# Resource Management issues in Future Wireless Multimedia Networks

Youssef Iraqi and Raouf Boutaba

*University of Waterloo, Waterloo, Ont. N2L 3G1, Canada*
E-mail: {iraqi, rboutaba}@bbcr.uwaterloo.ca

Broadband multimedia services and wireless services are becoming very popular. They are presently the two major drivers in the telecommunication industry. It is expected that wireless mobile users' demands for multimedia mobile services will rapidly increase in the future. The paper starts by presenting the evolution of wireless networks to support future integrated multimedia wireless personal services and highlight the need for efficient resource management. The limited resources in wireless systems, such as spectrum resource and transmitter power also stress the need for efficient resource management. Emphasis is given in the paper to the discussion of the channel allocation, mobility management, and bandwidth distribution problems. A review of some approaches proposed in the literature to solve these problems is presented. Finally, the paper discusses some of the key resource management issues that need to be addressed in the context of future wireless mobile networks that support multimedia communications while ensuring quality of service guarantees.

## 1. Introduction

The confluence of users demand for personal communication systems and technological advances in hardware design, RF circuitry, rechargeable batteries, and spread spectrum communications has led world-wide to a phenomenal explosion and deployment of mobile wireless communications. Compared to traditional wired telecommunication networks, wireless communications have evolved at a rapid pace from simple first-generation analogue to second-generation digital networks. The last decade has particularly witnessed the preparation for future wireless personnel communication. The latter aims at providing communication services from any person to any person in any place at any time without any delay in any form through any medium by using one pocket-sized unit at minimum cost with acceptable quality and security [1]. The deployment of third-generation mobile wireless networks is expected for year 2001 via Universal Personal Communications systems to provide universal speech and local multimedia services. Moreover, fourth-generation mobile wireless networks are emerging to offer capacities up to 150 Mb/s to fully mobile users in various environments [2]. In such rapidly evolving environment, resource management remains an important issue in the near and distant future.

Efficient resource management is of paramount importance due to: the rapid increase in size of the wireless mobile community; its demand for high-speed multimedia communications; and the limited resources. This survey article outlines some of the crucial resource management problems in wireless and mobile networks. In particular, the paper reviews channel allocation, mobility management, and bandwidth distribution mechanisms. Approaches proposed in the literature are summarized and discussed. Furthermore, the requirements of multimedia applications on the surveyed resource management issues are highlighted.

The most common resource management issue faced by wireless operators is channel allocation encompassing frequency planning and channel reuse methods, as well as the optimization of network capacity. Channel allocation is crucial as it influences the network performance. Several techniques have been proposed to improve the network capacity through efficient utilization of channel resources. However, most of these techniques involved compromises in terms of service quality. Channel allocation techniques can typically be classified into fixed, dynamic and hybrid. Fixed cellular allocation strategies permanently assigns specific channels to specific cells. On the other hand, dynamic channel allocation schemes have been developed to use the channel resource more efficiently by dynamically assigning channels to cells, on demand basis, to satisfy call requests. Hybrid channel allocation strategies combine fixed and dynamic approaches by supporting both permanent and dynamic channel allocations. Also, variations of the fixed channel allocation technique that exploit the overlapping of adjacent cells are used to improve the wireless system performance. Directed retry and selective handoff schemes are examples of such variations.

The previous techniques are used to assign frequency channels to the wireless network cells, which use these channels to set up mobile calls. The cells have also to deal with handoffs to ensure call continuity, while the allocated radio channel changes either within one cell, intra-cell handoff, or during cell crossing, inter-cell handoff. Therefore, channel allocation strategies in the event of a handoff are highly desirable. Such strategies have to be reliable in order for wireless mobile networks to provide seamless communications under fluctuating radio channel behavior, and unpredictable radio resource conditions. They have a significant impact on both system capacity and performance. For example, channel assignment schemes are developed for macrocell/microcell overlays in such a way to increase overall system capacity and to reduce the number of handoffs. Power control is another important issue that is strongly related to the channel assignment problem and which is used to increase the overall system capacity by reducing carrier-to-interference ratio.

Another key issue in mobile wireless networks is mobility management, which can be categorized into two areas: Location management procedures that enable the network to track the current location of every user and deliver incoming calls; Handoff processes that ensure the continuity of a mobile user connection.

Location management schemes are essentially based on both users' mobility and incoming call rate characteristics. It involves two basic procedures: location and paging. The paging process consists of sending paging messages in all cells where the mobile terminal could be located. The location procedure implements mobile terminal registration and tracking. It allows the system to keep the user's location knowledge up to date and to bring the user's service profile near its location. In case of an incoming call, the system is then capable to locate the users and to provide them rapidly with their customized services. Several location updating procedures have been proposed which can be either as manual, performed by the user when requested, or automatic, e.g., geographic when the mobile enters a new location area, periodic every average time period, etc. Although the previous basic procedures of network mobility, i.e., location and paging, are antagonist, it is highly desirable to select location updating and terminal-paging policies that minimize the total cost in terms of wireless bandwidth and terminal power. The previous basic procedures of mobility management become more complex when dealing with inter-systems mobility. The requirement for inter-systems mobility refers to the ability of a mobile user to make and receive calls in places other than the home system. Roaming between different systems may not be possible unless many critical issues are solved like location-awareness services, roaming between similar and dissimilar systems, call detail recording, privacy and authentication.

The handoff process is related to access, radio resources, and network control. It successively involves a measurement phase, a handoff initiation/control phase, a channel assignment phase and a change in the connection within the network involving possibly rerouting and buffering. Several handoff schemes have been proposed, which can be classified as hard, seamless or soft according to the way the new path is set up and the connection shifted from the actual to the new path. The handoff is already a key process in current systems and it is foreseen to gain increasing importance in third and fourth generation cellular systems as cells radius is decreasing, while the number of users is expected to grow dramatically. In particular, appropriate multimedia handoff schemes are essential to overcome the additional complexity due to the nature of multimedia applications involving many mobile users and distributed media sources/sinks. The wireless mobile system has to support the reestablishment and the rerouting of distributed multimedia streams in the event of a handoff.

Broadband multimedia services and wireless services are presently the two major drivers in the telecommunication industry. Considering the explosion of the World Wide Web and multimedia-based applications, it is expected that wireless mobile user demands for integrated multimedia mobile services will continue increasing for the foreseeable future. This trend will raise the requirement for mechanisms to control the access to the wireless network to avoid congestion and to ensure user quality of service (QoS). QoS provisioning in wireless mobile networks is more challenging than in fixed networks. Indeed, characteristics of wireless mobile networks such as channel fading, mobility and handoff lead to fluctuations in network resources availability, and hence, bring more complexity to the QoS provisioning problem. Issues such as traffic characterization, bandwidth partitioning, call admission control, and resource reservation are posing new challenges, and little work has been done, so far, to address these issues in wireless mobile networks. For instance, call admission control (CAC) has to be provided when the mobile station initially enters the network, but also whenever the mobile station roams in the wireless network and hands off from one cell to another. This imposes temporal constraints on the CAC mechanism as it necessitates a continuous control due to users' mobility and the dynamic nature and fluctuations of resources availability at a wireless network access point, typically the base station within a cell. Bandwidth in a wireless network is the most precious and scarce resource of the entire communication system. Base stations must find the best compromise between maintaining maximum bandwidth utilization and reserve enough bandwidth resources so that to reduce the rate of unsuccessful incoming handoffs due to insufficient available resources. Yet the requirement for QoS end-to-end in wireless mobile networks still need to be addressed.

This paper is organized into 5 main sections. Section 2 introduces a typical wireless mobile network architecture, and discusses the evolution of wireless personal communications. Section 3 gives a brief overview of existing multiple access and allocation strategies. Performances of the basic channel allocation strategies are discussed together with a number of variations of these strategies and related issues such as handoff handling and power control. Section 4 is dedicated to the mobility management issue in wireless networks. It discusses two major problems related to mobility management: location; and handoff procedures. It also presents a survey of existing approaches to solve these problems. Section 5 addresses the requirement for bandwidth partitioning and allocation mechanisms for the provisioning of QoS in future integrated multimedia mobile communications. Section 6 concludes this survey paper.

## 2. Wireless Mobile Networks

### 2.1. Architecture and Evolution

Wireless mobile networks consist of a fixed network and a large number of mobile terminals. These terminals include telephones, portable computers, and other devices that exchange information with remote terminals through the fixed network. The wireline network can be the telephone network in use today or the ATM network in the future. In order to effectively utilize the very limited wireless bandwidth to support an increasing number of mobile subscribers, current wireless mobile networks are designed based on a cellular architecture. The network coverage area is divided into a large number of smaller areas called cells. Terminals within a cell communicate with the wireline network through a base station (BS) that is installed inside the cell. This base station serves as the network access point for all the terminals within the cell. When a terminal enters another cell, if the new base station has enough resources to accept the mobile, this later will switch to the new base station. In the other case, many mechanisms can be used to accommodate the user (like load balancing strategies, directed retry, power assignment, auxiliary pilot...) if no other solution is possible the call is dropped. A number of adjacent cells grouped together form an area and the corresponding BSs communicate through a so-called Mobile Switching Center (MSC). The MSC may be connected to other MSCs on the same network or to the wireline network.

A typical architecture of a cellular mobile network contains mainly the following components (see Figure 1): Mobile Switching Center (MSC), Base Stations (BSs), and Mobile Stations (MSs). These components interact via wired links and air interfaces. The MS is the physical communication equipment of the user (mobile service subscriber). As part of the MSC also the HLR (Home Location Register) and the VLR (Visitor Location Register) are the support databases where information on the subscribers are stored (e.g., their home and current locations). Two other databases are also used within the MSC: the Equipment Identity Register (EIR) and the Authentication Center (AuC). The BS handles radio traffic within its cell (from and to a MS). A Base Station Controller (BSC), which controls a number of BSs continually collect statistics on the number of calls, successful or unsuccessful handoffs, traffic per cell, etc.

The future wireless personnel communications, defined as being the ultimate goal of today's communication engineers, will provide communication services from any person to any person in any place at any time without any delay in any form through any medium by using one pocket-sized unit at minimum cost with acceptable quality and security through the use of a personal telecommunication reference number [1]. The objectives of the research and development of these systems are focused in three technological platforms [3]: universal mobile telecommunication system (UMTSs), mobile broadband systems (MBSs), and wireless customer premises networks (WCPNs).

Third-generation systems are expected to be deployed by the year 2001 via Universal Personal Communications systems, which will provide universal speech and local multimedia services. They are in the process of development world-wide by the ITU (International Telecommunications Union) within the framework of the International Mobile Telecommunications 2000 (IMT-2000). The European Telecommunications Standards Institute (ETSI) has started with Universal Mobile Telecommunication System (UMTS), which will be a member of the IMT-2000 "family of systems" within the responsibility of the ITU.

With the emergence of unlimited networks in places of congregation such as airports, conferences and hotels, and with the increase reliance on the easy availability of multimedia information on the Internet,

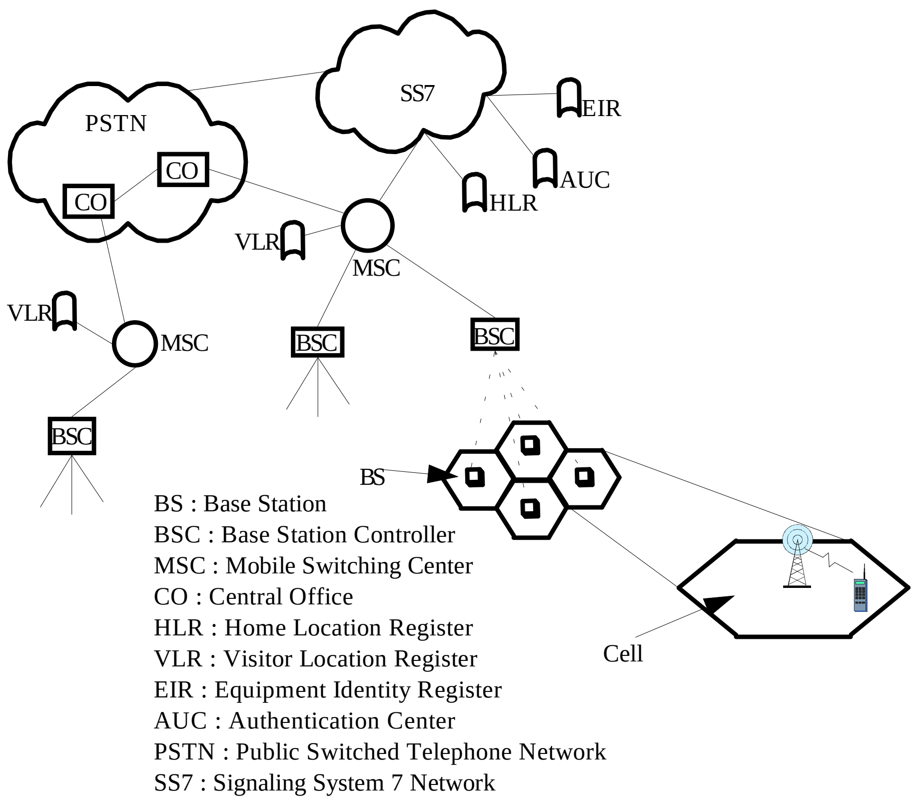


Figure 1. A typical architecture of a cellular network

we are surely moving toward widespread use of handheld wireless multimedia communications.

UMTS is intended to provide a wide range of mobile services to users via a range of mobile terminals that enable the use of the pocket telephone in almost any location, indoor or outdoor, in home, office, or street. The future access scheme will be either CDMA or one of the hybrid multiple access schemes based mainly on the combination of CDMA and TDMA/FDMA [1,4,5]. Figure 2 shows three wireless environments; PCS/Cellular, Third generation and Wireless LANs. They are classified according to cell size and data rates supported. Readers interested in a recent discussion of the evolution of wireless communications can refer to [6].

Market studies show that multimedia services will have in the year 2001 a penetration of more than 60 percent in most industrialized countries. Emerging fourth generation mobile wireless systems will allow the provision of seamless services over an increasing number of distinct and heterogeneous fixed and wireless networks operating across different frequency bands.

### *2.2. Resource Management : An overall view*

The evolution of wireless mobile networks is tributary of the availability of more sophisticated resource management mechanisms capable to handle the growing number of mobile users; multimedia applications; and the limited wireless resources, e.g., spectrum resource and transmitter power. Figure 3 illustrates the most common resource management problems encountered in the wireless mobile environment.

Figure 3 also shows where the enumerated resource management problems must be dealt with in the network.

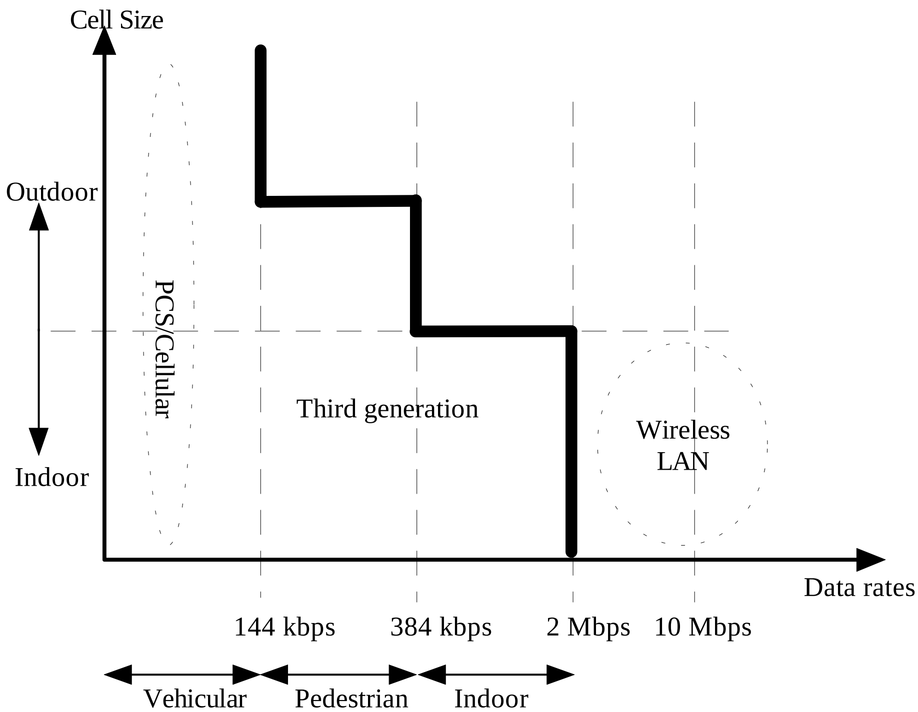


Figure 2. Wireless environments

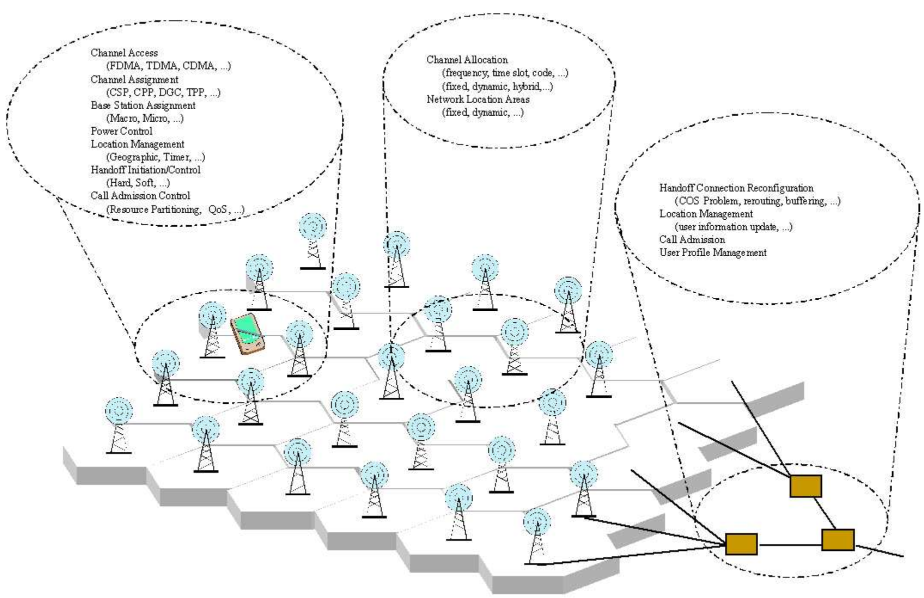


Figure 3. An overall view

Supporting multimedia communications and advanced services, such as bit-hungry high-resolution multimedia ones, over wireless networks is a hard problem because of several factors: scarcity of bandwidth, time-varying error characteristics of the transmission channel, power limitations of the wireless devices, and mobility of the users. The provision of wireless multimedia services to mobile users with a certain quality of service imposes stringent requirements on the design of resource management mechanisms. Multimedia applications are often of a bursty nature and require a dynamic amount of bandwidth.

Guaranteeing quality of service for such multimedia applications while optimizing resource utilization necessitates specific resource management mechanisms. Figure 4 zoom on the left area circled in Figure 3. It illustrates some of the resource management problems that may occur while the user roam the network.

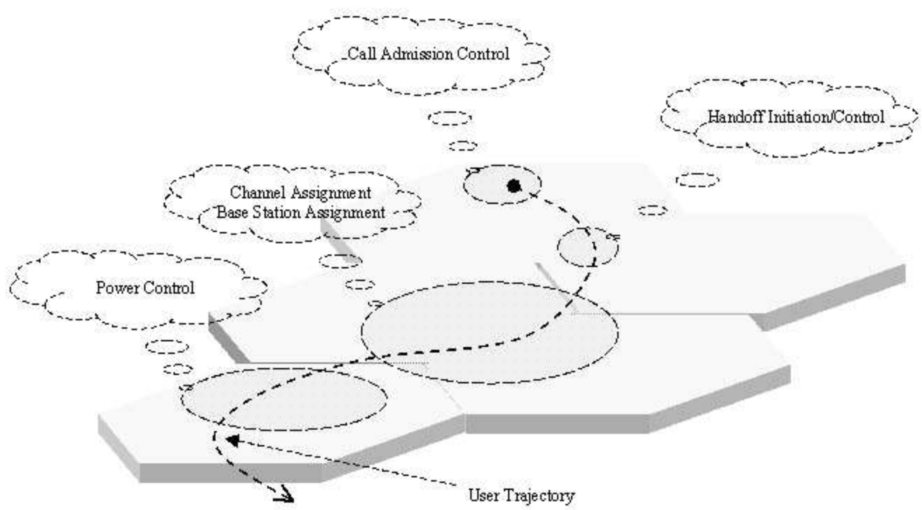


Figure 4. Some resource management issues

The next section reviews resource allocation schemes that encompass frequency planning, power control, and channel reuse methods, as well as the optimization of network capacity.

## 3. Multiple access and Channel allocation schemes

FDMA (Frequency Division Multiple Access), TDMA (Time Division Multiple Access) and CDMA (Code Division Multiple Access) are the three multiple access schemes commonly used. These techniques are used to divide a given radio spectrum into a set of disjoint or non-interfering radio channels. FDMA divides radio channels into a range of radio frequencies and is used in the traditional analog cellular system. With FDMA, only one subscriber is assigned to a channel at a time. Other users can access this channel only after the subscriber's call has terminated or after the original call is handed off to a different channel by the system. TDMA is a common multiple access technique employed in digital cellular systems. It divides conventional radio channels into time slots to increase the capacity. As with FDMA, no other users can access an occupied TDMA channel until the channel is vacated. CDMA uses a radically deferent approach. It assigns each subscriber a unique code to put multiple users on the same wideband channel at the same time. The codes, called "pseudo-random code sequences", are used by both the mobile station and the base station to distinguish between connections. The FDMA, TDMA, and CDMA techniques can be combined in various ways to design more elaborated schemes. A system such as GSM is a combination of TDMA/FDMA system achieved by dividing each frequency band of an FDMA system into time slots.

Both channel derivation and allocation methods will influence the network performance. In FDMA or TDMA the resource usage is based on the channel reuse concept applied throughout the cells and microcells. Channel reuse implies that distinct terminals use the same channel in different cells, with the only constraint of meeting a given interference threshold.

The basic prohibiting factor in radio spectrum reuse is interference caused by the environment or other mobiles. Co-channel interference caused by channel reuse is the most restraining factor on the overall system capacity in the wireless network. There are a number of other problems related to the operation of the cellular network in a real world. Adjacent channel interference is one problem that occurs when there is a strong signal being emitted from a channel adjacent to the channel of concern. The environment in which the cellular network operates is another source of problems. Mountains, hills, and valleys act like huge reflectors, causing multipath problems. The atmosphere can cause propagation loss. The buildings can cause reception problems as well as logistic problems like impeding the placement of base stations.

The main idea behind channel allocation algorithms is to make use of radio propagation path loss ([7,8]) characteristics in order to minimize the carrier-to-interference ratio (CIR). The major driving factor in determining the number of channels with certain quality that can be used for a given wireless spectrum is the level of received signal quality that can be achieved in each channel.

The channel allocation scheme is an essential feature in personal communications systems and impacts the network performance.

Channel allocation schemes can be implemented in centralized or distributed fashion. Channel allocation schemes can be divided into a number of different categories depending on the comparison basis. For example, when channel allocation algorithms are compared based on the manner in which co-channels are separated, they can be divided into fixed channel allocation (FCA), dynamic channel allocation (DCA), and hybrid channel allocation (HCA).

The diagram below shows the sub-strategies existing within the three major allocation strategies stated above.

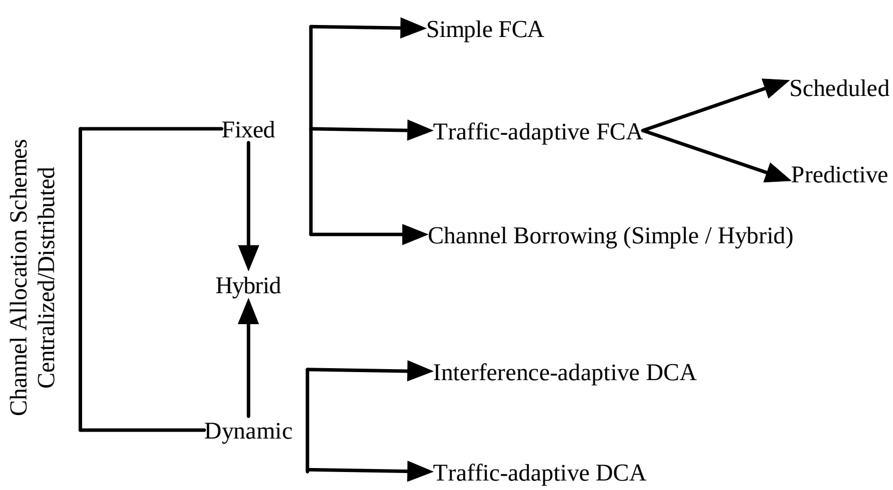


Figure 5. Variants of channel allocation strategies

### *3.1. Fixed allocation strategies*

In fixed channel allocation schemes [9–11] a number of channels are assigned to each cell according to some reuse pattern, depending on the desired signal quality. The common underlying theme in all fixed allocation strategies is the permanent allocation of a set of channels to each cell. The same set of radio frequencies is reused by another cell at some distance away. Here a definite relationship is assumed between each channel and each cell, in accordance to co-channel reuse constraints.

Fixed allocation strategies can be classified into simple fixed, traffic-adaptive and borrowing schemes.

### *3.1.1. Simple fixed channel allocation*

In the simple fixed channel allocation strategy, the same number of nominal channels is allocated to each cell. So a call attempt at a cell site can only be served by unoccupied channels from the predetermined set of channels at that cell site (see an example in Figure 6).

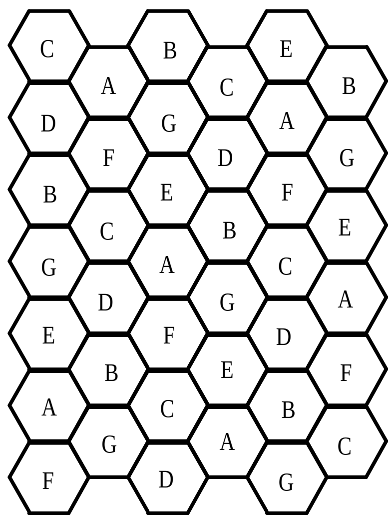


Figure 6. Basic fixed allocation strategy. A-G denote different sorts of channels permanently assigned to cells

### *3.1.2. Traffic-adaptive fixed channel allocation*

Because traffic in cellular systems can be nonuniform with temporal and spatial fluctuations, simple fixed channel allocation scheme may result in poor channel utilization. Non-uniform channel allocation [12,13], static borrowing [14,15] or flexible channel allocation strategy [16] were proposed to solve this problem.

In non-uniform channel allocation the number of nominal channels allocated to each cell depends on the expected traffic profile in that cell. Thus heavily loaded cells are assigned more channels than lightly loaded ones. In [12] an algorithm, namely nonuniform compact pattern allocation, is proposed for allocating channels to cells according to the traffic distribution in each of them. Simulation results in [12] show that the blocking probability using nonuniform compact pattern is always lower than the blocking probability of uniform channel allocation.

In the static borrowing schemes, unused channels from lightly loaded cells are reassigned to heavily loaded ones at distances higher than the minimum reuse distance. The number of nominal channels assigned in each cell may be reassigned periodically on a scheduled or predictive manner.

In the flexible channel allocation schemes, the set of available channels is divided into fixed and flexible sets. Each cell is assigned a set of fixed channels that typically suffices under a light traffic load. The flexible channels are assigned to those cells whose channels have become inadequate under increasing traffic loads. The assignment of these emergency channels among the cells is done in either a scheduled or predictive manner. In the literature proposed fixed allocation techniques differ according to the time at which and the basis on which additional channels are assigned.

In the predictive strategy, the traffic intensity or, equivalently, the blocking probability is constantly measured at every cell site so that the reallocation of the flexible channels can be carried at any point in time.

If the flexible channels are assigned on a scheduled basis, it is assumed that the variations of traffic are estimated a priori. The change in assignment of flexible channels is then made at the predetermined peaks of traffic change [17].

#### *3.1.3. Channel borrowing*

In channel borrowing schemes, if all permanent channels of a cell are busy, a channel is borrowed from a neighboring cell, only if this channel does not interfere with the ongoing calls. When a channel is borrowed, additional cells are prohibited from using it. The MSC supervises the borrowing procedure, favoring channels of cells with the most unoccupied channels to be borrowed. It keeps track of free, serving and borrowed channels and informs all the base stations (BSs) about locked channels.

The proposed channel borrowing schemes differ in the way a free channel is selected from a cell to be borrowed by another cell. They can be categorized into simple and hybrid schemes.

In simple channel borrowing schemes, any nominal channel in a cell can be borrowed by a neighboring cell for temporary use. Examples of these schemes are Simple Channel Borrowing [14,15,18,19], Borrow from the Richest [14], Basic Algorithm [14,15], Basic Algorithm with Reassignment [15] and Borrow First Available [14]. A general conclusion reached by most performance studies of simple channel borrowing schemes is that adopting a simple test for borrowing yields performance results quite comparable to systems which perform an exhaustive and complex search method to find a candidate channel.

In hybrid channel borrowing strategies, the set of channels assigned to each cell is divided into two groups, local channels and borrowable channels. The first group is for use only in the nominally assigned cell, while the second group is allowed to be lent to neighboring cells. The ratio of the numbers of channels in the two groups is pre-determined, depending on the estimation of the traffic conditions.

Examples of hybrid channel allocation schemes are: Simple Hybrid Channel Borrowing strategy [20], Borrowing with Channel Ordering [12,18,19] and Borrowing with Directional Channel Locking [12].

Borrowing with ordering strategy elaborates on the idea of hybrid allocation by dynamically varying the local-to-borrowable channel ratio according to the changing traffic conditions. Each channel has a different adjustable probability of being borrowed and is ranked with respect to this probability. This way, channels with high probability are more likely to be borrowed. The MSC determines and updates each channel's probability of being borrowed based on the traffic conditions, by using an adaptive algorithm.

In [18], directional channel locking is used to increase the number of channels available for borrowing. In [19], rearrangements are used when a local channel becomes available.

In [21], a channel sharing method for cellular communications, called Channel Borrowing Without Locking (CBWL), is presented. In CBWL, a channel can be borrowed only from an adjacent cell. The Borrowed channel is used with reduced transmitted power such that co-channel interference caused by channel borrowing is no worse than that of a non-borrowing scheme. Borrowed channels can be accessed only in part of the cell. The authors in [22] present CBWL with cut-off priority for calls that arise in the cell. This scheme discourages excessive channel lending and borrowing at high traffic load and promotes a more uniform grade of service throughout the service area.

### *3.2. Dynamic allocation strategies*

Fixed Channel Allocation schemes are simple, however they do not adapt to traffic conditions and user distribution. In order to overcome these deficiencies, Dynamic Channel Allocation (DCA) strategies [23] have been introduced.

Dynamic channel allocation policies provide system operators with flexible planning procedure capable of adapting the allocation of resources available at base stations according to environmental conditions.

In DCA, cells have no channels themselves but refer all call attempts to the MSC, which manages all channel allocation in its region. All channels are kept in a central pool and are assigned dynamically to radio cells as new calls arrive in the system. After a call is completed, its channel is returned to the central pool.

In general, more than one channel might be available to be assigned to a cell that requires a channel, the MSC must apply some strategy to select the assigned channel. The MSC evaluates the cost of using each candidate channel and the channel with the minimum cost, provided that certain interference constraints are satisfied, is selected. The selection of the cost function is what differentiates dynamic channel allocation schemes. The cost function depends on several criteria such as:

1. the future blocking probability in the vicinity of the cell,
2. the usage frequency of the candidate channel,
3. the reuse distance,
4. channel occupancy distribution under current traffic conditions,
5. radio channel measurements of individual mobile users, or
6. the average blocking probability of the system.

In the literature numerous adaptive resource allocation schemes have been proposed [24–26] which can be classified into two broad categories, namely traffic-adaptive and interference-adaptive algorithms, according to their adaptability to either traffic or interference.

The basic principle of dynamic channel allocation algorithms, both traffic and interference adaptive, is to make radio channels available everywhere for every call by means of smart rules capable of minimizing mutual interference among all different active calls.

Traffic-adaptive schemes are based on the a-priori knowledge of the mutual interference among cells; so an off line process to evaluate the statistics of carrier to interference ratio is necessary. This process is feasible only in macrocellular environment whereas with small cells it is difficult due to deep and rapid changes in the propagation conditions. Furthermore this off line process must be repeated if a new cell has to be added due to network development.

DCA can be classified either as call-by-call DCA or adaptive DCA schemes. In the call-by-call DCA, the channel assignment is based only on current channel usage conditions in the service area, while in adaptive DCA the channel assignment is adaptively carried out using information on the previous as well as present channel usage conditions. DCA schemes can be also divided into centralized and distributed schemes with respect to the type of control they employ. Examples of centralized DCA are First Available [27] and Locally Optimized Dynamic Assignment [12,18].

In [28] a dynamic channel assignment algorithm with rearrangement for cellular mobile communication systems is suggested. They claim that their DCA algorithm is both traffic and interference adaptive.

Simulation and analysis results show that under low traffic intensity, DCA strategies performs better [20,27,29]. However, FCA schemes become superior at high offered traffic, especially in the case of uniform traffic.

### *3.3. Hybrid allocation strategies*

DCA schemes provide flexibility and traffic adaptability. However, DCA strategies are less efficient than FCA under high load conditions. To overcome this drawback, hybrid allocation [20] techniques were designed by combining FCA and DCA schemes.

In HCA, the total number of channels available for service is divided into fixed and dynamic sets. The fixed set contains a number of nominal channels that are assigned to cells as in the FCA schemes and, in all cases, are to be preferred for use in their respective cells. The second set of channels is shared by all users in the system to increase flexibility. When a call requires service from a cell and all of its nominal channels are busy, a channel from the dynamic set is assigned to the call. The channel assignment procedure from the dynamic set follows any of the DCA strategies presented in the previous subsection.

The ratio of fixed to dynamic channels is a significant parameter that defines the performance of the system. In general, the ratio of fixed to dynamic channels is a function of the traffic load and would vary over time according to offered load distribution estimations.

Another hybrid scheme is the ”fixed and dynamic channel assignment”. It is a combination of FCA and DCA that tries to achieve the lowest blocking rate of each technique depending on traffic intensity. In low traffic intensity the DCA scheme is used; in heavy traffic situations the FCA strategy is used.

### *3.4. Exploiting overlapping cells*

In a real wireless cellular network, cells are not disjoined. There is an inevitable overlap among neighboring cells due to their non-regular shapes. The schemes presented so far do not exploit this fact. In this section, we present schemes that take advantage of cells' overlapping.

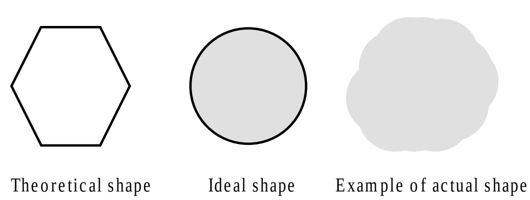


Figure 7. Theoretical, ideal and actual cell coverage shapes

Eklundh [30] proposed a variation of FCA called directed retry (DR). It is a channel-sharing scheme that exploits the overlap that exists among cells. In directed retry if a user is in an overlapping area, and finds its first-attempt cell has no free channels, it can look for free radio channels in more than one base if the BS can provide sufficient signal quality.

Directed retry preserves the merits of FCA and at the same time, reduces the blocking probability of a cell by increasing its channel utilization. The improvement is accomplished at the expense of an increased number of handoffs and an increased level of co-channel interference. [31] stated that the use of directed retry, is expected to cause only a minimum amount of additional load in handoff processing and has only a minimal effect on the probability of handoff failure.

In [32], the authors present a hybrid scheme (CBWL-DR) that has the advantages of CBWL (see section 3.1.3) and directed retry such as high channel capacity and simple channel management. The scheme also overcomes their disadvantages such as high system processing loads and intracell handoff. The simulation results presented in [32] show that in comparison with FCA, the proposed scheme can significantly improve system performance.

In [33], the authors propose a dynamic load-balancing scheme for the channel allocation problem in a cellular mobile environment. Their load-balancing scheme migrates unused channels from underloaded cells to an overloaded one. Detailed simulation experiments are carried out and the performance of the proposed load balancing scheme is compared with the fixed channel assignment, simple borrowing, and two existing strategies with load balancing (e.g. directed retry and CBWL), and a significant improvement of the system behavior is noted.

Another selective handoff scheme for traffic balancing, called directed handoff (DH), based on the concept of FCA and overlapping cells is proposed in [34]. If the traffic of a cell increases temporarily such that all or almost all of its channels are in use, those schemes may direct some of the calls currently in progress in the cell to attempt handoff to the appropriate adjacent cells. The schemes attempt to redistribute calls in heavily loaded cells to lighter loaded cells. Therefore, in the case of a temporarily traffic increase calls can be distributed to adjacent cells which share the overlapping area.

Simulation results in [34] show that selective handoff scheme improves traffic performance under the condition of uniformly distributed traffic, and enhances the spectrum utilization in the time domain through the handoff of mobiles in the overlapped areas of the cell. In [35] the performance of both the DR and DH schemes was compared with the MP dynamic scheme that provides an upper bound in the performance of DCA schemes. The conclusions reached by simulations in [35] were that both schemes improve the efficiency of the system.

The above schemes are expected to improve system performance. This improvement depends on the percentage of overlapping between adjacent cells. The wider the cell overlapping, the more traffic performance is expected to improve.

### *3.5. Channel Allocation in the Event of Handoff*

In a cellular radio network, a cell must deal with two types of calls, new calls and handoff calls. New calls refer to calls that originate within the cell, while handoff calls refer to ongoing calls that are transferred from one cell to another due to the mobility of the portables. The handoff procedure has an important effect on the performance of the system.

From the point of view of a mobile user forced termination of an ongoing call is less desirable than blocking a new call. Consequently, it is important to limit the probability of forced call termination. For this purpose, the system must reduce the chances of unsuccessful handoffs by reserving some channels explicitly for handoff calls. Many channel assignment strategies that allocate channels to handoff requests more readily than new calls have been proposed in previous works [13,36–39]. Those schemes provide improved performance at the expense of a reduction in the total admitted traffic and an increase in the blocking probability of new calls.

It is important to study the new call and handoff blocking probabilities because these two quantities affect the QoS in cellular networks. The first determines the fraction of new calls that are blocked, while the second is closely related to the fraction of admitted calls that terminate prematurely due to dropout. A good evaluation of the measures of performance can help a system designer to make its strategic decisions concerning cell size and frequency planing.

In [40] the authors present a model that captures the differences between new call blocking and handoff blocking. They demonstrated that the usual assumptions made in the literature which do not differentiate between the new call blocking probability and the handoff blocking probability may be incorrect. They found that the difference between the two kinds of blocking is particularly significant when the users move fast or when the cells are very small.

Under Complete Sharing Policy (CSP), no distinction is made between new calls and handoff calls for channel assignment. However, from a subscriber's perspective, abrupt termination of an ongoing call is more disruptive than getting a busy signal. Therefore, several policies with priority to handoff calls have been proposed [13,36,38,39,41].

One of the priority-oriented policies that is quite well known is the Cut-off Priority Policy (CPP) [13,38,42,43] or the guard channel scheme [39]. Reserving a certain number of channels, also known as guard channels ensures priority to handoff calls. Under CPP, a new call is accepted only if the total number of calls in progress, regardless of their type, is below a cut-off value and a free channel is available.

While the effectiveness of a fixed number of guard channels has been demonstrated under stationary traffic conditions, with non stationary call arrival rates in a practical system, the achieved handoff call blocking probability may deviate significantly from the desired objective. The authors in [44] propose a dynamic guard channel scheme, which adapts the number of guard channels in each cell according to the current estimate of the handoff call arrival rate. The latter is derived from the current number of ongoing calls in neighboring cells and the mobility pattern. The pursued goal is to keep the handoff call blocking probability close to a given target while constraining the new call blocking probability under given level.

Another policy that gives priority to handoff calls is the Threshold Priority Policy (TPP) [45]. Under both TPP and CPP, a handoff call is accepted as long as a channel is free. However, under TPP, a new call is accepted only if the number of new calls in progress is below a threshold value and a free channel is available. [46] compares TPP with CSP and CPP using major performance measures including new call blocking probability, handoff call blocking probability, forced termination probability, and the call non-completion probability. Several numerical experiments are done for various values of offered load and portable mobility. The results are compared using performance metrics which mainly take into account the trade-off between new call blocking probability and the forced termination probability. The numerical results show that when the offered load is high, TPP may be better alternative to CPP with respect to the considered performance metrics.

Other prioritizing schemes allow either the handoff to be queued or new calls to be queued until new channels are obtained in the cell. Several variations of the basic cutoff priority scheme, with queuing of handoff requests or of new call requests, have also been discussed in the literature [39,44]. The guard channel concept can be used in FCA or DCA schemes.

The basic queuing discipline in queuing handoff requests is first-in-first-out (FIFO). Other researchers tried to improve the performance of the handoff queuing scheme by modifying the queuing discipline. In [43], a non-preemptive priority queuing discipline based on a mobile's subscriber measurement was used for queuing handoffs. A handoff request is ranked according to how close the mobile stands to, and possibly how fast it is approaching, the receiver level. The simulation and analysis results in [43] indicate that the proposed scheme offers a better performance in terms of quality of service and spectrum efficiency.

In [47] the authors use a fuzzy logic controller to generate proper guard channels according to the fluctuating handoff traffic pattern so as not to waste or not to lack the valuable resources.

The limited availability of radio frequency spectrum will require future wireless systems to use more efficient and sophisticated resource allocation methods to increase network capacity and support multimedia users. A way to achieve this goal is to combine channel assignment with power assignment and base station assignment. The next subsections will deal with those aspects.

### *3.6. Macrocell/Microcell overlays*

In microcellular systems, frequent handoffs are very common. In overlaying macrocellular scheme a cluster of microcells are grouped together and covered by a macrocell. The total wireless resource is divided between the macrocell and all the microcells in its domain.

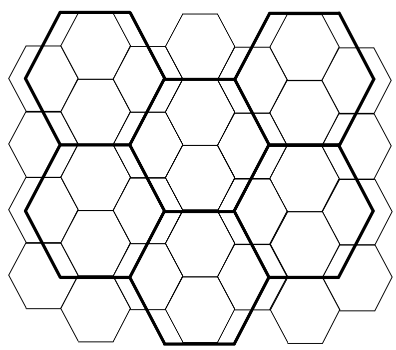

Figure 8. An example of macrocells and microcells

Fast moving users are generally encouraged to join macrocells whereas slow users typically join microcells. This jointly reduces the number of handoffs and increases the total system capacity. Also, in case of congestion, if there are not enough microcell channels for handoff calls, then macrocell channels can be used. Later, if an appropriate channel becomes available in a microcell, the macrocell channel is released and the call is handed off to the microcell channel. Several channel assignment schemes for macrocell/microcell overlays system have been proposed [48,49]. The macro and micro cells can use either the same or different channel access schemes [50]. For instance, the macro cells can use CDMA while the micro cells are using TDMA. Combining Macrocells and microcells increases the capacity of the system without increasing the handoff rates. It is, however, achieved at the expense of an increased complexity of the channel allocation and assignment schemes.

### *3.7. Power Management and Control*

The radio connection comprises three dimensions: the allocated channel (channel allocation), allocated power (power assignment procedure), and allocated location (base-station assignment procedure). The three problems are interrelated, and can be considered jointly.

Power control schemes play an important role in spectrum and resource allocation in cellular networks. Since the CIR (Carrier-to-Interference Ratio) at a wireless terminal is directly proportional to the power level of the desired signal and inversely proportional to the sum of the power of co-channel interferers, one can use power control schemes to achieve the required CIR level. By increasing the transmitted power of the desired signal and/or decreasing the power level of interfering signals the CIR level can be accommodated.

Power control schemes try to reduce the overall CIR in the system by measuring the received power and increasing (or decreasing) the transmitted power in order to maximize the minimum CIR in a given channel allocation of the system. This can result in a dramatic increase of overall system capacity measured in terms of the number of mobiles that can be supported.

Power control can be done in either centralized or distributed fashion. Centralized power control schemes require a central controller that has complete knowledge of all radio links and their power levels in the system [51,52]. In the distributed approach [53,54], each wireless terminal adjusts its transmitter's power level based on local measurements.

In [55] the authors propose a joint resource allocation algorithm that makes the channel, base station and power assignment attempting to minimize the number of channels needed to provide each user in the system with an acceptable radio connection. The algorithm operates in a wireless network with an arbitrary number of base stations and mobiles. It is compared, in terms of the achievable traffic capacity, with some bounds on the performance of the maximum packing, clique packing, and reuse partitioning techniques. The latter bounds are usually used as benchmarks on the capacity that can be achieved by any traffic-adaptive dynamic channel assignment strategy, where the quality is guaranteed by the reuse distance. The performance results confirm the improvement expected from the integration of the channel, base station and power assignments.

The channel assignment problem in cellular networks (when the powers and base stations are preassigned and fixed) has been shown to be equivalent to a generalized graph-coloring problem [17,27], which is known to be NP complete. Since no efficient algorithm that solves this problem exists, many heuristic channel assignment algorithms with varying complexity have been proposed and evaluated in the literature.

Existing studies on the channel assignment problem mostly focus on developing efficient heuristic algorithms [56,57], owing to its NP-completeness [58]. The solutions thereby generated are often less than satisfactory and usually there is no way of determining how far they are from the optimum [59]. Studies, as [60], on deriving lower bounds on the numbers of frequencies required for frequency assignment problem's have thus arisen out of their capacity of indirectly checking the quality of the assignment solutions at hand.

Power management is a more general problem than power control. One objective of current research is to develop efficient and low power consumption algorithms and protocols. Power management in wireless mobile networks is expected to gain more importance when multimedia applications will be deployed due to the extra processing needed such as data encoding/decoding and transcoding.

### *3.8. Code assignment*

As wireless communications evolve towards third generation, wideband CDMA (W-CDMA) appears to be the de facto network access scheme. This sub-section overviews code spreading and code assignment schemes adopted within UMTS/IMT-2000 framework.

CDMA techniques allow many users to transmit simultaneously in the same band without substantial interference by using approximately orthogonal spread-spectrum waveforms. A CDMA spread spectrum signal is created by modulating the radio frequency signal with a spreading sequence known as a pseudo-noise digital signal. The receiver uses a locally generated replica pseudo noise code to separate the desired coded information from all possible signals.

If the spreading signals are orthogonal then the interference of any other signal on the wanted signal is theoretically zero after despreadings. Some good examples of orthogonal codes are Hamadard/Walsh codes [61,62]. However, when these codes are misaligned, the cross-correlation of these codes will be non-zero, that is, they are no longer orthogonal.

For CDMA to eliminate the other channels' interference, the local despreading code is required to be accurately aligned with the arriving wanted code. This is a costly requirement for practical systems where the communication terminals are constantly moving.

For this reason non-orthogonal codes are used. However, to reduce the interference from other channels to a minimum, the cross-correlation between any pair of codes of the code set at any time shift should be small. Examples of non-orthogonal codes are Gold codes and Kasami codes [62–64].

UMTS/IMT-2000 adopts multiple spreading or two-layered spreading in such a way that whenever synchronization between channels can be easily maintained orthogonal codes are used, and when asynchronous transmission is favored in terms of system implementation and deployment, non-orthogonal codes are used.

Multiple spreading is realized in the following way: all the channels originated from a single transmitter can be readily synchronized; therefore orthogonal codes are used to separate these channels. These channels are then linearly combined and multiplied by a transmitter-specific non-orthogonal code.

Since CDMA uses codes instead of frequency bands or time slots to distinguish between multiple wireless conversations, code assignment is an important issue.

Almost all the channel allocation schemes presented in the above sub-sections can be applied to do code allocation taking into consideration that two users can not use the same code, at least in the same cell, and that the codes are not infinite, so spatial reuse of codes is utilized.

The provision of wireless multimedia applications to mobile users has an impact on channel access and allocation schemes. Indeed, multimedia applications such as a video application often require a dynamic amount of bandwidth. For example, compressed video sources produce a variable bit rate (VBR) traffic with a high degree of burstiness. Also a multimedia user can ask for more than one channel, which further complicates the channel assignment problem. Guaranteeing quality of service for such VBR applications over wireless links necessitates specific resource management mechanisms.

## 4. Mobility Management

Mobility Management is a key issue in wireless mobile networks. It encompasses the functionalities that enable the network to track the current location of every user and deliver incoming calls. Another key functionality that is closely related to mobility management is the handling of handoffs in order to ensure the continuity of a mobile user connection.

### *4.1. Location Management*

When an incoming call is received for a mobile station, the call has to be routed to the cell in which the mobile is located so it can be set-up. One way to locate the mobile would be to transmit a calling message (page) for the mobile on every cell site in the network (as was the case for the first generation wireless networks such as the North American Advanced Mobile Phone Service (AMPS) [9]). Clearly with several hundreds of cells and many thousands of mobiles, the required signaling capacity would be too high. Because of the radio bandwidth inefficiency of such a policy, second generation networks, such as Pan-European Digital Cellular (GSM) [65] and North American Digital Cellular [66], employ a different approach. For example, the GSM [67] and IS-41 [68] cellular standards use Home Location Registers (HLR) and Visitor Location Registers (VLR) to implement mobile registration and tracking. Moreover, the cellular network is usually split up into a number of location areas in order to further reduce the signaling traffic.

The network location areas are either fixed or dynamically determined. In fixed-location area schemes, the mapping between location areas and cells is fixed (based on traffic and mobility patterns). In dynamic or adaptive location tracking schemes, the mapping of location areas to cells is adjusted according to terminal movement, as well as in response to other events such as call pattern changes or system failures.

It is desirable to select location-updating and terminal-paging policies that can minimize the total cost (wireless bandwidth and terminal power). Registrations are only performed in the border cells of a location area. When location areas are large the location update traffic is reduced but the paging message load is increased. Inversely small location areas save on paging cost at the expense of increased location update traffic. A method for computing the optimal location area size given the respective costs for location update and terminal paging is introduced in [69].

The following subsection enumerates different mobile users location update methods. Since the geographic location update method is used in most evolving second-generation cellular systems, it will be studied in section 4.1.2.

#### *4.1.1. Mobile users location update methods*

There are four basic location update methods:

1. Geographic: A user updates the system when it enters a new location area.
2. Timer: The user updates the location periodically every given average time period.
3. Stimulus: The user performs a location update only when requested.
4. ON/OFF: A location update occurs only after the mobile powers on and before it powers down.

Three other sophisticated location update schemes are studied in [70]: Time-based, movement-based and distance-based. Under these three schemes, location updates are performed based respectively, on the time elapsed, the number of movements performed, and the distance traveled, since the last location update. A distance based location update scheme is introduced in [71] where an iterative algorithm generates the optimal threshold distance resulting in a minimum cost. However, the number of iterations required for the algorithm to converge varies widely depending on the considered mobility and call arrival probabilities. The time required to locate a mobile terminal is directly proportional to the distance traveled by the mobile terminal since its last location update. A dynamic location update mechanism is introduced in [72] where the location update time based on data obtained on-line is dynamically determined. In [73], paging subject to delay constraints is considered. In [74], a location management scheme that combines a distance based location update mechanism with a paging scheme that guarantees a pre-defined maximum delay requirement is presented.

*4.1.2. Geographic-based location update of mobile users*

In this type of location update methods, the cellular network is split up into a number of location areas, each with its own area identity number. This number is then transmitted regularly from all base stations in the area as part of the system's control information. A mobile station when not engaged in a call, will lock onto the control channel of the nearest base station upon switching on and, as it moves in the network, will from time to time select a new base station to lock onto. The mobile will check the area identity number transmitted by the base station, and when it detects a change indicating that it has moved to a new location area, it will automatically inform the network of its new location by means of a signaling exchange with the base station. This way, the network can keep a record (registration) of the current location area of each mobile, and hence will call the mobile (find) within that area only.

Several studies [75–77] indicated that the signaling traffic due to find and registration operations is significant. When the frequency of incoming calls is high compared to terminal mobility (i.e., the rate that a terminal moves to new registration areas), a location cache scheme [78,79] has been proposed to reduce the traffic (i.e., to reduce the number of find operations). When the call frequency is low compared to terminal mobility, the number of registration operations increase, which significantly degrades the performance of the location cache scheme. Several algorithms such as the pointer forwarding algorithm [77,80] and the alternative location algorithm [81] have been proposed to reduce the registration cost. Xie and Goodman [82] propose the use of a gateway VLR to limit mobility management traffic to the metropolitan area where a mobile currently resides.

While the HLR/VLR concept reduces the signaling traffic on the radio link over first generation approaches, numerous studies [75,76] indicate that it incurs a tremendous increase of signaling traffic in the fixed network. [76] estimates that around 50% of the time people don't have their phones on. Since at least 50% of all mobile-terminated call attempts fail there is inefficiency in the HLR/VLR based approach. As an alternative, the reverse virtual call set-up (RVC) algorithm is proposed in [83]. In this algorithm a mobile-terminated call is routed from the fixed phone to a switch with the capability to query the HLR. However, the switch does not set-up the call in the usual manner, but instead, waits for a signal to arrive from the called party and then bridges the two connections. Bridging functionality has already been implemented in a public cordless telephone system trial [84]. But it is necessary to consider the billing implications of such a strategy. The authors claimed that the fixed network signaling load and the call set-up delay under RVC are reduced except when the location area is smaller than a VLR area, which is unlikely to happen as they propose the use of an overlaid paging in conjunction with RVC.

The locality of reference patterns is exploited in [85]. The notion of working set for mobile hosts is introduced. Nodes in a mobile host's working set communicate with the mobile host more frequently than nodes that are not in the working set. A location management scheme has been described in [85] in which a mobile host can dynamically determine its working set depending on the call-to-mobility ratio between network node and mobile host pairs. Nodes in the working set are informed about the location update when a mobile host moves, while other nodes have to search for the mobile host when they wish to communicate with it.

In [86], some base stations are designated as reporting centers (similar to location servers). Location update is done when a mobile host moves into the cells corresponding to the reporting centers. When a mobile host has to be located, it is searched for in vicinity of the reporting center at which the last update was made. However, an issue that needs to be addressed is how such a reporting center is determined. One simple solution would be to probe all reporting centers to determine the one with the latest update. However, this imposes high communication overhead on the fixed network.

### *4.2. Multimedia handoff schemes*

With the spectacular growth of mobile telephony and the success of second-generation mobile cellular systems paralleled by the staggering growth of the Internet, mobile multimedia services are expected to be in high demand by mobile wireless users on a global scale. It is anticipated that the traffic in next generation high-speed wireless networks will be mostly generated by personal multimedia applications. Examples of such applications are: fax, video-on-demand, news-on-demand, WWW browsing, and traveler information systems. Therefore, next generation wireless networks are expected to provide adequate support for multimedia services and support user mobility with extended geographical coverage.

Multimedia applications can require a dynamic amount of bandwidth. To guarantee quality of service for such bandwidth-greedy applications when used over a wireless link, some of the traditional schemes used for supporting voice services should be reviewed and specific resource management solutions must be proposed. Handoff is a time-critical feature in wireless mobile communications that has to be addressed to provide seamless multimedia communications under changing radio resource conditions.

Handoff ensures the continuity of a call, or any relationship between the mobile terminal and the network, while the dedicated radio resource changes within one cell (intra-cell handoff) or during cell crossing (inter-cell handoff), within or outside of a switch control area. Handoff is related to access, radio resources, and network control. It has a significant impact on system capacity and performance.

Effective and reliable handoff is highly desirable from the subscriber's point of view. The handoff is already a key process in current systems and it is foreseen to gain increasing importance in third and fourth generation cellular systems as cells radius is going to decrease and the number of users is expected to grow dramatically.

A prevalent underlying theme is the techniques used to control the handoff of users as they move between shrinking cells, at greater speeds, and with stricter requirements on both the QoS delivered to the user and the operational costs associated with a connection. Minimizing the expected number of handoffs minimizes the switching load. Another concern is delay. If handoff does not occur quickly, the QoS may degenerate below an acceptable level. Combining macrocells and microcells further complicates the control of a handoff.

A handoff process is either mobile station-triggered or network-triggered. Basically, it involves four successive phases:

1. measurement;
2. handoff initiation/control;
3. channel assignment; and
4. network connection reconfiguration possibly involving rerouting and buffering.

Measurements concern mainly the signal strength and allow deciding the initiation of a handoff. Among the handoff initiation approaches analyzed in the literature, there are:

1. Relative signal strength: chooses the strongest received base station at all times.
2. Relative signal strength with threshold: initiates a handoff only if the current signal is sufficiently weak (less than a threshold) and the other is the stronger of the two.

3. Relative signal strength with hysteresis: initiates a handoff only if the new base station is sufficiently stronger (by a hysteresis margin) than the current one.

4. Relative signal strength with hysteresis and threshold: initiates a handoff only if the current signal level drops below a threshold and the target BS is stronger than the current one by a given hysteresis margin.

5. Prediction techniques that base the handoff decision on the expected future value of the received signal strength.

Huang [87] introduces a factor called glue-point, in the design of a handoff protocol in order to keep data dependency and presentation continuity. The glue-point delimits the boundary of two consecutive media units. Using the concept of glue-point, handoff can occur at glue-points but no where else.

Tabbane [88] proposes three target cell decision algorithms for handoffs, which are based on the combination of three criteria. The first criterion is the path loss criterion [89] that is used in GSM to choose the target BS. The second criterion is the capacity criterion, which allows to take into account resource availability. The last criterion used when selecting the target cell is the cell type criterion. Both the type of the current cell and the type of target cell (pico, micro or macro cell) are taken into account during the decision making process.

Channel assignment in the event of a handoff has been introduced in section 3 as part of the review of existing channel allocation schemes.

The last phase of the handoff process dealing with network connection reconfiguration is described in the following together with a classification of handoff schemes with respect to connection reconfiguration in both connection-oriented and connectionless networks. Wireless Asynchronous Transfer Mode (WATM) based networks are particularly emphasized for the potential they offer to support future mobile multimedia communications.

The handoff process can be classified according to the way the new path is set up and the connection shifted from the actual to the new path. Figure 9, Figure 10 and Figure 11 illustrate the three main classes of handoff, namely hard handoff, seamless handoff, and soft handoff.

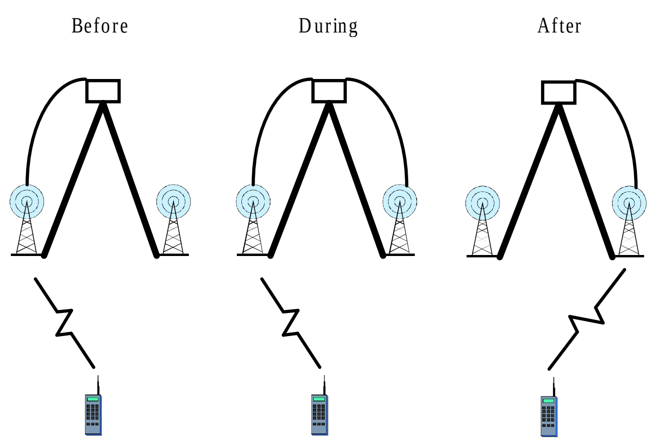


Figure 9. Hard handoff

In hard handoff, the mobile terminal has to change radio channel (frequency) to the new path with possibly a short interruption of the connection in progress. The new path is constructed in advance through the network in such a way that interruption is as short as possible. Switching to the new path and rerouting of the carried information are performed simultaneously.

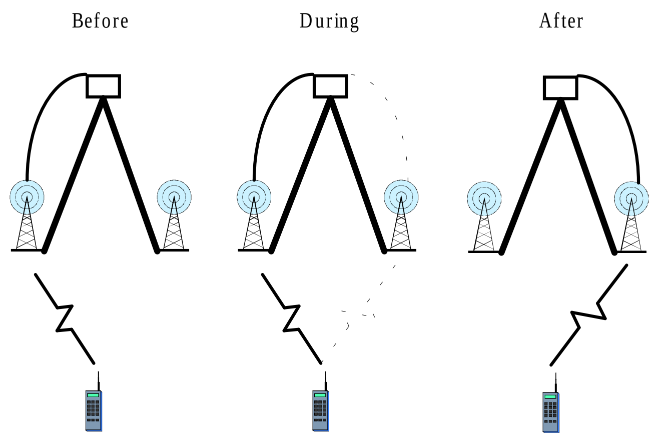


Figure 10. Seamless Handoff

In a seamless handoff process, the new path is established in parallel with the old one and the mobile terminal transmits the information flow on both paths. The active path remains the old one for a while. Then, the new stream is activated (through a switching action in the network), the old one is stopped and the old links are released.

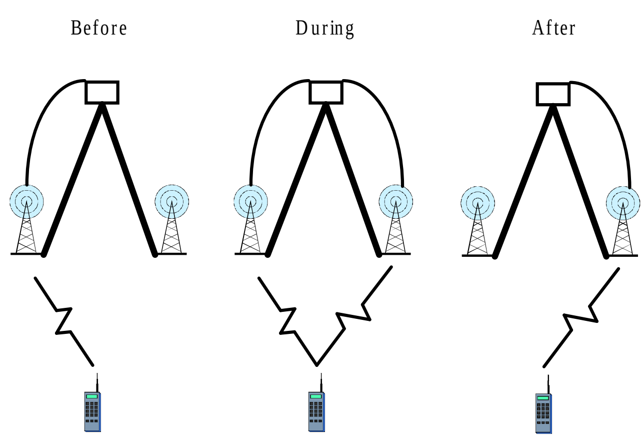


Figure 11. Soft handoff

In a soft handoff there are two active paths and two corresponding streams, at least for a while. The asynchronous behavior of the handoff process makes time a non-critical element. The mobile terminal is simultaneously connected with two (or more) base stations and both streams are considered on the mobile and network side to recover a single information flow. CDMA based wireless networks are examples of systems using soft handoff.

Hierarchical and low-latency handoff schemes have been proposed for connection-oriented networks such as WATM networks. One of the main objectives of WATM is to support multimedia communications characterized by varying connection bandwidth and Quality of Service (QoS) requirements. Some connections mainly require low delay and delay variation (e.g., video traffic) while others require very low cell loss (e.g., data traffic). The 'goodness' of a handoff impacts on the end-to-end QoS that can be provided to mobile users.

Handoff in a Wireless ATM network requires changes in virtual connections. Several handoff schemes have been proposed, such as connection extension [90], full establishment [91], partial re-establishment [91] and multicast establishment [91,92].

The full establishment approach requires the setup of a completely new connection between the end users. Figure 12 represents the path evolution for a mobile terminal that moves within three cells in a network adopting the full establishment approach.

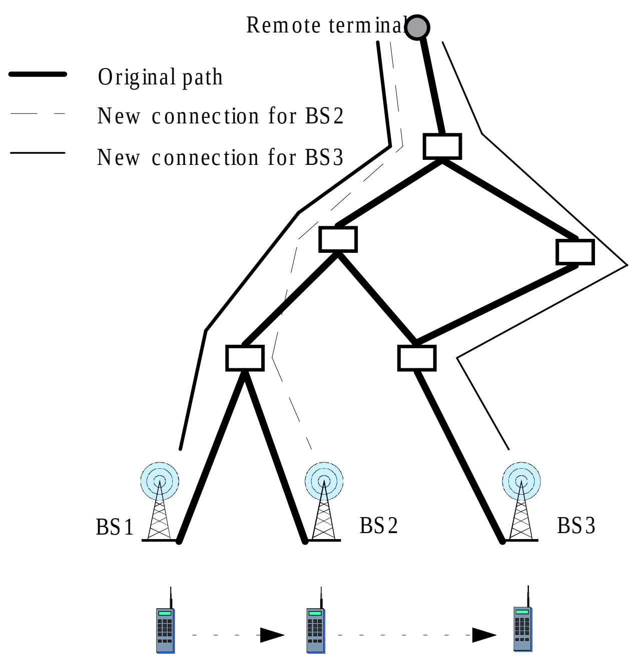


Figure 12. path evolution example adopting full establishment approach

When the handoff occurs, the connection extension scheme extends the path between the users by adding one hop that provides the connection from the old base station to the new base station through the fixed network. Figure 13 shows the path modifications needed to follow the roaming terminal.

The multicast establishment approach preallocates resources in the network portion surrounding the cell where the user is located. When a new mobile connection is established, a set of virtual connections, named a virtual connection tree, is created. The latter reaches all BSs managing the cells toward which the mobile terminal might move in the future. Thus, the mobile user can freely roam in the area covered by the tree without invoking the network call admission control during handoff. The allocation of the virtual connection tree may be static [93] or dynamic [94] during the connection lifetime. Figure 14 shows a multicast establishment, assuming that the mobile terminal roams through three cells.

Acampora et al. [93] propose building a virtual connection tree covering base stations in a local area, with virtual circuits pre-established from the root of the tree to each base station. In this case, a handoff only involves the switching to an already established virtual circuit.

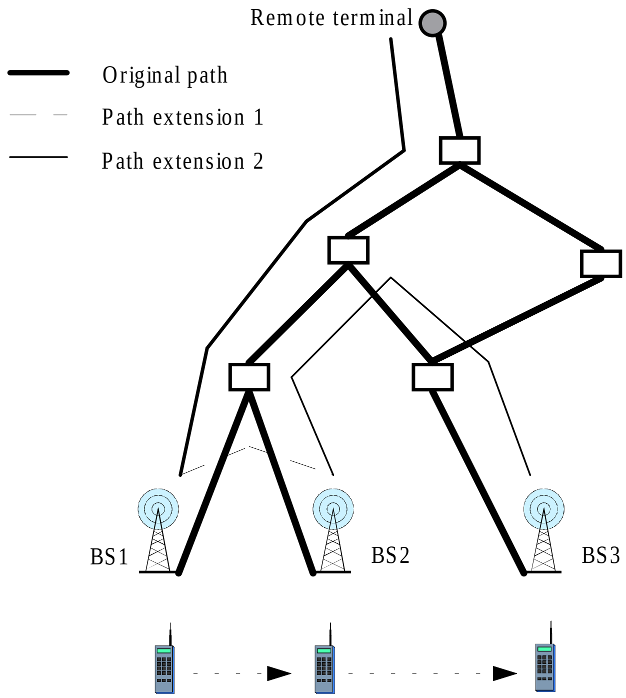


Figure 13. path evolution example adopting connection extension approach

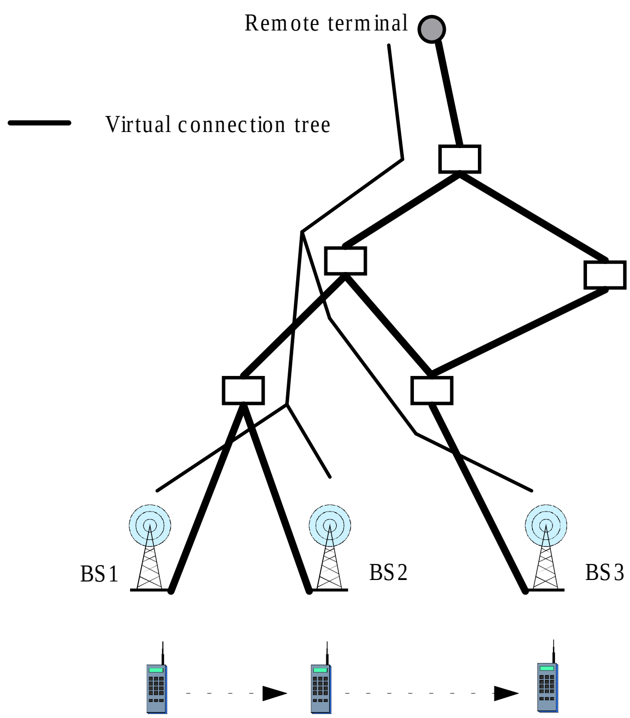


Figure 14. path evolution example adopting multicast establishment approach

In the partial re-establishment handoff scheme (see Figure 15), a new path is established from the new base station to a node in the original connection path. Hence this method requires the discovery of the crossover switch (COS), the setting up of the new partial path and the tearing down of the old partial path. The purpose of crossover switch discovery is to locate a suitable COS so that a new partial path can be established from the new base station (BS) to the COS.

Five COS discovery schemes (loose select, prior path knowledge, prior path resultant optimal, distributed hunt and backward tracking) have been presented and evaluated in [95]. Another COS discovery scheme is presented in [96].

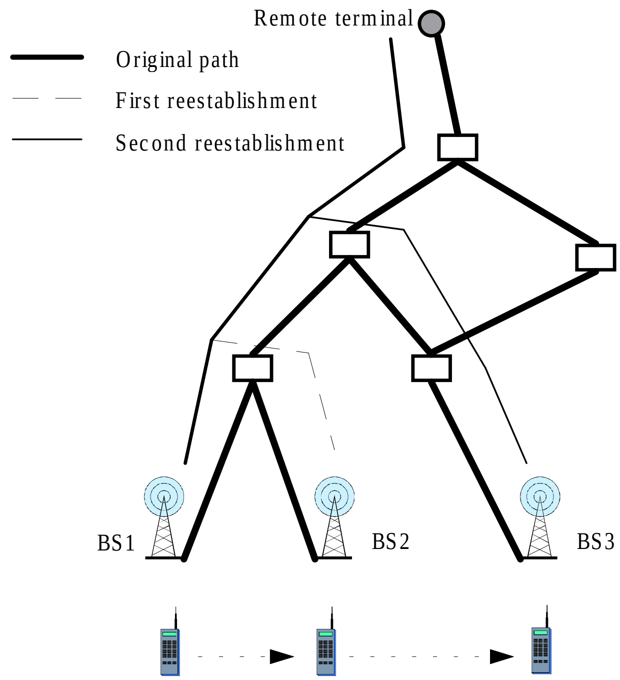


Figure 15. path evolution example adopting partial reestablishment approach

In loose select discovery, the new BS establishes a new path towards the destination node with no regard to existing path nodes. If the new path 'intersects' with the old path, the intersecting node is the crossover node. In prior path knowledge discovery, information about existing path nodes (i.e. possible COS) is obtained by querying the connection server. With this information, the new BS derives all possible partial paths and selects the partial path that fulfills QoS requirements.

The prior path resultant COS discovery is derived from prior path knowledge discovery with the exception that the considered nodes are only those supporting new paths either shorter or equal to the existing path prior to handoff. In distributed hunt COS discovery, the new BS initiates a broadcast to locate all possible COSs within the network. Again, among all possible partial paths, the partial path that fulfills QoS requirements is selected. Finally, in backward tracking discovery, each node in the existing path will progressively verify if it can reach both old and new BSs. Backtracking starts at the closest node to the old BS and continues along the existing path until a crossover node is found.

A route chaining technique is used in [97] to extend virtual circuits to the mobile's new location. To obtain a more efficient route, the previous chains are collapsed into a new virtual circuit. Another scheme based on extending and collapsing virtual circuits is proposed by Agrawal et al. in [98].

Several works used clustering to structure the wireless environment and hence to facilitate the handling of handoffs. For instance, Toh [95] groups into clusters adjacent cells in a wireless network. Movement within a cluster is handled in a single ATM switch that is connected to all the BSs in the cluster, while movement between clusters is handled by rerouting virtual circuits at the last switch on the route common to both clusters.

Akyol [99] proposes a procedure called "Nearest Common Node Rerouting (NCNR)" to perform the rerouting of user connections due to handoff event. The NCNR attempts to perform the rerouting for a handoff at the closest ATM network node that is common to both base stations involved in the handoff transaction. The rerouting is then performed starting from this common node.

Hierarchical and low-latency handoffs have been also proposed for connectionless networks. For instance, DeSimone and Nanda describe in [100] the scheme used in Cellular Digital Packet Data (CDPD). In this scheme, base stations are grouped and connected to a single Mobile Data Intermediate Station (MDIS), which handles the roaming between base stations within the corresponding group.

A scheme involving a mobile's home agent and multicasting is proposed by Seshan in [101]. The mobile's home agent encapsulates data destined for the mobile in multicast packets, and sends these packets to multiple BSs in close vicinity of the mobile. Among the previous BSs, only one actively forwards packets to the mobile. The other BSs buffer recent packets and quickly forward them to the mobile should a handoff occur. This scheme allows reducing handoff latency and packet losses at the expense of increased use of buffer space in the BSs. The use of multicast relieves the home agent of detailed knowledge of the mobile's current location, but incurs the complexity of managing multicast groups as mobiles move.

Subir Kumar et al. [102] propose a concept of per mobile software entity, known as mobile representative. A representative acts as an intermediate agent between its client mobile unit and other remote peer entities. Representatives can dynamically migrate themselves, following their client mobile users. A connection handoff is dealt with by a caching scheme where a bank of connection segments are dynamically created and destroyed, following the mobility pattern of an end user. During a handoff those segments are appropriately switched in and out to maintain the desired connectivity.

Some of the above schemes attempt to anticipate handoffs by using measurements of signal strength and knowledge of previous mobility patterns. The performance metrics usually used to evaluate handoff schemes are: Call blocking probability, Handoff blocking probability, Handoff probability, Call dropping probability, Probability of an unnecessary handoff, Rate of handoff, Duration of interruption and Delay.

Jiang et al. [103] propose a handoff support approach which uses message between the user and the network to handle handoff. By classifying the user's requirements of mobility support according to user states, this scheme asks the user to declare the type of mobility support in the call request and requires the network to inform the user in advance if a handoff fails. Simulation results presented in [103] show that this approach accepts more calls and reduces call-dropping probability thus improving resource utilization in comparison with the traditional approach.

Iraqi et al. [104] propose a configurable multi-agents architecture for QoS control in WATM. The agents dynamically manage the buffer space at the level of a switch and interact to reduce the cell loss ratio while guaranteeing a bounded transit delay. A dynamic reconfiguration of the agents is performed during partial reestablishment handoffs in order to continue meeting user QoS requirements end-to-end.

## 5. QoS-sensitive resource management

Following the enormous trend in wireless communication, there is a growing interest in how to deploy the system network infrastructure to achieve maximum capacity, coverage and flexibility in the most cost-efficient manner. To maximize the number of subscribers within a cellular network while maintaining overall quality of service, one can apply a range of techniques, each smoothly boosting the total network capacity a little further to meet the market demand.

Creating a multi-layered network by adding micro- and picocells is one such technique that allows for capacity optimization. Methods such as power control, efficient channel allocation, and traffic control between the layers are also employed to exploit the full potential of the network. In [105], the authors discuss some important aspects of creating high-capacity cellular networks that operate with a limited amount of frequency spectrum. Macrocells are initially used to build a network with cost-effective wide-area coverage. By decreasing site-to-site distance and tightening frequency reuse, the capacity of the macro network can be increased substantially. Subsequently, micro- and picocells are added, creating a multi-layered cellular network. Picocells are introduced to efficiently handle even higher traffic in indoor environments such as offices.

For QoS-sensitive communications such as multimedia communications, the goal of an efficient resource management technique is to guarantee the QoS of ongoing connections, while at the same time the available radio spectrum is used efficiently. This goal addressed by network operators and service providers into two main steps: firstly by partitioning the resources; and secondly by controlling the access and usage of these resources.

The bandwidth in a wireless network is perhaps the most precious and scarce resource of the entire communication system. This resource should be used in the most efficient manner. Several bandwidth partitioning strategies that allocate bandwidth "fairly" for different traffic classes while attempting to achieve maximum network throughput have been proposed in the literature [106,4,107]. Previous studies of these techniques in wireless networks have focused on the co-existence of data and voice traffic, while video traffic has generally been ignored.

At the two extremes of resource partitioning strategies are the Complete Sharing (CS) and Complete Partitioning (CP) (also called Mutually Restricted Access or MRA) strategies, and in between are the rest, generally referred as hybrid strategies.

In CS, all traffic classes share the entire bandwidth. Although trivial to enforce, the main drawback of this strategy is that a temporary overload of one traffic class results in degrading the connection quality of all other classes. In CP, bandwidth is divided into distinct portions with each portion corresponding to a particular traffic class. CP is wasteful of bandwidth if the predicted bandwidth demand for a particular traffic class is greater than the actual bandwidth demand. A compromise between CS and CP is a strategy in which bandwidth is allocated dynamically to match the varying traffic load. One such technique is called Priority Borrowing (PB). A moving boundary exists between the bandwidth allocated for the various traffic classes and priority users (usually voice traffic) are allowed to borrow bandwidth from non-priority users (data traffic). It has been shown that this hybrid scheme provides better performance than both CS and CP, over a range of offered loads both in micro-cellular and macro-cellular environments [4]. The reader is referred to [106,108] for a survey of such schemes.

In [109], the authors have extended Priority Borrowing to include static bandwidth reservation with a moving boundary. They call it Priority Sharing with Restrictions (PSR). In this scheme only real-time video connections are allowed to make lifetime (or static) reservations; voice and data connections make reservations dynamically. The amount of bandwidth reserved for static reservations is determined by their "Smart Allocate Algorithm".

Finding the optimal partitioning point is a very difficult task since it can be directly modeled by an NP-complete graph-coloring problem. Because of the algorithmic intractability of finding the exact optimum, various sub-optimal solutions have been proposed in the literature [108]. In fact, the intractability of finding the optimum is present already in the simplest situation, when the traffic consists of voice calls only and the statistics of the offered traffic class are completely known. The problem becomes even more difficult when the wireless network carries integrated non-homogeneous traffic, a situation occurring naturally in wireless multimedia networks.

Wireless mobile networks are characterized by user/terminal mobility. Therefore, access control to resources (referred to as call admission control) must be implemented both at the initial call set-up and during handoffs. Hence, a base station may need to reserve resources, even if this means denying access to a mobile terminal requesting admission to the network, in order to keep enough resources to support active users currently outside of its coverage area, but who may soon emigrate to its cell. Base stations must maintain a balance between the two conflicting requirements:

1. Maintain maximum resource (bandwidth) utilization.
2. Reserve enough bandwidth resources so that the maximum rate of unsuccessful incoming handoffs (due to insufficient resources) is kept below an acceptable level.

The probability of unsuccessful handoffs can be established in terms of a quality-of-service (QoS) metric, e.g., call dropping probability that the network agrees to maintain.

A method for a base station to maintain a certain call dropping probability is to reserve some resources. An accurate determination of the amount of resources that a base station must reserve is likely to become an issue of extreme importance in future wireless networks. In contrast to current wireless mobile systems, future ones will support a wide range of applications with diverse bandwidth requirements. Also, in future systems the demand on wireless bandwidth within a cell may change abruptly in a short period of time, as for example, when several video or high data rate users enter or leave a cell at the same time. In contrast, in current systems the bandwidth demand usually varies gradually, and hence it is much easier to handle. Moreover, future wireless networks may provide customized QoS parameters on a per call and/or on a service base, enabling users to select a level of service according to the pricing plan.

In order to maintain an acceptable call dropping probability, several schemes [17,110–112] have been proposed to dynamically organize the allocation of bandwidth resources. These schemes consider only limited information from neighboring cells, and do not specifically consider admission control policies as means to prevent congestion. Issues and relationships between resource reservation, channel allocation, call admission, and traffic intensity have been studied previously [16,113]. Admission control policies, which determine the number of new voice or data users for acceptance in a packet radio network, are given in [113]. For these policies, voice users are accepted only if a long-term blocking probability is not exceeded, while data users are accepted only if the mean packet delay and the packet loss probability are maintained below certain levels. In [16], a flexible channel allocation scheme is proposed based on the analysis of offered traffic distributions or blocking probabilities. A distributed call admission control procedure is proposed in [114]. It takes into consideration the number of calls in adjacent cells as well as in the cell where a new call request is being made, in addition to the knowledge of the mean call arrival, call departure, and call handoff rates. None of these schemes consider the individual trends of the users in the wireless network, e.g., position, speed, direction, and bandwidth demands.

David A. Levine et al. [115] use the shadow cluster concept [116] to estimate future resource requirements and to perform call admission decisions in wireless networks. The framework of a shadow cluster system can be viewed as a message system where mobile terminals inform the base stations in their neighborhood about their requirements, position, and movement parameters. With this information, base stations predict future demands, reserve resources accordingly, and admit only those mobile terminals which can be supported adequately. Base stations determine the probabilities that a mobile terminal will be active in other cells at future times, define and maintain shadow clusters by using probabilistic information on the future position of their mobile terminals with active calls, and predict resource demands based on shadow cluster information.

The shadow cluster approach is an important work step towards a compromise between maintaining maximum resource utilization and satisfying user QoS requirements. However, the prediction capability assumes that mobile terminals use a fixed number of bandwidth units for all the call duration. This allows guaranteeing the desired QoS at the expense of non-optimal resources utilization in case of variable bit rate traffic. Also, as recognized by the authors, the proposed approach lacks fairness as it highly prioritarizes ongoing calls over new calls in the system.

In ATM networks, calls are accepted according to some call admission policy, but due to the bursty nature of non-voice traffic, the bursts from different traffic streams may coincide, and temporarily require bit rates higher than the network capability. Effective bandwidth [117–123] is one popular approach to this problem, where each type of call is assigned an effective bandwidth lying somewhere between the mean rate and the peak rate.

In [124], the authors investigate how effective bandwidth can be utilized in the modeling of CDMA cellular mobile networks. They allocate an effective bandwidth to each mobile depending on its class and its location relative to a target cell. The important contribution of this work is that the resultant network model is of similar form to a circuit-switched network and the traffic analysis for the latter can be directly applied to the CDMA cellular model.

In general existing call admission control approaches defined for wireless mobile networks have not addressed the quality of service requirement end-to-end. This will necessitate interoperability between different QoS management mechanisms supported by the heterogeneous fixed and wireless interconnected networks.

## 6. Conclusion

The last decade has witnessed a fantastic growth of mobile personal communications systems. This growth encompasses several telecommunication applications, particularly at the access network level, such as cordless phones for residential lines, wireless PBX for corporate networks, paging systems and cellular systems. For example, it is easy to predict that cellular systems will continue growing to accommodate the increasing number of mobile subscribers and the volume of traffic generated by each mobile user. The traffic will grow significantly with the arrival of Internet accesses and multimedia services requiring, consequently, higher capacity network infrastructures and more sophisticated resource allocation and management mechanisms. In this article we have discussed some of the issues and approaches related to resource management in wireless and mobile multimedia networks. The discussed issues include Access and allocation strategies; their influence on network performance; mobility management aspects including location and handoff procedures; bandwidth allocation; and network access control.

The transport of multimedia traffic over wireless mobile networks introduces specific resource management requirements, some of which have been discussed throughout the paper such as channel access and assignment, call admission control, and handoff. There are many challenging and open research issues, and we do not claim to have addressed them all. One key issue in supporting future multimedia mobile networking and QoS provisioning end-to-end is the design of resource management mechanisms capable to adapt to the network state changes and to the nature of multimedia traffic. This is a research area currently attracting a lot of attention [44,125–128]. Another key issue is the trade-off between maximizing resource utilization and reserving enough resources so as to reduce the rate of unsuccessful incoming handoffs. This is an area where research is just in its infant stages. We have tried in the following table to assess the resource management issues on which multimedia applications have an impact and to identify for each of these issues the features suitable for supporting multimedia applications.

## References


[1] R. Prasad, *Code Division Multiple Access for Wireless Personal Communication Networks*, Boston-London: Artech House, 1996.

[2] M. Progler, C. Evci, and M. Umehira, "Air interface access schemes for broadband mobile systems," in *IEEE Communications Magazine*. September 1999.

[3] J. A. Schwartz da silva and B. E. Fernandes, "The european research program for advanced mobile radio systems," in *IEEE Personal Communications*, vol. 2, pp. 14–19. February 1995.

[4] M. Schwartz, "Network management and control issues in multimedia wireless networks," in *IEEE Personal Communications*, vol. 2, pp. 8–16. June 1995.

[5] P. W. Baier, P. Jung., and A. Klein, "Taking the challenge of multiple access for third generation cellular mobile radio systems - a european view," in *IEEE Communications Magazine*, vol. 34, pp. 82–89. February 1996.

[6] M. Zeng, A. Annamalai, and Vijay K. Bhargava, "Recent advances in cellular wireless communications," in *IEEE Communications Magazine*, vol. 37, pp. 128–138. September 1999.

[7] W. C. Jakes, *Microwave Mobile Communications*, IEEE Press, 1994.

[8] W. C. Y. Lee, "New cellular schemes for spectral efficiency," *IEEE Transactions on Vehicular Technology*, vol. 6, pp. 188–92, 1987.

[9] W. C. Lee, *Mobile Cellular Telecommunications Systems*, New York: McGraw-Hill, 1989.

[10] V. H. McDonald, "The cellular concept," *Bell System Tech.*, vol. 58, no. 1, pp. 15–41, 1979.

[11] K. N. Sivarajan, *Spectrum efficient frequency assignment for cellular radio*, Ph.D. thesis, California Institute of Technology, Pasadena, May 1990.

[12] M. Zhang and T.-S. P. Yum, "The nonuniform compact pattern allocation algorithm for cellular mobile systems," *IEEE Transactions on Vehicular Technology*, vol. 40, no. 2, pp. 387–391, May 1991.

[13] S-H. Oh and D-W. Tcha, "Prioritized channel assignement in a cellular radio network," *IEEE Transactions on Communications*, vol. 40, pp. 1259–1269, July 1992.

[14] L. Anderson, "A simulation study of sonle dynamic channel assignment algorithms in high capacity mobile telecommunications system," *IEEE Transactions on Vehicular Technology*, vol. 22, pp. 210, 1973.

[15] J. S. Engel and M. Peritsky, "Statistically optimum dynamic server assignment in systems with interfering servers," *IEEE Transactions on Vehicular Technology*, vol. 22, pp. 203–209, 1973.

[16] Tajima J. and Imamura K., "A strategy for flexible channel assignment in mobile communication systems," *IEEE Transactions on Vehicular Technology*, vol. 37, 1988.

[17] S. Tekinay and B. Jabbari, "Handover and channel assignments in mobile cellular networks," in *IEEE Communications Magazine*, vol. 29. November 1991.

[18] M. Zhang and T.-S. P. Yum, "Comparisons of channel-assignment strategies in cellular mobile telephone systems," *IEEE Transactions on Vehicular Technology*, vol. 38, no. 4, pp. 211–215, November 1989.

[19] S. M. Elnoubi, R. Singh, and S.C. Gupta, "A new frequency channel assignment algorithm in high capacity mobile communication systems," *IEEE Transactions on Vehicular Technology*, vol. 31, no. 3, pp. 125–131, August 1982.

[20] Kanwa T.J. and Georganas N.D., "A hybrid channel assignment scheme in large-scale cellular structured mobile communication systems," *IEEE Transactions on Vehicular Technology*, vol. 26, April 1988.

[21] Hua Jiang and S. S. Rappaport, "Cbwl: a new channel assignment and sharing method for cellular communication systems," *IEEE Transactions on Vehicular Technology*, pp. 313–322, May 1994.

[22] Hua Jiang and S. S. Rappaport, "Prioritized channel borrowing without locking: a channel strategy for cellular communications," *IEEE/ACM Transactions on Networking*, vol. 4, no. 2, pp. 163–172, April 1996.

[23] D.C. Cox and D. O. Reudink, "Increasing channel occupancy in large-scale mobile radio systems: Dynamic channel reassignment," *IEEE Transactions on Vehicular Technology*, vol. 22, pp. 218–223, November 1973.

[24] R. Steele, J. Whitehead, and W.C. Wong, "System aspects of cellular radio," in *IEEE Communications Magazine*. January 1995.

[25] Lustin C.-I. Chuang, "Performance issues and algorithms for dynamic channel assignment," *IEEE Journal on Selected Areas in Communications*, vol. 11, no. 2, August 1993.

[26] S-H. Oh and D-W. Tcha, "Performance analysis of cellular networks," *IEEE Transactions on Communications*, vol. 39, no. 12, December 1991.

[27] D.C. Cox and D. O. Reudink, "Dynamic channel assignment in two dimension large-scale mobile radio systems," *Bell Syst. Tech.*, vol. 51, pp. 1611–28, 1979.

[28] Y. Han S.L. Kim and S.H. Hwang, "traffic and interference adaptive dca algorithm with rearrangement in microcellular systems," in *VTC'97*, 1995, vol. 1, pp. 130–134.

[29] "A comparison of some non-uniform spatial demand profiles on mobile radio system performance," *IEEE Transactions on Communications*, vol. 20, December 1972.

[30] B. Eklundh, "Channel utilization and blocking probability in a cellular mobile telephone system with directed retry," *IEEE Transactions on Communications*, vol. 34, no. 4, April 1986.

[31] T.S. P. Yum and K. L. Yeung, "Blocking and handoff performance analysis of directed retry in cellular mobile systems," *IEEE Transactions on Vehicular Technology*, vol. 44, no. 3, August 1995.

[32] Hua Jiang and S. S. Rappaport, "Hybrid channel borrowing and directed retry in highway cellular communications," *IEEE/ACM Transactions on Networking*.

[33] Sajal K. Das and Sanjoy K. Sen, "A dynamic load balancing strategy for channel assignment using selective borrowing in cellular mobile environment," *Wireless Networks*, vol. 3, pp. 333–347, 1997.

[34] T. Fujii, "Selective handover for traffic balance in mobile communications," in *ICC'92*, 1992, vol. 4, pp. 212.3.1–212.3.7.

[35] D. Everitt, "Traffic capacity of cellular mobile communications systems," *Computer Networks ISDN Sys.*, vol. 20, pp. 447–454, 1990.

[36] D. Cox and D. Reudink, "A comparison of some channel assignment strategies in large mobile communication systems," *IEEE Transactions on Communications*, vol. 20, pp. 190–195, December 1972.

[37] A. S. Acampora and M. Naghshineh, "Control and quality of service provisioning in high-speed micro- cellular networks," in *IEEE Personal Communications*, vol. 1. 1994.

[38] D. Hong and S. Rappaport, "Traffic modeling and performance analysis for cellular mobile radio telephone systems with prioritized and nonprioritized handoff procedures," *IEEE Transactions on Vehicular Technology*, vol. 35, pp. 77–92, 1986.

[39] R. Guerin, "Queueing blocking system with two arrival streams and guard channels," *IEEE Transactions on Communications*, vol. 36, pp. 153–163, 1988.

[40] M. Sidi and D. Starobinski, "New call blocking versus handoff blocking in cellular networks," *Wireless Networks*, vol. 3, pp. 15–27, 1997.

[41] I. Katzela and M. Naghshineh, "Channel assignment schemes for cellular mobile telecommunication systems," in *IEEE Personal Communications*, pp. 10–31. June 1996.

[42] Y-B. Lin, S. Mohan, and A. Noerpel, "Pcs channel assignment strategies for handoff and initial access," in *IEEE Personal Communications*, pp. 47–56. 1994.

[43] S. Tekinay, "A measurement-based prioritization scheme for handovers in mobile cellular networks," *IEEE Journal on Selected Areas in Communications*, vol. 1, pp. 1343–1350, 1992.

[44] Yu O.T.W. and Leung V.C.M., "Adaptive resource allocation for prioritized call admission over an atm-based wireless pcn," *IEEE Journal on Selected Areas in Communications*, vol. 15, no. 7, pp. 1208–1225, September 1997.

[45] B. Gavish and S. Sridhar, "Priority-oriented control policies for channel assignment in cellular networks," Tech. Rep. 96-SR-1, November 1996.

[46] B. Gavish and S. Sridhar, "Threshold priority policy for channel assignment in cellular networks," *IEEE Transactions on Communications*, vol. 46, no. 3, March 1997.

[47] C. Cho, Y. Ko, and H. Lee-Kwang, "Fuzzy adaptive guard channel assignment strategy for handoff in pcs system," in *IEEE International Conference on Fuzzy Systems*, 1997, vol. 3, pp. 1511–1517.

[48] K. Ivanov and G. Spring, "Mobile speed sensitive handover in a mixed cell environment," in *VTC'95*, Jully 1995, pp. 892–896.

[49] M. Benveniste, "Cell selection in two-tier microcellular/macrocellular systems," in *GLOBECOM'95*, November 1995.

[50] G. P. Pollini, "Trends in handover design," in *IEEE Communications Magazine*, pp. 82–90. March 1996.

[51] S. Grandhi et al., "Centralized power control in cellular radio systems," *IEEE Transactions on Vehicular Technology*, vol. 42, no. 4, 1993.

[52] J. Zander, "Performance of optimum transmitter power control in cellular radio systems," *IEEE Transactions on Vehicular Technology*, vol. 41, no. 1, 1992.

[53] J. Zander, "Distributed cochannel interference control in cellular radio systems," *IEEE Transactions on Vehicular Technology*, vol. 41, no. 1, 1992.

[54] D. Mitra, "An asynchronous distributed algorithm for power control in cellular radio systems," in *WINLAB Wksp. on Third Generation Wireless Info. Networks*, October 1993.

[55] S. Papavassiliou and L. Tassiulas, "Improving the capacity in wireless networks through integrated channel base station and power assignment," *IEEE Transactions on Vehicular Technology*, vol. 47, no. 2, May 1998.

[56] F. Box, "A heuristic technique for assigning frequencies to mobile radio nets," *IEEE Transactions on Vehicular Technology*, vol. 27, no. 2, pp. 57–64, 1978.

[57] K. N. Sivarajan, R. J. McEliece, and J. K. Ketchurn, "Channel assignment in cellular radio," in *VTC'89*, 1989.

[58] W. K. Hale, "Frequency assignment: Theory and applications," in *IEEE*, 1980, vol. 68, pp. 1497–1514.

[59] A. Gamst, "Some lower bounds for a class of frequency assignment problems," *IEEE Transactions on Vehicular Technology*, vol. 35, no. 1, pp. 8–14, 1986.

[60] D.W. Tcha, Y.J. Chung, and T.J. Choi, "A new lower bound for the frequency assignment problem," *IEEE/ACM Transactions on Networking*, vol. 5, no. 1, February 1997.

[61] F. Adachi, M. Sawahashi, and K. Okawa, "Tree-structured generation of orthogonal spreading codes with different lengths for forward link ds-cdma mobile radio," in *Electronic Letters*, vol. 33, pp. 27–28. January 1997.

[62] E.H. Dinan and B. Jabbari, "Spreading codes for direct sequence cdma and wideband cdma cellular networks," in *IEEE Communications Magazine*, pp. 48–54. September 1998.

[63] R. Gold, "Optimal binary sequences for spread spectrum multiplexing," *IEEE Transactions on Information Theory*, vol. 13, pp. 619–621, October 1967.

[64] D.V. Sarwate and M.B. Pursley, "Crosscorrelation properties of pseudorandom and related sequences," in *IEEE*, May 1980, vol. 68, pp. 598–619.

[65] "Digital cellular telecommunications system (phase 2+); principles of telecommunication services supported by a gsm public land mobile network (plmn)," Tech. Rep. GTS GSM 02.01 V5.0.0 (1995-12), 1995, DGTS/SMG-010201Q.

[66] "Electronics industry association/ telecommunications industry association (eia/tia) interim standard 54," Tech. Rep., 1992.

[67] M. Rahnema, "Overview of the gsm system and protocol architecture," in *IEEE Communications*, vol. 31. April 1993.

[68] "Electronics industry association/ telecommunications industry association, cellular radio-telecommunications inter-systems operations," Tech. Rep. IS-41, January 1993.

[69] H. Xie, S. Tabbane, and D. Goodman, "Dynamic location area management and performance analysis," in *IEEE VTC'93*, May 1993, pp. 536–539.

[70] A. Bar-Noy, I. Kessler, and M. Sidi, "Mobile users: To update or not to update?," *ACM-Baltzer Journal of Wireless Networks*, April 1995.

[71] U. Madhow, M. L. Honig, and K. Steigliz, "Optimization of wireless resources for personal communications mobility tracking," in *IEEE INFOCOM'94*, June 1994, pp. 577–584.

[72] I. F. Akyildiz and J. S. M. Ho, "Dynamic mobile user location update for wireless pcs networks," *ACM-Baltzer Journal of Wireless Networks*, April 1995.

[73] C. Rose and R. Yates, "Paging cost minimization under delay constraints," *ACM-Baltzer Journal of Wireless Networks*, April 1995.

[74] I. F. Akyildiz and J. S. M. Ho, "A mobile user location update and paging mechanism under delay constraints," in *ACM SIGCOMM'95*, 1995.

[75] C. N. Lo, R. S. Wolff, and R. C. Bernhardt, "An estimate of network database transaction volume to support universal personal communication services," in *1st Int. Conf. Universal Personal Communications*, September 1992.

[76] K. S. Meier-Hellstern, E. Alonso, and D. R. O'Neil, "The use of ss7 and gsm to support high density personal communications," in *ICC'92*, June 1992.

[77] T. Imielinski and B. R. Badrinath, "Querying in highly mobile distributed environment," in *18th VLDB*, 1992.

[78] R. Jain, Y.-B. Lin, C. N. Lo, and S. Mohan, "A caching strategy to reduce network impacts of pcs," *IEEE JSAC*, vol. 12, no. 8, pp. 1434–1445, 1994.

[79] Y.-B. Lin, "Determining the user locations for personal communications networks," *IEEE Transactions on Vehicular Technology*, vol. 43, pp. 466–473, 1994.

[80] R. Jain and Y.-B. Lin, "Perfomance modeling of an auxiliary user location strategy in a pcs network," *ACM-Baltzer Wireless Networks*, vol. 1, no. 2, pp. 197–210, 1995.

[81] G. P. Pollini and S. Tabbane, "The intelligent network signaling and switching costs of an alternate location strategy using memory," *IEEE Transactions on Vehicular Technology*, vol. 42, 1993.

[82] H. Xie and D. J. Goodman, "Signaling system architecture based on metropolitan area networks," Tech. Rep. WINLAB-TR-39, Rutgers University, August 1992.

[83] Chih-Lin, G. P. Pollini, and R. D. Gitlin, "Pcs mobility management using the reverse virtual call setup algorithm," *IEEE/ACM Transactions on Networking*, vol. 5, no. 1, February 1997.

[84] C. Banerjee and J. J. O'Neil, "Radio frequency coverage and network management for cordless charge-a-call service market trial," in *IEEE Int. Conf. Networks for Personal Communications (NPC 94)*, March 1994.

[85] S. Rajagopalan and B. R. Badrinath, "An adaptive location management strategy for mobile ip," in *1st ACM Mobicom 95*, November 1995.

[86] A. Bar-Noy and I. Kessler, "Traking mobile users in wireless communication networks," in *IEEE INFOCOM*, 1993, pp. 1232–1239.

[87] C.M. Huang and M.Y. Jang, "Handoff architectures and protocols for transmitting compressed multimedia information in mobile pcss," *IEEE Transactions on Consumer Electronics*, vol. 43, no. 3, pp. 784–794, August 1997.

[88] S. Tabbane, "Evaluation of handover target cell determination algorithms for heterogeneous cellular environments," in *ICC'97*, 1997, pp. 396–400.

[89] S. T. S. Chia and R. J. Warburton, "Handover criteria for city microcellular radio systems," in *40th IEEE VTC'90*, May 1990.

[90] P. P. Mishra and M. B. Srivastava, "Call establishment and re-routing in mobile computing networks," Tech. Rep. TM 11384-940906-13, AT&T Bell Labs., September 1994.

[91] K. Keeton, B. Mah, S. Seshan, R. Katz, and D. Ferrari, "Providing connection-oriented network services to mobile hosts," in *USENIX Symposium on Mobile and Location Independent Computing*, August 1993, pp. 83–102.

[92] R. Ghai and S. Singh, "A protocol for seamless communication in a picocellular network," in *IEEE Supercomm/ICC*, May 1994, pp. 192–196.

[93] A. S. Acampora and M. Naghshineh, "An architecture and methodology for mobile-executed handoff in cellular atm networks," *IEEE JSAC*, vol. 12, no. 8, October 1994.

[94] O. Yu and V. Leung, "Extending b-isdn to support user terminal mobility over an atm-based personal communications network," in *GLOBECOM '95*, 1995, pp. 2289–2293.

[95] C. K. Toh, "Crossover switch discovery for wireless atm lans," *ACM Journal on Mobile Networks and Applications - Special Issue on Routing in Mobile Communication Networks*, vol. 1, no. 2, pp. 141–165, October 1996.

[96] R. R. Bhat and R. Gupta, "Framework for dynamic cos discovery in wireless atm," Tech. Rep. 98-0005, ATM Forum, 1998.

[97] K. Y. Eng, M. Karol, M. Veeraraghavan, E. Ayanoglu, C. Woodworth, and R. A. Valenzuela, "A wireless broadband ad-hoc atm local-area network," *ACM-Baltzer Journal of Wireless Networks*, May 1995.

[98] P. Agrawal, P. P. Mishra, and M. B. Srivastava, "Network architecture for mobile and wireless atm," in *16th IEEE Int. Conf. on Distributed Computing Systems*, May 1996.

[99] Bora A. Akyol and Donald C. Cox, "Rerouting for handoff in a wireless atm network," in *IEEE Personal Communications*, pp. 26–33. October 1996.

[100] A. DeSimone and S. Nanda, "Wireless data: Systems, standards, services," *ACM-Baltzer Journal of Wireless Networks*, October 1995.

[101] S. Seshan, *Low-Latency Handoffs for Cellular Data Networks*, Ph.D. thesis, University of California at Berkley, March 1996.

[102] S. K. Biswas and A. Hopper, "A representative based architecture for handling mobility in connection oriented radio networks," in *ICUPC'95*, November 1999.

[103] S. Jiang, D. H.K. Tsang, and B. Li, "Subscriber-assisted handoff support in multimedia pcs," *Mobile Computing and Communications Review*, vol. 1, no. 3, September 1997.

[104] Y. Iraqi, R. Boutaba, and A.Mehaoua, "Configurable multi-agent system for qos control in watm," in *GLOBE-COM'98*, November 1998.

[105] M Madfors, K. Wallstedt, S. Magnusson, H. Olofsson, P.O. Backman, and S. Engstrm, "High capacity with limited spectrum in cellular systems," in *IEEE Communication Magazine*. August 1997.

[106] Y. H. Kim and C. K. Un, "Analysis of bandwidth allocation strategies with access restrictions in broadband isdn," *IEEE Transactions on Communications*, vol. 41, no. 5, pp. 771–781, May 1993.

[107] B. Kraimeche and M. Schwartz, "Bandwidth allocation strategies in wideband integrated networks," *IEEE JSAC*, vol. 4, pp. 869–878, September 1986.

[108] "Special issue on bandwidth allocation in atm networks," in *IEEE Communications Magazine*, vol. 35. May 1997.

[109] P. Bahl, I. Chlamtac, and A. Farago, "Optimizing resource utilization in wireless multimedia networks," in *ICC'97*, 1997, pp. 1432–1437.

[110] S. Nanda and D. J. Goodman, "Dynamic resource acquisition: Distributed carrier allocation for tdma cellular systems in third generation wireless information networks," pp. 99–124. 1992.

[111] G. Falciasecca, M. Frullone, G. Riva, M. Sentinelli, and A. M. Serra, "Investigation on a dynamic channel allocation for high capacity mobile radio systems," in *38th IEEE VTC'88*, 1988, pp. 176–181.

[112] H. Panzer and R. Beck, "Adaptive resource allocation in metropolitan area cellular mobile radio systems," in *40th IEEE VTC'90*, 1990, pp. 156–160.

[113] W.-B. Yang and E. Geraniotis, "Admission policies for integrated voice and data traffic in cdma packet radio networks," *IEEE JSAC*, vol. 12, pp. 654–664, May 1994.

[114] M. Naghshineh and M. Schwartz, "Distributed call admission control in mobile/wireless networks," in *PIMRC'95*, September 1995.

[115] David A. Levine, Ian F. Akyildiz, and Mahmoud Naghshineh, "A resource estimation and call admission algorithm for wireless multimedia networks using the shadow cluster concept," *IEEE/ACM Transactions on Networking*, vol. 5, no. 1, February 1997.

[116] D. A. Levine, I. F. Akyildiz, and M. Naghshineh, "The shadow cluster concept for resource allocation and call admission in atm-based wireless networks," in *ACM MOBICOM'95*, November 1995, pp. 142–150.

[117] D. Mitra and J. A. Morrison, "Erlang capacity and uniform approximations for shared unbuffered resources," *IEEE/ACM Transactions on Networking*, vol. 2, pp. 558–570, December 1994.

[118] J. Y. Hui, "Resource allocation for broadband networks," *IEEE JSAC*, vol. 6, pp. 1598–1608, December 1988.

[119] R. Guerin, H. Ahmadi, and M. Naghshineh, "Equivalent capacity and its application to bandwidth allocation in high-speed networks," *IEEE JSAC*, vol. 9, pp. 968–981, September 1991.

[120] F. P. Kelly, "Effective bandwidth at multi-class queues," *Queuing Systems*, vol. 9, pp. 5–16, 1991.

[121] A. I. Elwalid and D. Mitra, "Effective bandwidth of general markovian traffic sources and admission control of high speed networks," *IEEE/ACM Transactions on Networking*, vol. 1, pp. 329–343, June 1993.

[122] G. Kesidis, J. Walrand, and C. Chang, "Effective bandwidths for multiclass markov fluids and other atm sources," *IEEE/ACM Transactions on Networking*, vol. 1, pp. 424–428, August 1993.

[123] R. Griffiths and P. Key, "Adaptive call admission control in atm networks," in *International Teletraffic Congress*, June 1994, pp. 1089–1098.

[124] J. S. Evans and D. Everitt, "Effective bandwidth based admission control for multi-service cdma cellular networks," *IEEE Transactions on Vehicular Technology*, vol. 48, no. 1, pp. 36–46, January 1999.

[125] P. Bahl, "Supporting digital video in a managed wireless network," in *IEEE Communications Magazine*, pp. 94–102. June 1998.

[126] N. Davies, J. Finney, A. Friday, and A. Scott, "Supporting adaptive video applications in mobile environments," in *IEEE Communications Magazine*, pp. 138–143. June 1998.

[127] M. Naghshineh and M. W. LeMair, "end-to-end qos provisioning in multimedia wireless/mobile networks using an adaptive framework," in *IEEE Communications Magazine*. November 1997.

[128] N. Morinaga, M. Nakagawa, and R. Kohno, "New concepts and technologies for achieving highly reliable and high-capacity multimedia wireless communications systems," in *IEEE Communications Magazine*. January 1997.

| **Problem** | **affected by multimedia** | **suitable characteristics** | | | |
|---|---|---|---|---|---|
| | | **Adaptive** | **Fast** | **Scalable** | **Others** |
| Channel access | Yes | - | - | - | Flexible, high bit rate, low error rate, maximize resource utilization |
| Channel allocation | Yes | Yes | Yes | Yes | Simple |
| Channel assignment | Yes | Yes | Yes | - | Low handoff blocking probability |
| Power management and control | Yes | Yes | - | Yes | QoS oriented |
| Location management | No | - | - | - | - |
| Handoff | Yes | Yes | Yes | Yes | Low/without loss, efficient buffer space management, simple |
| Call admission control | Yes | Yes | Yes | Yes | Maximize resource utilization, guarantee QoS |