile. With MAHO, the mobile sends the pilot signal strengths received from the APs as feedback information via the serving AP to the MSC for handoff decision making. The AP also computes and supplies the distance information to the MSC. This distance information is used by the MSC to determine the direction of movement of the mobile. To facilitate the construction of a feedback-based handoff algorithm, we assume that the mobile can always measure the strength of the pilot signals broadcast by the serving and adjacent APs.

General Design Goals. The design goals of an effective handoff scheme include

a. low handoff delay,
b. low cell loss,
c. small buffer required, and
d. efficient use of resources.

Based on the received signal strengths from the surrounding APs, the mobile creates a profile of signal strengths and sends the profile as feedback information to the MSC, via the serving AP.

Let the cell topology be a regular hexagon. Since each hexagon has six neighbors, for simplicity we consider a tagged cell and its six nearest neighbors as a cluster of seven cells. The base stations of the seven cells are connected to the MSC by wirelines. In practice, the cluster size may be much larger than seven.

The profiles to be fed back to the MSC through the serving AP contain control signals to aid decision making. Thus, feedback profiles represent an overhead on system performance. It is, therefore, desirable that the feedback profiles be as small as possible. To make the feedback profile sufficiently small, we further make the following assumptions, taking the location of the mobile within the cellular array into consideration.

Assumptions

a. A mobile receives equal signal strengths from at most three APs. This occurs when the mobile is located at the intersection point of three neighboring cells, as shown in Figure 7.3.
b. The feedback profile contains the signal strengths from the three APs with the highest signal strength being sufficient for decision making.
c. If the mobile is located in the middle of the boundary between two cells, as shown in Figure 7.4, the signals from A and B are stronger than those from C and D.

The profile needs to be formatted to provide identification of the MSCs.

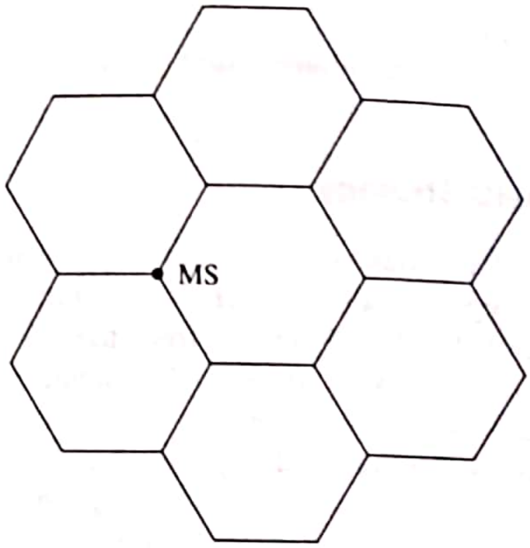

Figure 7.3 Mobile located at the vertex of three cells.

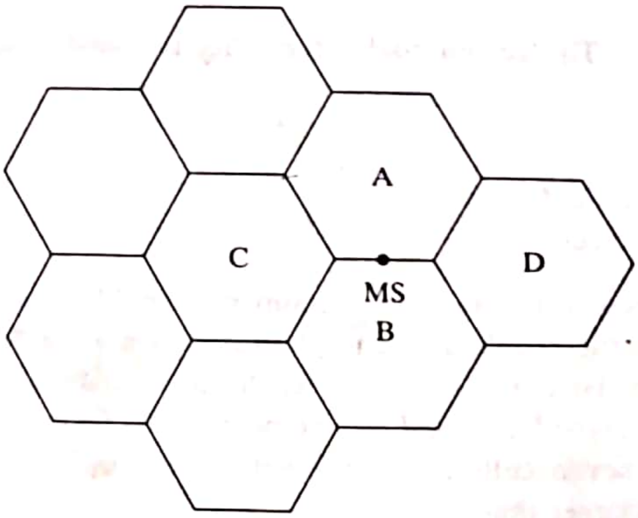

Figure 7.4 Mobile located at the boundary of two cells.

7.3.5 AP/MSC Identification

Consider the situations shown in Figures 7.3 and 7.4. For convenience, we will use the term *cluster* here to denote a group of APs served by the same MSC. There are three scenarios in which the three APs are located closest to the mobile:

a. The three APs are located within the same cluster.
b. The three APs are divided between two clusters. This happens when the mobile is located at the boundary shared by two clusters.
c. The APs are divided among three clusters (i.e., one from each cluster). This occurs when the mobile is located at the boundary shared by three clusters.

Therefore, it is only necessary to identify a single cluster or at most three adjacent clusters uniquely to the current serving MSC. Each MSC (or cluster) has six adjacent MSCs. Let MSC$_r$

denote the MSC that is currently serving the tagged mobile. To uniquely identify the adjacent MSCs, their identities have to be different. Since there are seven MSCs altogether, three bits are sufficient to uniquely identify all adjacent MSCs. The 3-bit codeword shall be referred to as a *3-bit identity* [101].

Example 7.1 3-bit AP/MSC Identification

Describe how a 3-bit codeword can be used to uniquely identity a cluster of seven MSCs.

Solution The clustering of MSCs is shown in Figure 7.5. Here, a 3-bit codeword is used to uniquely identify a supercluster of 7 MSCs. The *3-bit identity* (000) is used to identify MSC_{curr}, and the 6 adjacent MSCs are identified by the code vector {(001),(010),(011),(100),(101),(110)}. Therefore, if MSC (000) maintains a list of all adjacent MSCs, it can identify an adjacent MSC based on a *3-bit identity*.

To see that a 3-bit identity is necessary and sufficient, consider the group of tier 2 clusters. There are a total of 12 clusters. Based on the identification system indicated above, MSC (001) will be surrounded by 6 adjacent MSCs represented by the code vector {(000),(010),(011),(100),(101),(110)}. It is important to note that there are two occurrences of

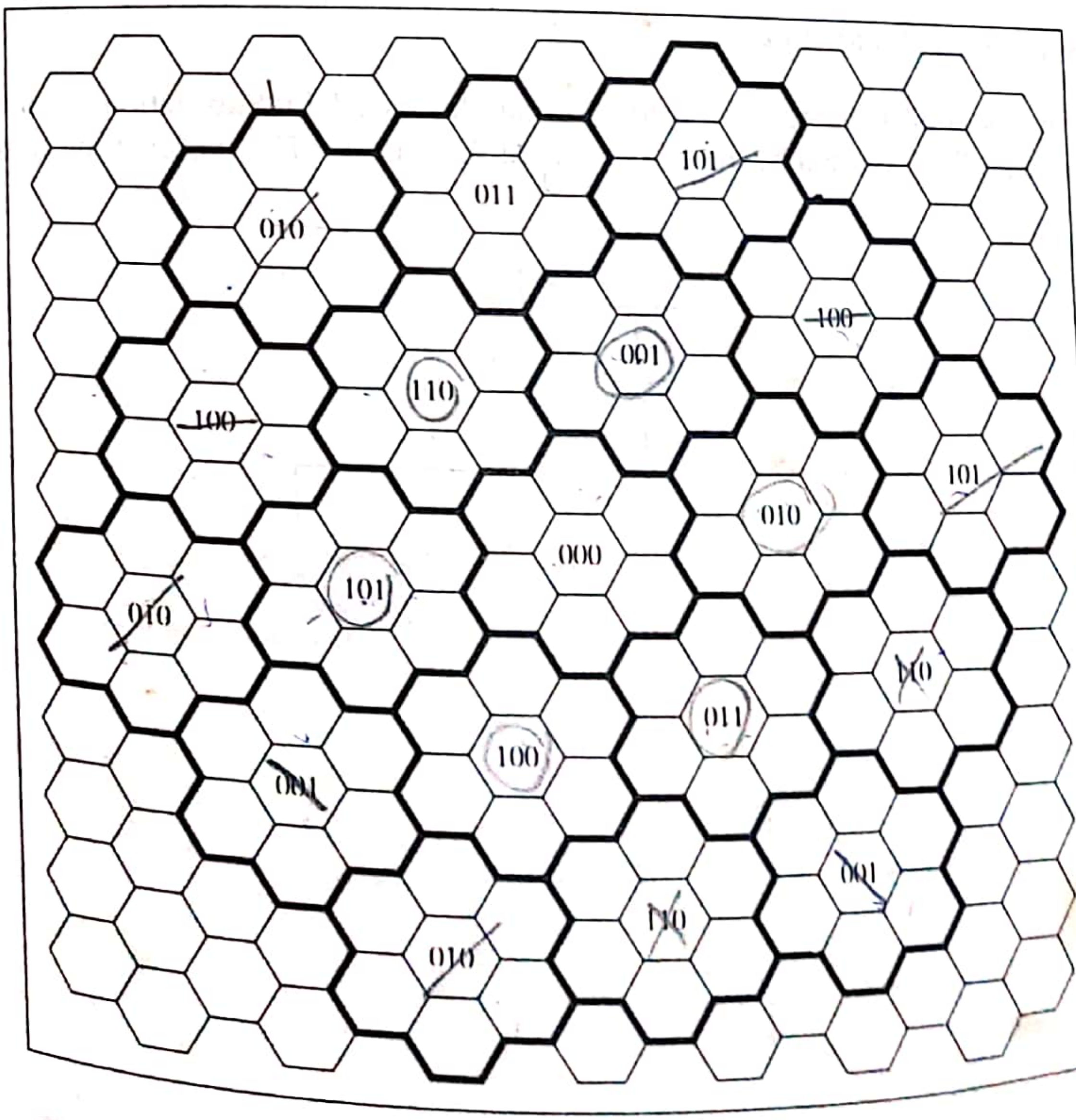

Figure 7.5 MSC identification.

each of {(001),(010),(011),(100),(101),(110)} in tier 2. To uniquely identify MSCs in tier 2 to the network, an additional bit is needed for tier 2 identification.

This identification process can continue until all the tiers are exhausted. In order to have equal size identities, if MSCs in tier L (the highest tier) have a K-bit identity (3 of the K bits identifying the MSCs and the other $(K-3)$ bits for tier identification), with 7 APs per MSC, an additional 3 bits are needed for AP identification, so that the profile would contain $(K+3)$ bits.

$\underbrace{\underbrace{\text{tier id}}_{(k-3)} + \underbrace{\text{MSC id}}_{(3)}}_{k} + \underbrace{\text{AP id}}_{(3)}$ = k+3

7.3.6 Profile

In addition to the AP/MSC identification, the profile has to contain bits that represent the signal strengths received by the mobile from the nearest APs. With the possibility of the strongest signals from three APs (see Figure 7.3), only three power fields are needed to carry the information about the signal strengths. Suppose eight bits are allocated to digitally represent each power level. Then, the profile should allocate 24 bits (three bytes) to represent the power levels. One of the three APs should be the serving AP, residing in the serving MSC. The profile does not need fields to identify the serving AP or MSC, but must contain fields to identify the other two APs and their corresponding MSCs.

7.3.7 Capability of the Mobile

The flowchart for implementing the profile format discussed in Subsection 7.3.6 is shown in Figure 7.6. The profile is created by the transceiver at the mobile. The mobile transceiver processes

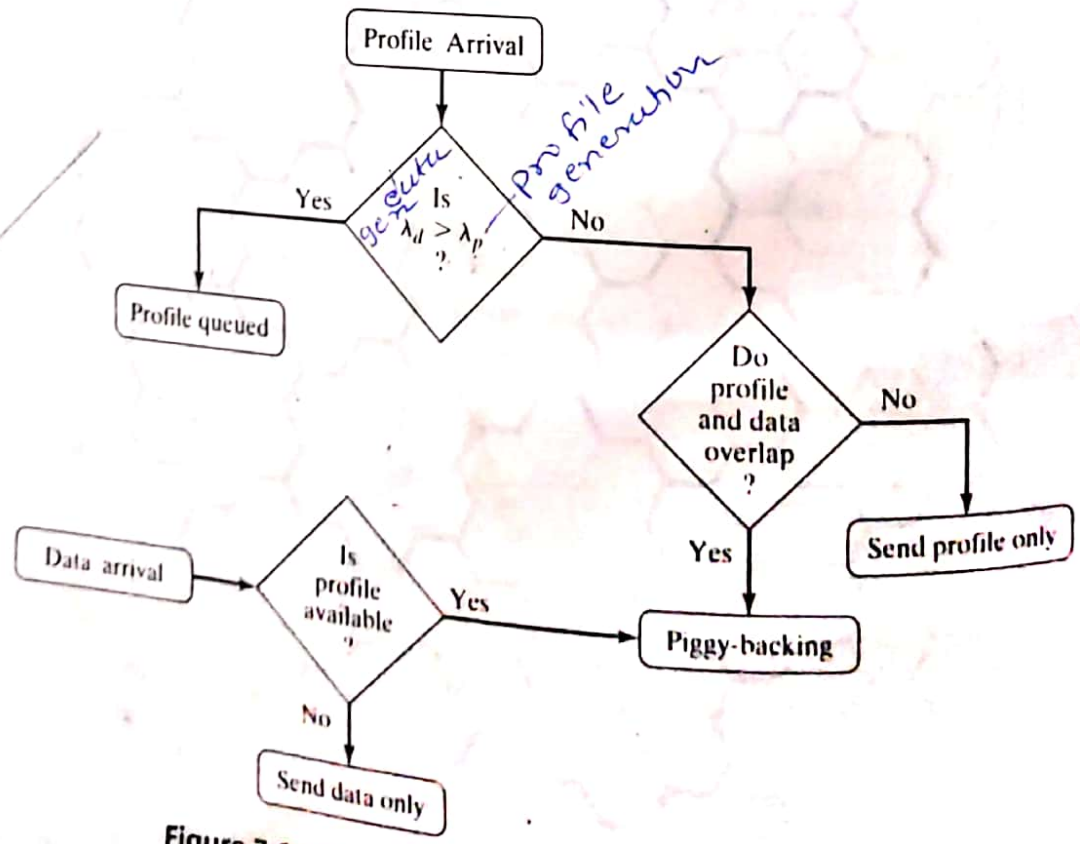

Figure 7.6 Implementation of piggy-backed profile.

the beacon signals received from the APs to create a profile, and then transmits the profile and the generated data packet according to the implementation shown in Figure 7.6. The procedure works as follows.

Let λ_d and λ_p be, respectively, the data generation rate and the profile generation rate. The transceiver compares the data rate, λ_d, with the profile rate, λ_p. If $\lambda_d > \lambda_p$, the profile is queued in the profile buffer. Upon arrival of the next data packet, the profile is dequeued and piggy-backed onto the data before transmission. Since the data packets will be generated at a rate faster than that of the profiles, there is only one profile in the queue at any time. On the other hand, if $\lambda_d < \lambda_p$, the profile is transmitted separately if there is no overlap with data transmission; otherwise it is piggy-backed and transmitted with the data packet.

Based on the preceding description of the generation of profiles, it is clear that, when a data packet is generated, the profile queue and also any overlaps must be checked for piggy-backing. Thus, if there is a profile in the queue, or if there is an overlap, the profile is piggy-backed and transmitted with the data. On the other hand, if none of these two possibilities exist, the data packet is sent separately.

It is important to note that the utilization of this scheme is significantly affected by the relationship between λ_d and λ_p. If $\lambda_d > \lambda_p$, the utilization is governed by the data rate. However, for $\lambda_d < \lambda_p$, the profile rate heavily influences the utilization and can result in significant degradation if the data rate is small. This is because most of the profiles will be sent separately.

Assuming that the profile is properly formatted, the performance measures include

a. buffer requirement,
b. mean handoff delay,
c. feedback interval, and
d. system resource utilization efficiency.

The listed performance measures are influenced by the movement patterns of the mobile. To facilitate performance assessment, a suitable mobility model is needed.

7.3.8 Mobility Model

Knowledge of mobility information is critical to the design of handoff algorithms. The mobility pattern of a mobile is defined by the speed and the direction of the mobile's movement. Accurate information about speed and direction of movement can only be obtained through monitoring and measurement. An alternative is the construction of a mobility model, based on assumed statistics.

In general, the mobility pattern is a function of the mobile's speed and the direction of movement (i.e., the mobility pattern is a function of two variables). The change in direction of movement relative to the mobile's current direction of movement can be any value in the $[0, 2\pi]$ range. Also, the speed of travel can be variable. Therefore, in general, mobility patterns can be quite complex. Of particular interest is the expected time interval between the time a mobile initiates a call in a cell and the time the mobile reaches the cell boundary. We refer to this as the sojourn time[2] and denote it by t_s. For simplicity, assume that the mobile travels in a straight line. This reduces the mobility pattern to be a function of the mobile's speed only. This assumption is

[2] If the mobile immediately hands off to the neighboring cell when it reaches a cell boundary, the sojourn time is the same as the channel holding time discussed in Section 5.5.

appropriate for motor vehicles that travel at relatively high speeds. For a low speed mobile (e.g., a pedestrian), direction changes can be made easily. In this case, the mobility patterns will be a function of both speed and direction.

Note that mobility models are not unique. We will consider the following mobility model and determine the expected sojourn time of the mobile's movement.

Mobility Model 1. Let X be the random variable representing the distance traveled by the mobile before reaching the cell boundary. Let Y be the random variable representing the velocity of the mobile's movement. The sojourn time t_s depends on both X and Y. To simplify the modeling problem, assume further that

a. X is uniformly distributed between 0 and D_{max}, and
b. Y is uniformly distributed between V_{min} and V_{max},

where D_{max} is the maximum distance traveled by the mobile, and V_{min} and V_{max} are, respectively, the minimum and maximum mobile speeds.

The sojourn time t_s is related to X and Y by $t_s = X/Y$. From the performance consideration point of view, the expected sojourn time is related to the rate at which the mobile needs to send profile information to the MSC in the MAHO strategy. To determine the expected sojourn time, we need to first find the probability density function (pdf) of t_s. The parameters in this model are D_{max}, V_{max} and V_{min}. For notational convenience, let

$$\beta = \frac{D_{max}}{V_{max}}, \quad \alpha = \frac{D_{max}}{V_{min}}, \quad \frac{\beta}{\alpha} = \frac{V_{min}}{V_{max}}$$

and

$$u(t) = \begin{cases} 1 & t \geq 0 \\ 0 & t < 0 \end{cases}.$$

It can be derived that the pdf of t_s is

$$f_{t_s}(t) = \frac{(\alpha+\beta)}{2\alpha\beta}[u(t) - u(t-\beta)] + \left[\frac{\alpha\beta}{2t^2(\alpha-\beta)} - \frac{\beta}{2\alpha(\alpha-\beta)}\right][u(t-\beta) - u(t-\alpha)]. \quad (7.3.1)$$

Using the preceding pdf, the expected value of the sojourn time is given by

$$E[t_s] = \frac{\alpha\beta}{2(\alpha-\beta)} \ln\left(\frac{\alpha}{\beta}\right). \quad (7.3.2)$$

This mean has the following asymptotic results:

$$\lim_{\frac{\alpha}{\beta}\to 1} E[t_s] = \alpha/2 \quad (7.3.3)$$

and

$$\lim_{\frac{\alpha}{\beta}\to\infty} E[t_s] = \beta/2. \quad (7.3.4)$$

The performance of a feedback-based handoff scheme, in which the mobile sends profile information to the MSC, depends on the frequency with which the mobile feeds information back to the

MSC. Profiles contain control signals, but do not carry information in a messaging sense. Sending feedback information too frequently reduces efficiency; on the other hand, sending feedback information too infrequently degrades effectiveness. Thus, the selection of the feedback interval is important. Let I_f denote the average feedback interval. The rate of transmitting profiles is $r_p = 1/I_f$. Let N_p denote the mean number of profiles sent by the mobile before the mobile moves to the cell boundary. N_p is approximately given by $E[t_s] \times r_p$. To make a good compromise between resource utilization efficiency and handoff performance, it may be appropriate to decide on the mean number of profiles to be sent and determine the rate of sending profiles from $r_p \approx N_p/E[t_s]$, from which we can get a measure of the feedback interval, I_f.

7.3.9 Intraswitch Handoff Algorithm

Handoff from one AP to another AP under the control of the same MSC is referred to as intraswitch handoff; handoffs between AP's under the control of different MSC's is referred to as interswitch handoff. Figure 7.7 depicts a situation of intraswitch handoff where all handoff decisions are under the control of the MSC. In the situation shown in Figure 7.7, the fixed terminal and AP_0 and AP_1 are each connected to the MSC by wirelines. As shown, the fixed terminal communicates with the mobile via the MSC.

As mentioned earlier, the MSC collects data sent to it by the APs within the cluster. These data include profile information sent by the mobile and measurements taken by the candidate AP to facilitate estimation of the current location of the mobile within the current cell. These data, together with the accumulated information about the channel occupancy of each of the cells in the cluster, enable the MSC to take the following items into consideration in making handoff decisions:

a. Based on the profile information from the mobile and measurements taken by the candidate AP, the MSC makes handoff decisions;
b. The MSC determines the readiness of the new AP to accommodate the handoff; and
c. The MSC ensures that the handoff algorithm maintains packet sequencing after handoff.

A general design consideration is that the mobile must hand off upon receipt of a command from the MSC. Depending on the operational procedure associated with a handoff scheme, there

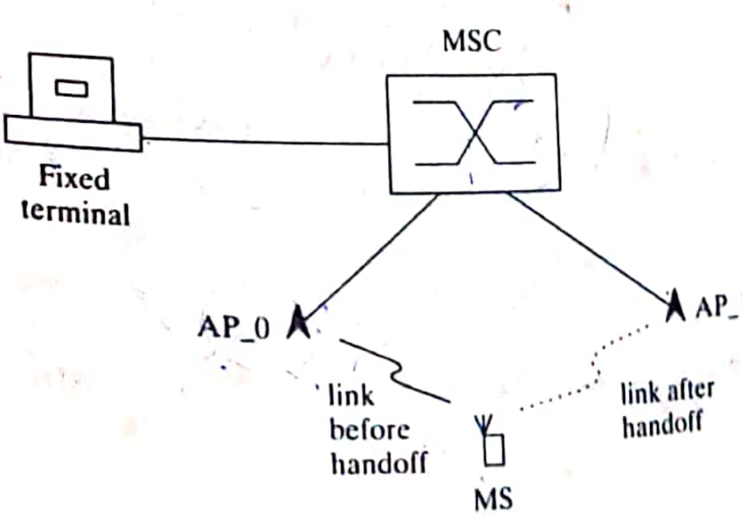

Figure 7.7 Intraswitch handoff scenario.

may be different approaches to constructing the handoff scheme. Here, we consider one design example [101].

Example 7.2 Intraswitch Handoff Design

Consider Mobile Assisted Handoff for intraswitch operation. The system shown in Figure 7.7 consists of five components: MSC, AP_0, AP_1, fixed terminal, and mobile station. In this scenario, the fixed terminal is used to generate data packets destined for the mobile, and acts as a sink for data packets arriving from the mobile via the MSC. The other components perform the following functions:

a. The APs provide access points for the mobile via a radio link;
b. All packets to/from the mobile must transit through the serving AP;
c. The MSC ensures packet ordering.

Consider the situation in which the tagged mobile being served by access point AP_0 is moving toward the neighboring cell, with access point AP_1. Both AP_0 and AP_1 are connected to an MSC, which acts as the central controller to initiate and execute handoff algorithms. Using an MAHO approach, design an intraswitch handoff scheme to hand the ongoing connection of the tagged mobile off from AP_0 to AP_1.

Solution The handoff scenario is shown in Figure 7.8. The mobile is currently located in the cell being served by AP_0, and is moving toward the cell being served by AP_1. When the

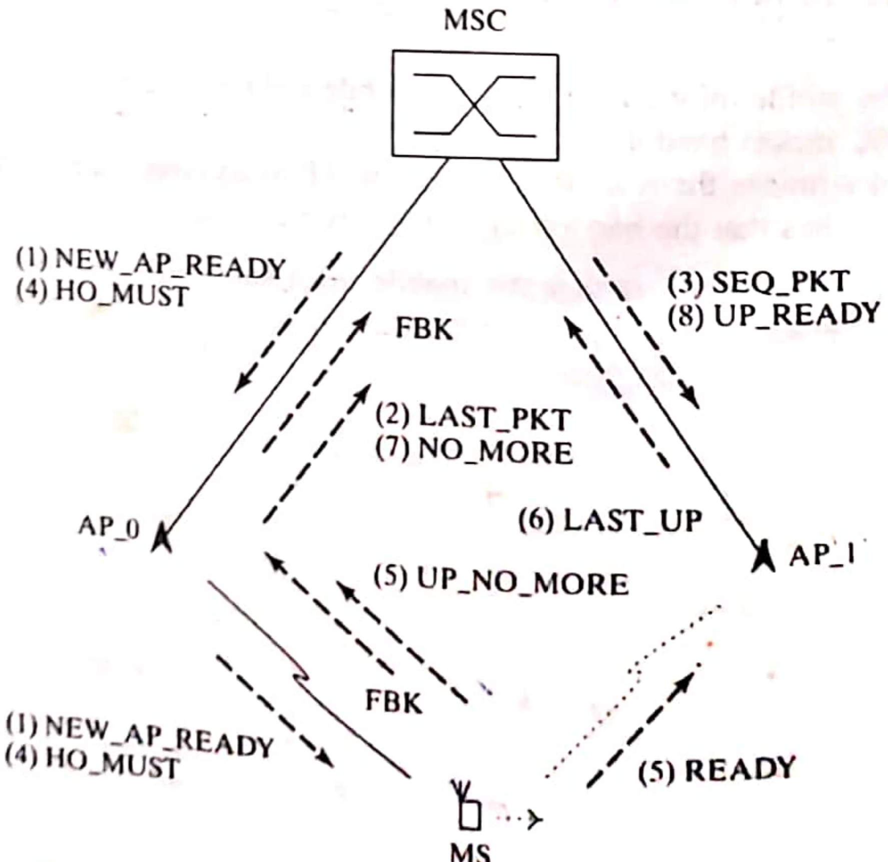

Figure 7.8 Signaling sequence for intraswitch handoff.

mobile reaches the cell boundary of AP_0, the MSC initiates and executes the handoff algorithm to be designed.

The sample handoff design is portrayed in Figure 7.8. This is an MAHO strategy. Germane to the intraswitch mobile assisted handoff scheme shown in Figure 7.8 is a set of control signals. The numbered control messages indicate the order and actions of the signal flows, under the control of the MSC.

Suppose the MSC, based on information collected, knows that the mobile (currently served by AP_0) is now at the cell boundary between AP_0 and AP_1 (Figure 7.8) and that AP_1 has a channel available to accept the handoff. The MSC now directs the mobile to handoff to AP_1. The handoff procedure, as indicated in Figure 7.8, is as follows.

Steps of handoff procedure for the scheme in Figure 7.8:

1. Send the message

   MSC *(1) NEW_AP_READY*

   to the mobile via AP_0. This message indicates to the mobile that the candidate AP_1 is ready and it can handoff whenever it desires to do so. The identity of the candidate AP_1 is contained in this message.

2. Upon receipt of the *NEW_AP_READY* message, AP_0 responds with the message

   (2) LAST_PKT

   which contains the sequence number of the packet sent (or waiting to be sent) to the mobile prior to receiving the *NEW_AP_READY* message. When the MSC receives the sequence number, it calculates the sequence number of the last downlink packet transmitted to AP_0 by the MSC. Then the MSC sends the next number as the sequence number of the following downlink packet to AP_1 using the message

   MSC calculates next downlink packet sequence no. send to AP-1

   (3) SEQ_PKT.

3. The message

   (4) HO_MUST last downlink packet from MSC. end of uplink packet from Mobile.

   implicitly indicates the last downlink packet from the MSC to AP_0. This message indicates the end of uplink packets from the mobile. When AP_0 receives this message, it sends all uplink packets to the MSC, if any, and flags the termination of the connection by sending the message sent

   (5) UP_NO_MORE

   to the MSC.

4. After sending the *UP_NO_MORE* message, the mobile switches its operating frequency and sends the message

   (6) READY to AP-1 & continues uplink transmission

   to AP_1, and continues uplink transmission via the new connection. This message contains the sequence number of the last packet correctly received by the mobile. Sending *UP_NO_MORE* in parallel with the *READY* message speeds up the approval process of uplink transmission by the MSC.

5. As soon as AP_1 receives the *READY* message, it starts the downlink transmission and, at the same time, buffers all uplink packets arriving from the mobile. It also sends the message

   (7) *LAST_UP* to MSC

   to the MSC requesting for the approval of uplink transmission.

6. Upon receipt of the *LAST_UP* message, the MSC waits for the message

   (8) *NO_MORE* from AP0

   from AP_0 (if it has not yet been received). Once this message is received, the MSC switches the uplink connection from AP_0 to AP_1 and sends the message

   (9) *UP_READY*

   to indicate that the uplink flow can resume. In this way, uplink packet sequence is also maintained.

The handoff procedure is complete when AP_1 receives the *UP_READY* message from the MSC.

7.4 LOCATION MANAGEMENT FOR CELLULAR NETWORKS

To provide a wide geographical coverage, a wireless mobile cellular network is normally configured as an interconnection of many regional wireless subnetworks. The interconnection mechanism may be an intelligent router or a backbone network. The intelligent router may be in the form of an MSC, while the backbone network may be a wireline network (e.g., the Internet), or a satellite network.

The cellular network is a message transport platform for point-to-point, point-to-multipoint, or multipoint-to-multipoint communications. The end user can be a mobile station or a fixed terminal. For message delivery purposes, a mobile is identified with a home network (i.e., one of the subnetworks). The home network must know the current location of the mobile at all times for message delivery purposes. The association between the mobile and its home network is made through a *registration* process. The identity of the mobile is kept in a database of a *location register*. The location register that belongs to the mobile's home network is called the *home location register* (HLR). When the mobile moves into a different regional subnetwork, it still has to maintain an association with its home network through the registration process. Thus, whenever the mobile moves away from its home network, it must update its registration with its HLR. The regional subnetwork where the mobile now resides is called the *foreign* or *visitor network*. The mobile has to register its current location with a *visitor location register* (VLR) in the visitor network. When the mobile moves away from its home network, if the mobile does not update its registration with its HLR, the HLR has no way of knowing the current location of the mobile. All messages destined for the mobile will end up at the mobile's home network. Since the HLR has no knowledge of the mobile's current location, it is unable to deliver the messages. To facilitate message delivery, the mobile must update its registration with the HLR. Before the registration update process can take place, the HLR needs to know that a mobile attempting to update registration is the rightful owner of the ID residing in the database of the HLR. This is a security measure to protect the network from fraudulent attacks. The process of ascertaining

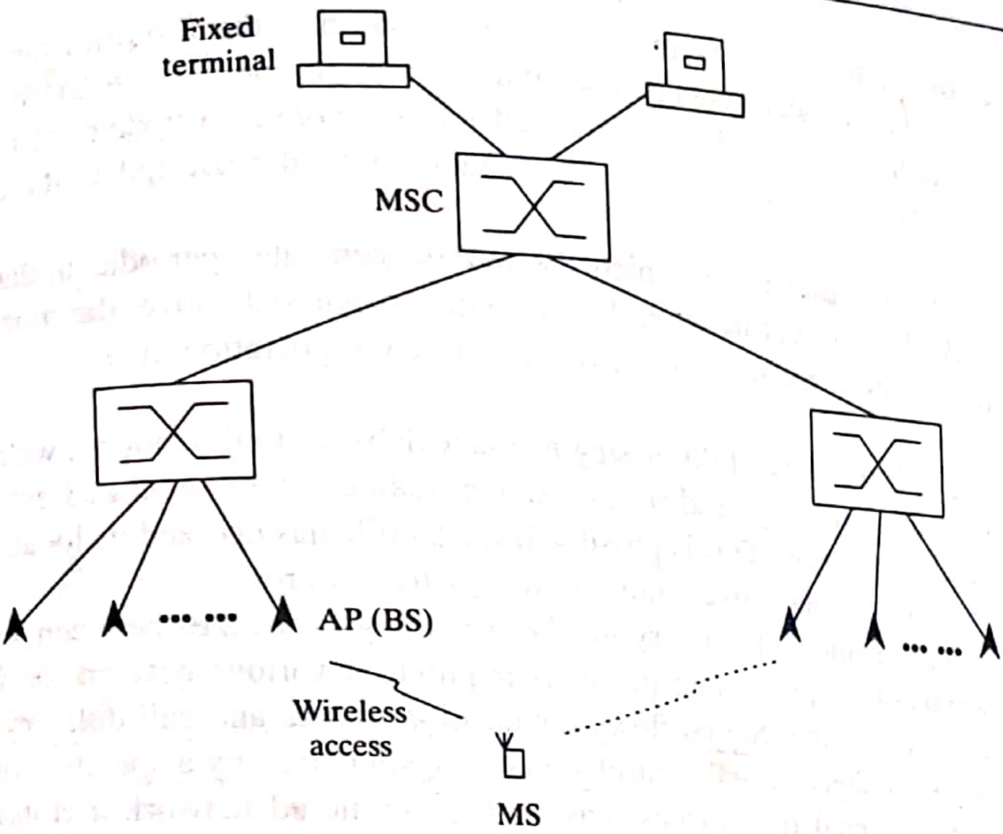

Figure 7.9 Cellular network architecture.

that the mobile attempting to update its registration is, in fact, the rightful owner of the ID in the HLR's database is referred to as an *authentication* process.

For the purpose of discussing location management issues, we will use the cellular network architecture shown in Figure 7.9. Here, the cellular network consists of BSs connected to MSCs by wirelines, MSCs connected to other MSCs by wirelines, and fixed terminals connected to MSCs by wirelines. The system shown in Figure 7.9 portrays a mobile station in motion. For the purpose of discussing location management issues, we focus attention on the processes needed for the mobile to reestablish an association with its HLR and for delivery of messages from a fixed terminal to the mobile terminal. In this regard, location management consists of three distinct components: authentication, location update, and call delivery.

Authentication. An essential function to achieve secured communication in cellular networks is reliable authentication of communicating parties and network components. Thus, authentication is a process to ensure that any party that claims to be the owner of a particular permanent address is, in fact, the rightful owner. Authentication typically depends on exchanges of cryptographic messages between the communicating parties. Encryption of a message is tantamount to putting the message in an indestructible box and locking the box. A key is needed to open the lock in order to read the message. The mobile's identity is its permanent address; only the mobile and its HLR know the mobile's permanent address (i.e., the identity of the address is locked in a box and the mobile and the HLR are the only parties who have a key to open the box).

In this text, we do not consider the design of data encryption algorithms or authentication protocols. Location management will be discussed with respect to location update and call delivery on the assumption that authentication between the mobile and the HLR has already taken place.

Location Update. For discussion purposes, we assume that the footprint covered by all the cells in a subnet of the cellular network constitutes a registration area (RA). When the mobile moves within a registration area, there is no need for the mobile to update its registration with its HLR. When the mobile moves into a new registration area, it must update its registration with the HLR.

Location update is a process in which the mobile generates periodic update messages to inform the network of its current location. In the context discussed above, the mobile informs the HLR of its current location whenever it moves into a new registration area.

Call Delivery. For the purpose of discussing message delivery to the mobile, we assume that the fixed terminal generates a message and sends it to the mobile's permanent address for delivery to the mobile in its current location. This is possible if the mobile has updated its location information with the HLR when the former moves into a new registration area.

Both location update and call delivery involve signaling exchanges between various network entities. To accomplish this, processing power is required at various network nodes to properly handle the location information. Signal flows for location update and call delivery are supported by a specific network architecture and control signals are supported by a specific control signaling network. We will base signal flow discussions on the two-tiered network architecture of IS-41 and the handling of control signals by the SS7 signaling network.

7.4.1 Two-Tiered Architecture of IS-41

The second generation systems have been operational for sometime, whereas the third generation systems are still evolving. In second generation mobile wireless networks, there are two standards currently available for location management. These are

a. IS-41, which is a companion standard to IS-54 (or the new version IS-136), and
b. Mobile Applications Part (MAP) of GSM.

Both IS-41 and GSM's MAP are two-tiered architectures, which maintain a database in the HLR and the VLR. The two-tiered architecture of IS-41 is illustrated in Figure 7.10. In this chapter, we will use the two-tiered architecture of IS-41 as the basis for discussions of registration updates.

The cluster of cells connected to an MSC, as shown in Figure 7.10, is referred to as an RA. No registration is required when the mobile roams within a given RA, but the mobile must register with the HLR when it moves to a new RA. The IS-41 architecture has the following characteristics:

a. Each RA consists of one or more radio cells.
b. Cells belonging to the same RA are connected to the same MSC.
c. The MSC is connected to both the backbone network and the signaling network, and provides typical network switching functions in addition to coordinating the location registration and call setup.
d. A VLR is associated with each RA.
e. The VLR contains an entry for all mobiles which are currently visiting its RA.
f. There is an HLR associated with each mobile in the cellular network. In IS-41, as shown in Figure 7.10, there is a central HLR associated with each cellular network.

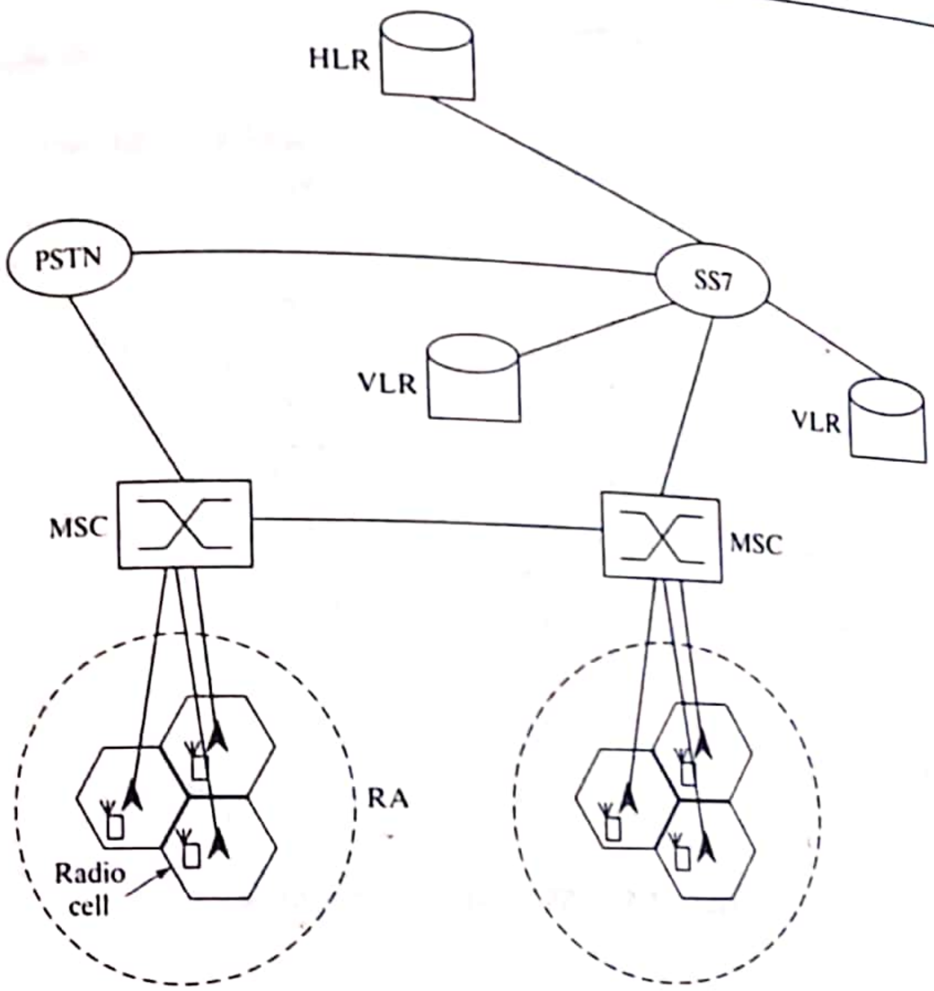

Figure 7.10 IS-41 network architecture.

g. The HLR contains an entry for each user in the network.
h. Each HLR entry indicates, for its corresponding mobile, the current VLR at which the mobile is registered.

In IS-41, network signaling exchange is carried out through the SS7 signaling network using common channel signaling.

7.4.2 SS7 Network and Common Channel Signaling

Although all signal transmissions between mobile stations and base stations propagate through wireless channels, MSCs are connected to base stations, to other MSCs, and to the backbone wireline network by wire links. All control signals in the wireline domain are handled by the SS7 signaling network. A functional block diagram of the SS7 signaling network is shown in Figure 7.11. For conciseness, acronyms are used as labels in Figure 7.11. In the following, we describe the signal flows in the signaling network using the acronyms to portray each of the service points. The acronyms are defined as follows:

STP - signaling transfer point
SCP - service control point
SSP - service switching point

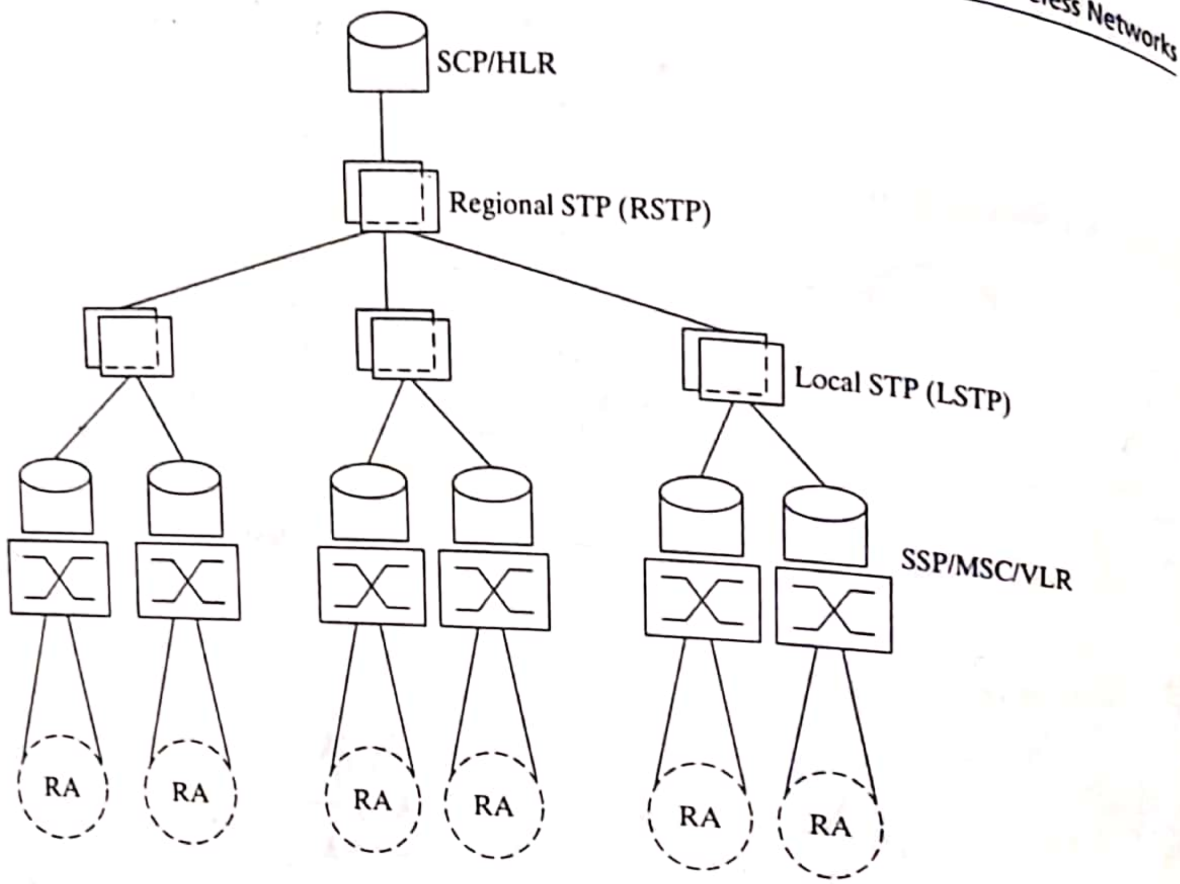

Figure 7.11 SS7 signaling network.

LSTP - local STP
RSTP - regional STP.

As shown in Figure 7.11, the SSP, MSC and VLR are all functionally colocated. Also, the SCP and the HLR are functionally colocated. From the registration update point of view, all base stations connected to the same MSC represent one registration area. Signaling between the mobile in the current RA and the HLR passes through the SSP, LSTP, RSTP and then the SCP.

The base stations of an RA are connected via a wireline network to an end-office switch, or an SSP. The SS7 signaling network has the following features:

a. Each SSP serves one RA;
b. SSPs of different RAs are, in turn, connected to a two-level hierarchy of signaling transfer points, comprising a regional STP connected to all local STPs;
c. The STPs, which are installed in pairs to ensure robustness, perform message routing and other SS7 functions;
d. The RSTP is also connected to an SCP which contains the functionality associated with an HLR.

As mentioned earlier, the SS7 network carries out network signaling exchange operations using common channel signaling (CCS). Briefly, the salient features of CCS are as follows. CCS is a digital communications technique that provides simultaneous transmission of user data, signaling data, and other related traffic throughout a network. Using out-of-band signaling channels that logically separate the network control data from the user information (e.g., voice or data) on the same channel, CCS is implemented in a time division multiplexing (TDM) format for serial data transmissions, even though the concept of CCS may imply the use of dedicated, parallel channels.

7.4.3 Location Update, Call Setup, and Paging

For second generation wireless communications systems, CCS is used to pass user data and control/supervisory signals between the mobile and the BS, between the BS and the MSC, and between MSCs.

Once the mobile has identified its current registration area, location management is mainly concerned with location update and call setup. Once the registration update is complete, call delivery is then a very simple procedure. Whenever the mobile moves into a new RA, it must update the registration with its HLR.

A mobile is identified by a mobile identification number (MIN). A registration request sent by the mobile should contain its MIN. The base stations in each RA periodically broadcast beacon signals. A mobile listens to the pilot signals from the base stations in the RA and uses these to identify its current RA. A location registration occurs when a mobile enters a new RA. The signaling flow diagram for location registration in IS-41 is shown in Figure 7.12. Registration update is a query (request) and response (ACK) process. The steps involved in the location update procedure trace out the signal flow shown in Figure 7.12.

Example 7.3 Location Update Procedure

Consider the two-tier network architecture of IS-41, shown in Figure 7.10, and the signal flow diagram shown in Figure 7.12. Describe the steps for implementing the location update procedure.

Solution The location update by the mobile with its HLR can be implemented using the following six steps:

(1) By monitoring the beacon signals from the base stations in the RA, the mobile senses that it is in a new RA and sends a registration request to the base station of the current cell. The registration request contains the mobile's MIN.

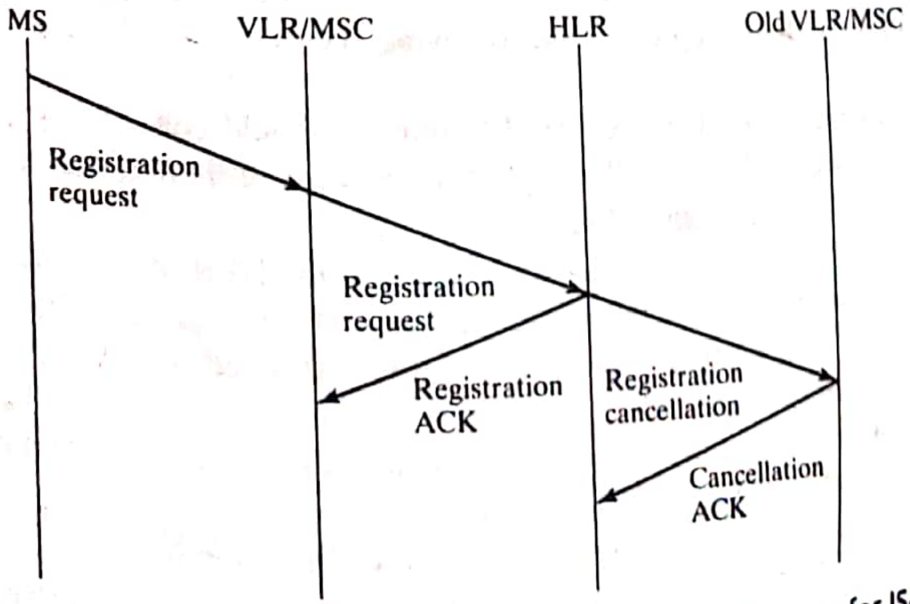

Figure 7.12 Location registration signaling flow diagram for IS-41.

(2) The base station forwards the request to the MSC which, in turn, launches a registration query to its associated VLR.
(3) The VLR adds an entry in its records on the location of the mobile and forwards the registration request to the HLR. The HLR of the mobile is determined from the MIN of the mobile through a global title translation.
(4) The HLR performs the necessary authentication routine and records the identity of the new serving VLR in its database entry for the mobile. The HLR then sends a registration acknowledgment message to the new VLR.
(5) The HLR sends a registration cancelation message to the old serving VLR of the mobile.
(6) The old VLR removes the record of the mobile from its location database and returns a cancelation acknowledgment message to the HLR.

The mobile's registration with its HLR is now complete. With the dismantling of the association between the HLR and the *old* VLR, the HLR now has an association with the *new* VLR.

When the network receives a call for a mobile, the network access point, which receives the request, must locate the called mobile and deliver the call to it. The procedure to set up the call is as follows. The calling mobile sends the call request to the MSC in the registration area via its serving base station. The MSC which handles the call request is referred to as the calling MSC. The calling MSC locates the called MSC/VLR combination through its association with the HLR. The HLR routes the call request to the called MSC which, in turn, pages the called mobile to set up the call. The call setup is illustrated by the signaling flow diagram shown in Figure 7.13, where TLDN stands for temporary local directory number.

Example 7.4 Call Setup Procedure

Consider the call setup signaling flow diagram shown in Figure 7.13. Describe the steps involved in setting up the call between the calling MSC and the called MSC.

Solution The steps involved in setting up the call should trace out the signal flows shown in Figure 7.13. The call can be set up using the following six steps:

(1) The calling BS receives a call request. The request should contain the MIN of the called mobile. The calling BS forwards the request to the VLR of its RA. The VLR then forwards the request to its corresponding MSC.
(2) The calling MSC sends a location lookup request to the HLR of the called mobile.
(3) The HLR of the called mobile determines the current VLR of the called mobile (from its database) and sends a routing request message to the VLR. The VLR then forwards the message to its corresponding MSC.
(4) The called MSC allocates a temporary local directory number to the called mobile and sends a reply to the HLR, together with the TLDN.
(5) The HLR forwards the TLDN to the MSC of the calling mobile.
(6) Using the TLDN, the MSC of the calling mobile initiates a connection request to the called MSC through the SS7 network.

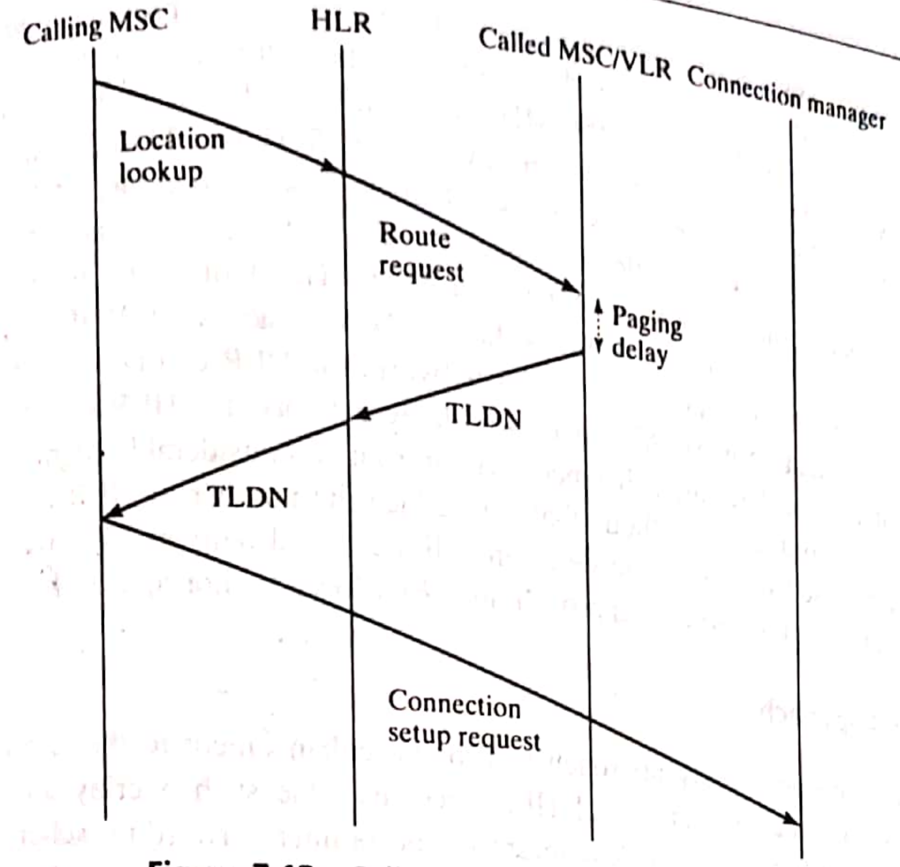

Figure 7.13 Call setup signaling flow diagram.

The above call setup procedure establishes a linkage between the calling MSC and the called MSC. In general, the called MSC only knows that the called mobile is in the RA under its jurisdiction, but does not know in which cell the called mobile is located. A *paging* procedure is needed to locate the called mobile.

Paging Procedure. Paging is a means of locating a mobile within a registration area. There may be different approaches to performing paging. A commonly used approach is *polling*. To locate the called mobile using the polling mechanism, the called MSC broadcasts a polling message, which contains the MIN (mobile identification number) of the called mobile, to all cells within the RA. The BS of each cell relays the paging message to the mobile terminals. Upon receiving the paging message, the called mobile responds to the called MSC, through its currently serving BS, with the base station ID of its current cell. The called MSC then knows where to forward the call.

The second generation wireless systems mainly support voice communications. It is anticipated that future wireless systems, called *personal communication service* (PCS) systems, will support multimedia services at transmission rates that are orders of magnitude larger than that for voice. The two-tiered architecture of IS-41 and GSM's MAP may not be adequate.

7.5 LOCATION MANAGEMENT FOR PCS NETWORKS

A PCS network is expected to allow communications by anyone, from anywhere, and at anytime. Compared with the second generation cellular networks such as IS-136 and GSM, future PCS networks (e.g., third generation wireless networks) are expected to provide higher transmission rates and operate at higher regions of the frequency spectrum (e.g., in the spectral range of 1.85–1.99 GHz as compared with the 850 MHz range in second generation cellular systems). In this case, the two-tiered location management system of IS-41 and MAP, which is suitable for second generation cellular systems, may not be adequate to support future PCS networks. Also, future PCS networks will have smaller cell sizes to enlarge the system capacity (see Chapter 5 for detailed discussions on the capacity expansion issue). This will result in smaller registration areas and an increase in the number of times that the HLR is accessed. With more mobile users in the PCS than in a second generation cellular network, the HLR can become the bottleneck of the location management network. Moreover, in a large network, the HLR may be very far away from its point of access. The large distances will introduce considerable signaling delays. It is therefore desirable to introduce methods that can reduce the number of HLR accesses. Different approaches may be used to reduce the delay in call setup and delivery. We will discuss below two such approaches: (*a*) an overlay approach and (*b*) a local anchor approach.

7.5.1 Overlay Approach

One strategy is to use an overlay approach to provide enhancement to the two-tiered architecture of IS-41 by reducing the number of HLR accesses. One such overlay approach is called pointer forwarding. A functional block diagram of the pointer forwarding scheme is illustrated in Figure 7.14. In the pointer forwarding scheme, instead of reporting to the HLR on every RA crossing, the pointer forwarding strategy sets up pointers from the old VLR to the new VLR, thus eliminating signaling exchanges between the HLR and the VLR's. To locate a mobile during call delivery, the system must traverse the link chain from the initial VLR to the mobile's current VLR. If the length of the chain is overly long, the delay in call setup and message delivery may

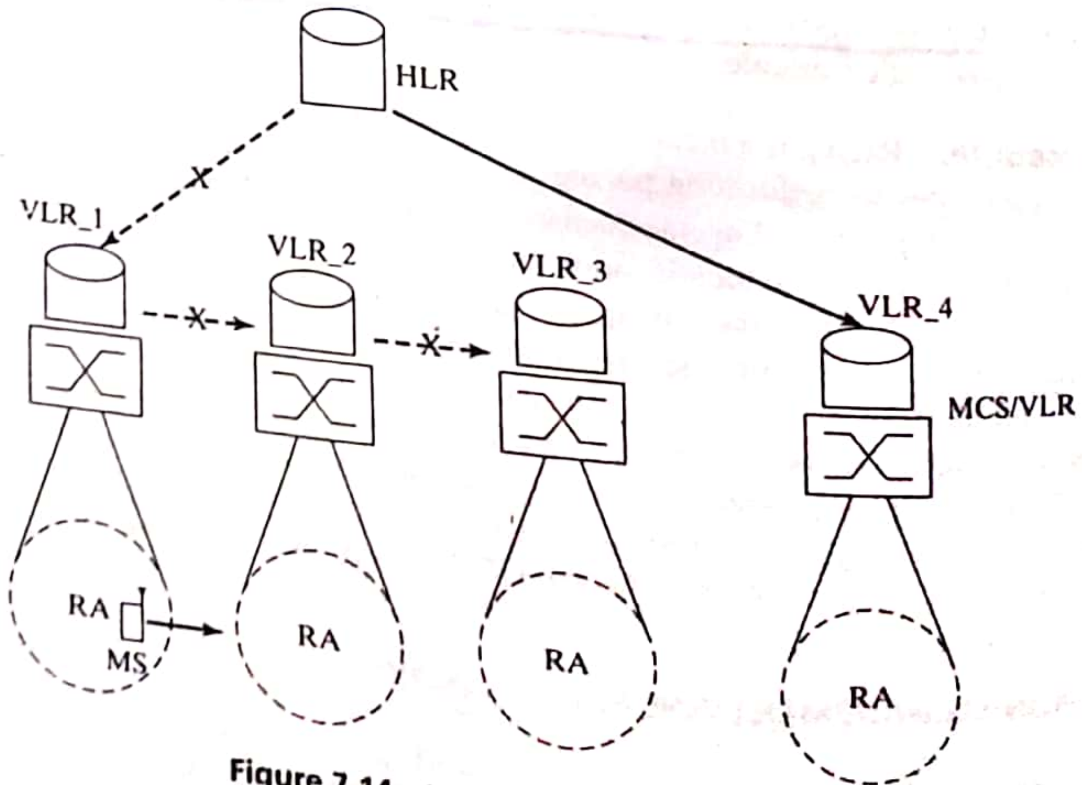

Figure 7.14 Pointer forwarding scheme.

be intolerable. Thus, in this form of overlay, an upper bound on the cost of call setup needs to be incorporated.

To place an upper bound on the cost of call setup, a limit of value K can be placed on the length of the link chain. When the length of the link chain exceeds K, an update to the HLR is forced and the link chain is reset to NULL. As an example, in Figure 7.14, consider the case of $K = 2$. When the mobile moves from VLR_3 to VLR_4, the link chain length limit is exceeded, and the HLR is notified of the location change.

The rate of call arrivals and the rate of RA crossings have an impact on overlay approaches. The call arrival rate and the RA crossing rate can be combined into a single call-to-mobility ratio (CMR), which is defined as

$$\text{Call-to-mobility ratio} \triangleq \frac{\text{rate of call arrivals}}{\text{rate of RA crossings}}.$$

The pointer forwarding scheme works well if the CMR is relatively low. Also, after K moves, the mobile is forced to update the HLR even if the mobile is still very close to its previously reported location.

7.5.2 Local Anchor Approach

Another approach to reducing the frequency of accessing the HLR is to create a virtual HLR in the form of a local anchor (LA) [5]. When the mobile crosses an RA boundary, it registers with the LA. The LA has an association with the HLR, and plays the role of the HLR as far as registration updating is concerned. This is the reason why we call the LA a virtual HLR. In the local anchor scheme, the MSC of the newly entered RA registers the mobile's location at the colocated VLR. This colocated VLR is called the LA. The associated LSTP region (see Figure 7.11) is the anchor LSTP region. Note that each mobile may have a different LA and the LA for a mobile may change from time to time. Once a VLR is selected as the LA for a particular mobile, an entry is set up in a table at this VLR indicating the current serving MSC/VLR for the mobile. A question of how to select the LA arises. One strategy is to select the serving VLR of the mobile during the last call arrival as the LA. A diagram that portrays the signal flow of reporting location change to the LA is shown in Figure 7.15. The numerals shown in Figure 7.15 indicate the steps involved in the location registration.

Example 7.5 Location Registration Procedure

Consider the numbered signal flow diagram for reporting location change shown in Figure 7.15. The steps used to implement the location registration procedure should correspond to the numbered signal flows shown in the diagram. Describe the steps needed to implement the registration procedure.

Solution Following the numbered signal flows, we can establish the following six steps to implement the location registration procedure:

(1) Start by considering a mobile moving into the new RA and sending a location update message to the nearby BS.
(2) The BS forwards this update message to the new MSC/VLR.

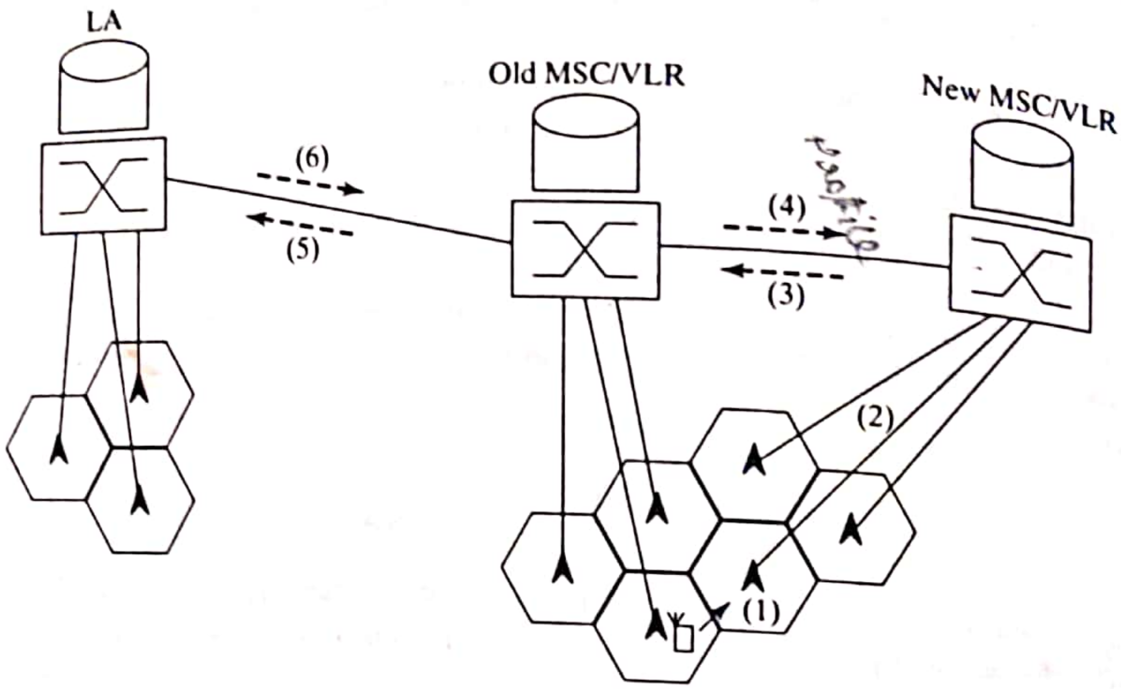

Figure 7.15 Signaling flow for reporting location change to the local anchor.

(3) The new MSC sends a message to inform the old MSC that the mobile has moved out of its associated RA.
(4) The old MSC sends an acknowledgment message to the new MSC together with a copy of the mobile's profile.
(5) The old MSC removes the record of the mobile station in its associated VLR and sends a message to inform the LA of the location change.
(6) The LA updates its record indicating the new location of the mobile and sends an acknowledgment message to the old MSC.

After executing step (6), the reporting process is complete.

Consider a mobile-to-mobile communication scenario. The calling mobile's message is to be delivered to the called mobile. The calling mobile sends the message to the calling MSC via its serving base station. Since there is no direct association between the calling MSC and the called MSC, communication between these two MSCs has to hop through the central HLR. The signal flow for call delivery is illustrated in Figure 7.16. The numerals in the figure indicate the actions to be executed in the call delivery process.

Example 7.6 Call Delivery Procedure

Consider the call delivery problem shown by the signal flows in Figure 7.16. Describe the steps needed to implement the call delivery procedure.

Solution The call delivery process involves seven steps that correspond to the numbered signal flows shown in Figure 7.16.

(1) A call is initiated by a mobile which sends a call initiation signal to its serving MSC through the nearby BS.

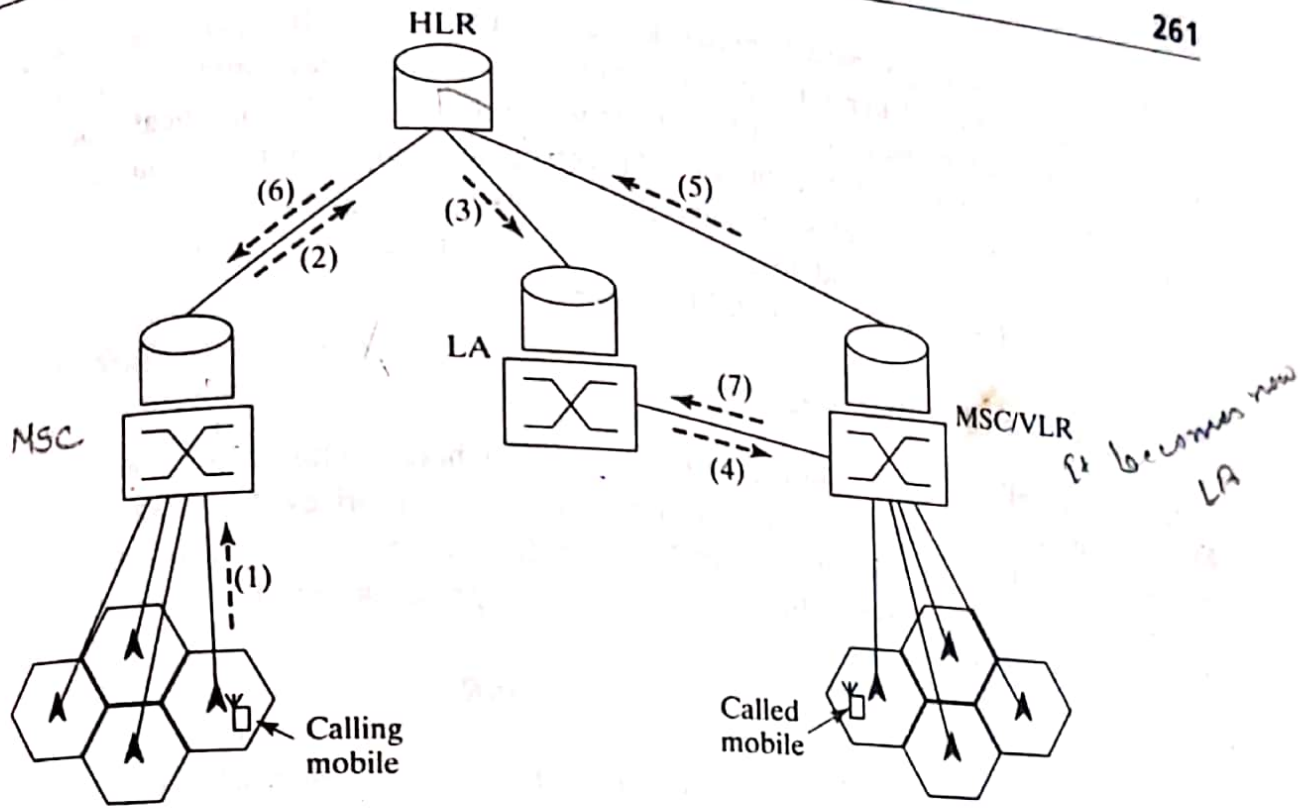

Figure 7.16 Signaling flow for call delivery in LA-based scheme.

(2) The MSC of the calling mobile sends a location request message to the HLR.
(3) The HLR sends a location request message to the LA of the called mobile.
(4) The LA forwards the location request message to the MSC serving the called mobile.
(5) The called MSC allocates a TLDN to the mobile and sends this TLDN to the HLR. The HLR updates its record such that the current serving VLR of the called mobile becomes the new LA.
(6) The HLR forwards the TLDN to the MSC of the calling mobile terminal. The calling MSC can now set up a connection to the called MSC using this TLDN.
(7) The serving MSC of the called mobile sends an acknowledgment message to the old LA. The old LA removes its record of the called mobile terminal.

In the preceding call delivery procedure, it is assumed that the LA is different from the serving VLR. If they are the same, the messages between the LA and the serving VLR are not needed. Step (7) can take place concurrently with steps (5) and (6).

7.6 TRAFFIC CALCULATION

The rate of boundary crossing by the mobile is an important parameter for performance assessment of both handoff and location management schemes. Handoff can be intraswitch or interswitch with different handoff rates. In location management, a location update is necessary whenever the mobile crosses the boundary of a registration area. Knowledge of the handoff rate and location update rate can provide a good indication of system performance and a basis for designing handoff schemes.

For a given mobility model, network architecture, and traffic parameters, it is possible to calculate the intraswitch handoff rate, the interswitch handoff rate, and the location update rate. The traffic calculation can be model-based or measurement-based. Analytical models are normally constructed based on known distributions. Therefore, an analytical model may not portray the real scenario very well. Nevertheless, results obtained using analytical models can provide guidance for the design of handoff and location management strategies. The traffic calculations in this section are based on a universal mobility model.

7.6.1 System and Traffic Parameters

Consider the cellular layout shown in Figure 7.17, where a cluster of cells is connected to an MSC. Assuming a hexagonal cell topology, a cluster comprises N hexagonal cells. Although in general the shape of a cluster is not hexagonal, from the modeling point of view, we use a hexagon (the large hexagon in Figure 7.17) as an approximation. The area of a hexagonal cell, A_{cell}, is approximately given by

$$A_{cell} = 2.598R^2,$$

where R is the radius of the small hexagon in Figure 7.17, measured from the center of the cell to the vertex (it is also the length of each side of the hexagon). The area of the cluster is approximately given by

$$A_{cluster} \approx N \times 2.598R^2 \approx 2.598R^2_{cluster}, \tag{7.6.1}$$

where $R_{cluster}$ is the radius of the large hexagon. From Eq. (7.6.1), we have

$$R_{cluster} \approx \sqrt{N}R. \tag{7.6.2}$$

To calculate the intracluster handoff rate, we need a knowledge of the number of boundary crossings per cell per unit time. To calculate the intercluster handoff rate, we need a knowledge of the number of cell crossings per cluster per unit time. Handoff rate calculation requires information on the network parameters and traffic parameters, and the mobile's movement pattern. Note that a mobile is in a dormant state unless it is powered on. Also, there is no traffic loading from a mobile unless it is active (i.e., either transmitting or receiving messages). Taking these factors into

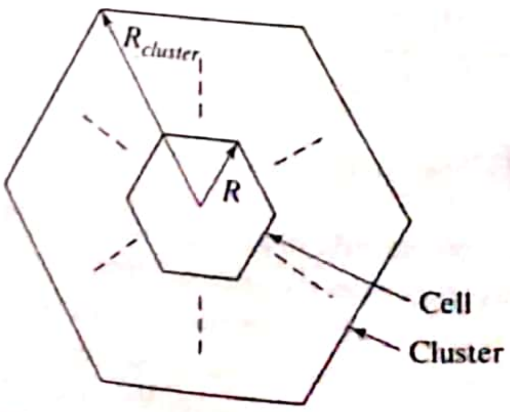

Figure 7.17 Cellular layout of clustering.

consideration, the following is a list of the relevant system and traffic parameters that contribute to the formulation of an expression for the handoff rate:

1 Population density of mobiles = ρ (number of mobiles/km^2)
7 Hexagonal cell radius = R (km)
6 Number of cells per cluster = N
8 Speed of mobile: V (km/hr)
9 Traffic load per mobile station = λ_{ar} (Erlangs/MS)
2 Percentage of powered stations = δ
3 Percentage of active mobiles among the powered stations = ϵ
4 Number of crossings per cell per unit time = μ_{cell}
5 Number of crossings per cluster per unit time = $\mu_{cluster}$

Note that a mobile can be in motion even if it is not powered on or not active. When a mobile terminal is powered on, it is in a listening mode. When the mobile is active, it has an ongoing connection. Therefore, a mobile is active only if it is powered on. Thus, the percentage of the total population that is active is given by $\epsilon \cdot \delta$ and the number of handoffs per unit time is considerably smaller than the number of boundary crossings per unit time.

In formulating Mobility Model 1 in Subsection 7.3.8, we focused attention on the movement of a single mobile. To consider handoff rates in general, we need to account for all active mobiles within the cluster. We call the model that characterizes the movements of all mobiles Mobility Model 2.

Mobility Model 2. Mobility Model 2 is governed by the following assumptions:

a. The mobiles are uniformly distributed in the cell;
b. The direction of travel of all mobiles relative to the cell boundary is uniformly distributed on $[0, 2\pi]$;
c. Once a mobile has been admitted into the cell, it assumes a direction to travel and then travels along that direction with constant velocity V.

7.6.2 Handoff Rate Calculation

With Mobility Model 2, the rate of boundary crossing can be determined by the rate at which a randomly chosen mobile crosses the cell boundary times the number of mobiles per cell. The rate of boundary crossing is a function of the cell size, the population density, and the speed of mobile movement. Let L_{cell} be the perimeter of the cell. For a hexagon, $L_{cell} = 6R$. The rate of mobiles departing the cell (i.e., the number of mobiles departing a cell per unit time), μ_{cell}, is given by

$$\mu_{cell} = \frac{\rho V L_{cell}}{\pi} \tag{7.6.3}$$

$$28.64 = \frac{6\rho V R}{\pi}.$$

By the flow conservation law, the number departing a cell per unit time equals the number entering the cell per unit time. With the area of the cluster approximated by a hexagon of radius, $R_{cluster}$, given by Eq. (7.6.2), the rate of mobiles departing the cluster, $\mu_{cluster}$, can also be expressed as

$$\mu_{cluster} = \frac{\rho V L_{cluster}}{\pi}, \tag{7.6.4}$$

where $L_{cluster}$ is the perimeter of the hexagon used to approximate the area of the cluster. From Eq. (7.6.2), we can approximate the perimeter of this large hexagon as

$$L_{cluster} \approx 6R_{cluster} \approx 6\sqrt{N}R.$$

Hence,

$$\mu_{cluster} \approx \frac{6\rho V\sqrt{N}R}{\pi}. \quad \mu_{cell} = \frac{6\rho V R}{\pi} \tag{7.6.5}$$

Let $\lambda_{ho\_cluster}$ be the total handoff rate of a cluster (i.e., the number of handoffs in the cluster, both within the cluster and to neighboring clusters, per unit time). Then, the total handoff rate is given by

$$\begin{aligned}\lambda_{ho\_cluster} &= \text{number of cells in the cluster} \\ \propto N \quad & \times \text{ number of crossings per cell per unit time} \times \text{\% active mobiles} \\ & \times \text{\% powered stations} \times \text{traffic load per MS} \\ &= N \times \mu_{cell} \times \epsilon \times \delta \times \lambda_{ar}.\end{aligned} \tag{7.6.6}$$

The total handoff rate of a cluster includes both intracluster handoff rate and intercluster handoff rate. The intercluster handoff rate, $\lambda_{ho\_intercluster}$, is given by

$$\begin{aligned}\lambda_{ho\_intercluster} &= \text{number of crossings per cluster per unit time} \times \text{\% active mobiles} \\ \propto \sqrt{N} \quad & \times \text{\% powered stations} \times \text{traffic load per MS} \\ &= \mu_{cluster} \times \epsilon \times \delta \times \lambda_{ar}.\end{aligned} \tag{7.6.7}$$

It follows that the intracluster handoff rate, $\lambda_{ho\_intracluster}$, is given by

$$\lambda_{ho\_intracluster} = \lambda_{ho\_cluster} - \lambda_{ho\_intercluster}. \tag{7.6.8}$$

Example 7.7 Intracluster Handoff Rate

Consider a cellular network with the following traffic and system parameter values:

$\rho = 45{,}000$ mobiles/km^2

$V = 6$ km/hr

$R = 1.0$ km

$$N = 16$$

$$\epsilon = 10\%$$

$$\lambda_{ar} = 0.06 \text{ Erlangs/mobile station}$$

$$\delta = 50\%$$

Find the intracluster handoff rate, $\lambda_{ho\_intracluster}$.

Solution We first determine the average number of crossings/cell/hr, μ_{cell}, and the average number of crossings/cluster/hr, $\mu_{cluster}$, using Eqs. (7.6.3) and (7.6.4), respectively. From Eq. (7.6.3), we have

$$\mu_{cell} = \frac{6 \times 45{,}000 \times 6 \times 1}{\pi} = 515{,}595 \text{ crossings/cell/hr.}$$

From Eq. (7.6.4), we have

$$\mu_{cluster} = \frac{6 \times 45{,}000 \times 6 \times \sqrt{16} \times 1}{\pi} = 2{,}062{,}381 \text{ crossings/cluster/hr.}$$

The total handoff rate can then be obtained using Eq. (7.6.6) as

$$\lambda_{ho\_cluster} = 16 \times 515{,}595 \times 0.1 \times 0.5 \times 0.06 = 24{,}748.56/\text{hour.}$$

From Eq. (7.6.7), we get

$$\lambda_{ho\_intercluster} = 2{,}062{,}381 \times 0.1 \times 0.5 \times 0.06 = 6{,}187.14/\text{hour.}$$

Therefore, from Eq. (7.6.8), we get

$$\lambda_{ho\_intracluster} = 24{,}748.56 - 6{,}187.143 = 18{,}561.42/\text{hour.}$$

From Example 7.7, it is noted that the intracluster handoff rate, $\lambda_{ho\_intracluster}$, is much larger than the intercluster handoff rate, $\lambda_{ho\_intercluster}$. This result can be inferred from the fact that the total handoff rate in a cluster, $\lambda_{ho\_cluster}$, is proportional to the number of cells, N, in the cluster, while the intercluster handoff rate, $\lambda_{ho\_intercluster}$, is proportional to $\sqrt{N}$. The limiting case is $\lambda_{ho\_intracluster} = \lambda_{ho\_intercluster}$, which occurs when $N = 1$.

Example 7.8 Effect of Cell Splitting on Handoff Rates

Consider the effect of cell splitting on handoff rates. Suppose a cell of radius R_{cell} is split into microcells, each with radius $R_{microcell} = R_{cell}/4$. If the traffic and system parameters are those given in Example 7.7, what is the impact of cell splitting on handoff rates?

Solution Since the total handoff rate is proportional to the number of cells in the cluster, the effect of cell splitting on handoff rates can be obtained by first calculating the number of cells in the cluster after cell splitting. Let K be the number of microcells per original cell after cell splitting. The number of microcells per large cell can be determined by taking the ratio of the area of the large cell to the area of the microcell. Let A_{cell} be the area of the large cell and $A_{microcell}$ be the area of the microcell. We have

$$A_{microcell} = 2.598 R^2_{microcell}$$

and

$$A_{cell} = 2.598 R^2_{cell} = K \times 2.598 R^2_{microcell}.$$

Then

$$K = \frac{R^2_{cell}}{R^2_{microcell}} = \frac{R^2_{cell}}{(R_{cell}/4)^2} = 16.$$

The number of cells per cluster is, therefore,

$$N \times K = 16 \times 16 = 196.$$

The total handoff rate in the cluster is increased by $K = 16$-fold, while the intercluster handoff rate is increased by $\sqrt{K} = 4$-fold. This means that most of the increase in the total handoff rate is with the intracluster handoff.

SUMMARY

The salient feature of wireless communication and networking is the provision of an information transport platform that allows the mobile users to roam. To facilitate roaming, effective mobility management procedures need to be in place. Mobility management consists of two components: handoff management and location management. This chapter describes and discusses the fundamental aspects of handoff management and location management. The latter is illustrated by the two-tier architecture of the IS-41 standard.

ENDNOTES

1. For a general discussion of network issues in wireless communications, see the paper by Jabbari, Colombo and Kulkarni [68] and the book by Lin and Chlamtac [87]. For fundamentals of communication networks, see the book by Bertsekas and Gallager [13].
2. For a general discussion of handoff issues in wireless communications, see the papers by Pollini [117] and by Noerpel and Lin [105]. For feedback-based mobile assisted handoff, see the paper by Mukhi and Mark [101]. For soft handoff in CDMA systems, see the paper by Wong and Lim [159].
3. For location management, see the papers by Akyildiz and Ho [5], Tabbane [149], and Zhang and Mark [164].
4. For details of IS-41, see EIA/TIA interim standard [44]. For details of the Mobile Application Part of GSM, see ESTI technical report recommendation [47]. For an overview of signaling system no. 7, see the paper by Modarressi and Skoog [98].