Rajagopalan, B. & Reininger, D. “Wireless ATM: QoS and Mobility Management”
Mobile Communications Handbook
Ed. Suthan S. Suthersan
Boca Raton: CRC Press LLC, 1999

Wireless ATM: QoS and Mobility Management

35.1 Introduction
35.2 QoS in Wireless ATM
ATM QoS Model • QoS Approach in Wireless ATM • MAC Layer Functions • Network and Application Layer Functions
35.3 Mobility Management in Wireless ATM
Location Management in Wireless ATM • Network Entities Involved in LM • Location Management Functions and Control Flow • Connection Handover in Wireless ATM
35.4 Summary and Conclusions
References

Bala Rajagopalan
NEC C&C Research Laboratories

Daniel Reininger
NEC C&C Research Laboratories

35.1 Introduction

Wireless ATM (WATM) refers to the technology that enables ATM end-system mobility as well as tetherless access to ATM core networks. Wireless ATM has two distinct components: the radio access technology, and enhancements to existing ATM technology to support end-system mobility. The latter component is referred to as "MATM" (mobility-enhanced ATM) and it is independent of the radio access technology. The rationale for wireless ATM has been discussed at length elsewhere [1, 2]. In this chapter, we restrict our discussion to two challenging issues in wireless ATM: the provisioning of ATM quality of service (QoS) for connections that terminate on mobile end systems over a radio link, and the protocols for mobility management in the MATM infrastructure.

Figure 35.1 illustrates the WATM reference model considered in this chapter [3]. "W" UNI in this figure indicates the ATM user-network interface established over the wireless link. "M" NNI refers to the ATM network-node interface supporting mobility management protocols. The figure depicts the scenario where an MATM network has both end system mobility-supporting ATM switches (EMAS) and traditional ATM switches with no mobility support. Thus, one of the features of mobility management protocols in an MATM network is the ability to work transparently over switches that do not implement mobility support.

35.2 QoS in Wireless ATM

The type of QoS guarantees to be provided in wireless ATM systems is debatable [4]. On the one hand, the QoS model for traditional ATM networks is based on fixed terminals and high quality links.

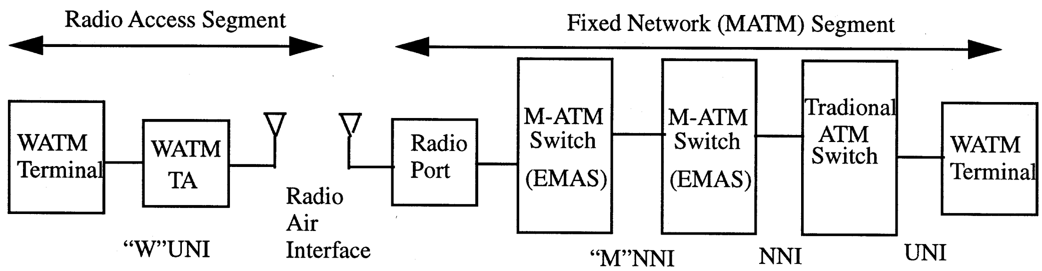

FIGURE 35.1: WATM reference model.

Terminal mobility and error-prone wireless links introduce numerous problems [5]. On the other hand, maintaining the existing QoS model allows the transparent extension of fixed ATM applications into the domain of mobile networking. Existing prototype implementations have chosen the latter approach [6]–[8]. This is also the decision of the ATM Forum wireless ATM working group [9]. Our discussion, therefore, is oriented in the same direction, and to this end we first briefly summarize the existing ATM QoS model.

35.2.1 ATM QoS Model

Five service categories have been defined under ATM [10]. These categories are differentiated according to whether they support constant or variable rate traffic, and real-time or non-real-time constraints. The service parameters include a characterization of the traffic and a reservation specification in the form of QoS parameters. Also, traffic is policed to ensure that it conforms to the traffic characterization, and rules are specified for how to treat nonconforming traffic. ATM provides the ability to tag nonconforming cells and specify whether tagged cells are policed (and dropped) or provided with best-effort service.

Under UNI 4.0, the service categories are constant bit rate (CBR), real-time variable bit rate (rt-VBR), non-real-time variable bit rate (nrt-VBR), unspecified bit rate (UBR) and available bit rate (ABR). The definition of these services can be found in [10]. Table 35.1 summarizes the traffic descriptor parameters and QoS parameters relevant to each service category in ATM traffic management specifications version 4.0 [11]. Here, the traffic parameters are peak cell rate (PCR), cell delay variation tolerance (CDVT), sustainable cell rate (SCR), maximum burst size (MBS) and minimum cell rate (MCR). The QoS parameters are cell loss ratio (CLR), maximum cell transfer delay (max CTD) and cell delay variation (CDV). The explanation of these parameters can be found in [11].

Functions related to the implementation of QoS in ATM networks are usage parameter control (UPC) and connection admission control (CAC). In essence, the UPC function (implemented at the network edge) ensures that the traffic generated over a connection conforms to the declared traffic parameters. Excess traffic may be dropped or carried on a best-effort basis (i.e., QoS guarantees do not apply). The CAC function is implemented by each switch in an ATM network to determine whether the QoS requirements of a connection can be satisfied with the available resources. Finally, ATM connections can be either point-to-point or point-to-multipoint. In the former case, the connection is bidirectional, with separate traffic and QoS parameters for each direction, while in the latter case it is unidirectional. In this chapter, we consider only point-to-point connections for the sake of simplicity.

TABLE 35.1 ATM Traffic and QoS Parameters

	ATM Service Category				
Attribute	CBR	rt-VBR	nrt-VBR	UBR	ABR
Traffic Parameters					
PCR and CDVT	Yes	Yes	Yes	Yes	Yes
SCR and MBS	N/A	Yes	Yes	N/A	N/A
MCR	N/A	N/A	N/A	N/A	Yes
QoS Parameters					
CDV	Yes	Yes	No	No	No
Maximum CTD	Yes	Yes	No	No	No
CLR	Yes	Yes	Yes	No	No

35.2.2 QoS Approach in Wireless ATM

QoS in wireless ATM requires a combination of several mechanisms acting in concert. Figure 35.2 illustrates the various points in the system where QoS mechanisms are needed:

- *At the radio interface:* A QoS-capable medium access control (MAC) layer is required. The mechanisms here are resource reservation and allocation for ATM virtual circuits under various service categories, and scheduling to meet delay requirements. Furthermore, an error control function is needed to cope with radio link errors that can otherwise degrade the link quality. Finally, a CAC mechanism is required to limit access to the multiple access radio link in order to maintain QoS for existing connections.
- *In the network:* ATM QoS mechanisms are assumed in the network. In addition, a capability for QoS renegotiation will be useful. This allows the network or the mobile terminal (MT) to renegotiate the connection QoS when the existing connection QoS cannot be maintained during handover. Renegotiation may also be combined with *soft* QoS mechanisms, as described later. Finally, mobility management protocols must include mechanisms to maintain QoS of connections rerouted within the network during handover.
- *At the MT:* The MT implements the complementary functions related to QoS provisioning in the MAC and network layers. In addition, application layer functions may be implemented to deal with variations in the available QoS due to radio link degradation and/or terminal mobility. Similar functions may be implemented in fixed terminals communicating with MTs.

In the following, we consider the QoS mechanisms in some detail. We focus on the MAC layer and the new network layer functions such as QoS renegotiation and soft QoS. The implementation of existing ATM QoS mechanisms have been described in much detail by others (for example, see [12]).

35.2.3 MAC Layer Functions

The radio link in a wireless ATM system is typically a broadcast multiple access channel shared by a number of MTs. Different multiple access technologies are possible, for instance frequency, time, or code division multiple access (FDMA, TDMA, and CDMA, respectively). A combination of FDMA and *dynamic* TDMA is popular in wireless ATM implementations. That is, each radio port

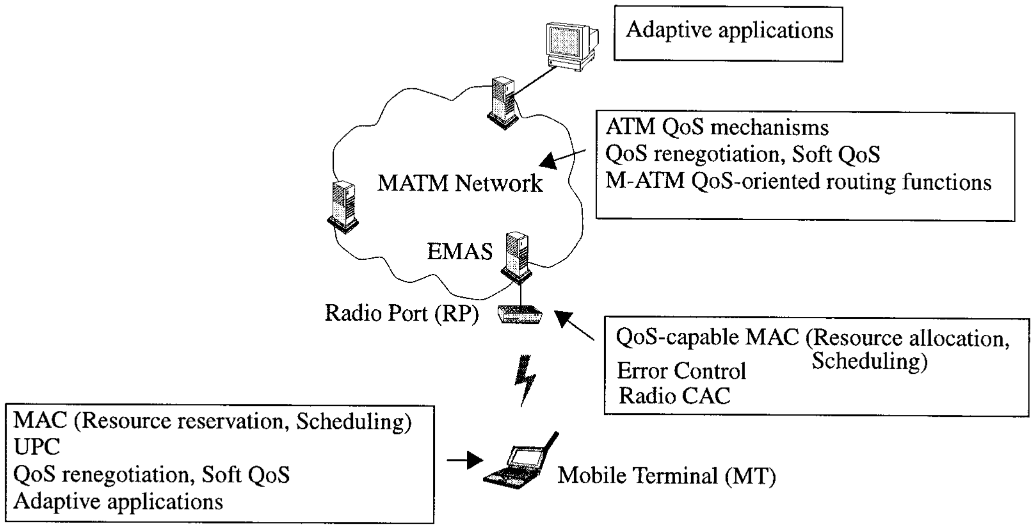

FIGURE 35.2: QoS mechanisms in wireless ATM.

(RP) operates on a certain frequency band and this bandwidth is shared dynamically among ATM connections terminating on multiple MTs using a TDMA scheme [13]. ATM QoS is achieved under dynamic TDMA using a combination of a resource reservation/allocation scheme and a scheduling mechanism. This is further explained using the example of two wireless ATM implementations: NEC's WATMnet 2.0 prototype [13] and the European Union's Magic WAND (Wireless ATM Network Demonstrator) project [8].

Resource Reservation and Allocation Mechanisms (WATMnet 2.0)

WATMnet utilizes a TDMA/TDD (time division duplexing) scheme for medium access. The logical transmission frame structure under this scheme is shown in Fig. 35.3. As shown, this scheme allows the flexibility to partition the frame dynamically for downlink (from EMAS to MTs) and uplink (from MTs to EMAS) traffic, depending on the traffic load in each direction. Other notable features of this scheme are:

- A significant portion of each slot is used for forward error control (FEC)
- A separate contention region in the frame is used for MTs to communicate with the EMAS
- 8-byte control packets are used for bandwidth request and allocation announcements. An MT can tag request packets along with the WATM cells it sends or in the contention slots
- WATM cells are modified ATM cells with data link control (DLC) and cyclic redundancy check (CRC) information added

In the downlink direction, the WATM cells transported belong to various ATM connections terminating on different MTs. After such cells arrive at the EMAS from the fixed network, the allocation of TDMA slots for specific connections is done at the EMAS based on the connections' traffic and QoS parameters. This procedure is described in the next section. In the uplink direction, the allocation is based on requests from MTs. For bursty traffic, an MT makes a request only after each burst is generated. Once uplink slots are allocated to specific MTs, the transmission of cells from multiple

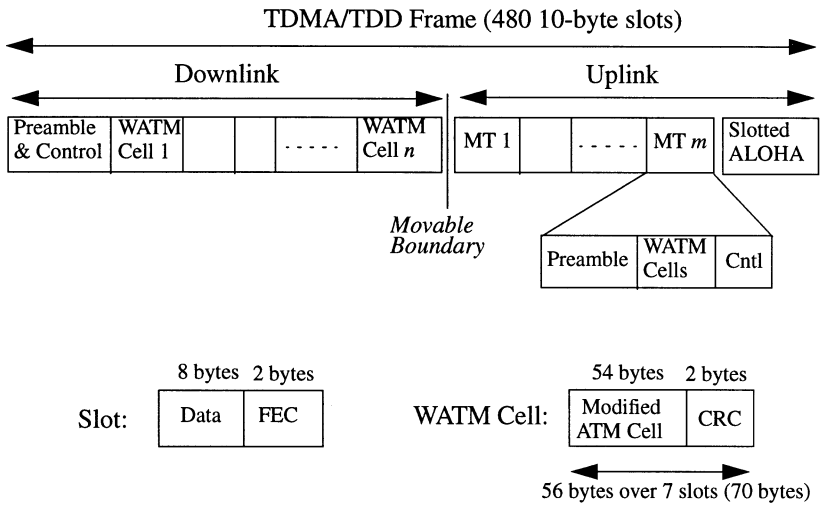

FIGURE 35.3: TDMA logical frame format.

active connections at an MT is again subject to the scheduling scheme. Both the request for uplink slot allocation from the MTs and the results from the EMAS are carried in control packets whose format is shown in Fig. 35.4. Here, the numbers indicate the bits allocated for various fields. The sequence number is used to recover from transmission losses. Request and allocation types indicate one of four types: CBR, VBR, ABR or UBR. The allocation packet has a start slot field which indicates where in the frame the MT should start transmission. The number of allocated slots is also indicated.

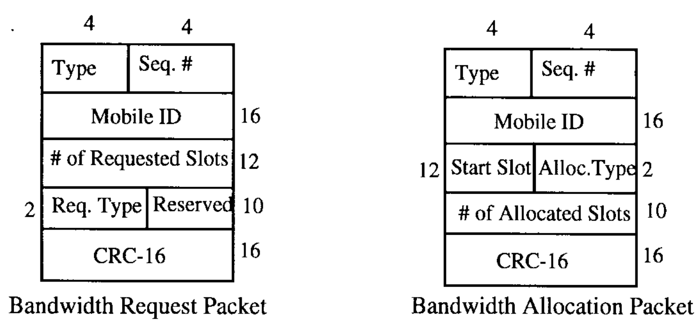

FIGURE 35.4: Bandwidth control packet formats.

The DLC layer implementation in WATMnet is used to reduce the impact of errors that cannot be corrected using the FEC information. The DLC layer is responsible for selective retransmission of cells with uncorrectable errors or lost cells. Furthermore, the DLC layer provides request/reply control interface to the ATM layer to manage the access to the wireless bandwidth, based on the instantaneous amount of traffic to be transmitted. The wireless ATM cell sent over the air interface is a modified version of the standard ATM cell with DLC and CRC information, as shown in Fig. 35.5. The same figure also shows the acknowledgment packet format for implementing selective retransmission. In this packet, the VCI field specifies the ATM connection for which the acknowledgment is being sent. The sequence number field indicates the beginning sequence number from which the 16-bit acknowledgment bitmap indicates the cells correctly received (a “1” in the bit map indicates correct reception).

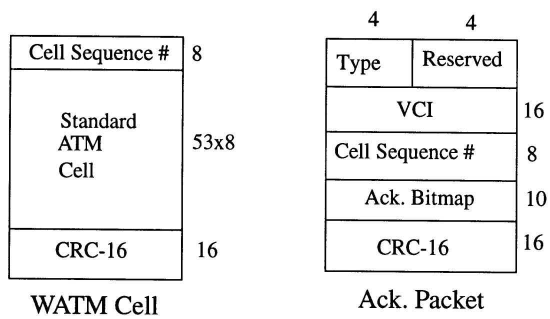

FIGURE 35.5: WATM formats at the DLC layer.

The TDMA/TDD scheme thus provides a mechanism for dynamic bandwidth allocation to multiple ATM connections. How active connections are serviced at the EMAS and the MTs to maintain their QoS needs is another matter. This is described next using Magic WAND as the example.

Scheduling (Magic WAND)

The Magic WAND system utilizes a TDMA scheme similar to that used by the WATMnet. This is shown in Fig. 35.6. Here, each MAC Protocol Data Unit (MPDU) consists of a header and a sequence of WATM cells from the same MT (or the EMAS) referred to as a *cell train*. In the Magic WAND system, the scheduling of both uplink and downlink transmissions is done at the EMAS. Furthermore, the scheduling is based on the fact that the frame length is variable. A simplified description of the scheduling scheme is presented below. More details can be found in [14, 15].

At the beginning of each TDMA frame, the scheduling function at the EMAS considers pending transmission requests, uplink and downlink, from active connections. The scheduler addresses two issues:

1. The determination of the number of cells to be transmitted from each connection in the frame, and
2. The transmission sequence of the selected cells within the frame

The objective of the scheduler is to regulate the traffic over the radio interface as per the declared ATM traffic parameters of various connections and to ensure that the delay constraints (if any) are met for these connections over this interface.

The selection of cells for transmission is done based on the service categories of active connections as well as their traffic characteristics. First, for each connection, a priority based on its service category is assigned. CBR connections are assigned the highest priority, followed by rt-VBR, nrt-VBR, ABR. In addition, for each active connection that is not of type UBR, a token pool is implemented. Tokens for a connection are generated at the declared SCR of the connection and tokens may be accumulated

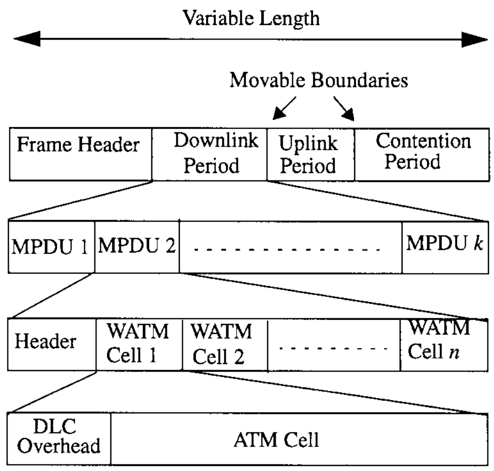

FIGURE 35.6: Magic WAND TDMA frame structure.

in the pool as long as their number does not exceed the declared MBS for the connection. The scheduler services active connections in two passes: in the first pass, connections are considered in priority order from CBR to ABR (UBR is omitted) and within each priority class only connections with a positive number of tokens in their pools are considered. Such connections are serviced in the decreasing order of the number of tokens in their pools. Whenever a cell belonging to a connection is selected for transmission, a token is removed from its pool. At the end of the first pass, either all the slots in the downlink portion of the frame are used up or there are still some slots available. In the latter case, the second pass is started. In this pass, the scheduler services remaining excess traffic in each of CBR, rt and nrt-VBR and ABR classes, and UBR traffic in the priority order. It is clear that in order to adequately service active connections, the mean bandwidth requirement of the connections cannot exceed the number of downlink slots available in each frame. The CAC function is used to block the setting up of new connections over a radio interface when there is a danger of overloading. Another factor that can result in overloading is the handover of connections. The CAC must have some knowledge of the expected load due to handovers so that it can limit new connection admissions. Preserving the QoS for handed over connections while not degrading existing connections at a radio interface requires good network engineering. In addition, mechanisms such as QoS renegotiation and soft QoS (Section 35.2.4) may be helpful.

Now, at the end of the selection phase, the scheduler has determined the number of cells to be transmitted from each active connection. Some of these cells are to be transmitted uplink while the others are to be transmitted downlink. The scheduler attempts to place a cell for transmission within the frame such that the cell falls in the appropriate portion of the frame (uplink or downlink, Fig. 35.6) and the delay constraint (CDT) of the corresponding connection is met. To do this, first the delay allowed over the radio segment is determined for each connection with a delay constraint (it is assumed that this value can be obtained during the connection routing phase by decomposing the path delay into delays for each hop). Then, for downlink cells, the arrival time for the cell from the fixed network is marked. For uplink cells, the arrival time is estimated from the time at which the request was received from the MT. The deadline for the transmission of a cell (uplink or downlink) is computed as the arrival time plus the delay allowed over the radio link.

The final placement of the cells in the frame is based on a three-step process, as illustrated by an example with six connections (Fig. 35.7). Here, Dn and Un indicate downlink and uplink cells with deadline = slot n, respectively, and D^i_j and U^i_j indicate downlink and uplink cells of the ith connection with deadline = slot j, respectively. In the first step, the cells are ordered based on their deadlines [Fig. 35.7(a)]. Several cells belonging to the same connection may have been selected for transmission in a frame. When assigning a slot for the first cell of such a "cell train" the scheduler positions the cell such that its transmission will be before and as close to its deadline as possible. Some cells may have to be shifted from their previously allocated slots to make room for the new allocation. This is done only if the action does not violate the deadline of any cell. When assigning a slot for another cell in the train, the scheduler attempts to place it in the slot next to the one allocated to the previous cell. This may require shifting of existing allocations as before. This is shown in Figs. 35.7(b)–35.7(d). Here, the transmission frame is assumed to begin at slot 5.

At the end of the first step, the cell sequencing may be such that uplink and downlink cells may be interleaved. The second step builds the downlink portion of the frame by first shifting all downlink cells occurring before the first uplink cell as close to the beginning of the frame as possible. In the space between the last such downlink cell and the first uplink cell, as many downlink cells as possible are packed. This is illustrated in Fig. 35.7(e). A slot for *period overhead* (PO) is added between the downlink and uplink portions. Finally, in the last step, the uplink cells are packed, by moving all uplink cells occurring before the next downlink cell as shown in Fig. 35.7(f). The contention slots are added after the last uplink cell and the remaining cells are left for the next frame.

Thus, scheduling can be a rather complicated function. The specific scheduling scheme used in the Magic WAND system relies on the fact that the frame length is variable. Scheduling schemes for other frame structures could be different.

35.2.4 Network and Application Layer Functions

Wireless broadband access is subject to sudden variations in bandwidth availability due to the dynamic nature of the service demand (e.g., terminals moving in and out of RPs coverage area, variable bit-rate interactive multimedia connections) and the natural constraints of the physical channel (e.g, fading and other propagation conditions). QoS control mechanisms should be able to handle efficiently both the mobility and the heterogeneous and dynamic bandwidth needs of multimedia applications. In addition, multimedia applications themselves should be able to adapt to terminal heterogeneity, computing limitations, and varying availability of network resources [16].

In this section, the network and application layer QoS control functions are examined in the context of NECs WATMnet system [6]. In this system, a concept called *soft-QoS* is used to effectively support terminal mobility, maintain acceptable application performance and high network capacity utilization. Soft-QoS relies on a QoS control framework that permits the allocation of network resources to dynamically match the varying demands of mobile multimedia applications. Network mechanisms under this framework for connection admission, QoS renegotiation, and handoff control are described. The soft-QoS concept and the realization of the soft-QoS controller based on this concept are described next. Finally, experimental results on the impact of network utilization and soft-QoS provisioning for video applications are discussed.

A Dynamic Framework for QoS Control

The bit-rates of multimedia applications vary significantly among sessions and within a session due to user interactivity and traffic characteristics. Contributing factors include the presence of heterogeneous media (e.g., video, audio, and images) compression schemes (e.g., MPEG, JPEG), presentation quality requirements (e.g., quantization, display size), and session interactivity (e.g.,

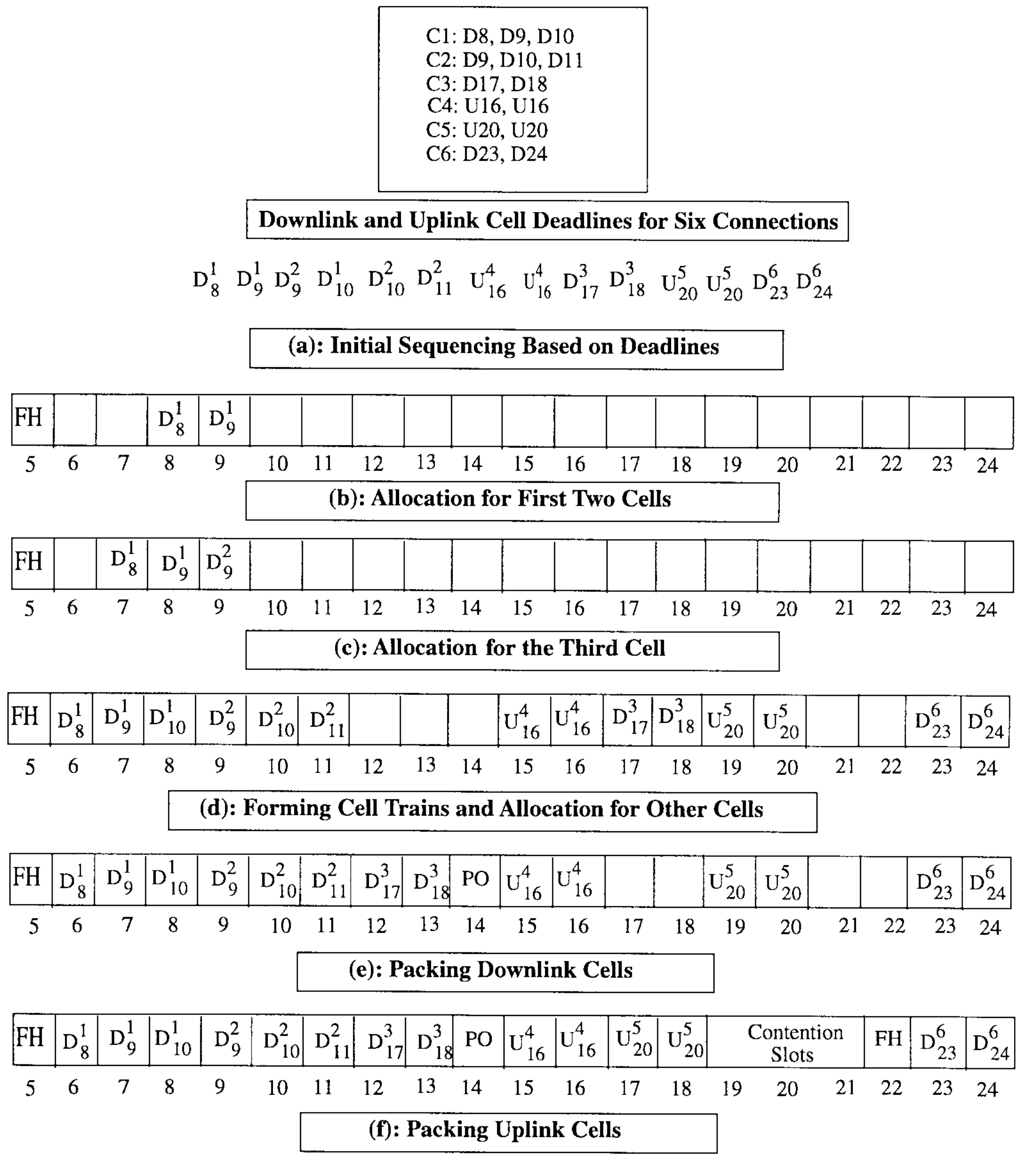

FIGURE 35.7: Scheduling example.

image scaling, VCR-like control). Consider, for example, a multimedia application using several media components or media objects (such as a multiwindow multimedia user interface or future MPEG-4 encoded video) which allows users to vary the relative importance-of-presence (IoP) of a given media object to match the current viewing priorities. In this case there would be a strong dependency of user/application interaction on the bandwidth requirements of individual media components. Figure 35.8 shows the bit-rate when the user changes the video level-of-detail (LoD) during a session. A suitable network service for these applications should support bandwidth renegotiation to simultaneously achieve high network utilization and maintain acceptable performance. For this purpose, an efficient network service model should support traffic contract renegotiation during a session. It has been experimentally verified that bandwidth renegotiation is key for efficient QoS support of network-aware adaptive multimedia applications, and that the added implementation complexity is reasonable [17, 18].

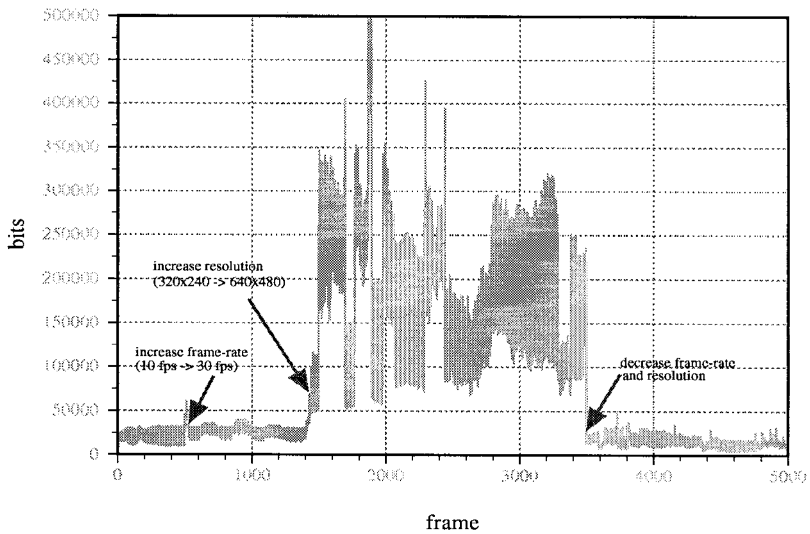

FIGURE 35.8: Video bit-rate changes on an interactive multimedia session.

In the mobile multimedia communication scenario, bandwidth renegotiation is particularly important. Conventional network services use static bandwidth allocation models that lack the flexibility needed to cope with multimedia interactivity and session mobility. These session properties enlarge the dynamic range of bandwidth requirements and make dynamic bandwidth management protocols a requirement for effective end-to-end QoS support. Renegotiation may be required during handover, as well as when resource allocation changes are warranted due to instantaneous application needs and sudden changes in network resource availability.

Figure 35.9 shows the system and API model for QoS control with bandwidth renegotiation. The application programming interface (API) between the adaptive application and the QoS control module is dynamic, i.e., its parameters can be modified during the session. For example, the Winsock 2 API under Microsoft Windows [19] allows the dynamic specification of QoS parameters suitable for the application. In addition, the API between the QoS controller and the network allows the traffic descriptor to be varied to track the bit-rate requirements of the bitstream. A new network service, called VBR$^+$ allows renegotiation of traffic descriptors between the network elements and the terminals [20]. VBR$^+$ allows multimedia applications to request "bandwidth-on-demand" suitable for their needs.

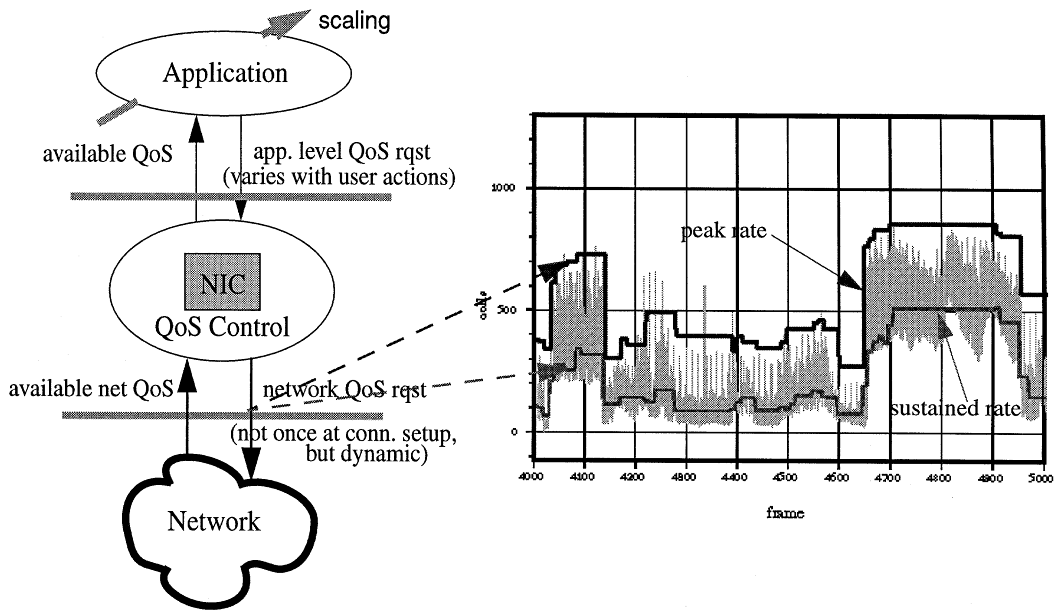

FIGURE 35.9: System and API model for QoS control with bandwidth renegotiation.

Soft-QoS Model

Although multimedia applications have a wide range of bandwidth requirements, most can gracefully adapt to sporadic network congestion while still providing acceptable performance. This graceful adaptation can be quantified by a softness profile [17]. Figure 35.10 shows the characteristics of a softness profile. The softness profile is a function defined on the scales of two parameters: satisfaction index and bandwidth ratio. The satisfaction index is based on the subjective mean-opinion-score (MOS), graded from 1 to 5; a minimum satisfaction divides the scale in two operational regions: the acceptable satisfaction region and the low satisfaction region. The bandwidth ratio is defined by dividing the current bandwidth allocated by the network to the bandwidth requested to maintain the desired application performance. Thus, the bandwidth ratio is graded from 0 to 1; a value of 1 means that the allocated bandwidth is sufficient to achieve the desired application performance. The point indicated as B is called the critical bandwidth ratio since it is the value that results in minimum acceptable satisfaction. As shown in Fig. 35.10, the softness profile is approximated by piecewise linear "S-shaped" function consisting of three linear segments. The slope of each linear segment represents the rate at which applications performance degrades (satisfaction index decreases) when the network allocates only a portion of the requested bandwidth: the steeper the slope is, the "harder" the corresponding profile is.

The softness profile allows efficient match of application requirements to network resource availability. With the knowledge of the softness profile, network elements can perform soft-QoS control—QoS-fair allocation of resources among contending applications when congestion arises. Applications can define a softness profile that best represents their needs. For example, the softness profile for digital compressed video is based on the nonlinear relationship between coding bit-rate and quality, and the satisfaction index is correlated to the user perception of quality [21, 22]. While video-on-demand (VoD) applications may, in general, tolerate bit-rate regulations within a small dynamic range, appli-

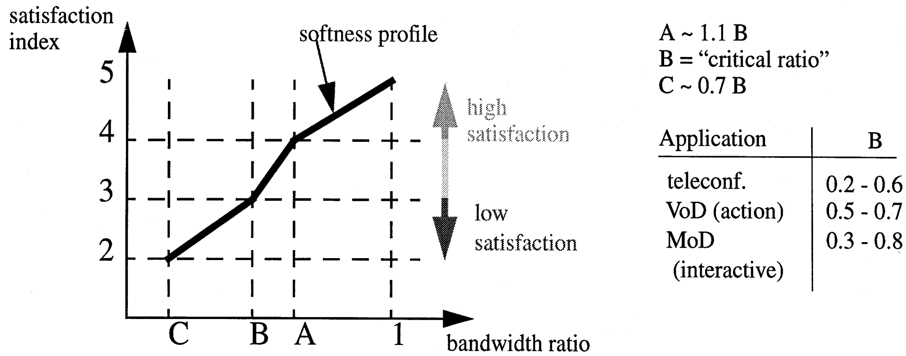

FIGURE 35.10: Example softness profile.

cations such as surveillance or teleconference may have a larger dynamic range for bit-rate control. Other multimedia applications may allow a larger range of bit-rate control by resolution scaling [18]. In these examples, VoD applications are matched to a "harder" profile than the other, more adaptive multimedia applications. Users on wireless mobile terminals may select a "softer" profile for an application in order to reduce the connection's cost, while a "harder" profile may be selected when the application is used on wired desktop terminal. Thus, adaptive multimedia applications able to scale their video quality could specify their soft-QoS requirements dynamically to control the session's cost.

Figure 35.11 conceptually illustrates the role of application QoS/bandwidth renegotiation, service contract, and session cost in the service model. The soft-QoS service model is suitable for adaptive multimedia applications capable of gracefully adjusting their performance to variable network conditions. The service definition is needed to match the requirements of the application with the capabilities of the network. The service definition consists of two parts: a usage profile that specifies the target regime of operation and the service contract that statistically quantifies the soft-QoS service to be provided by the network. The usage profile, for example, can describe the media type (e.g., MPEG video), interactivity model (e.g., multimedia browsing, video conference), mobility model (indoors, urban semi-mobile, metropolitan coverage area), traffic, and softness profiles. The service contract quantifies soft-QoS in terms of the probability that the satisfaction of a connection will fall outside the acceptable range (given in its softness profile), the expected duration of "satisfaction outage," and the new connection blocking probability.

Network resource allocation is done in two phases. First, a connection admission control procedure, called soft-CAC, checks the availability of resources on the terminals coverage area at connection set-up time. The necessary resources are estimated based on the service definition. The new connection is accepted if sufficient resources are estimated to be available for the connection to operate within the service contract without affecting the service of other ongoing connections. Otherwise the connection is blocked. Second, while the connection is in progress, dynamic bandwidth allocation is performed to match the requirements of interactive variable bit-rate traffic. When congestion occurs, the soft-QoS control mechanism (re)-allocates bandwidth among connections to maintain the service of all ongoing connections within their service contracts. The resulting allocation improves the satisfaction of undersatisfied connections while maintaining the overall satisfaction of other connections as high as possible [17]. Under this model, connections compete for bandwidth in a "socially responsible" manner based on their softness profiles. Clearly, if a cost model is not in place, users would request the maximum QoS possible. The cost model provides feedback on session

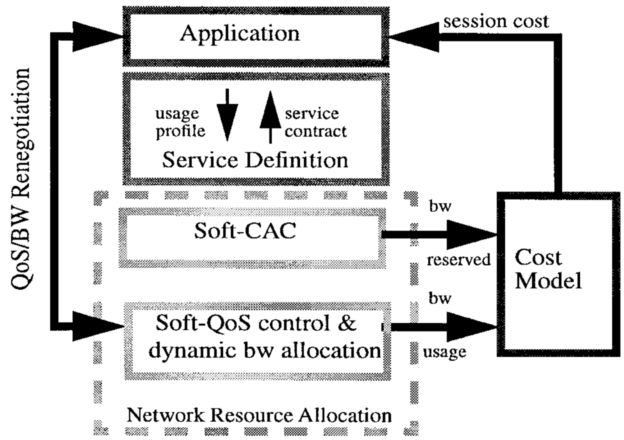

FIGURE 35.11: Service model for dynamic bandwidth allocation with soft-QoS.

cost to the applications; the user can adjust the long-term QoS requirements to maintain the session cost within budget.

Soft-QoS Control in the WATMnet System

In the WATMnet system, soft-QoS control allows effective support of mobile multimedia applications with high network capacity utilization. When congestion occurs, the soft-QoS controller at the EMASs allocates bandwidth to connections based on their relative robustness to congestion given by the applications softness profiles. This allocation improves the satisfaction of undersatisfied connections while maintaining the overall satisfaction of other connections as high as possible. Within each EMAS, connections compete for bandwidth in a "socially responsible" manner based on their softness profiles.

ATM UNI signalling extensions are used in the WATMnet system to support dynamic bandwidth management. These extensions follow ITU-T recommendations for ATM traffic parameter modification while the connection is active [23]. Although these procedures are not finalized at the time of this writing, an overview of the current state of the recommendation is given next with an emphasis on its use to support soft-QoS in the mobile WATM scenario.

ITU-T Q.2963 allows all three ATM traffic parameters, PCR, SCR, and MBS, to be modified during a call. All traffic parameters must be increased or decreased; it is not possible to increase a subset of the parameters while decreasing the others. The user who initiates the modification request expects to receive from the network a new set of traffic parameters that are greater than or equal to (or less than) the existing traffic parameters if the modification request is an increase (or decrease). Traffic parameter modification is applicable only to point-to-point connections and may be requested only by the terminal that initiated the connection while in the active state.

The following messages are added to the UNI:

- MODIFY REQUEST message is sent by the connection owner to request modification of the traffic descriptor; its information element (IE) is the ATM traffic descriptor.
- MODIFY ACKNOWLEDGE message is sent by the called user or network to indicate that the modify request is accepted. The broadband report type IE is included in the message when the called user requires confirmation of the success of modification.

- CONNECTION AVAILABLE is an optional message issued by the connection owner to confirm the connection modification performed in the addressed user to requesting user direction. The need for explicit confirmation of modification is indicated by the "modification confirmation" field in the MODIFY ACKNOWLEDGE broadband report IE.
- MODIFY REJECT message is sent by the called user or network to indicate that the modify connection request is rejected. The cause of the rejection is informed through the cause IE.

Figures 35.12, 35.13, and 35.14 show the use of these messages for successful, addressed user rejection and network rejection of modification request, respectively.

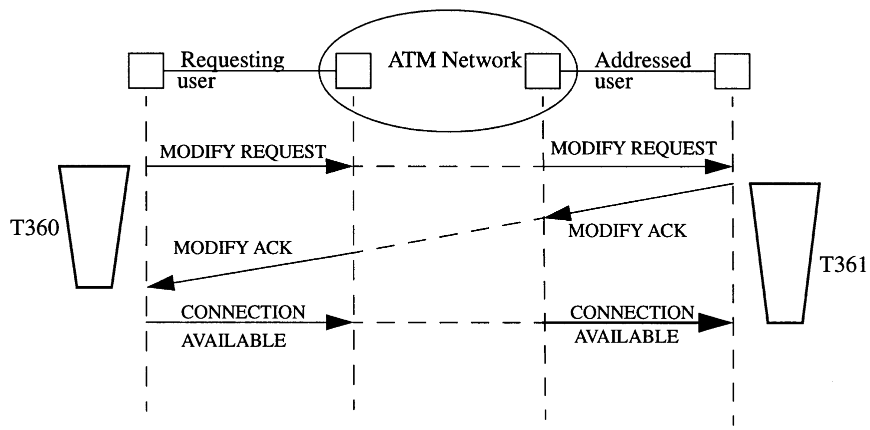

FIGURE 35.12: Successful Q2963 modification of ATM traffic parameters with (optional) confirmation.

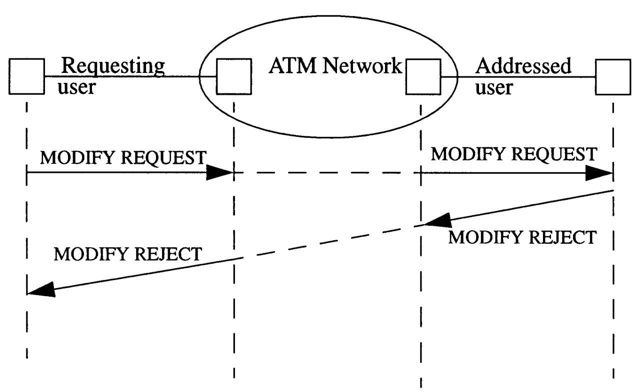

FIGURE 35.13: Addressed user rejection of modification.

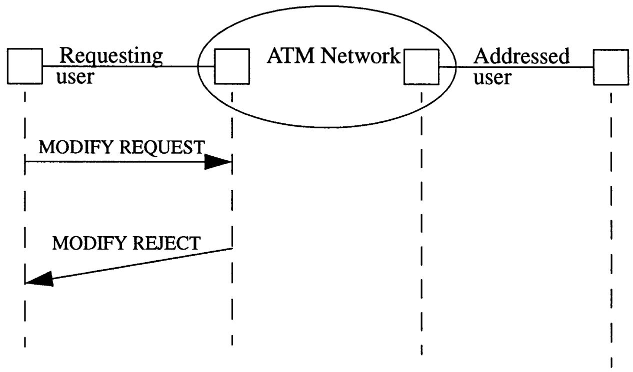

FIGURE 35.14: Network rejection of modification request.

Additionally, the soft-QoS control framework of the WATMnet system uses the following modifications to the Q2963 signalling mechanisms:

- BANDWIDTH CHANGE INDICATION (BCI) message supports network-initiated and called user-initiated modification. The message is issued by the network or called user to initiate a modification procedure. The traffic descriptor to be used by the connection owner when issuing the corresponding MODIFY REQUEST message is specified in the BCIs ATM traffic descriptor IE. Figure 35.15 illustrates the use of BCI for called user-initiated modification. Timer T362 is set when issuing the BCI message and cleared when the corresponding MODIFY REQUEST message is received; if T362 expires, the terminal and/or network element can modify the traffic policers to use the ATM traffic descriptor issued in the BCI message.
- Specification of softness profile and associated minimum acceptable satisfaction level (sat_{min}) in the MODIFY REQUEST message. The softness profile and sat_{min} are used for QoS-fair allocation within the soft-QoS control algorithm.
- Specification of available bandwidth fraction (ABF) for each ATM traffic descriptor parameter. ABF is defined as the ratio of the available to requested traffic descriptor parameter. This results in ABF-PCR, ABF-SCR, and ABF-MBS for the peak, sustained, and maximum burst size, respectively. These parameters are included in the MODIFY REJECT message. Using the ABF information, the connection owner may recompute the requested ATM traffic descriptor and reissue an appropriate MODIFY REQUEST message.

Two additional call states are defined to support modification. An entity enters the modify request state when it issues a MODIFY REQUEST of BCI message to the other side of the interface. An entity enters the modify received state when it receives a MODIFY REQUEST of BCI message from the other side of the interface.

Soft-QoS control is particularly useful during the handover procedure as a new MT moves into a cell and places demands on resources presently allocated to connections from other MTs. In the present WATM baseline handover specification [9], an MT sends a prioritized list of VCs to the network during handover, but there is no specification as to what the MT or the network should do if not all these VCs can be accommodated at the new cell. A flexible way of prioritizing the bandwidth allocation to various session VCs is through their softness profiles. If a mobile terminal faces significant drop

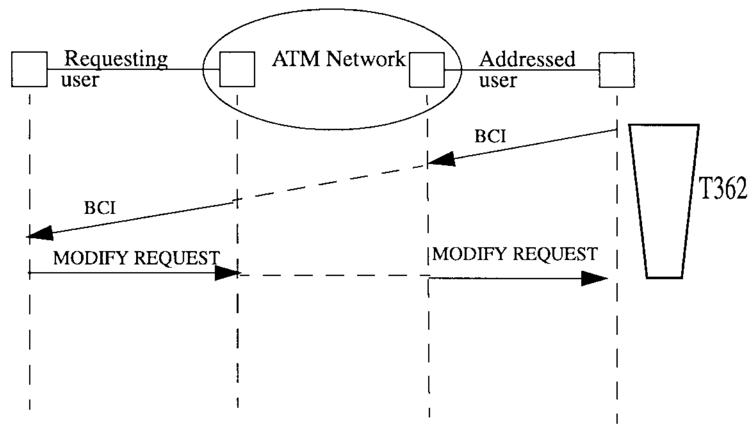

FIGURE 35.15: Use of BCI message for addressed user-initiated modification.

in bandwidth availability as it moves from one cell to another, rather than dropping the handover connections, the EMAS might be able to reallocate bandwidth among selected active connections in the new cell.

Within the soft-QoS framework, the soft-QoS controller selects a set of connections, called donors, and changes their bandwidth reservation so as to ensure satisfactory service for all [17]. This process is called network-initiated renegotiation. Network-initiated renegotiation improves the session handover success probability since multiple connections within and among sessions can share the available resources at the new EMAS, maintaining the satisfaction of individual connections above the minimum required. This mechanism allows multimedia sessions to transparently migrate the relative priority of connections as the MT moves across cells without a need to further specify details of the media session's content.

Figure 35.16 shows an MT moving into the coverage area of a new RP under a new EMAS and issuing a MODIFY REQUEST message (MRE). As a result, the EMAS might have to reallocate bandwidth

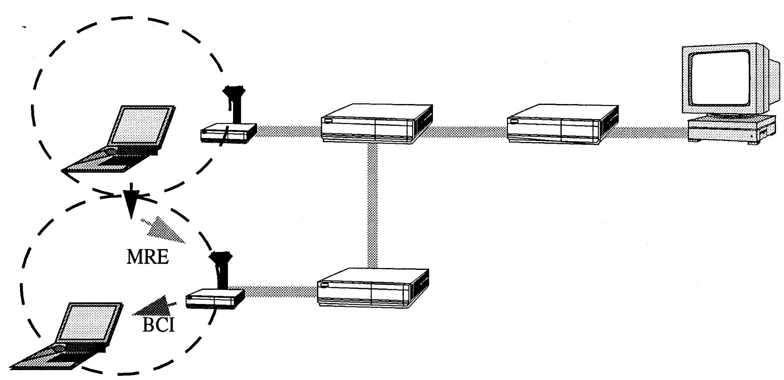

FIGURE 35.16: Handover procedure with soft-QoS in the WATMnet system.

of other connections under the RP to successfully complete the handover. This is accomplished by issuing BCI messages to a selected set of connections, called donors. At the time of receiving the first MRE message for a connection being handed over, no state exists for that connection within the new EMAS. This event differentiates the MRE messages from ongoing connections and connections being handed over. Different algorithmic provisions can be made to expedite bandwidth allocation to MRE messages of connections being handed over, reducing the probability of handover drop. For example, upon identifying a MRE from such connection, the soft QoS controller can use cached bandwidth reserves to maintain the satisfaction of the connection above the minimum. The size of the bandwidth cache is made adaptive to the ratio of handed-over to local bandwidth demand. The bandwidth cache for each RP can be replenished off-line using the network-initiated modification procedure. In this way, a handed-over connection need not wait for the network-initiated modification procedure to end before being able to use the bandwidth. The outcome of the reallocation enables most connections to sustain a better than minimum application performance while resources become available. Short-term congestion may occur due to statistical multiplexing. If long-term congestion arises due to the creation of a hot spot, dynamic channel allocation (DCA) may be used to provide additional resources. It is also possible that if connection rerouting is required inside the network for handover, the required resources to support the original QoS request may not be available within the network along the new path. Renegotiation is a useful feature in this case also.

An important performance metric for the soft-QoS service model is the low satisfaction rate (LSR). LSR measures the probability of failing to obtain link capacity necessary to maintain acceptable application performance. Figure 35.17 compares the LSR with and without network-initiated modification over a wide range of link utilization for an MPEG-based interactive video application. The softness profiles for MPEG video were derived from empirical results reported in [21, 22]. The

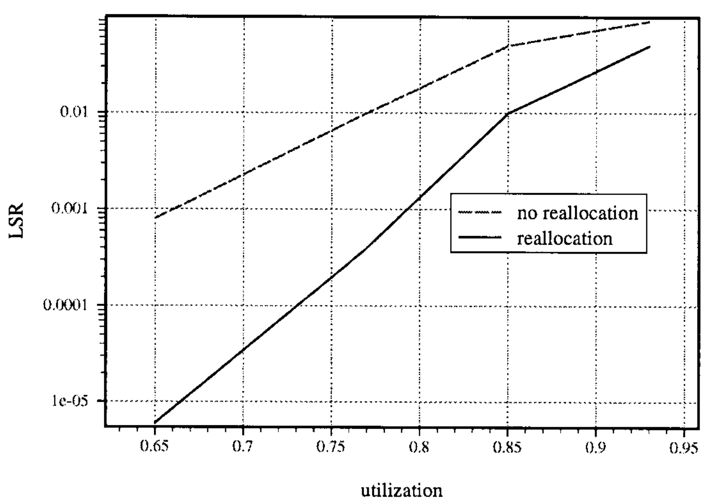

FIGURE 35.17: Effect of soft-QoS control with and without network-initiated modification.

figure shows that network-initiated modification has an important contribution to soft-QoS control performance: robust operation (LSR $< 10^{-3}$) is achievable while maintaining 70 to 80% utilization. In the WATM scenario, the handoff success probability with soft-QoS control is related to the LSR by Prob(handoff success) = $P(sat^+ > sat_{min}) > 1 - LSR >> 1 - P_b$, where sat^+ represents the satisfaction after handover to the new AP completes. Since it is better to block a new connection than to drop an existing connection for lack of capacity, the condition $LSR \ll P_b$ is used, where P_b is the connection blocking probability. The operating goal for the system is to maintain utilization > 70%, $LSR \sim 10^{-5}$, $P_b \sim 10^{-3}$.

Although the results presented are based on softness profiles for video, the definition of soft-QoS is appropriate for adaptive multimedia applications in general. Various representative softness profiles can be defined and refined as users' experience with distributed multimedia applications grows. New profiles can easily be incorporated within the framework as they become available.

35.3 Mobility Management in Wireless ATM

Allowing end system mobility in ATM networks gives rise to the problem of mobility management, i.e., maintaining service to end systems regardless of their location or movement. A fundamental design choice here is whether mobility management deals with user mobility or terminal mobility. When a network supports user mobility, it recognizes the user as the subscriber with an associated service profile. The user can then utilize any MT for access to the subscribed services. This results in flexibility in service provisioning and usage, but some extra complexity is introduced in the system implementation, as exemplified by the GSM system [24]. Support for user mobility implies support for terminal mobility. A network may support only terminal mobility and not recognize the user of the terminal, resulting in a simpler implementation. In either case, the mobility management tasks include:

- *Location Management:* Keeping track of the current location of an MT in order to permit correspondent systems to set up connections to it. A key requirement here is that the correspondent systems need not be aware of the mobility or the current location of the MT.
- *Connection Handover:* Maintaining active connections to an MT as it moves between different points of attachment in the network. The handover function requires protocols at both the radio layer and at the network layer. The issue of preserving QoS during handovers was described earlier and this introduces some complexity in handover implementations.
- *Security Management:* Authentication of mobile users (or terminals) and establishing cryptographic procedures for secure communications based on the user (or terminal) profile [25].
- *Service Management:* Maintaining service features as a user (or terminal) roams among networks managed by different administrative entities. Security and service management can be incorporated as part of location management procedures [24].

Early wireless ATM implementations have considered only terminal mobility [6, 7]. This has been to focus the initial effort on addressing the core technical problems of mobility management, i.e., location management and handover [26]. Flexible service management in wide-area settings, in the flavor of GSM, has not been an initial concern in these systems. The mobility management protocol standards being developed by the ATM Forum WATM working group may include support for user mobility. But the details are yet to be specified. In the following, therefore, we concentrate on the location management and handover functions required to support terminal mobility in wireless ATM. Our description follows along the lines of the ongoing ATM Forum specifications [9] with examples from wireless ATM implementations.

35.3.1 Location Management in Wireless ATM

Location management (LM) in WATM networks is based on the notions of permanent and temporary ATM addresses. A permanent ATM address is a location-invariant, unique address assigned to each MT. As the MT attaches to different points in a WATM network, it may be assigned different temporary ATM addresses. As all ATM end system addresses, both permanent and temporary addresses are derived from the addresses of switches in the network, in this case EMASs. This allows connection set-up messages to be routed towards the MT, as described later. The EMAS whose address is used to derive the permanent address of an MT is referred to as the home EMAS of that MT. The LM function in essence keeps track of the current temporary address corresponding to the permanent address of each MT. Using this function, it becomes possible for correspondent systems to establish connections to an MT using only its permanent address and without knowledge of its location.

35.3.2 Network Entities Involved in LM

The LM functions are distributed across four entities:

- *The Location Server (LS):* This is a logical entity maintaining the database of associations between the permanent and temporary addresses of mobile terminals. The LS responds to query and update requests from EMASs to retrieve and modify database entries. The LS may also keep track of service-specific information for each MT.
- *The AUthentication Server (AUS):* This is a logical entity maintaining a secure database of authentication and privacy related information for each MT. The authentication protocol may be implemented between EMASs and the AUS, or directly between MTs and the AUS.
- *The Mobile Terminal:* The MT is required to execute certain functions to initiate location updates and participate in authentication and privacy protocols.
- *The EMAS:* Certain EMASs are required to identify connection set-up messages destined to MTs and invoke location resolution functions. These can be home EMASs or certain intermediate EMASs in the connection path. All EMASs in direct contact with MTs (via their RPs) may be required to execute location update functions. Home EMASs require the ability to redirect a connection set-up message. In addition, all EMASs may be required to participate in the redirection of a connection set-up message to the current location of an MT.

There could be multiple LSs and AUSs in a WATM network. Specifically, an LS and an AUS may be incorporated with each home EMAS, containing information on all the MTs that the EMAS is home to. On the other hand, an LS or an AUS may be shared between several EMASs, by virtue of being separate entities. These choices are illustrated in Fig. 35.18, where the terms “integrated” and “modular” are used to indicate built-in and separated LS and AUS. In either case, protocols must be implemented for reliably querying and updating the LS, and mechanisms to maintain the integrity and security of the AUS. NEC’s WATMnet [6] and BAHAMA [7] are examples of systems implementing integrated servers. The modular approach is illustrated by GSM [24] and next-generation wireless network proposals [27].

35.3.3 Location Management Functions and Control Flow

At a high level, the LM functions are registration, location update, connection routing to home or gateway EMAS, location query, and connection redirect.

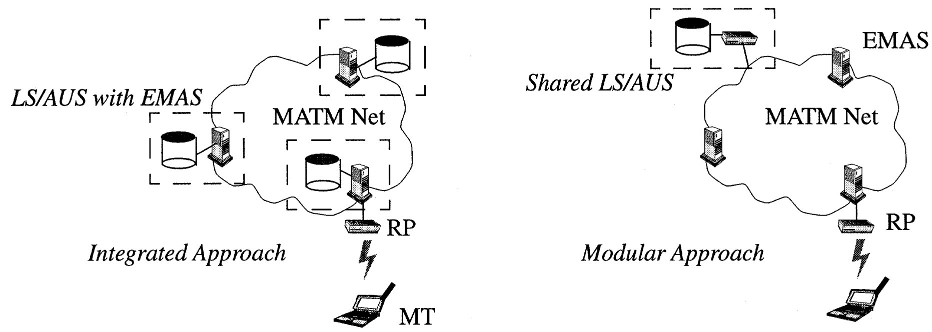

FIGURE 35.18: Server organizations.

Registration and Location Update

When an MT connects to a WATM network, a number of resources must be instantiated for that mobile. This instantiation is handled by two radio layer functions: association, which establishes a channel for the MT to communicate with the edge EMAS, and registration, which binds the permanent address of the MT to a temporary address. In addition, the routing information pertaining to the mobile at one or more location servers must be updated whenever a new temporary address is assigned. This is done using location updates.

The authentication of a mobile terminal and the establishment of encryption parameters for further communication can be done during the location updating procedure. This is illustrated in Fig. 35.19 which shows one possible control flow when an MT changes location from one EMAS to another. Here, the Broadcast_ID indicates the identity of the network, the location area, and the current radio port. Based on this information, e.g., by comparing its access rights and the network ID, the MT can decide to access the network. After an association phase, which includes the setting up of the signalling channel to the EMAS, the MT sends a registration message to the switch. This message includes the MT's home address and authentication information. The location update is initiated by the visited EMAS and the further progression is as shown. The LS/AUS are shown logically separate from the home EMAS for generality. They can be integrated with the home EMAS. Details on the implementation of a similar location update scheme can be found in [28].

Now, there are other possible configurations of LSs that give rise to different control message flow. For example, a two-level hierarchical LS arrangement can be used. Under this organization, the LS in the visiting network is updated as long as the MT remains in this network, and the home LS is updated only when the MT moves to a different WATM network. The information kept in the home LS must, therefore, point to a gateway EMAS in the visited network, since precise location in the visited network will not be available at the home LS. GSM location management is an example of this scheme [24].

Connection Forwarding, Location Query, and Connection Redirect

After a location update, a location server handling the MT has the correct association between its permanent and temporary ATM addresses. When a new connection to the MT is established, the set-up message must be routed to some EMAS that can query the LS to determine the current address of the MT. This is the connection forwarding function. Depending on how MT addresses are assigned, the location query can occur very close to the origination of the connection or it must progress to the home EMAS of the MT. For instance, if MT addresses are assigned from a separately

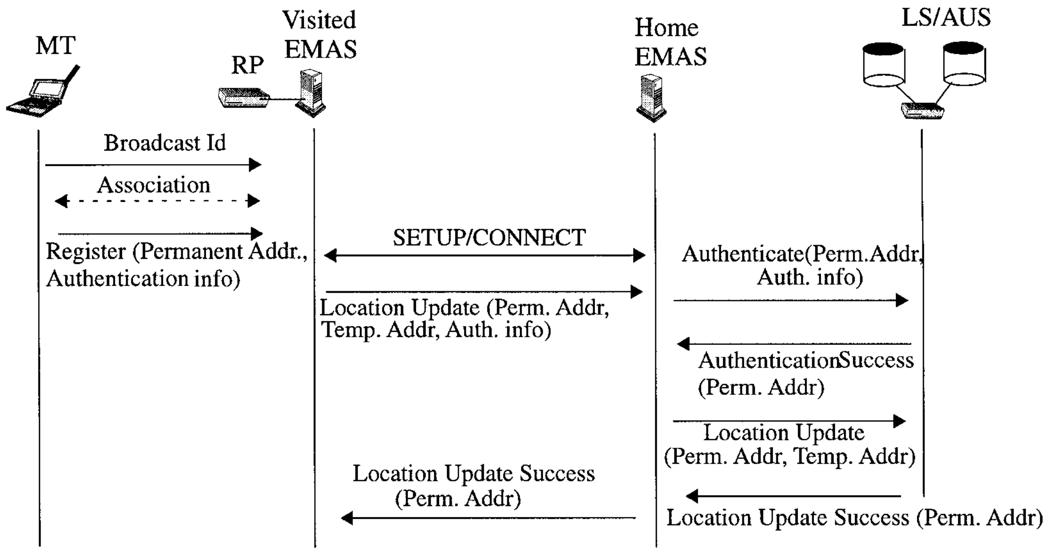

FIGURE 35.19: Location update control flow.

reserved ATM address space within a network, a gateway EMAS in the network can invoke location query when it processes a set-up message with a destination address known to be an MT address. To reach some EMAS that can interpret an MT address, it is sufficient to always forward connection set-up messages towards the home EMAS. This ensures that at least the home EMAS can invoke the query if no other EMAS enroute can do this. The location query is simply a reliable control message exchange between an EMAS and an LS. If the LS is integrated with the EMAS, this is a trivial operation. Otherwise, it requires a protocol to execute this transaction.

The control flow for connection establishment when the MT is visiting a foreign network is shown in Fig. 35.20. The addresses of various entities shown have been simplified for illustration purposes. Here, a fixed ATM terminal (A.1.1.0) issues a SETUP towards the MT whose permanent address is C.2.1.1. The SETUP message is routed towards the home EMAS whose address is C.2.1. It is assumed that no other EMAS in the path to the home EMAS can detect MT addresses. Thus, the message reaches the home EMAS which determines that the end system whose address is C.2.1.1 is an MT. It then invokes a location query to the LS which returns the temporary address for the MT (B.3.3). The home EMAS issues a redirected SETUP towards the temporary address. In this message, the MT is identified by its permanent address thereby enabling the visited EMAS to identify the MT and proceed with the connection SETUP signalling.

It should be noted that in the topology shown in Fig. 35.20, the redirection of the connection set-up does not result in a nonoptimal path. But, in general, this may not be the case. To improve the overall end-to-end path, redirection can be done with partial teardown in which case a part of the established path is released and the connection set-up is redirected from an EMAS that occurs further upstream of the home EMAS. This is shown in Fig. 35.21. Here, the EMAS labelled COS (Cross Over Switch) occurs in the original connection path upstream of the home EMAS. To redirect the set-up to B.3.3, the connection already established to the home EMAS is torn down up to COS, and new segment is established from the COS to B.3.3. This requires additional signalling procedures.

Finally, in Fig. 35.22, the case of hierarchical location servers is illustrated. Here, as long as the MT

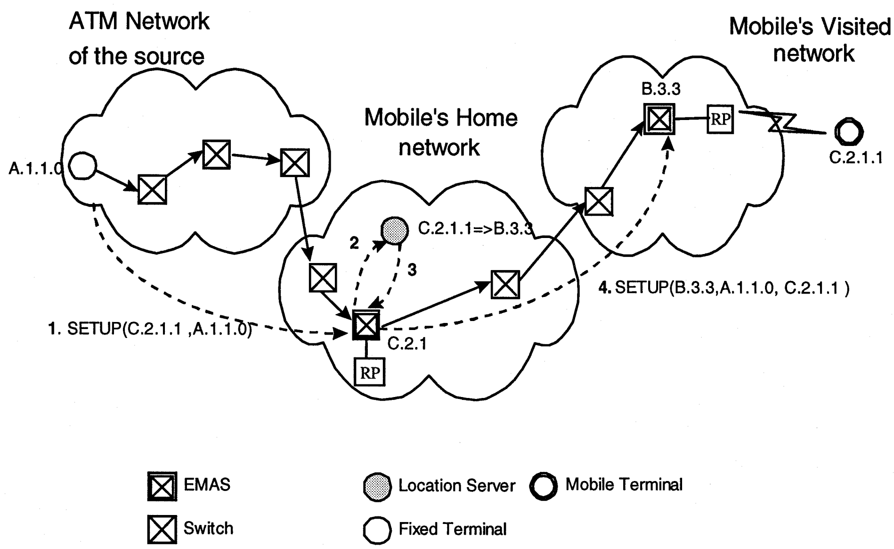

FIGURE 35.20: MT in foreign network.

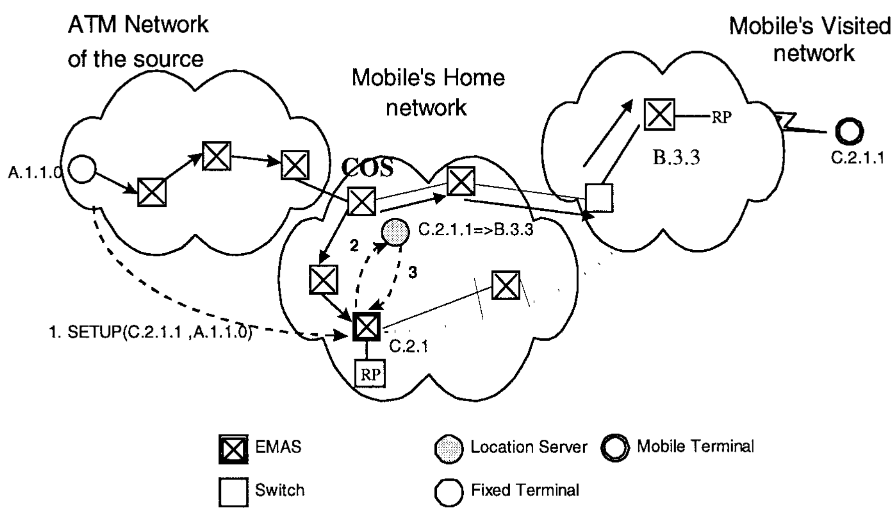

FIGURE 35.21: Connection redirect.

is in the visited network, the address of the gateway EMAS (B.1.1) is registered in its home LS. The connection set-up is sent via the home EMAS to the gateway EMAS. The gateway then queries its local LS to obtain the exact location (B.3.3) of the MT. It is assumed that the gateway can distinguish the MT address (C.2.1.1) in a SETUP message from the fact that this address has a different network prefix (C) than the one used in the visited network (B).

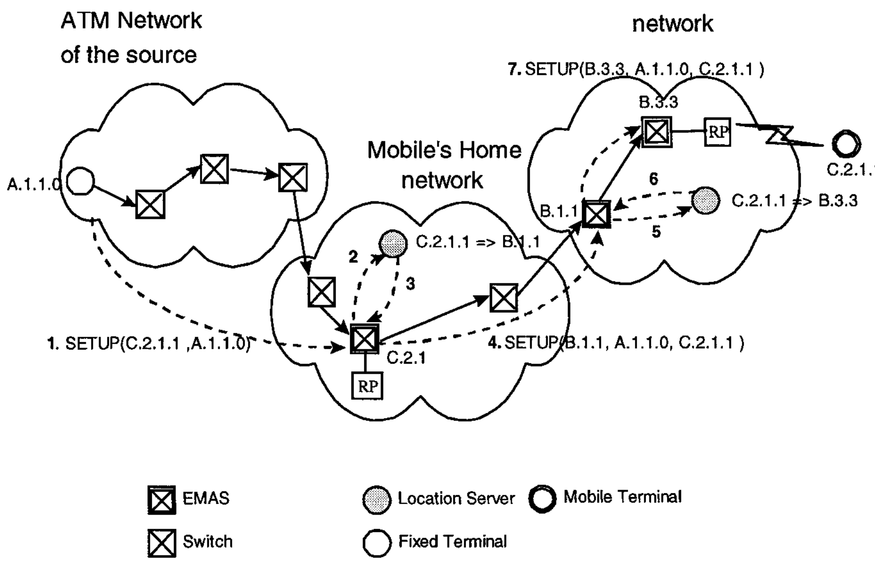

FIGURE 35.22: Hierarchical LS configuration.

Signalling and Control Messages for LM

The signalling and control messages required can be derived from the scenarios above. Specifically, interactions between the EMAS and the LS require control message exchange over a VC established for this purpose. This is described in [9]. The ATM signalling support needed for connection set-up and redirection is described in [28].

35.3.4 Connection Handover in Wireless ATM

Wireless ATM implementations, as well as the standards being developed by the ATM Forum, rely on mobile-initiated handovers whereby the MT is responsible for monitoring the radio link quality and decides when to initiate a handover [9, 26]. A handover process typically involves the following steps:

1. *Link quality monitoring:* When there are active connections, the MT constantly monitors the strength of the signal it receives from each RP within range.
2. *Handover trigger:* At a given instance, all the connections from/to the MT are routed through the same RP, but deterioration in the quality of the link to this RP triggers the handover procedure.
3. *Handover initiation:* Once a handover is triggered, the MT initiates the procedure by sending a signal to the edge EMAS with which it is in direct contact. This signal indicates to the EMAS the list of candidate RPs to which active connections can be handed over.
4. *Target RP selection:* The edge EMAS selects one RP as the handover target from the list of candidates sent by the MT. This step may make use of network-specific criteria for spreading the traffic load among various RPs and interaction between the edge EMAS and other EMASs housing the candidate RPs.

5. *Connection rerouting:* Once the target RP is selected, the edge EMAS initiates the rerouting of all connections from/to the MT within the MATM network to the target RP. The complexity of this step depends on the specific procedures chosen for rerouting connections, as described next. Due to constraints on the network or radio resources, it is possible that not all connections are successfully rerouted at the end of this step.
6. *Handover completion:* The MT is notified of the completion of handover for one or more active connections. The MT may then associate with the new RP and begin sending/receiving data over the connections successfully handed over.

Specific implementations may differ in the precise sequence of events during handover. Furthermore, the handover complexity and capabilities may be different. For instance, some systems may implement lossless handover whereby cell loss and missequencing of cells are avoided during handover by buffering cells inside the network [29]. The handover control flow is described in detail below for two types of handovers:

- *Backward handover:* The MT initiates handover through the current RP it is connected to. This is the normal scenario.
- *Forward handover:* The MT loses connectivity to the current RP due to a sudden degeneration of the radio link. It then chooses a new RP and initiates the handover of active connections.

Our description is based on the handover model being considered for standardization by the ATM Forum [9]. This model presently allows only hard handovers, i.e., active connections are routed via exactly one RP at a given instance, as opposed to soft handovers in which the MT can receive data for active connections simultaneously from more than one RP during handover.

Backward Handover Control Flow

Figure 35.23 depicts the control sequence for backward handover when the handover involves two different EMASs. Here, “old” and “new” EMAS refer to the current EMAS and the target EMAS, respectively. The figure does not show handover steps (1) and (2), which are radio layer functions, but starts with step (3). The following actions take place:

1. The MT initiates handover by sending an *HO_REQUEST* message to the old EMAS. With this message, the MT identifies a set of candidate RPs. Upon receiving the message, the old EMAS identifies a set of candidate EMASs that house the indicated RPs. It then sends an *HO_REQUEST_QUERY* to each candidate EMAS, identifying the candidate RP as well as the set of connections (including the traffic and QoS parameters) to be handed over. The connection identifiers are assumed to be unique within the network [28].
2. After receiving the *HO_REQUEST_QUERY* message, a candidate EMAS checks the radio resources available on all the candidate RPs it houses and selects the one that can accommodate the most number of connections listed. It then sends an *HO_REQUEST_RESPONSE* message to the old EMAS identifying the target RP chosen and a set of connections that can be accommodated (this may be a subset of connections indicated in the *QUERY* message).
3. After receiving an *HO_REQUEST_RESPONSE* message from all candidate EMASs, the old EMAS selects one target RP, based on some local criteria (e.g., traffic load spreading). It then sends an *HO_RESPONSE* message to the MT, indicating the target RP. At the same time, it also sends an *HO_COMMAND* to the new EMAS. This message identifies the

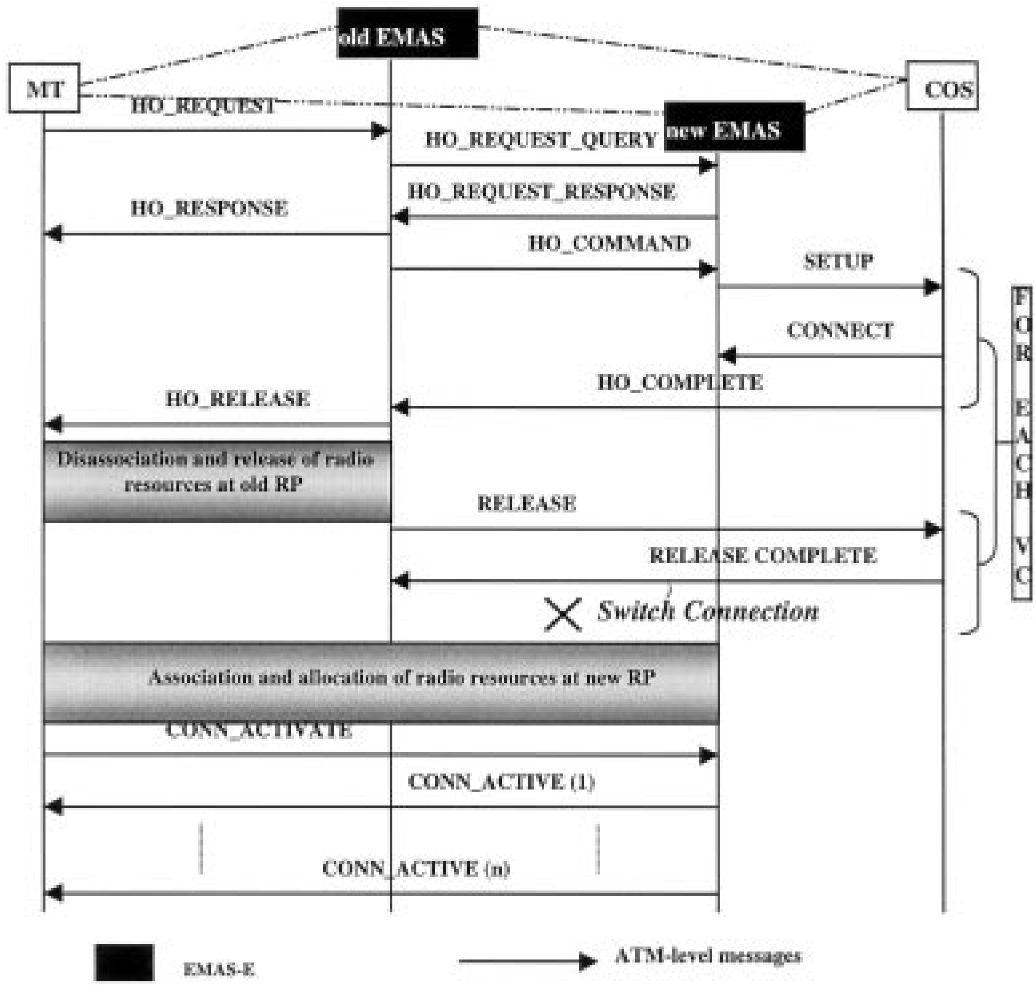

FIGURE 35.23: Backward handover control flow.

target RP and all the connections to be handed over along with their ATM traffic and QoS parameters. This message may also indicate the connection rerouting method. Rerouting involves first the selection of a cross-over switch (COS) which is an EMAS in the existing connection path. A new connection segment is created from the new EMAS to the COS and the existing segment from the COS to the old EMAS is deleted. Some COS selection options are:

VC Extension: The old EMAS-E itself serves as the COS.

Anchor-Based Rerouting: The COS is determined apriori (e.g., a designated EMAS in the network) or during connection set-up (e.g., the EMAS that first served the MT when the connection was set up). The selected COS is used for all handovers during the lifetime of the connection.

Dynamic COS Discovery: The COS is dynamically determined during each handover.

These procedures are illustrated in Fig. 35.24. While anchor-based rerouting and VC extension result in the same COS being used for all the connections being handed over, dynamic COS selection may result in different COSs for different connections. The *HO_COMMAND* message indicates which COS selection method is used, and if VC extension or anchor-based rerouting is used, it also includes the identity of the COS. If dynamic COS selection is used, the message includes the identity of the first EMAS in the connection path from the source (this information is collected during connection set-up [28]). For illustrative purposes, we assume that the dynamic COS selection procedure is used.

4. Upon receiving the *HO_COMMAND* message, the new EMAS allocates radio resources in the target RP for as many connections in the list as possible. It then sends a *SETUP* message towards the COS of each such connection. An EMAS in the existing connection path that first processes this message becomes the actual COS. This is illustrated in Fig. 35.25. This action, if successful, establishes a new segment for each connection from the new EMAS to the COS.
5. If the set-up attempt is not successful, new EMAS sends an *HO_FAILURE* message to the old EMAS, after releasing all the local resources reserved in step (4). This message identifies the connection in question and it is forwarded by the old EMAS to the MT. What the MT does in response to an *HO_FAILURE* message is not part of the backward handover specification.
6. If the COS successfully receives the *SETUP* message, it sends a *CONNECT* message in reply, thereby completing the partial connection establishment procedure. It then sends an *HO_COMPLETE* message to old EMAS-E. The *HO_COMPLETE* message is necessary to deal with the situation when handover is simultaneously initiated by both ends of the connection when two MTs communicate (for the sake of simplicity, we omit further description of this situation, but the reader may refer to [30] for further details).
7. The old EMAS-E waits to receive *HO_COMPLETE* messages for all the connections being handed over. However, the waiting period is limited by the expiry of a timer. Upon receiving the *HO_COMPLETE* message for the last connection, or if the timer expires, the old EMAS sends an *HO_RELEASE* message to MT. Waiting for the *HO_RELEASE* message allows the MT to utilize the existing connection segment as long as possible threreby minimizing data loss. However, if the radio link deteriorates rapidly, the MT can switch over to the new RP without receiving the *HO_RELEASE* message.
8. The old EMAS initiates the release of each connection for which an *HO_COMPLETE* was received by sending a *RELEASE* message to the corresponding COS.
9. Upon receiving the *RELEASE* message, the COS sends a *RELEASE COMPLETE* to the previous switch in the path (as per regular ATM signalling) and switches the data flow from the old to the new connection segment.
10. Meanwhile, after receiving the *HO_RELEASE* message from the old EMAS or after link deterioration, the MT dissociates from the old RP and associates with the new RP. This action triggers the assignment of radio resources for the signalling channel and user data connections for which resources were reserved in step (4).
11. Finally, the MT communicates to the new EMAS its readiness to send and receive data on all connections that have been handed over by sending a *CONN_ACTIVATE* message.

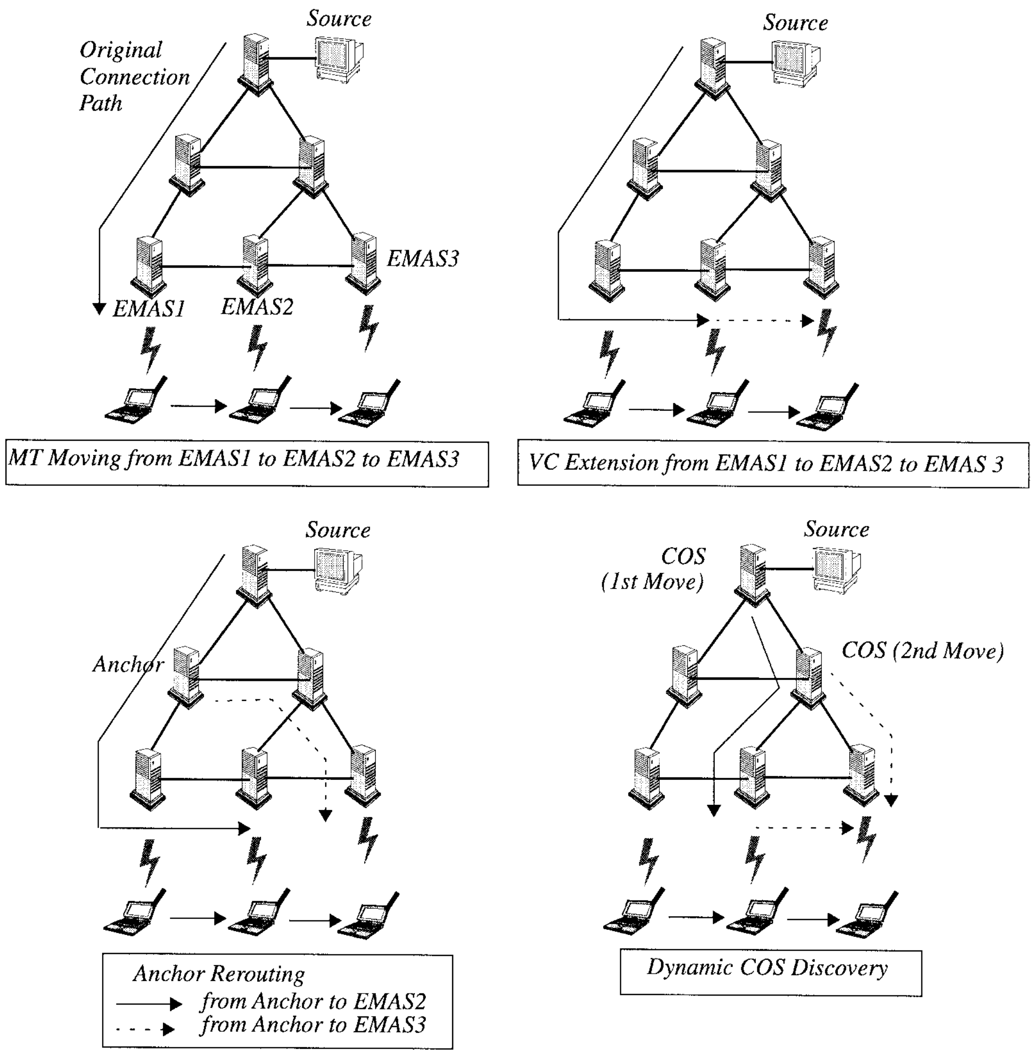

FIGURE 35.24: COS discovery.

12. Upon receiving the *CONN_ACTIVATE* message from the MT, new EMAS responds with a *CONN_ACTIVE* message. This message contains the identity of the connections that have been handed over, including their new ATM VC identifiers. Multiple *CONN_ACTIVE* messages may be generated, if all the connections have not been handed over when the *CONN_ACTIVATE* message was received. However, handover of remaining connections and the subsequent generation of *CONN_ACTIVE* signals are timer-bound: if the MT does not receive information about a connection in a *CONN_ACTIVE* message before the corresponding timer expires, it assumes that the connection was not successfully handed over. The recovery in this case is left up to the MT.

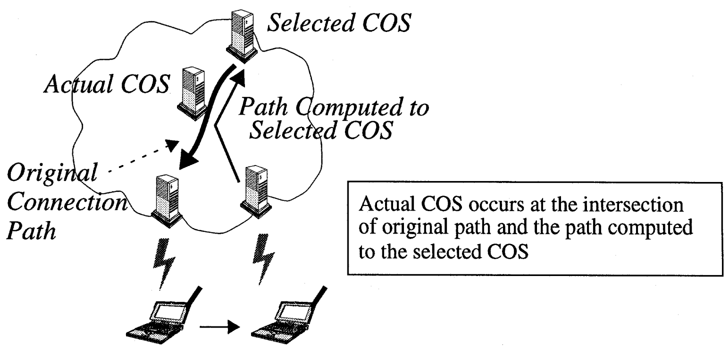

FIGURE 35.25: COS selection when MT moves.

The description above has left open some questions. Among them: what mechanisms are used to reliably exchange control messages between the various entities that take part in handover? What actions are taken when network or radio link failures occur during handover? How can lossless handover be included in the control flow? What effect do transient disruptions in service during handover have on application behavior?, and what are the performance impacts of signalling for handover? The short answers to these questions are: reliability can be incorporated by implementing a reliable transfer protocol for those control messages that do not already use such a transport (*SETUP*, *RELEASE*, etc., do, but *HO_REQUEST_QUERY* for example, requires attention). Actions taken during network failures require further analysis, but forward handover can be used to recover from radio link failures during handover. Lossless handover requires inband signalling within each connection and buffering in the network. Details on this can be found in [29]. The effect of transient disruptions on applications can be minimal, depending on how rerouting is implemented during handover. This is described in detail in [31]. Finally, some of the performance issues related to mobility management are investigated in [32].

Forward Handover

Forward handover is considered as the measure of last resort, to be invoked when the current radio link deteriorates suddenly. The forward handover procedure is simpler than backward handover, as illustrated in Fig. 35.26. Here,

1. The radio link with the old EMAS degrades abruptly. This results in the MT being dissociated with the old RP. The MT then chooses a new RP and associates with it. In the example shown, the new RP is housed by a different EMAS (the new EMAS).
2. The MT initiates forward handover by sending a *FW_HO_REQUEST* message to the new EMAS. This message indicates the active connections by their network-wide unique identifiers, along with their ATM traffic and QoS parameters, the identity of the previous EMAS (the “old” EMAS), and the COS information (this information may be obtained when the connection is initially set up).
3. The new EMAS sends an *HO_NOTIFY* message to the old EMAS indicating the initiation of handover. This serves to keep the old EMAS from prematurely releasing the existing connections.
4. The new EMAS reserves radio resources for as many listed connections as possible on the radio port to which the MT is associated. It then sends a *FW_HO_RESPONSE* message

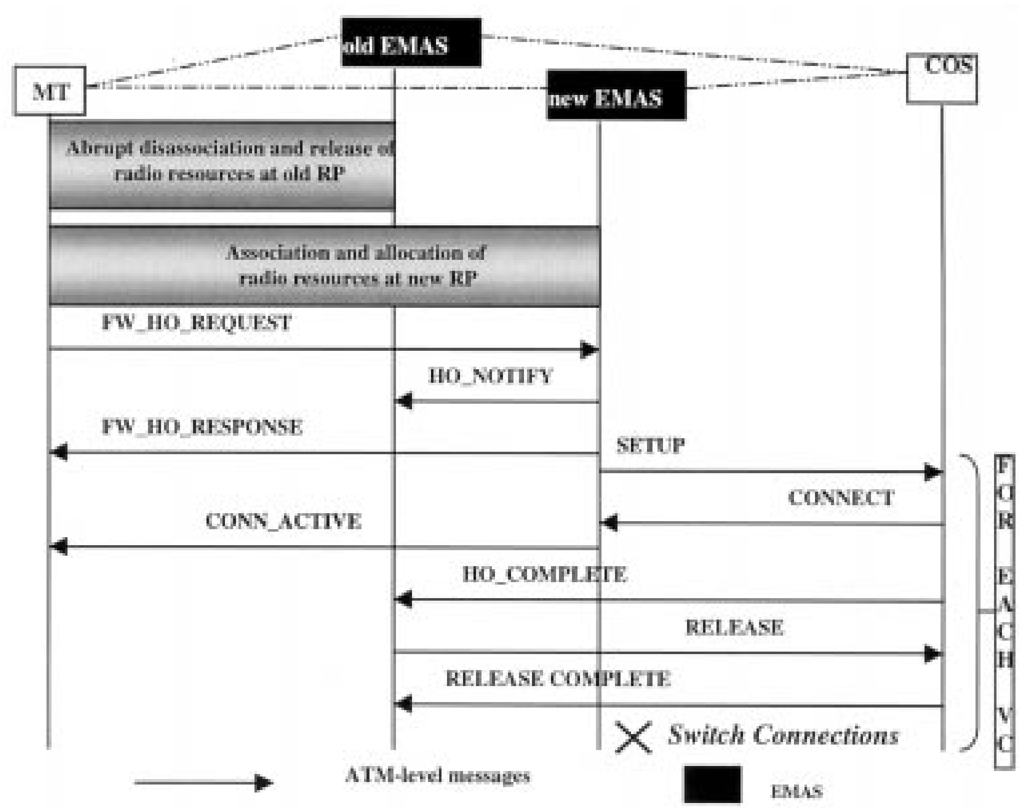

FIGURE 35.26: Forward handover control flow.

to the MT identifying the connections that can be handed over.

5. For each such connection, the new EMAS generates a SETUP message towards the COS to establish the new connection segment. This message includes the identity of the new EMAS. An EMAS in the existing connection path that first processes this message becomes the COS (Fig. 35.25).
6. Upon receiving the *SETUP* message, the COS completes the establishment of the new connection segment by sending a *CONNECT* message to the new EMAS.
7. After receiving the *CONNECT* message, new EMAS sends a *CONN_ACTIVE* message to the MT, indicating the connection has become active. Reception of *CONN_ACTIVE* by the MT is subject to a timer expiry: if it does not receive information about a connection in any *CONN_ACTIVE* message before the corresponding timer expires, it may initiate any locally defined recovery procedure.
8. If the new connection segment cannot be setup, the new EMAS sends an *HO_FAILURE* message to the old EMAS and the MT, after releasing all the local resources reserved for the connection. Recovery in this case is left up to the MT.
9. If the COS did send a *CONNECT* in step 7, it switches the connection data to the new segment and sends an *HO_COMPLETE* message to old EMAS. As in the case of backward handover, the *HO_COMPLETE* message is necessary to resolve conflicts in COS selection

when handover is simultaneously initiated by both ends of the connection when two MTs communicate.

10. Upon receiving the *HO_COMPLETE* message, the old EMAS releases the existing connection segment by sending a *RELEASE* message to the COS. In response, the COS sends a *RELEASE COMPLETE* to the previous switch.

35.4 Summary and Conclusions

In this chapter, the QoS and mobility management aspects of wireless ATM were described. WATM implementations, as well as the WATM standards being developed, allow the same ATM service categories in WATM networks as found in fixed ATM networks. The support for these classes of service in wireless ATM requires a variety of QoS control mechanisms acting in concert. The implementation of QoS in the wireless MAC layer, as well as the new QoS control mechanisms in the network and application layers, were described. The role of QoS renegotiation during handover was also described.

While mobility management in wide area involves various service-related aspects, the wireless ATM implementations and the standards efforts have so far focused on the core technical problems of location management and connection handover. These were described in some detail, along the lines of the specifications being developed by the ATM forum with examples from WATM implementations. In conclusion, we note that wireless ATM is still an evolving area and much work is needed to fully understand how to efficiently support QoS and mobility management.

References

[1] Raychaudhuri, D. and Wilson, N., ATM based transport architecture for multiservices wireless personal communication network, *IEEE J. Selected Areas in Commun.*, 1401–1414, Oct. 1994.

[2] Raychaudhuri, D., Wireless ATM: An enabling technology for personal multimedia communications, in *Proc. Mobile Multimedia Commun. Workshop*, Bristol, U.K., Apr. 1995.

[3] Raychaudhuri, D., Wireless ATM networks: Architecture, system design and prototyping, *IEEE Personal Commun.*, 42–49, Aug. 1996.

[4] Singh, S., Quality of service guarantees in mobile computing, *Computer Communications*, 19, 359–371, 1996.

[5] Naghshineh, M., Schwartz, M., and Acampora, A.S., Issues in wireless access broadband networks, in *Wireless Information Networks, Architecture, Resource Management, and Mobile Data*, Holtzman, J.M., Ed., Kluwer, 1996.

[6] Raychaudhuri, D., French, L.J., Siracusa, R.J., Biswas, S.K., Yuan, R., Narasimhan, P., and Johnston, C.A., WATMnet: A prototype wireless ATM system for multimedia personal communication, *IEEE J. Selected Areas in Commun.*, 83–95, Jan. 1997.

[7] Veeraraghavan, M., Karol, M.J., and Eng, K.Y., Mobility and connection management in a wireless ATM LAN, *IEEE J. Selected Areas in Commun.*, 50–68, Jan. 1997.

[8] Ala-Laurila, J. and Awater, G., The magic WAND: Wireless ATM network demonstrator, in *Proc. ACTS Mobile Summit 97*, Denmark, Oct. 1997.

[9] Rauhala, K., Ed., ATM Forum BTD-WATM-01.07, *Wireless ATM Baseline Text*, Apr. 1998.

[10] The ATM Forum Technical Committee, *ATM User-Network Signalling Specification*, Version 4.0, AF-95-1434R9, Jan. 1996.

[11] The ATM Forum Technical Committee, *Traffic Management Specification*, Version 4.0, AF-95-0013R11, Mar, 1996.
[12] Liu, K., Petr, D.W., Frost, V.S., Zhu, H., Braun, C., and Edwards, W.L., A bandwidth management framework for ATM-based broadband ISDN, *IEEE Communications Mag.*, 138–145, May 1997.
[13] Johnston, C.A., Narasimhan, P., and Kokudo, J., Architecture and implementation of radio access protocols in wireless ATM networks, in *Proc. IEEE ICC 98*, Atlanta, Jun. 1998.
[14] Passas, N., Paskalis, S., Vali., D., and Merakos, L., Quality-of-service-oriented medium access control for wireless ATM networks, *IEEE Communications Mag.*, 42–50, Nov. 1997.
[15] Passas, N., Merakos, L., and Skyrianoglou, D., Traffic scheduling in wireless ATM networks, in *Proc. IEEE ATM 97 Workshop*, Lisbon, May 1997.
[16] Raychaudhuri, D., Reininger, D., Ott, M., and Welling, G., Multimedia processing and transport for the wireless personal terminal scenario, *Proceedings SPIE Visual Communications and Image Processing Conference*, VCIP95, May 1995.
[17] Reininger, D. and Izmailov, R., Soft Quality-of-Service with VBR$^+$ video, *Proceedings of 8th International Workshop on Packet Video (AVSPN97)*, Aberdeen, Scotland, Sept. 1997.
[18] Ott, M., Michelitsch, G., Reininger, D., and Welling, G., An architecture for adaptive QoS and its application to multimedia systems design *Computers and Communications*, 1997.
[19] Microsoft Corporation., Windows Quality of Service Technology, White paper available on-line at `http://www.microsoft.com/ntserver/`
[20] Reininger, D., Raychaudhuri, D., and Hui, J., Dynamic bandwidth allocation for VBR video over ATM networks, *IEEE Journal on Selected Areas in Communications*, 14(6), 1076–1086, Aug. 1996.
[21] Lourens, J.G., Malleson, H.H., and Theron, C.C., Optimization of bit-rates, for digitally compressed television services as a function of acceptable picture quality and picture complexity, *Proceedings IEE Colloquium on Digitally Compressed TV by Satellite*, 1995.
[22] Nakasu, E., Aoi, K., Yajima, R., Kanatsugu, Y., and Kubota, K., A statistical analysis of MPEG-2 picture quality for television broadcasting, *SMPTE Journal*, 702–711, Nov. 1996.
[23] ITU-T, ATM Traffic Descriptor Modification by the connection owner, ITU-T Q.2963.2, Sept. 1997.
[24] Mouly, M. and Pautet, M-B., *The GSM System for Mobile Communications*, Cell & Sys, Palaiseau, France, 1992.
[25] Brown, D., Techniques for privacy and authentication in personal communication systems, *IEEE Personal Comm.*, Aug. 1985.
[26] Acharya, A., Li, J., Rajagopalan, B., and Raychaudhuri, D., Mobility management in wireless ATM networks, *IEEE Communications Mag.*, 100–109, Nov. 1997.
[27] Tabbane, S., Location management methods for third-generation mobile systems, *IEEE Communications Mag.*, Aug. 1997.
[28] Acharya, A., Li, J., Bakre, A., and Raychaudhuri, D., Design and prototyping of location management and handoff protocols for wireless ATM networks, in *Proc. ICUPC*, San Diego, Nov. 1997.
[29] Mitts, H., Hansen, H., Immonen, J., and Veikkolainen, Lossless handover in wireless ATM, *Mobile Networks and Applications*, 299–312, Dec. 1996.
[30] Rajagopalan, B., Mobility management in integrated wireless ATM networks, *Mobile Networks and Applications*, 273–286, Dec. 1996.

[31] Mishra, P. and Srivastava, M., Effect of connection rerouting on application performance in mobile networks, in *Proc. IEEE Conf. on Distributed Computing Syst.*, May 1997.
[32] Pollini, G.P., Meier-Hellstern, K.S., and Goodman, D.J., Signalling traffic volume generated by mobile and personal communications, *IEEE Communications Mag.*, Jun. 1995.