Abstract

This article deals with the problem of virtual circuit (VC) management in wireless ATM (W-ATM) networks with mobile user terminals. In W-ATM networks, a VC terminating at a mobile user may require dynamic reestablishment during the short time span necessary for terminal handover due to its movement from one (macro)cell to another. The VC reestablishment procedure has to ensure in-sequence and loss-free delivery of the ATM cells containing user data. After a classification of the solutions proposed so far in the literature, a novel technique for the dynamic reestablishment of VCs in W-ATM networks is described, and its performance is evaluated through simulation. The proposed technique allows for a progressive upgrade of the fixed part of the ATM network and for the incremental introduction of user terminal mobility.

Local and Global Handovers for Mobility Management in Wireless ATM Networks

**Marco Ajmone Marsan, Carla-Fabiana Chiasserini,
Renato Lo Cigno, and Maurizio Munafò
Politecnico di Torino
Andrea Fumagalli, University of Texas at Dallas**

Research in mobile telecommunications in Europe has traditionally been quite active: the North European countries were the first to introduce cellular telephony services, and their early experiences led to the development of the Global System for Mobile Communications (GSM) standard, which today is the most widely accepted approach for mobile digital telephony worldwide, not only in European countries, but also as far as Australia.

The growing number of digital mobile telephony terminals is paving the way to the introduction of mobile *data* communications, which are already possible at lower speeds, but will need increasingly high data rates for the introduction of multimedia services.

Several alternatives are being considered for the provision of integrated multimedia services to mobile users. One of the most promising long-term solutions is believed to be based on the adoption of the asynchronous transfer mode (ATM), typical of wired broadband integrated services digital networks (B-ISDN) also in wireless networks. This approach is usually termed *wireless ATM* (W-ATM). Evidence of this trend can be found in several projects within the research programs of the European Community (EC). An example is the RACE MBS Project [1, 2], which focuses on the design of the Mobile Broadband System (MBS), a wireless cellular network fully integrated into B-ISDN. The same approach is also being investigated at the international level within the ATM Forum, which recently established a Working Group on Wireless ATM, and by standardization bodies like the International Telecommunication Union — Telecommunication Standardization Sector (ITU-T) and the European Telecommunications Standards Institute (ETSI).

The design of W-ATM networks raises a number of challenges, among which we mention just two. First of all, integrated multimedia services require quite high user data rates (from 2 Mb/s up), which are nowadays common in high-speed wired networks but still a challenge over a radio link. Second, the integration of mobility within B-ISDN implies the dynamic reestablishment of the ATM virtual circuits (VCs) within the short time span of the mobile terminal handover from one macrocell (which may result from the collection of several microcells) to another. In addition, an important goal of the VC reestablishment procedure is to ensure in-sequence and loss-free delivery of the ATM cells containing user data in order to guarantee the quality of service (QoS) requirements on the connections.

This article focuses on the latter issue, and is organized as follows. In the next section we provide the basic definitions concerning the various network elements in the W-ATM scenario and the W-ATM architecture. We then define the VC management problem. In the third section we concisely discuss the contributions in this field that previously appeared in the literature, and we present a novel technique that deals with the VC reestablishment problem, emphasizing its innovative aspects which permit a progressive upgrade of the fixed part of the ATM network for the incremental introduction of user terminal mobility. In the fourth section we illustrate some numerical performance results obtained with the detailed simulation of a simple W-ATM scenario at the ATM cell level. Finally, the fifth section concludes the article. A detailed description of our proposal for the management of VCs in W-ATM networks is included in Appendix A.

Architecture Overview

An illustration of the W-ATM scenario is presented in Fig. 1, where we have adopted the same terminology used in [2].

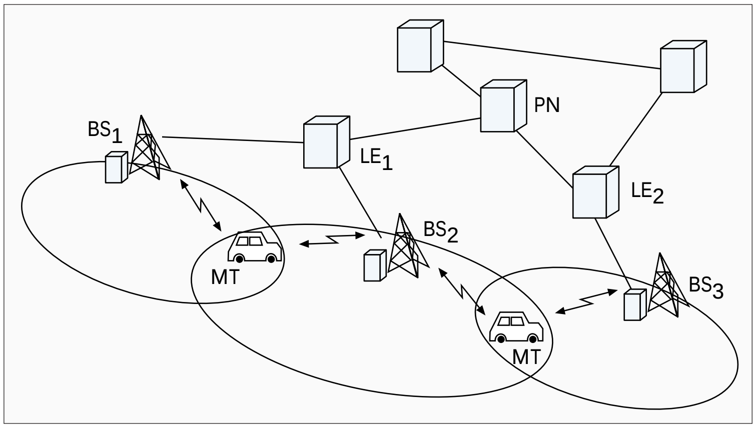
The W-ATM scenario comprises mobile terminals (MTs), base stations (BSs), and ATM switches (ATM nodes). ATM switches and BSs belong to the fixed network segment, whereas MTs can directly communicate only with BSs, which provide the interface between the wired and wireless portions of the network.

This work was supported in part by the EC through Copernicus project 1463 ATMIN, and in part by the Italian Ministry for University and Research.

1070-9916/97/$10.00 © 1997 IEEE

Each BS controls one macrocell of the W-ATM network (further subdivided into microcells or picocells depending on the radio transmission technology used), and communicates with all the MTs in the macrocell. A number of BSs can be connected to the same ATM switch. The ATM switch connected to one or more BSs is termed the *local exchange* (LE), with respect to the MTs that communicate with the BSs.

According to several proposals, during VC reestablishment, one ATM switch must perform special functions to manage the rerouting of the connection; this switch is called the *pivot node* (PN) in this article; other authors call such an ATM node the *cross-over switch* (COS).



■ **Figure 1.** *W-ATM scenario.*

The protocol stack within the fixed and wireless network segments is illustrated in Fig. 2. The radio interface between the BS and MT comprises the radio physical layer (RPHY), radio multiple access (RMAC) layer, and radio logical link control layer (RLLC), which are globally referred to as the radio access layer (RAL). Similar to fixed ATM networks, end-to-end connections between mobile/fixed terminals are provided by the ATM adaptation layer (AAL) present in the end-user protocol stack. Conversely, the protocol stacks at BSs and ATM switches reach only the ATM layer.

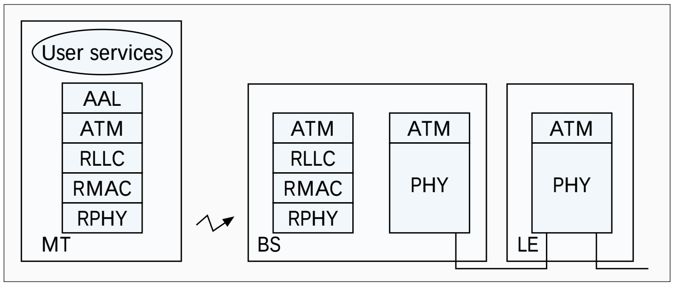
It can be observed that, in this architecture, ATM cells are exchanged all the way between end-user terminals, thus implying that ATM cells are carried over the radio interface. This approach yields a simple and homogeneous network architecture with end-to-end ATM cell delivery through standard ATM service access points.

Approaches to Network Handovers

User mobility requires different types of handover procedures. First of all, a *radio handover* procedure is necessary to manage the user terminal movements from one microcell (or picocell) to another within the same macrocell. This handover procedure involves only the BS to which the MT is connected, and can be limited to the RAL without reaching the ATM layer, so as long as the ATM cell sequence is preserved, any radio handover protocol, from the simple ones used today in cellular telephony to the more sophisticated alternatives exploiting spatial, or macro, diversity, can be adopted. Thus, only a change of radio channel is necessary in this case.

Second, a *network handover* procedure is necessary to manage the user terminal movements from one macrocell (managed by the *source* BS, say BS_1) to another macrocell (managed by the *destination* BS, say BS_2). Two scenarios are possible: both BS_1 and BS_2 are connected to the same LE, originating the *local* handover, or conversely, the two BSs are connected to different LEs, originating the *global* handover.

Besides the RAL management of the radio handover, the network handover also requires VC reestablishment at the ATM layer. From the MT viewpoint there is no difference in handling either type of network handover. Conversely, from the system viewpoint, a local handover involves a smaller number of network nodes than does a global handover.



■ **Figure 2.** *Protocol stack of the fixed and wireless network segments.*

As indicated in [3], the different approaches proposed to handle network handovers can be broadly subdivided into four categories, which have completely different characteristics, performance, and impact on the ATM standards:

- Full establishment
- Connection extension
- Incremental reestablishment
- Multicast establishment

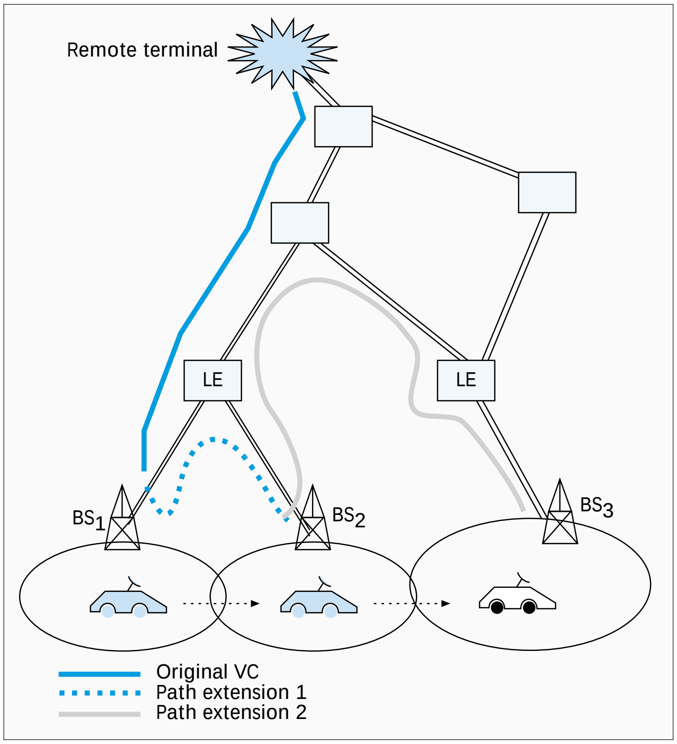
The *full establishment* approach requires the setup of a completely new connection between the end terminals. This is one of the earliest proposals, and it has a minor impact on the fixed network architecture. However, this procedure may not be sufficiently fast to guarantee that handovers do not cause timeouts to expire and connections to be abruptly terminated. In addition, both terminals must be involved in the path reestablishment operation. Figure 3 represents the VC evolution for a mobile terminal that roams through three macrocells in a network adopting the full establishment approach.

At each handover, the *connection extension* technique prolongates the VC between the terminals by adding one hop that provides the connection from the source BS to the destination BS through the fixed network. As proposed in [4], this path extension can be performed by the source BS, as shown in Fig. 4 (if BSs have switching capabilities), or, as proposed in [5], by the LE. The advantage of this approach is twofold: simple and reasonably fast execution, and intrinsic preservation of ATM cell sequence. Since no rerouting is performed some inefficiency may arise, especially when the mobile user circulates in a limited area, possibly returning to previously visited BSs. In this case closed loops may arise in the connection path. Figure 4 shows the VC modifications needed to follow the roaming terminal. It is quite evident that, although no closed loop arises in the illustrated situation, the resource waste is remarkable.

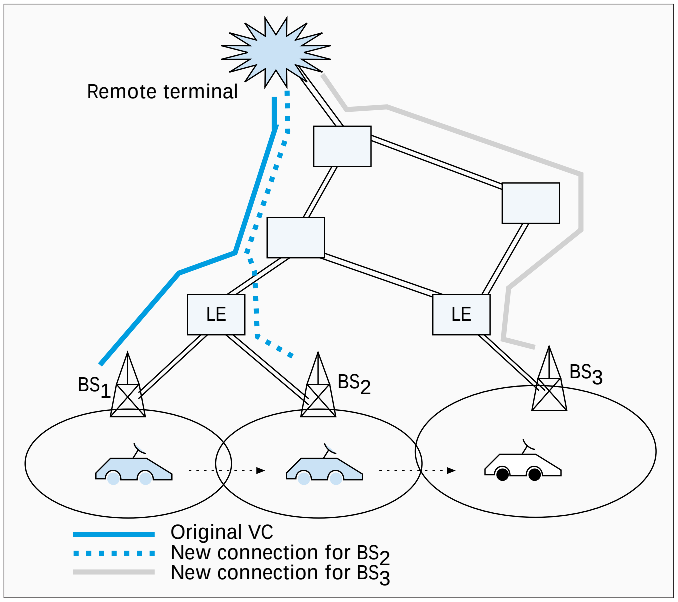
Incremental reestablishment is the handover category where our proposal can be classified. This technique is appealing because it requires only the establishment of a new partial path (without the involvement of the remote terminal and network entities) which connects to a portion of the original connection path, therefore allowing VCs to be partly reused [3, 6, 7]. Note that, because of spatial locality in movement, it is very likely that the reestablished path to the new location of the mobile user shares most of the virtual paths (VPs) in the original path. As a consequence, this technique is expected to be fast, efficient, and transparent, so it can be imagined that the end user does not perceive the network handover as a service interruption. Figure 5 shows the path rerouting performed while the terminal moves through the network. At each handover the optimal path is established, thus avoiding resource waste. The figure also indicates the PN, that is, the ATM switch that connects the original path to the incremental path for the handover occurring from BS_2 to BS_3.

Finally, the *multicast establishment* approach preallocates resources in the network portion surrounding the macrocell where the mobile user is located. When a new mobile connection is established, a set of virtual connections, named a *virtual connection tree*, is created, reaching all BSs managing the macrocells toward which the MT might move in the future. Thus, the mobile user can freely roam in the area covered by the tree without invoking the network call acceptance capabilities during handover. The allocation of the virtual connection tree may be static [8] or dynamic [9] during the connection lifetime. This approach is fast and statistically guarantees the QoS contract in case of network handover, since the QoS negotiation is executed only once, at connection establishment, allocating resources in the entire area where the mobile user is expected to roam. However, this approach may not be efficient in terms of network bandwidth utilization, since it introduces the possibility of refusing a connection because of lack of resources that may never be needed, and high signaling overheads, especially in the case of dynamic tree allocation. Figure 6 shows a multicast establishment, assuming that the MT moves within three macrocells. All the paths to the macrocells where the mobile is likely to roam are open all the time, although only one is used at any given time.

A two-phase handover was recently proposed in [10], which combines the advantages of both connection extension and incremental reestablishment. The rationale behind this hybrid approach is the use of a fast procedure to handle the connection extension during handover, followed by the optimal VC reestablishment procedure, which is activated once the MT is already connected to the destination BS.



■ Figure 4. *Path evolution of the VC during the MT roaming through three macrocells: the connection extension case.*



■ Figure 3. *Path evolution of the VC during the MT roaming through three macrocells: the full-establishment case.*

A Network Handover Approach Based on In-Band Signaling

The proposed network handover management technique exploits the incremental VC reestablishment procedure. Unlike some previous work on mobility management [3, 6, 11], the presented handover protocol is based on in-band signaling, thus resulting in a solution that requires marginal modifications of the ATM signaling protocol standards and recommendations.

In-band signaling is implemented through the use of dedicated resource management ATM cells, similar to those used in the available bit rate (ABR) and ATM block transfer (ABT) ATM transfer capabilities [12 , 13]. These ATM cells, termed *mobility enhancement signaling* (MES) cells, are inserted in the data flow of a connection when the handover takes place.

MES cells transparently cross ATM switches of the fixed network that are not upgraded to handle dynamic VC reestablishment. This allows a gradual introduction in the fixed network of ATM switches capable of interpreting MES cells and thus participating in the VC reestablishment procedure. In an initial phase, when only a (small) fraction of ATM switches have VC reestablishment capabilities, the PN, which must be chosen among upgraded ATM switches, may be quite far from LEs. This adversely affects performance with respect to a situation in which all ATM switches have VC reestablishment capabilities, so the PN can be much closer to LEs.

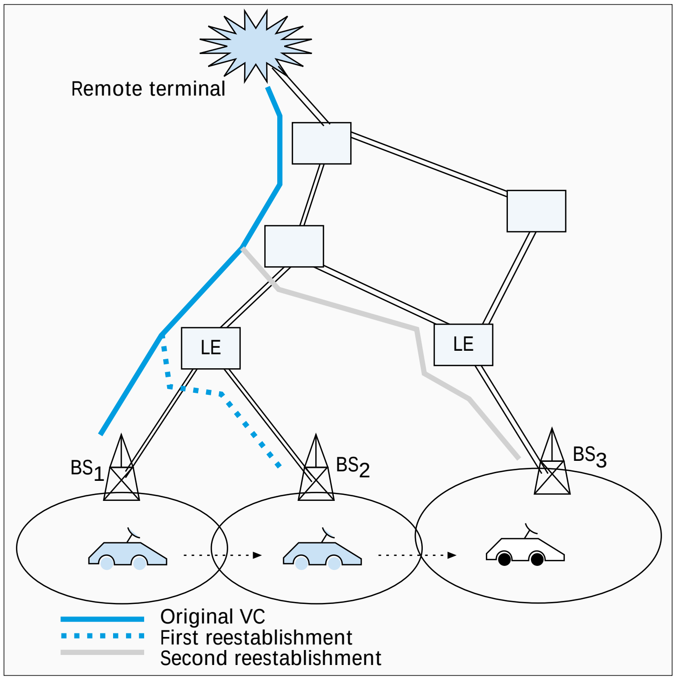
The VC reestablishment protocol is based on a few different types of MES cells to be exchanged between the MT and PN. The protocol is designed to perform both local and global handovers in a way that is transparent to the MT which does not need to be aware of the type of handover being executed.

The protocol procedure is explained in detail in Appendix A. In the following we give only a description of the key protocol features, to allow the reader to follow the performance discussion of the fourth section.

The handover protocol is the same in both local and global handover procedures, but the greater number of network entities involved in the case of global handover and the greater network span may induce different behaviors, so the two cases will be analyzed separately.

Without restriction, we discuss here the case when the handover protocol is triggered by the MT. More generally, protocols based on in-band signaling have been found suitable for handling handover in other network configurations currently under investigation within the standardization bodies [14]. The objective of the protocol is to ensure that ATM cells are delivered to end terminals in sequence and without any loss, in spite of the interruption of the radio transmission[1] and the interruption of the fixed part of the connection due to network reconfiguration. This requires that ATM cells temporarily be buffered along the connection path while the VC is being reestablished. Indeed, the basic workout of the procedure is as described below for all handover configurations, even if some minor changes are needed in some cases; the BS handover buffer requirements are influenced mainly by the time the network needs to perform the reconfiguration, not by the specific network setup or what entity triggers the handover.

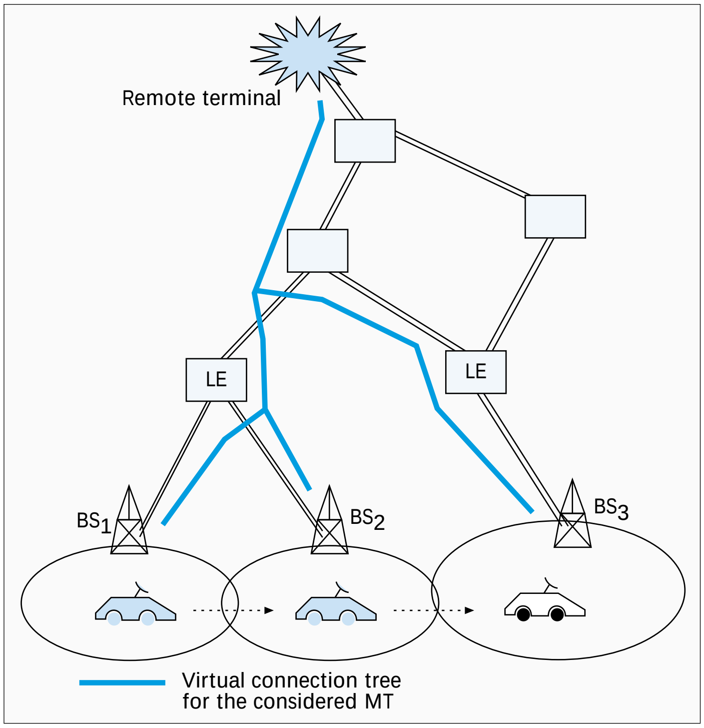
The considered network handover procedure is based on a *make-break* approach and, in order to reduce the ATM cell flow disruption, the rerouting of the downstream[2] connection is executed independent of the rerouting of the upstream connection; that is, the two rerouting operations possibly occur at different time instants. The incremental reestablishment of the connection is managed by the PN, selected in the early stages of the handover procedure.[3] When the PN commutes the downstream traffic to the new connection path, ATM cells are stored at the destination BS in the *downstream handover buffer* until the MT connects to the destination BS. Likewise, if the MT connects to the destination BS before the PN switches the upstream connection, the destination BS has to buffer data in the *upstream handover buffer* until the uplink connection to the network is ready. Also, buffering of the upstream traffic has to be performed at the MT during the radio handover latency (i.e., the period of absence of transmission and reception at the MT) due to the fact that it is disconnected from the source BS and not yet connected to the destination BS.



■ **Figure 5.** *Path evolution of the VC during the MT roaming through three macrocells: the incremental reestablishment case.*

The advantage of hosting the handover buffers at the destination BS is twofold:

- ATM nodes need not allocate additional buffers to handle terminal mobility.
- Traffic fluctuations originated by concurrent handovers can be dealt with at the destination BS using a shared buffer pool whose overall size can be designed to accommodate a maximum number of concurrent handovers, a parameter easily monitored by the BS itself.



■ **Figure 6.** *Virtual tree path of the VC during the MT roaming through three macrocells: the musticast establishment case.*

Buffer Requirements at the Destination Base Station

What we deem one of the critical issues in base station design is the buffering capacity necessary to handle multiple MTs simultaneously requesting handover toward the same destination BS. On the one hand, the number of ATM cells buffered at the MT is a function of both the MT offered load and the radio handover latency; on the other hand, the amount of cells stored in the upstream and downstream buffers at the destination BS is determined by the duration of the network handover and by the number of concurrent handovers being processed by the BS.

BSs, on the one hand, should be kept as simple and

---

[1] *We assume that the MT and BS do not exploit spatial diversity to allow the MT to simultaneously communicate with a number of BSs.*

[2] *Downstream connections go from the remote terminal to the MT, upstream connections go from the MT to the remote terminal.*

[3] *In critical applications, the PN selection can be handled off-line, considering all neighboring BSs together with the connection route and listing the corresponding PN IDs in a table.*

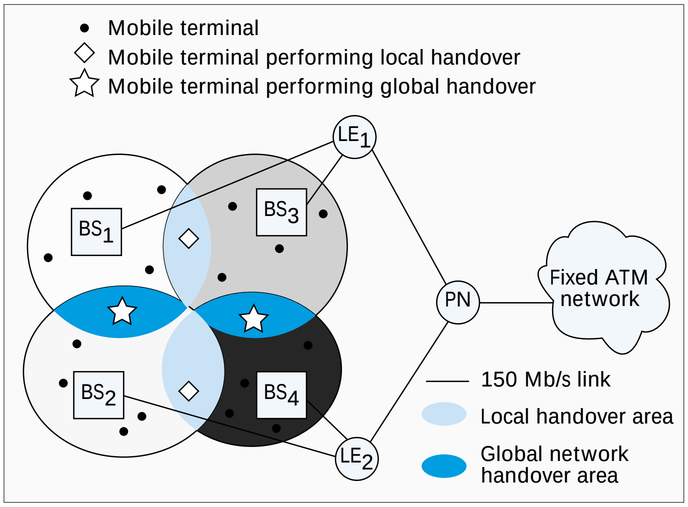


Figure 7. *The simulated W-ATM network.*

inexpensive as possible; on the other, they should not limit the QoS offered to mobile users. Therefore, it is important to determine the correlation between the maximum number of concurrent handovers that can be carried out at the destination BS and the buffer space necessary to guarantee loss-free ATM cell delivery.

The buffering requirements for the management of multiple simultaneous handovers were studied with detailed simulations at the ATM cell level for a simple W-ATM scenario.

The proposed protocol was implemented within CLASS[4] [15], a cell-level ATM network simulator developed at Politecnico di Torino in cooperation with CSELT (Centro Studi e Laboratori Telecomunicazioni, the research center of Telecom Italia). In the simulation model [16], the opening of the incremental path and the closing of the original unused part are carried out without the need for additional signaling functions, so all the additional signaling for mobility management is currently limited to the in-band exchange of MES cells.

Figure 7 shows the W-ATM network architecture that was simulated. As described in the appendix, the handover procedure must be executed simultaneously on both the upstream and downstream connections. The destination BS must provide buffer space for both upstream and downstream ATM cells; however, for the sake of brevity we limit our analysis to the upstream handover buffer. Similar results also apply to the downstream handover buffer.

The simulated W-ATM network scenario comprises four macrocells, each managed by a base station; BS_1 and BS_3 are connected to LE_1, while BS_2 and BS_4 are connected to LE_2; a further ATM node (marked PN) connects the two LEs. Both local and global network handovers can take place in this scenario. A handover is local when an MT moves either from BS_2 to BS_4 (or vice versa) or from BS_1 to BS_3 (or vice versa). A handover is global when an MT moves from one of the upper macrocells (BS_1 and BS_3) to one of the lower macrocells (BS_2 and BS_4). When a global handover occurs, the ATM switch denoted PN serves as the pivot node. The mobility model implemented in our simulation experiments allows the inhibition of either handover type to allow an independent analysis of the two cases.

All links connecting ATM nodes and BSs have a data rate of 150 Mb/s. The one-way propagation delay between the BS and LE is equal to 0.25 ms, while between each LE and PN the propagation delay is a variable denoted D_p. The propagation delay on the air interface is considered quite small, since each macrocell is assumed to cover a rather small area; in our simulation a value equal to 8.5 µs was chosen.

Each MT in the model represents the endpoint of one VC that reaches a (remote) user connected to the fixed network segment. The radio bandwidth allocated to each MT is 2 Mb/s for both upstream and downstream transmissions (thus 4 Mb/s total).

Traffic sources are supposed to alternate busy and idle periods such that the average offered load L_o of each MT is equal to 1.9 Mb/s. The traffic is generated in the form of higher-layer protocol data units (PDUs) which are transmitted at the allowed peak cell rate so that the overall user behavior corresponds to an On-Off source, but the stochastical description of the on and off periods is not trivial. With the assumption that the average offered load is 1.9 Mb/s, considering that during the radio handover data are not transmitted but accumulate at the MT, when the transmission resumes the MT behaves like a persistent source for quite a long time. In this way we conduct our study under conservative conditions.

At the MT MES cells are transmitted with priority over data cells to avoid unnecessary delay determined by the queuing time of the data transmission buffer. The cell rate on the VC between the BS and LE is governed by a shaping device that controls the sustainable cell rate (SCR), peak cell rate (PCR), and maximum burst size (MBS). The results presented in this article are obtained assuming SCR = 2 Mb/s, PCR = 2.1 Mb/s, and the MBS high enough (a few tens of cells) to empty the handover buffer when the fixed part of the connection is reestablished at the end of the handover procedure.

A handover request can be rejected (blocked) due to two reasons:

- The destination BS is already serving the maximum number of MTs' connections allowed by the capacity of the link toward the LE.
- The destination BS is already handling the maximum allowed number of concurrent handovers.

We assume that the maximum number of MTs served by one BS is 70 (thus leading to a 93 percent maximum nominal load of the 150 Mb/s link capacity); the average number of MTs per BS is 60. The distribution of the time interval between two consecutive handovers of the same MT is assumed to be uniform with mean chosen to originate approximately an overall handover blocking probability of 2 percent. Practically speaking, this means that the average value of the time interval varies between roughly 155 ms and 2700 ms depending on the maximum number of concurrent handovers and the propagation delay between the BSs and PN, while the standard deviation is constant and equal to $9/(11\sqrt{3})$.

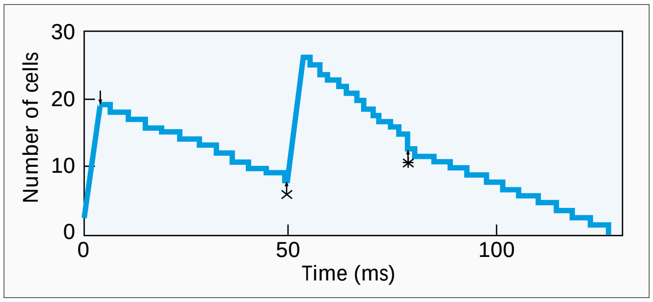
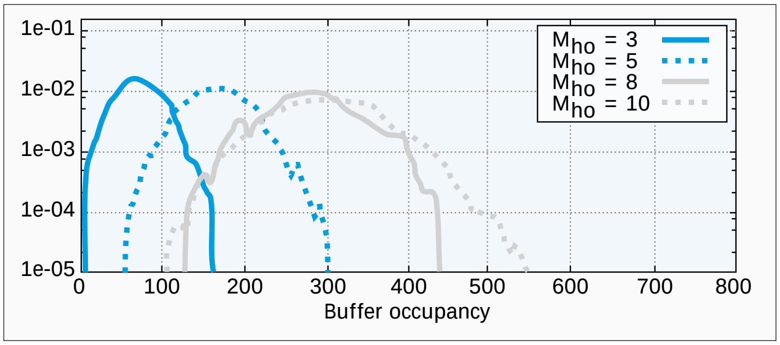


Figure 8. *Transient upstream handover buffer behavior during two overlapping handovers.*

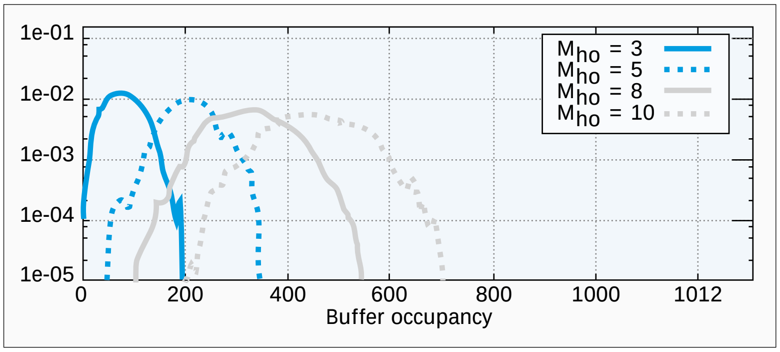
---

[4] *CLASS is a freeware software tool for the simulation of ATM networks; details about CLASS, together with citations of research works performed with it, can be found on the World Wide Web at http://www.tlc.polito.it/class.html.*

ATM nodes are assumed to be nonblocking with fixed crossover delay. As shown in Fig. 12 in Appendix A, the buffering time lapse at the destination BS is affected only by the update time of routing tables at the PN, while during the establishment of the incremental part of the path, the MT continues transmitting and receiving over the old part of the connection, and the time needed to perform this operation does not affect the buffer occupancy. The time needed by the PN to update the routing tables, for both downstream and upstream connections, is assumed to be about 4 ms, a value measured on small-size switches [7, 17]. The radio handover latency is a random variable with uniform distribution between two extremes: twice the radio propagation delay (17 µs), and twice the cell transmission time at the radio interface (425 µs).



■ Figure 9. *Probability density function of the number of cells in the upstream handover buffer for different values of the maximum allowed number of simultaneous network handovers* M_{HO}*; only local handovers are allowed.*



■ Figure 10. *Probability density function of the number of cells in the upstream handover buffer for different values of the maximum allowed number of simultaneous network handovers* M_{HO}*, with 0.5 ms propagation delay between LEs and PN; only global handovers are allowed.*

First of all, let us observe the upstream handover buffer occupancy versus time during two partially overlapping handovers performed by two different MTs, say MT_a and MT_b; the handover could be either local or global without affecting the transient behavior.

Figure 8 reports the total number of ATM cells in the BS handover buffer versus time expressed in milliseconds: time = 0 corresponds to the instant when the MT connects to the destination BS. The handover buffer is managed on a per-VC basis, but Fig. 8 reports the global overall buffer occupancy, adding the contributions of the two different procedures. After MT_a starts sending ATM cells to the destination BS, the buffer occupancy increases linearly until the connection with the LE is enabled; this time instant (δ_5 in Fig. 12) is highlighted by the downward arrow roughly at time 5 ms. When the uplink connection is ready, the BS begins to forward the data flow on the uplink, and the handover buffer starts emptying with a speed that depends on the ratio between PCR and SCR (unless the MT stops transmitting, but this is not the case in the handover under consideration). The upward arrow with a x marker points to the instant when MT_b connects to the destination BS: the buffer occupancy starts again to increase with a rate equal to the difference between the rate of ATM cell arrivals from MT_b and the rate at which ATM cells of MT_a leave the buffer (PCR > SCR). The downward arrow approximately at time 55 ms identifies the instant when MT_b starts forwarding data to the BS. After this time the buffer occupancy starts decreasing with a speed proportional to the difference between $PCR_a + PCR_b$ and $SCR_a + SCR_b$ until the transient of the first handover is over (upward arrow with * marker). From this instant on, the buffer continues emptying at a slower pace until the transient relative to the second handover also ends.

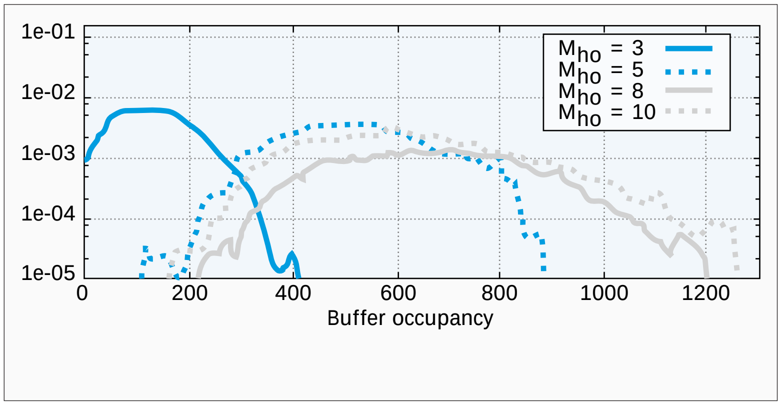
It is interesting to notice that, from the protocol viewpoint, the handover at the destination BS ends when the PN switches the upstream connection and the destination BS can start transmitting on the uplink, but the upstream handover buffer is not completely emptied for quite a long period after this time, depending on the allowed peak rate. This phenomenon may raise some concern about the buffer requirements when multiple handovers are managed simultaneously.

Figure 9 reports the probability density function of the number of ATM cells in the upstream handover buffer in the case of local handovers only, for different values of the maximum number of simultaneous network handovers M_{HO} allowed at each destination BS. M_{HO} only identifies handover procedures that are active from the protocol point of view, not considering the whole transient time needed to completely empty the upstream handover buffer. With reference to Fig. 12, a handover procedure is declared to be active at the destination base station for a time lapse starting halfway between T_1 and T_2 (when the incremental part of the path is opened) and ending at δ_5 (when the uplink incremental part of the path is enabled).

The distribution of the number of ATM cells is reported for values of M_{HO} = 3, 5, 8, and 10. As can be expected, the buffer occupancy grows as M_{HO} increases, and the growth of the maximum buffer occupancy is roughly linear. Although the buffer requirements increase almost linearly with the number of allowed simultaneous handovers, the overall requirements are not exceedingly large, and, regardless of M_{HO}, the decay of the tail of the probability density function is quite fast, so very low ATM cell loss probabilities are easily ensured.

Figure 10 reports similar results for the case of global network handovers, assuming that the propagation delay between each LE and the PN is 0.5 ms. Also in this case the distribution of the number of ATM cells is reported for values of M_{HO} = 3, 5, 8, and 10. The buffer occupancy distributions have shape similar to that obtained for local handovers, but indicate that somewhat larger buffers are necessary in this case due to the additional propagation delays necessary to propagate MES cells from one LE to the other via the PN.

Results shown in Fig. 11 are obtained assuming a propagation delay between the PN and the LEs of 5 ms. As expected, a further growth of the propagation delays involved in the network handover procedures negatively impacts the buffering requirements. Nevertheless, buffers of acceptable size are still sufficient to guarantee low buffer overflow probabilities even when quite a large number of concurrent handover procedures take place at the destination BS.



■ **Figure 11.** *Probability density function of the number of cells in the upstream handover buffer for different values of the maximm allowed number of simulataneous network handovers* M_{HO}*, with 5 ms propagation delay between LEs and PN; only global handovers are allowed.*

Conclusions

The problem of VC management within W-ATM networks with mobile terminals was discussed, illustrating a novel proposal for the dynamic reestablishment of VCs within the short time span of the mobile terminal handover from one macrocell to another.

It was shown that VC reestablishment with guaranteed in-sequence and loss-free ATM cell delivery can be achieved by using in-band signaling and reserving buffering resources at the destination base station, an approach that marginally affects the architecture of the ATM switches in the wired network.

In particular, the in-band signaling approach allows an incremental upgrade of the wired ATM network to serve mobile users, with a transition phase characterized by only a subset of ATM switches upgraded to handle VC reestablishment. In addition, limited buffering capability is sufficient at the destination base station in order to efficiently handle simultaneous handovers.

References

[1] L. Fernandes, "Developing a System Concept and Technologies for Mobile Broadband Communications," *IEEE Pers. Commun.*, Feb. 1995, pp. 54–59.
[2] M. Progler, "MBS Air Interface Principles," RACE Mobile Telecom Summit, Cascais, Portugal, Nov. 1995.
[3] C.-K. Toh, "A Hybrid Handover Protocol for Local Area Wireless ATM Networks," *ACM/Baltzer MONET Special Issue on Wireless ATM*, vol. 1, no. 3, Dec. 1996, pp. 313–34.
[4] M.J. Karol *et al.*, "BAHAMA: A Broadband Ad-Hoc Wireless ATM Local-Area Network," *Proc. ICC '95*, June 1995, pp. 1216–23.
[5] T. La Porta, "Distributed Processing for Mobility and Service Management in Mobile ATM Networks," *Wireless ATM Networking Wksp.*, New York, June 1996.
[6] C.-K. Toh, "The Design and Implementation of a Hybrid Handover Protocol for Multimedia Wireless LANs," *Proc. Mobicom '95*, Berkeley, CA, Nov. 13–15, 1995.
[7] J. Porter *et al.*, "The ORL Radio ATM System, Architecture and Implementation," Olivetti & Oracle Res. Lab., Tech. Rep. 96-5, Jan. 1996.
[8] A. Acampora and M. Naghshineh, "An Architecture and Methodology for Mobile-Executed Handoff in Cellular ATM Networks," *IEEE JSAC*, vol. 12, no. 8, Oct. 1994, pp. 1365–74.
[9] O. Yu and V. Leung, "Extending B-ISDN to Support User Terminal Mobility over an ATM-Based Personal Communications Network," *Proc. GLOBECOM '95*, 1995, pp. 2289–93.
[10] M. Veeraraghavan, M. Karol, and K. Y. Eng, "A Combined Handoff Scheme for Mobile ATM Networks," ATM Forum/WATM — ATM_Forum/96-1700, Vancouver, Canada, Dec. 1996.
[11] R. Yuan *et al.*, "A Signaling and Control Architecture for Mobility Support in Wireless ATM Networks," *ACM/Baltzer MONET Special Issue on Wireless ATM*, vol. 1, no. 3, Dec. 1996, pp. 287–98.
[12] ITU-TSS SG 13, Rec. I.371 "Traffic Control and Congestion Control in B-ISDN," Geneva, Switzerland, July 1995.
[13] ATM Forum/af-tm-0056.000, "Traffic Management Specification," v. 4.0, Apr. 1996.
[14] C.F. Chiasserini and R. Lo Cigno, "Wireless ATM Handover Protocols via In-Band Signaling," Tech. rep. DE/MAM/97/2, Dip. Elett., Politecnico di Torino, July 1997; e-mail requests to locigno@polito.it.
[15] M. Ajmone Marsan *et al.*, "ATM Simulation with CLASS," *Performance Evaluation*, vol. 24, 1995, pp. 137–59.
[16] C.F. Chiasserini, R. Lo Cigno, and M. Munafo, "On the Implementation of Mobility Services in CLASS: Architecture, Validation and Results," *Proc. 5th IFIP Wksp. Performance Modelling and Evaluation of ATM Networks*, Ilkley, U.K., July 1997.
[17] R. Ramjee *et al.*, "Performance Evaluation of Connection Rerouting Schemes for ATM-Based Wireless Networks," Tech. rep. UM-CS-1997-038, Univ. Massachusetts, July 1997.

Biographies

MARCO AJMONE MARSAN [SM] (AJMONE@rtlca2.polito.it) is a full professor at the Electronics Department of Politecnico di Torino, Italy. He was born in Torino, Italy, in 1951. He holds a Dr. Ing. degree in electronic engineering from Politecnico di Torino, and a Master of Science from the University of California, Los Angeles. From November 1975 to October 1987 he was at the Electronics Department of Politecnico di Torino, first as a researcher, then as an associate professor. From November 1987 to October 1990 he was a full professor at the Computer Science Department of the University of Milan, Italy. During the summers of 1980 and 1981 he was with the Research in Distributed Processing Group, Computer Science Department, UCLA. He has coauthored over 150 journal and conference papers in the areas of communications and computer science, as well as the books *Performance Models of Multiprocessor Systems*, MIT Press, and *Modelling with Generalized Stochastic Petri Nets*, John Wiley. He received the Best Paper award at the Third International Conference on Distributed Computing Systems in Miami, Florida, in 1982. His current interests are in the fields of performance evaluation of data communication and computer systems, communication networks, and queuing theory.

CARLA-FABIANA CHIASSERINI is a Ph.D. student at the Department of Electronics of Politecnico di Torino. She received the Dr. Ing. degree in electronic engineering from Universitá di Firenze, Italy, in 1996. In 1995 she was a visiting scholar with the ECE Department of the University of Massachusetts at Amherst. Her research activities focus on wireless and ATM networks.

ANDREA FUMAGALLI (fumagalli@polito.it) received the Dr. Ing. degree in electronic engineering and the Ph.D. degree in electronic engineering from Politecnico di Torino, Italy, respectively in 1987 and 1992. He did postdoctoral research in the field of all-optical networks at the University of Massachusetts at Amherst. Since 1992 he has been with the Department of Electronics as a researcher. Currently he is associate professor at the Erik Jonsson School of Engineering and Computer Science at the University of Texas at Dallas. His recent research activities concern photonic switches, all-optical networks, WDM, and ATM radio networks.

RENATO LO CIGNO is a research engineer in the Electronics Department of Politecnico di Torino. He received a Dr. Ing. degree in electronic engineering from Politecnico di Torino in 1988. Since then he has been with the telecommunication research group of the Electronics Department of Politecnico di Torino, first under various research grants and contracts, then as a staff member. His research interests are in communication networks simulation and performance analysis. He is coauthor of about 20 journal and conference papers in the areas of communications.

MAURIZIO MUNAFÒ is a research engineer in the Electronics Department of Politecnico di Torino. He holds a Dr.Ing. degree in electronic engineering since 1991 and a Ph.D. in telecommunications engineering since 1994, both from Politecnico di Torino. Since November 1991 he has been with the Electronics Department of Politecnico di Torino, where he has been involved in the development of an ATM networks simulator. His research interests are in simulation and performance analysis of communication systems. He has coauthored a few articles regarding the simulation of ATM networks.

Appendix A: The Handover Protocol

When a mobile terminal moves from one macrocell to another, all the connections originating or terminating at the mobile terminal must be rerouted from the old base station, named *source base station* or BS_1, to a new base station, named *destination base station* or BS_2. A B-ISDN mobile terminal may have several connections open at the same time, belonging either to different calls (i.e., with different remote users) or to the same call but supporting different applications. The handover of each connection can be carried out separately (possibly at the same time, but with different protocol instances). Highly correlated connections can also require a coordinated handover procedure. For the sake of simplicity here we describe only the handover procedure for a single connection.

Table 1 describes the proposed format of the MES cells, which are special-purpose resource management cells. The ID field of MES cells is set to a value consistent with the ITU-T and ATM Forum specifications [12 , 13], and identifies the use of these ATM cells for mobility-enhanced services. Field names and possible values are reported below. Notice that ATM end station address (AESA) addressing is adopted, and only the essential 13 bytes are considered.

Header: ATM cell header.

ID: Standard resource management cell's protocol identifier.

MES_TYPE: This field identifies the handover protocol message carried by the MES cell. Allowed values identify all the different messages required for the handover protocol:

- *HOR:* Handover request — This message is sent by the mobile terminal to start a network handover.
- *HOC:* Handover confirm — This message certifies that the network is ready to switch the connection to the mobile terminal from the source base station to the destination base station, enabling the handover procedure.
- *EDF:* End of data flow on the uplink connection through the source base station — This message is also an implicit acknowledgment from the mobile terminal that the handover procedure is taking place.
- SDF_{DOWN}: Start data flow on the new downlink connection — This message is the first MES cell sent by the PN on the incremental connection established to the destination base station.
- SDF_{UP}: Start data flow on the new uplink connection — This message is the first MES cell sent by the destination BS on the enabled incremental connection to the PN; when it reaches the PN it implicitly acknowledges the end of the handover protocol.
- *ULR:* New uplink connection ready — This message is sent by the PN to enable transmission on the incremental portion of the uplink connection.

FLOW_ID: This field can be used to manage the coordinated handover of several connections. If the value is 0, the handover of the connection to which MES cells belong is managed independent from all others.

PRI: 4-bit field indicating how badly the handover is needed by the mobile terminal; it is used to manage handover priorities if the network cannot manage all handover requests simultaneously.

QOS_MOD: 4-bit field indicating the possibility of a renegotiation of the QoS parameters for the connection; 0 indicates that no renegotiation is needed or permitted, other values are still to be defined.

PFL: 4-bit field used to manage the MES cell propagation depth: depending on its value the MES cell will be removed from the connection by a different entity (e.g., the PN or the destination BS).

S_BSA: Source BS address (13 bytes).

T1_BSA: First destination BS address (13 bytes).

T2_BSA: Second destination BS address (13 bytes).

CRC: Standard resource management cells' cyclic redundancy code.

A network handover is initiated when an MT enters an area belonging to two or more macrocells. Macrocells cover fairly large areas, and overlap to an extent that makes the probability of an MT being in an area covered by several BSs not marginal; from this consideration stems the need to indicate more than one destination BS when requesting a handover: the network will choose one, for instance, depending on the signal level or the BS load. Once the destination BS has been defined, the handover takes place toward this BS, and any further network handover is inhibited for this connection until this one has ended. If both the origin and destination BSs are under the jurisdiction of the same LE, the network handover is defined as *local*; otherwise, it is defined as *global*. In global handovers the system has to identify the PN, also called *cross-over switch*, that is, the node where the connection is rerouted, while for local handovers this role is played directly by the LE. From the point of view of the MT, there is no difference in handling either network handover.

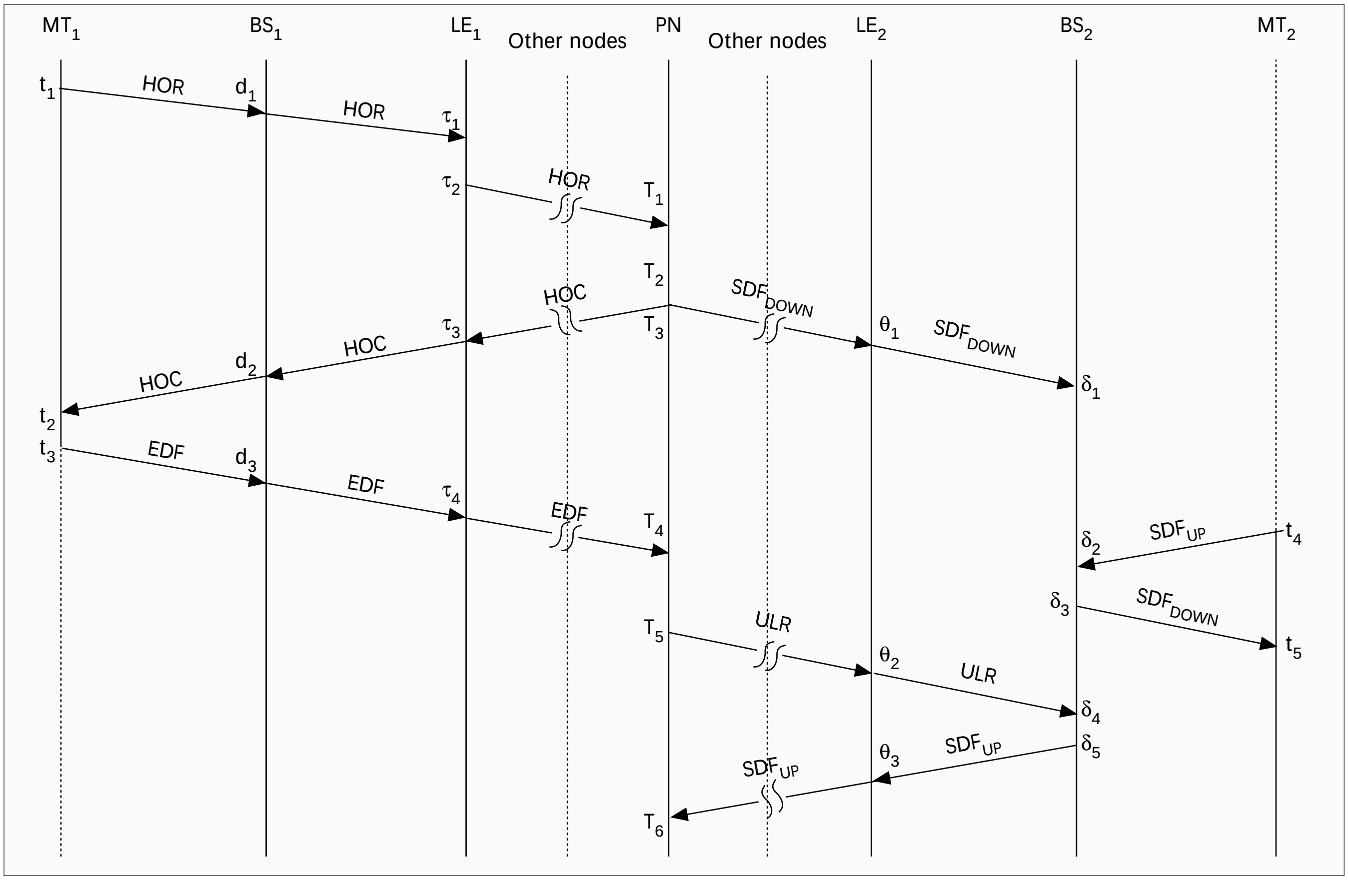
Figure 12 shows the temporal diagram of a network handover, assuming that the mobile terminal moves from the macrocell controlled by BS$_1$ to the one controlled by BS$_2$. Radio channel c_1 is used by the terminal to communicate with BS$_1$ while radio channel c_2 is used to communicate with BS$_2$; hence, in Fig. 12 MT is denoted by MT$_1$ and MT$_2$ before and after the radio handover, respectively.

The handover is driven by the mobile terminal, which starts the procedure when, for instance, the radio transmission conditions force it to switch from BS$_1$ to BS$_2$. In order to start the network handover procedure, the MT sends an HOR message over the connection. While waiting for the HOC message that confirms the acceptance to process the handover, the MT continues sending data cells using the original upstream connection. BS$_1$ copies the arriving MES cell containing the HOR message and immediately forwards it to the LE.

Upon reaching the LE, the HOR message is temporarily removed from the ATM cell flow, since the LE must parse it and decide whether the handover is local or global: in the former case the LE manages the handover; otherwise, the PN must be located, and the HOR message must be forwarded to the PN itself. The time lapse between τ_1 and τ_2 is indeed used

Octet	Field		Description	
	7 4	3 0		
1–5	Header		ATM header	
6	ID		Protocol identifier	
7	MES_TYPE		Message type	
8	FLOW_ID		Data flow identifier	
9	PRI	QoS_MOD	Priority	Quality of service modifier
10	PFL	N.D.	Pivot flag	Not defined
11–23	S_BSA		Source base station address	
24–36	T1_BSA		First destination base station address	
37–49	T2_BSA		Second destination base station address	
50	N.D.		Not defined	
51	N.D.		Not defined	
52	N.D.	CRC$_1$	Not defined	First part of the CRC code
53	CRC$_2$		Second part of the CRC code	

■ **Table 1.** *Proposed MES cell format.*



■ **Figure 12.** *Temporal diagram of a network handover using in-band signaling messages.*

by the network to identify the PN, be it the LE or any other upstream node. The details of this aspect of the handover procedure are not considered here, since the PN search algorithm is beyond the scope of this analysis. Details concerning algorithms that can be adopted for this purpose can be found in [3]; however, it can be argued that, given the connection route, the PN location depends only on the destination BS, not on the MT identity; hence, since the number of macrocells is limited, the PN search algorithm can be run off-line and a table with the PN location for each possible destination BS can be stored by LE, so the time lapse $\tau_1 - \tau_2$ can be kept small and almost constant.

The key functions in network handover procedures are performed by the PN upon receiving the HOR message; the interval $T_1 - T_2$ between HOR reception and HOC transmission is used by the PN to establish the incremental path toward BS_2. We assume that a make-break approach is used, so PN must establish both uplink and downlink connections toward BS_2, and verify that BS_2 has enough resources to perform the handover, at both the ATM and radio levels. Before sending the HOC message, the PN must update its downstream routing tables; the data transfer on the new route is initiated as soon as an ATM cell arrives at the PN on the downstream connection, and is preceded by the SDF_{DOWN} message; if, when the SDF_{DOWN} message reaches BS_2, the radio handover is not completed yet, ATM cells are stored at the BS. Notice that the proposed protocol limits the buffering of ATM cells during the handover at the destination BS only, while the source BS, and, most important, ATM nodes, never need to buffer ATM cells during network handovers. BS_2 stores ATM cells on the downlink in the *downstream handover buffer* during the interval δ_1–δ_3 .

Upon reception of the HOC message, the MT terminates the upstream transmission through BS_1 by sending the EDF message. This message informs the PN that no more upstream data cells will arrive through BS_1 and that the upstream routing table can be updated (the upstream connection from BS_2 has been opened together with the downstream one, so only the routing table update is needed); when the routing tables are updated, the PN sends the ULR message to BS_2, indicating that the upstream transmission on the wired segment can be resumed.

While the radio connection with the destination BS_2 is not yet established, upstream transmission is suspended at the MT that stores the generated data cells in the transmission buffer (during the time lapse between t_3 and t_4); when the radio connection with BS_2 is established, the MT sends the SDF_{UP} message and resumes its upstream transmission immediately, minimizing the buffering at the MT. Note, however, that if the MT radio interface has macrodiversity capabilities, the need for ATM cell buffering at the MT during handovers may become negligible.

BS_2 buffers SDF_{UP} and all the upstream data cells in the *upstream handover buffer* until the connection to the PN is established and enabled; this condition is signaled by means of the ULR message. When the uplink connection is ready, BS_2 forwards the SDF_{UP} message and immediately resumes the uplink transmission, so the maximum per-connection upstream handover buffer occupancy only depends on the interval between δ_2 and δ_5 . The handover buffers, both uplink and downlink, empty with a speed that depends on both the service rate of the wired and radio links, and the load offered by the mobile and remote terminals, respectively.

SDF_{UP} is also the implicit acknowledgment that the handover protocol is finishing; when it reaches the PN at time T_6 the procedure is over, and, if needed, a new network handover can be started for the same connection.