University of Surrey
School of Electronic Engineering, Information Technology and Mathematics
Centre for Communication Systems Research
Mobile Communications Research Group

Multiple Access Control

and

Mobility Management for
ATM Wireless LAN

by
Costas Apostolas

Thesis submitted to University of Surrey for the Degree of Doctor of Philosophy

© 27 February 1998
Dedicated to Dimitris and Yota
Summary

The concept of Wireless Asynchronous Transfer Mode (WATM) Networks has emerged recently and remains a topic of great research interest. The establishment of the WATM group [BASELINE] within ATM Forum [ATMForum] in 1996, the work carried out by the EU ACTS¹ projects such as WAND, SAMBA, MEDIAN [IPCOM'96], and the recent initiative on the Broadband Radio Access Network (BRAN) [BRAN] project within ETSI, justify the significance of deploying a wireless ATM network to carry multimedia traffic with support of user mobility or portability.

In Mobile Communications Research Group, CCSR, University of Surrey, the perspective of designing a wireless system based on the ATM protocols and mechanisms for bandwidth allocation, was adopted quite early [APOST'94]. The undergoing research was funded by the UK Department of Trade and Industry and EPSRC under the PCP-LINK project entitled ATM Wireless LAN. Part of the work presented in this thesis was carried out in the above mentioned project. Moreover, the results of the work were published in different publications and presented in several conferences and workshops [APOST'95], [APOST'96], [APOST'97], [IEEECOMMS],[BRUS'97].

In summary the major achievements of this work were:

• a set of proposed modifications on RIPERLAN PRM and its transmitted packet format to increase the transmission efficiency when ATM is supported on top of HIPERLAN,

• the definition of a new DECT DLC service optimised for transport of ATM traffic,

• a new dual mode PRM for ATM WLAN that provides support to the full range of ATM services in a cellular network configuration, but still enables the establishment of ad-hoc ATM WLAN,

• the design of a multiple access protocol for the MAC layer of ATM WLAN, capable to assign bandwidth on demand and support different types of services,

• the introduction of traffic policing functions in the MAC protocol to guarantee QoS to the established connections and prevent bandwidth violations by misbehaving users,

• the design of a radio hand-off protocol for terminals with multiple active connections and finally,

¹ Advanced Communication Technologies and Services
• the extension of the signalling protocols of core ATM network to support mobility management and in-call connection re-routing (network hand-off).

At this point the author would like to acknowledge the significant assistance that was provided by his supervisor Dr. Rahim Tafazolli and thank him for his advice, help and the opportunities he offered.
2.1 DECT overview

2.1.1 DECT DLC

2.1.2 DECT MAC

2.2 Proposed modifications on DECT for support of ATM WLAN concept

2.3 HIPERLAN overview

2.3.1 The physical layer of HIPERLAN

2.3.2 The Channel Access Control layer (CAC) of HIPERLAN

2.3.3 The MAC layer in HIPERLAN

2.4 Modifications to HIPERLAN for ATM support

2.4.1 Network Configuration for HIPERLAN based ATM WLAN

2.4.2 MAC scheme for ATM WLAN operation

2.4.3 Bandwidth allocation in HIPERLAN based ATM WLAN

2.5 Conclusion

CHAPTER 3

NEW ATM WLAN

3.1 Selection of MAC scheme

3.1.1 Candidate Multiple Access techniques for ATM WLAN

3.1.1.1 CSMA based mechanisms

3.1.1.2 FDMA

3.1.1.3 Fixed TDMA

3.1.1.4 CDMA

3.1.2 Advantages of a centralised approach

3.1.3 D-TDMA/FDD based MAC

3.1.3.1 Simulation Results

3.1.4 D-TDMA/TDD mechanism

3.1.5 Comparison between EY_NPMA and D-TDMA/TDD

3.2 Interfaces defined in ATM WLAN and corresponding PRM

3.2.1 Radio Transmission Specific Functions

3.2.2 Suitable Channel Access Cycle

3.2.3 Super-frame Delineation Function
CHAPTER 3

3.2.4 Packet Scheduling Function
3.2.5 Packet Transfer Function
3.2.6 Routing Information Exchange
3.2.7 Multiple UNI Controller
3.2.8 Air Interface for wireless M-NNI
3.2.9 ATM layer

3.3 ATM Adaptation Layer

3.4 Upper layers, Management layers

3.5 Conclusion

CHAPTER 4

MULTIPLE ACCESS SCHEME WITH TRAFFIC POLICING

FOR ATM WLAN

4.1 Introduction

4.2 Traffic Management in ATM Networks

4.2.1 Service Categories in ATM Networks

4.2.2 Traffic Policing functions in ATM Networks

4.2.2.1 Generic Cell Rate Algorithm

4.3 Traffic Policing in the MAC of ATM WLAN

4.3.1 QoS metrics in the ATM WLAN

4.3.2 Objectives of the proposed multiple access mechanism

4.3.3 Traffic patterns of connections

4.3.4 Exploiting silence and low bit rate periods of connections

4.3.5 Communication between terminals in the same cell

4.3.6 Re-negotiation of allocated bandwidth

4.3.7 Mechanism for QoS assurance

4.3.8 Air Interface to contain the QoS assurance mechanism
4.3.8.1 Description of the ATM WLAN model .................................................. 95
4.3.8.2 Simulation Results .............................................................................. 97

4.4 D-TDMA performance over the fading channel ................................. 98

4.5 Conclusion ......................................................................................... 101

CHAPTER 5 .................................................................................................. 102

SUPPORT OF TCP OVER ATM WLAN ............................................. 102

5.1 Introduction ..................................................................................... 102

5.2 TCP Principles ............................................................................. 103
  5.2.1 Measurement of Round Trip Time ............................................. 103
  5.2.2 Karn's Algorithm ......................................................................... 104
  5.2.3 Delayed ACKs ........................................................................... 104

5.3 TCP Congestion Control Scheme ...................................................... 105
  5.3.1 Slow Start ................................................................................... 105
  5.3.2 Congestion Avoidance ................................................................. 105
  5.3.3 Fast Retransmit ........................................................................... 106
  5.3.4 Fast Recovery ............................................................................ 106

5.4 Performance issues of TCP over ATM ............................................. 107

5.5 TCP Behaviour over Wireless Links ............................................... 108
  5.5.1 Indirect TCP ................................................................................ 109
  5.5.2 The Snoop Protocol .................................................................... 109
  5.5.3 Selective Acknowledgements ..................................................... 110

5.6 Simulation of TCP over ATM WLAN ............................................. 110
  5.6.1 Functions supported in the Simulation Model ................................. 111
    5.6.1.1 Application layer .................................................................... 111
    5.6.1.2 TCP layer ............................................................................. 111
    5.6.1.3 ATM Adaptation Layer .......................................................... 111
    5.6.1.4 ATM ................................................................................... 112
    5.6.1.5 WATM MAC ......................................................................... 112
    5.6.1.6 Tx/Rx Radio .......................................................................... 112
  5.6.2 Simulation Results ......................................................................... 112
    5.6.2.2 TCP vs. Cell Loss Ratio and Retransmission ......................... 114
List of Figures

Fig. 1.1 Study Areas ................................................................. 17
Fig. 1.2 Topology paradigm for private ATM LAN ........................................ 20
Fig. 1.3 Protocol Reference Model (PRM) of ATM LAN .......................... 22
Fig. 1.4 Standardised ATM cell formats ........................................... 22
Fig. 1.5 Topology paradigm for private ATM WLAN ........................... 29
Fig. 1.6 Protocol Reference Model (PRM) of ATM WLAN ....................... 30
Fig. 1.7 M-UNI as ATM switch .................................................... 31
Fig. 2.1 Possible configuration for DECT system .................................. 35
Fig. 2.2 DECT PRM .................................................................. 36
Fig. 2.3 PRM for ATM WLAN based on DECT ................................... 37
Fig. 2.4 System architecture for DECT based ATM WLAN .................... 38
Fig. 2.5 HIPERLAN Reference Model .................................................. 41
Fig. 2.6 Format of the transmitted packet in HIPERLAN ....................... 42
Fig. 2.7 Channel Access Cycle in HIPERLAN ..................................... 43
Fig. 2.8 Dropping Probability ($P_d$) for conforming and non-conforming connections in HIPERLAN versus amount of excessive bandwidth 
Fig. 2.9 Scenario for HIPERLAN based ATM WLAN configuration ............. 47
Fig. 2.10 PDU formats for the contention free period of the super-frame on the air interface of HIPERLAN based ATM WLAN 
Fig. 2.11 PRM for HIPERLAN based ATM WLAN applied between terminals and BSs .... 51
Fig. 2.12 Packet dropping probability for 1 Mbps VBR traffic .................... 53
Fig. 2.13 Mean access delays .................................................................. 53
Fig. 2.14 Dropping Probability ($P_d$) for conforming and non-conforming connections in HIPERLAN based ATM WLAN, versus amount of excessive bandwidth 
Fig. 3.1 Hidden terminal problem ...................................................... 56
Fig. 3.2 Modified ATM cell header for the ATM radio LAN .................. 61
Fig. 3.3 Uplink transmissions example ............................................... 62
Fig. 3.4 Traffic Models used in simulations ........................................... 64
Fig. 3.5 Simulation Results .................................................................. 65
Fig. 3.6 Structure of the air interface in the D-TDMA scenario .................. 66
Fig. 3.7 Simulation results for VBR sources ............................................ 67
Fig. 3.8 Network configuration for ATM WLAN based on fixed infrastructure 69
Fig. 3.9 Fully wireless ATM WLAN configuration .................................. 71
Fig. 3.10 Lower layers of ATM WLAN PRM ......................................... 72
Fig. 3.11 Frame structure of the channel between terminals and MUX ........................................75
Fig. 3.12 Frame structure of the channel between switch and MUX or between switches ..........76
Fig. 3.13 MAC-ATM layer ........................................................................................................77
Fig. 4.1 Two equivalent versions of the GCRA ........................................................................85
Fig. 4.2 Air Interface of ATM WLAN .....................................................................................95
Fig. 4.3 (a) Structure of sub-slots in the air interface of the ATM WLAN ...............................96
(b) structure of the transmitted packet
Fig. 4.4 Mean Access Delay and Packet Dropping Probability ..............................................98
in ATM WLAN for 2 Mbps VBR sources
Fig. 4.5 Mean Access Delay and Packet Dropping Probability in ATM WLAN ...................98
for conforming and on-conforming (2.5...4 Mbps) VBR sources
Fig. 4.6 Optimised Air interface for ATM WLAN to cope with packet re-transmissions .......99
Fig. 4.7 Performance of the ATM WLAN MAC scheme versus introduced BER ...............100
Fig. 4.8 Traffic slot utilisation for the ATM WLAN MAC scheme versus BER ..................101
Fig. 5.1 (a) Network Configuration for TCP over WATM .....................................................111
(b) Protocol Reference Model of Network Entities
Fig. 5.2 TCP segment profile ..................................................................................................113
Fig. 5.3 TCP layer throughput ...............................................................................................113
Fig. 5.4 TCP offered traffic .....................................................................................................113
Fig. 5.5 ATM layer throughput .............................................................................................113
Fig. 5.6 TCP throughput vs. CLR ..........................................................................................114
Fig. 5.7 TCP offered traffic vs. CLR ......................................................................................114
Fig. 5.8 ATM throughput vs. CLR ..........................................................................................114
Fig. 6.1 Introduced F5 OAM packets to assist the hand-off procedures in ATM WLAN .. 118
Fig. 6.2 Signalling for Intra-Cell Hand-off ...........................................................................120
Fig. 6.3 Signalling for Inter Cell Hand-off .............................................................................121
Fig. 6.4 Protocol Reference Model for WAM-net .................................................................124
Fig. 6.5 WATM-UNI configuration for address registration ..............................................126
Fig. 6.6 Address registration signalling involving one access switch .................................127
Fig. 6.7 Address registration signalling involving two access switches ..........................128
Fig. 6.8 Hierarchical organisation of the P-NNI routing domain and address reachability .129
Fig. 6.9 Example of WATM-NNI routing domain up to City Level ..................................133
Fig. 6.10 Location Update signalling example for a MT ......................................................136
Fig 6.11 Signalling for Hand-off and In-call connection re-routing .....................................141
### List of Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAL</td>
<td>ATM Adaptation Layer</td>
</tr>
<tr>
<td>ABR</td>
<td>Available Bit Rate</td>
</tr>
<tr>
<td>ACK</td>
<td>Acknowledgement</td>
</tr>
<tr>
<td>AP</td>
<td>Access Point</td>
</tr>
<tr>
<td>ATDM</td>
<td>Asynchronous Time Division Multiplexing</td>
</tr>
<tr>
<td>ATM</td>
<td>Asynchronous Transfer Mode</td>
</tr>
<tr>
<td>BER</td>
<td>Bit Error Rate</td>
</tr>
<tr>
<td>B-ISDN</td>
<td>Broadband Integrated Services Digital Network</td>
</tr>
<tr>
<td>BRAN</td>
<td>Broadband Radio Access Network</td>
</tr>
<tr>
<td>BS</td>
<td>Base Station</td>
</tr>
<tr>
<td>CAC</td>
<td>Channel Access Cycle</td>
</tr>
<tr>
<td>CBR</td>
<td>Constant Bit rate</td>
</tr>
<tr>
<td>CDT</td>
<td>Cell Delay Tolerance</td>
</tr>
<tr>
<td>CDVT</td>
<td>Cell Delay Variation Tolerance</td>
</tr>
<tr>
<td>CEPT</td>
<td>Conference Europenne des Postes et des Telecommunications</td>
</tr>
<tr>
<td>CLR</td>
<td>Cell Loss Rate</td>
</tr>
<tr>
<td>CPE</td>
<td>Customer Premises Equipment</td>
</tr>
<tr>
<td>DECT</td>
<td>Digital European Cordless Telecommunications</td>
</tr>
<tr>
<td>DLL</td>
<td>Data Link Layer</td>
</tr>
<tr>
<td>DTL</td>
<td>Designated Transit List</td>
</tr>
<tr>
<td>EPD</td>
<td>Early Packet Discard</td>
</tr>
<tr>
<td>ETSI</td>
<td>European Telecommunication Standard Institute</td>
</tr>
<tr>
<td>EY_NPMA</td>
<td>Elimination Yield Non Priemptive priority Multiple Access</td>
</tr>
<tr>
<td>FCFS</td>
<td>First Come First Served</td>
</tr>
<tr>
<td>FDD</td>
<td>Frequency Division Duplex</td>
</tr>
<tr>
<td>GCRA</td>
<td>Generic Cell Rate Algorithm</td>
</tr>
<tr>
<td>HIPERLAN</td>
<td>High Performance Radio Local Area Network</td>
</tr>
<tr>
<td>ILMI</td>
<td>Interim Local Management Interface</td>
</tr>
<tr>
<td>IMT2000</td>
<td>International Mobile Telecommunications 2000</td>
</tr>
<tr>
<td>IP</td>
<td>Internet Protocol</td>
</tr>
<tr>
<td>ISDN</td>
<td>Integrated Services Digital Network</td>
</tr>
<tr>
<td>ISI</td>
<td>Inter-Symbol Interference</td>
</tr>
<tr>
<td>LAN</td>
<td>Local Area Network</td>
</tr>
<tr>
<td>LMIB</td>
<td>Location Management Information Base</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>LN</td>
<td>Logical Node</td>
</tr>
<tr>
<td>MPEG</td>
<td>Moving Picture Expert Group</td>
</tr>
<tr>
<td>MT</td>
<td>Mobile Terminal</td>
</tr>
<tr>
<td>MUX</td>
<td>Multiplexer</td>
</tr>
<tr>
<td>NNI</td>
<td>Network-Network Interface</td>
</tr>
<tr>
<td>OAM</td>
<td>Operation Administration and Maintenance</td>
</tr>
<tr>
<td>PCS</td>
<td>Personal Communications System</td>
</tr>
<tr>
<td>PDU</td>
<td>Protocol Data Unit</td>
</tr>
<tr>
<td>PER</td>
<td>Packet Error Rate</td>
</tr>
<tr>
<td>PGL</td>
<td>Peer Group Leader</td>
</tr>
<tr>
<td>P-NNI</td>
<td>Private Network-Network Interface</td>
</tr>
<tr>
<td>PPD</td>
<td>Partial Packet Discard</td>
</tr>
<tr>
<td>PRM</td>
<td>Protocol Reference Model</td>
</tr>
<tr>
<td>PRMA</td>
<td>Packet Reservation Multiple Access</td>
</tr>
<tr>
<td>PSF</td>
<td>Packet Scheduling Function</td>
</tr>
<tr>
<td>PTSP</td>
<td>P-NNI Topology State Packet</td>
</tr>
<tr>
<td>PVC</td>
<td>Permanent Virtual Connection</td>
</tr>
<tr>
<td>QoS</td>
<td>Quality of Service</td>
</tr>
<tr>
<td>RTT</td>
<td>Round Trip Time</td>
</tr>
<tr>
<td>SAAL</td>
<td>Signalling ATM Adaptation Layer</td>
</tr>
<tr>
<td>SACK</td>
<td>Selective Acknowledgement</td>
</tr>
<tr>
<td>SAP</td>
<td>Service Access Point</td>
</tr>
<tr>
<td>SDU</td>
<td>Service Data Unit</td>
</tr>
<tr>
<td>SNMP</td>
<td>Small Network Management Protocol</td>
</tr>
<tr>
<td>SVC</td>
<td>Switched Virtual Connection</td>
</tr>
<tr>
<td>TCP</td>
<td>Transport Control Protocol</td>
</tr>
<tr>
<td>TDD</td>
<td>Time Division Duplex</td>
</tr>
<tr>
<td>UBR</td>
<td>Unspecified Bit Rate</td>
</tr>
<tr>
<td>UME</td>
<td>UNI Management Entity</td>
</tr>
<tr>
<td>UMTS</td>
<td>Universal Mobile Telecommunication System</td>
</tr>
<tr>
<td>UNI</td>
<td>User-Network-Interface</td>
</tr>
<tr>
<td>VBR</td>
<td>Variable Bit Rate</td>
</tr>
<tr>
<td>VBR-nrt</td>
<td>Variable Bit rate non real time</td>
</tr>
<tr>
<td>VBR-rt</td>
<td>Variable Bit Rate real time</td>
</tr>
<tr>
<td>WATM-UI</td>
<td>Wireless ATM User Identifier</td>
</tr>
<tr>
<td>WPBX</td>
<td>Wireless Private Branch Exchange</td>
</tr>
</tbody>
</table>
Introduction

ATM networks offer certain advantages when compared to traditional LAN technologies like Ethernet and FDDI, namely capability to provide Quality of Service (QoS) guarantees and efficient multiplexing of different application types over the same physical medium. Moreover, wireless access eases relocation of terminals and expansion of networks, and enables user mobility and equipment portability. Moreover, the wireless technology is the enabling factor for Universal Personal Communications. However, existing cellular systems [IS95], [GSM], [DCS1800] deal mainly with voice communication. Moreover, DECT does not offer the high transmission rate required for broadband multimedia communications, and standards like IEEE802.11 and HIPERLAN, support high bit rates but they do not provide for mobility. Furthermore, they do not support QoS. In that case the merging of wireless and ATM technologies could lead to Universal Personal Multimedia Communications. This concept is envisioned by systems like UMTS, IMT2000 and BRAN.

The objective of the work presented in this document is the provision of the migration path from existing wireless LAN standards to ATM Wireless LAN (ATM WLAN), which is envisioned as a system with small and densely populated wireless access parts and an ATM backbone network that could span in a geographical area justifying the term LAN. Of course, some of the presented solutions could be adopted in Wide Area Networks too.

The first chapter introduces the characteristics and requirements of the services supported in ATM networks. The same service characteristics must be supported in the ATM WLAN. Moreover, the concept of ATM LAN is outlined, its Protocol Reference Model (PRM) is presented and its main functions are highlighted. Furthermore, chapter 1 continues by identifying the required additions and modifications of the ATM LAN PRM, and setting the design objectives for the work related to ATM WLAN.

The solutions proposed in this work were influenced by the study of existing or proposed standards for wireless communications. DECT and HIPERLAN were investigated because they target at the same market as ATM WLAN, namely local area wireless networking. Their architecture and protocols are outlined in Chapter 2. Both DECT and HIPERLAN offer multiple access schemes that could be exploited to support the ATM WLAN. In Chapter 2 modifications to DECT PRM are proposed to support the ATM WLAN concept over DECT. Moreover, it is found that encapsulation of ATM cells (packets) in HIPERLAN PDUs combined with the HIPERLAN MAC scheme results in low transmission
Introduction

efficiency. Thus, in Chapter 2 necessary modifications are proposed on the HIPERLAN PRM and packet format, to increase its transmission efficiency for ATM WLAN operation.

ATM networks are characterised by point-to-point links leading to star based topologies, at least for the access part of the network. This issue affects the selection of the multiple access scheme for ATM WLAN and favours centralised adaptive schemes like these presented in Chapter 2. However, terminal portability and user mobility enable the establishment of ad-hoc networks. Then distributed, contention based schemes like these found in HIPERLAN and IEEE 802.11 could be deployed. Chapter 3 discusses the advantages and disadvantages of potential candidates for the ATM WLAN MAC and shows how the work on HIPERLAN (presented in Chapter 2) can be extended to define a dual mode system. A system that performs optimally in a cellular configuration and is able to emulate the ATM WLAN service within an ad-hoc network. Moreover, Chapter 3 compares the performance of HIPERLAN multiple access scheme and the performance of an adaptive TDMA based scheme proposed for the ATM WLAN.

The proposed adaptive TDMA scheme outperforms HIPERLAN MAC in terms of user capacity. Moreover, it is capable of applying traffic policing functions on the air-interface traffic. This means that QoS provisioning is enabled. Then the concept of traffic monitoring and QoS provision as included in ATM networks and described in chapter 4, is passed to ATM WLAN. Furthermore, Chapter 4 discusses the issues that affected the selection of the proposed MAC scheme and shows how ATM traffic classes are mapped onto ATM WLAN MAC traffic. Moreover, it verifies the capability of the scheme to preserve bandwidth of connections and guarantee QoS requirements.

Moreover, chapter 5 examines the performance of the Transport Control Protocol (TCP) over the adopted MAC scheme for the ATM WLAN. TCP is the main Transport Layer protocol to enable Internet applications. ATM WLAN should provide for TCP support, and deal with the problem of TCP throughput degradation. The latter arises because of congestion and ATM buffer overflows when TCP resides on top of the ATM PRM, and because of TCP segment losses when TCP is used over wireless links. In chapter 5 it is shown that by adopting an ATM cell retransmission policy, the throughput of the TCP layer can approach the maximum achievable. Moreover, by implementing Early Packet Discard in the MAC of ATM WLAN there is an increase of TCP throughput when the network is driven in congestion.

Furthermore, chapter 6 describes initially how the MAC of ATM WLAN is used to support seamless hand-off for a mobile terminal. The hand-off procedures are based on the Operation Administration and Maintenance (OAM) functions that exist already within the
Introduction

PRM of fixed ATM LANs. The above mentioned hand-off mechanism invokes procedures on the air interface of the ATM WLAN. However, it is highly likely that connections will have to be re-routed in the fixed backbone network. Chapter 6 describes a mechanism based on existing PNNI specifications to provide in call connection re-routing, and location management for mobile ATM users. The described algorithm which resides above the MAC layer of ATM WLAN requires minimum information by the lower layers enabling it to be used with different possible implementations of MAC for the ATM WLAN.

Finally, chapter 7 discusses the conclusions of this work and sets further research issues in the area of wireless ATM.
Chapter 1

Background

The functionality of the access mechanisms in ATM WLAN is based on the ATM concept and follows the principles governing ATM oriented networks. In this chapter the transition issues from fixed ATM networks to wireless/mobile ATM are presented, as well as the range of services expected to be supported in ATM WLAN. The fixed ATM protocol reference model, being the starting point in this study, is described and the protocols that were modified to suit the Wireless LAN concept are discussed. Moreover, an initial baseline MAC layer is presented to set the design objectives for the multiple access scheme of ATM WLAN.

1.1 ATM Combined benefits

ATM technology is characterised by four main principles:

1. point-to-point connections among the network entities and separation of all the functions into user and network part,

2. the existence of the call set-up phase for every connection, when traffic patterns and traffic limits of the particular connection are specified, and Quality of Service (QoS) parameters are assigned values appropriate for the supported service,

3. the allocation of bandwidth to established connections dynamically and according to the instantaneous traffic emission by sources, following the Asynchronous Time Division Multiplexing (ATDM) concept, and finally,
4. the introduction of functions monitoring the traffic per connection and based on the specified QoS values and traffic limits for the corresponding connection.

The combination of the above mentioned principles enables an ATM network to support and offer acceptable quality of service for different kinds of services over the same infrastructure, while avoiding congestion.

Moreover, regarding the exploitation of wireless technology in the Local Area Network (LAN) domain, the ease of reallocation of equipment and the capability of mobility within the office, factory or company environment are the main benefits to be gained.

Thus, combined benefits appear when introducing both ATM and wireless concept in the LAN domain. These are:

- provision of new powerful network technology with QoS objectives to support new applications in a LAN,
- maximisation of the bandwidth efficiency of the network because of the ATDM, and increase of the system capacity,
- feasible compatibility with the ATM oriented Broadband Integrated Services Digital Network (B-ISDN),
- minimisation of the costs for network reconfiguration, reallocation and expansion, and
- mobility and portability in the working environment and within B-ISDN.

1.2 Approach

In the migration path to ATM WLAN several study areas are defined as presented in Fig 1.1. The ATM WLAN concept has been the core of the related work. One of the objectives was to support as many characteristics as possible from the fixed ATM concept to the ATM WLAN concept. So, the main part of the effort was concentrated on the fixed ATM networks to understand the underlying techniques for traffic control and call handling. These techniques are described later in this chapter and in chapter 4. The result of this study was a set of requirements that the ATM WLAN should satisfy to justify its title.

However, aiming to use the radio as the transmission medium and deliver ATM services to mobile users, ATM WLAN must integrate solutions related to mobility management, call re-routing, and hand-off.

Moreover, within ETSI, there is a lot of activity related to the specification of wireless standards and there are two main representative outputs. One is Digital European Cordless Communications (DECT) standard which was approved in 1995 [DECT1] and the other is High Performance Local Area Network (HIPERLAN) standardised in 1997.
Chapter 1: Background

[HIPERLAN'95]. DECT, offering a comparatively low transmission rate focuses in the areas of Wireless Private Branch Exchanges (WPBX) for voice communications and file transmissions at low transport rates. Despite of its limited capabilities because of the low transmission rate, DECT is a complete proposal for indoor wireless voice and data communications and could be regarded as the first European wireless protocol architecture for the indoor environment. Moreover, HIPERLAN appears as the wireless substitution of Ethernet based fixed networks for data communications with transfer rates around 20 Mbps. Like Ethernet, HIPERLAN could be used to carry voice or video connections but, as shown in chapter 2, with limited capabilities. However, in both DECT and HIPERLAN cases the related specifications [DECT1], [DECT2], [DECT3], [HIPERLAN’95] were studied and the standards were modified to approach the ATM WLAN functionality.

1.3 Services to be supported by ATM WLAN

ATM WLAN is regarded as an autonomous ATM system or network and not only as a wireless extension of fixed ATM networks. However, as it will be shown later in this chapter, a solution was provided which integrates both fixed and wireless User-Network Interfaces (UNIs) in the same backbone ATM oriented network. Under this assumption it is expected to come across with a design that will support all the traffic classes as specified in [TM'96], and all the applications supported or being under deployment in fixed ATM networks.
Briefly, the aim is to support telephony, fast transfers of files, distributed computing and multimedia to the desktop. The bit rate requirements for these services combined with the expected population of ATM WLAN customers defined the transfer rates in the ATM WLAN. Moreover, the QoS requirements for the above mentioned application classes influenced the mechanisms related to the bandwidth allocation and error control.

Table 1.1 Services to be supported in ATM WLAN

<table>
<thead>
<tr>
<th>Application</th>
<th>Bit Rate (Kbps)</th>
<th>BER(^1)</th>
<th>PER(^2)</th>
<th>Call attempts in small business (50 users)</th>
<th>Call attempts in medium business (150 users)</th>
<th>Call attempts in large business (4700 users)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Telephony</td>
<td>32, 64</td>
<td>10(^{-7})</td>
<td>10(^{-3})</td>
<td>5</td>
<td>40</td>
<td>200</td>
</tr>
<tr>
<td>Videophone</td>
<td>64 - 2000</td>
<td>1.3 X 10(^{-6})</td>
<td>8 X 10(^{-6})</td>
<td>1.5</td>
<td>10</td>
<td>50</td>
</tr>
<tr>
<td>Video conference</td>
<td>5000</td>
<td>1.8 X 10(^{-6})</td>
<td>5 X 10(^{-6})</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>MPEG1</td>
<td>1500</td>
<td>2.5 X 10(^{-6})</td>
<td>9.5 X 10(^{-6})</td>
<td>0.5</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>MPEG2</td>
<td>10000</td>
<td>1.5 X 10(^{-6})</td>
<td>4 X 10(^{-6})</td>
<td>0.5</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>Hi-Fi distribution</td>
<td>400</td>
<td>10(^{-5})</td>
<td>10(^{-7})</td>
<td>0.9</td>
<td>0.9</td>
<td>0.9</td>
</tr>
<tr>
<td>Low speed file</td>
<td>64</td>
<td>10(^{-7})</td>
<td>10(^{-6})</td>
<td>1</td>
<td>10</td>
<td>50</td>
</tr>
<tr>
<td>High speed file</td>
<td>2000</td>
<td>10(^{-7})</td>
<td>10(^{-6})</td>
<td>3</td>
<td>20</td>
<td>100</td>
</tr>
</tbody>
</table>

In Table 1.1 [ONV'94] the average bit rate as well as the bit and packet error rate (BER, PER) requirements for some applications are presented. In this table, the number of call attempts per user during the busy hour is also shown. In the scope of the ATM LAN (fixed or wireless), the term call does not imply only a telephony service but an established connection (or set of connections) carrying traffic for some application. To have a complete view of the service requirements Table 1.2 shows service end-to-end delay requirements and their jitter requirements (i.e. the difference between the actual inter-arrival times of two subsequent packets of a connection at two points of the network compared to the expected ones). Furthermore, when multimedia services are considered in the network, a new problem arises called skew. It is defined as the relative delay between two packets that carry different

\(^1\) with single-bit error correction per packet and additional error correction in the Transport Layer (AAL) of B-ISDN
components of a multimedia connection (e.g. voice and video). Skew objectives for multimedia applications are presented in [ONV'94].

Table 1.2 Timing requirements of some of the supported services

<table>
<thead>
<tr>
<th>Application</th>
<th>End-to-end delay (msecs)</th>
<th>End-to-end packet delay variation (msecs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 Mbps HDTV video</td>
<td>0.8</td>
<td>1</td>
</tr>
<tr>
<td>Telephony with echo canceler</td>
<td>&lt;500</td>
<td>130</td>
</tr>
<tr>
<td>Videophone (64 kbps)</td>
<td>300</td>
<td>130</td>
</tr>
<tr>
<td>MPEG1 (NTSC)</td>
<td>5</td>
<td>6.5</td>
</tr>
</tbody>
</table>

1.4 Fixed ATM LAN concept

1.4.1 Topology and defined network interfaces

The protocol architecture and the topology of an ATM LAN is inherited from the B-ISDN protocol reference model and architecture [UNI3]. Fig. 1.2 presents a topology scenario of an ATM LAN and shows the different signalling protocols executed among the network entities. Three main interfaces\(^2\) are defined, namely Public User-to-Network Interface (public UNI), Private User-to-Network Interface (private UNI) and private Network-to-Network Interface (P-NNI). Public UNI is defined between a Customer Premises Equipment\(^3\) (CPE) and an ATM switch or multiplexer in a public domain ATM network. Private UNI is defined between a user terminal and an ATM switch of the same private ATM LAN (possibly an ATM oriented intra-net). Finally, P-NNI is defined among the switches that form the backbone part of a private ATM LAN.

Private and public UNI specifications cover:
- the electrical standards (cable types, transmission rates, modulation schemes, line coding, etc.) that must be used to interconnect two ATM devices,
- the error control functions, as well as the functions related to the scrambling of the packet payload

---
\(^2\) set of specifications covering the interoperability issues between two network entities (including physical medium, signalling and operations at the attached entities)
\(^3\) can be a user terminal or an ATM switch in a private ATM LAN
• the signalling and the functions that are related to call set-up (because in ATM networks every connection establishment is preceded by a call set-up phase), as well as, the addition of communicating parties in an established call or their rejection from it,

• the signalling, the data bases and the operations required for the management of an ATM LAN at the UNI scope (address allocation to terminals, definition of allowable number of established connections, specification of the electrical interface etc.),

• the functions carried out to monitor the traffic generated by the user terminals, avoid congestion, and allocate bandwidth efficiently while satisfying the QoS of the active calls.

Fig. 1.2 Topology paradigm for private ATM LAN

The key components of the UNI functionality will be presented later in this chapter. At this point, one could notice that the design of an ATM WLAN requires the modification of some UNI specifications (starting from the physical medium) to come across with a Wireless or Mobile UNI implementation (W-UNI or M-UNI).

Furthermore, P-NNI specifications cover:
the organisation of ATM switches into logical nodes and layers of an hierarchical (in terms of addressing and routing) backbone network structure,

- the functions and the signalling related to the topology update within the backbone part of the ATM LAN,

- the procedures for the call routing during the set-up phase and the exchanged signalling,

- the functions, data base specifications and signalling related to the backbone network management.

The P-NNI mechanisms that were studied and modified will be presented in chapter 6 where it is shown how P-NNI protocols could be adapted to support mobility management, in-call connection re-routing and call set-up re-routing because of mobility.

1.4.2 Protocol Reference Model (PRM)

Fig. 1.3 presents the protocol reference model (PRM) in the fixed ATM LAN which is adopted from the B-ISDN equivalent model [UNI31]. This model applies in both UNI and NNI within an ATM LAN. Layer Management deals with the control and maintenance of every horizontal layer independently, and for this purpose peer-to-peer messages have been defined in every layer, as well as the corresponding Operation Administration and Maintenance (OAM) functions of that layer. Plane Management is related to the control of the protocol reference model in a vertical manner, by providing to every layer information related to other layers. This happens by defining a higher layer protocol called Interim Layer Management Interface (ILMI) which is responsible for the maintenance of information elements related to all the layers of PRM, as well as for the exchange of these element values among the network entities. C-plane contains the protocols related to call establishment and termination, while U-plane is responsible for carrying and monitoring the traffic generated by user applications.
Chapter I: Background

Physical layer accepts from and delivers to ATM layer packets of fixed length and structure called cells. The cell format is one of the most representative characteristics of ATM networks. Cell formats across UNI and NNI are presented in Fig. 1.4.

Physical layer consists of two sub-layers:

- Transmission Convergence sub-layer dealing with error correction at the cell header, construction or recovery of frame

Fig. 1.3 Protocol Reference Model (PRM) of ATM LAN

Fig. 1.4 standardised ATM cell formats
structures that may exist in the transmission link (e.g. SDH), scrambling of the cell payload contents and cell delineation (i.e. identification of cell boundaries within transmitted frames),

- Physical Medium Dependent sub-layer which deals with the transmission issues (e.g. modulation, line coding, bit synchronisation).

All physical layer specifications define fixed transmission mediums (fibre optics, twisted pairs, coaxial cable [UNI31]). The links are point-to-point, as shown in Fig. 1.2 and there is no need for Multiple Access Sub-layer algorithms, as happens in traditional LANs (Ethernet, Token ring, FDDI) [LEE'93], [TANENB'88], because different transmission directions are physically separated (different cable is used in each direction). However, in the case of ATM WLAN the broadcast nature of radio transmission requires the implementation of a multiple access scheme to regulate the allocation of the radio channel. As the access of the link corresponds to bandwidth allocation we could initially state that the MAC sub-layer resides in the TC sub-layer. However, it also expands in the ATM layer of the ATM WLAN, because, as it will be shown, it uses mechanisms residing in ATM layer to control the bandwidth allocation and monitor the offered traffic.

1.4.2.2 ATM Layer

ATM layer is the heart of the PRM of an ATM LAN and B-ISDN in general. All the main functions that characterise an ATM oriented network reside in this layer. ATM exchanges with AAL fixed size (48 bytes) SDUs\(^4\) which form the ATM cell payload. The header of the ATM cell is added by the ATM layer and it contains information to:

- identify uniquely the connection where the cell belongs,
- indicate that the cell is redundant when congestion occurs in the network, and
- distinguish between cells that carry Plane Management or Layer Management messages, and C-plane or U-plane information.

ATM cells are exchanged between ATM and PHY layer where errors in the header are handled [UNI31].

Moreover, ATM layer performs bandwidth allocation and statistical multiplexing. Bandwidth allocation is based on the ATDM concept. Generally in ATM networks bandwidth allocation is performed in time. In every link there is a maximum transfer capability expressed in cells/sec which can be viewed as a continuous transmission of cell
slots. The filling of cell slots by actual cells is performed asynchronously and on demand basis. There are no predefined time slots allocated to a connection. The straightforward result is that the bandwidth allocation mechanisms can follow the traffic fluctuations of VBR connections. Moreover, bandwidth is not reserved for connections when it is not required, but it can be shared among them. Assuming that the instantaneous total required bandwidth does not exceed the total link capacity or the ATM switch capacity, compared to fixed bandwidth allocation mechanisms, ATM can support more connections having the same total bandwidth in the network. This is called statistical multiplexing gain⁵ and it means better bandwidth exploitation than fixed bandwidth allocation. Thus, several new applications can be supported in a bandwidth efficient way, increasing the network capacity and possibly reducing the service cost.

However, bandwidth allocation occurs not only in the links among network entities but also in the ATM switches in terms of buffer space and bus-bandwidth reservation. As ATM cells flow from the user terminals deep in the backbone network, the total traffic offered in a switch could overload it (or some of its ports). Then, the network is driven into congestion locally and neighbouring switches are affected too. To avoid such situations, traffic monitoring and control mechanisms have been introduced in the ATM layer to prevent [PERROS'96] and react to congestion [TM'96]. The main idea is that for every new connection there is a call set-up phase when the source specifies its QoS requirements and traffic pattern, and negotiates with the network about its traffic limits. In chapter 4, what is negotiated during the call set-up phase, is described in more details. To prevent congestion, the network executing the Call Admission Control algorithm, decides to accept or reject the new call (based on the expected satisfaction of the QoS parameters of the established calls and the new call) and defines the traffic limits of the sources of an accepted call. These traffic limits are fed into the traffic policing functions which decide for every generated cell in a connection if it could lead the network to congestion (because its source exceeds its traffic limits), and if yes the cell is tagged as redundant. When congestion occurs (or is about to occur) tagged cells are discarded.

In the case of ATM WLAN, the traffic monitoring functions are of great importance, because the radio channel is a limited and expensive resource. Moreover, it is very important to monitor the allocation of radio channel capacity as it has a direct impact on the QoS of the established connections. The access of the radio link by a source must take place according to the source traffic rate limits and without violating the bandwidth that should be allocated

⁴ Service Data Units
to other established connections. As presented in chapter 2, this mechanism is not adopted in HIPERLAN where bandwidth violations could occur. However, in chapter 4 it will be shown that the multiple access mechanism of ATM WLAN is based on the monitoring of the bandwidth allocation requests and on a negotiated traffic contract.

1.4.2.3 ATM Adaptation Layer (AAL)

Fig. 1.3 shows also the ATM Adaptation layer (AAL) which is the end-to-end transport protocol in the ATM networks and deals with the delivery of synchronisation information between the calling parties of a call and the error detection and possibly correction in the transmitted stream. Moreover, it segments if required, the information delivered by the higher layers to create the ATM cell payload which is delivered to the ATM layer. Different, AALs have been defined according to the service types [ONV'95]. This happens because ATM networks multiplex several service types with different traffic characteristics and different requirements regarding end-to-end delay, bit or cell error rate, and clock provision among the communicating parties at the application layer. The selection of the appropriate AAL is declared during call set-up phase. It should be noted that AAL is the last layer downwards in the ATM PRM which differentiates connections based on the corresponding service type. The information about the kind of the supported service is lost when a Service Data Unit passes to the ATM layer. However, during call set-up ATM layer learns about the traffic class and the QoS class for every established connection. In chapter 4 it is discussed how the traffic class and the QoS class of a new connection can be used to feed the required information in the mechanism that deals with the allocation of radio channel capacity to active sources.

1.4.2.4 C-Plane

In the higher layers the main component of this plane is the Q.2931 [UNI31] protocol adopted from the ISDN technology. It defines the signalling that takes place among network entities to initiate and terminate connections, to monitor the status of calls and increase or decrease the number of communicating parties in point-to-multipoint connections. Currently, Q.2931 supports point-to-point and point-to-multipoint connections only. Moreover, it supports the establishment of permanent or semi-permanent virtual connections (PVCs), as well as the establishment of switched virtual connections (SVCs).\(^6\) In the ATM Adaptation Layer (AAL) the C-plane functions are carried out by the Signalling AAL (SAAL), a connectionless transport protocol presented in [UNI4]. As specified in [UNI31], all the

---

5 the term statistical multiplexing reflects the expectation of having all the connections asking simultaneously for bandwidth or transmitting simultaneously at their peak rate.

6 the main differentiation among PVCs and SVCs considers the lifetime of the connection
signalling messages between an ATM terminal and an ATM switch or multiplexer are carried by ATM cells with VPI and VCI values that identify the control channel of the specific UNI [ONV'94].

1.4.2.5 Plane Management

While Q.2931 provides the required signalling for call control (set-up, termination, monitoring), Interim Local Management Interface (ILMI) defines the procedures required for the management of the UNI and the related signalling. ILMI defines a set of managed entities that describe a UNI and organises them into a data base called Management Information Base (MIB). This data base is distributed at the user side and network side of the UNI and provides configuration and status information related to the specific UNI. The protocol that is used to carry the required signalling between the user and network side of the UNI, and interrogate or modify MIB is the Simple Network Management Protocol (SNMP) residing on top of SAAL. SNMP messages [WASH'94] are carried by cells with well known VPI and VCI values (specified in [UNI31]). Moreover, ILMI specifies the address registration procedure for a terminal and in chapter 6, it is discussed how this procedure is used in the case of hand-off and location update procedures.

1.5 Differences between ATM LAN and ATM WLAN

The objective of this work was to define an indoor wireless LAN following the ATM concept i.e. employing the representative mechanisms of fixed ATM LANs, to come across with an efficient solution considering adaptive bandwidth allocation, congestion avoidance, preservation of QoS for established connections and support of different kinds of services and multimedia applications over the radio link. Since ATM WLAN, is expected to have similar architecture and protocol family and support the same set of services as ATM LAN, the above mentioned protocols and mechanisms should reside in ATM WLAN architecture, of course modified, to adapt to the wireless and mobility enhancing environment. However, this transition from ATM LAN to ATM WLAN is not free of problems. In consequent chapters the identified modifications of the ATM protocol family to support the ATM WLAN concept are described, as well as the required additions on the mechanisms related to bandwidth allocation, error handling and call control. The transition problems occur because of the requirements imposed by the nature of the radio link, the mobility factor and the lack of provision for these issues in existing ATM architecture.
1.5.1 Requirement for MAC layer definition in the ATM WLAN PRM

The basic difference between ATM LAN and ATM WLAN, is that ATM LAN does not contain any broadcast medium. All the entities are connected point-to-point using cables, thus there are no Medium Access Control (MAC) functions and no MAC layer in the PRM of ATM LAN. There is only one transmitter and one receiver for each direction at every interface and different cables are used for the different directions. However, in the case of radio, its broadcast nature imposes the necessity of a MAC layer within the PRM of ATM WLAN. Applying the MAC functions, logical links between communicating entities can be established. The MAC technique should enable the bandwidth allocation based on ATDM to enhance the bandwidth efficiency of the ATM WLAN and follow the traffic fluctuation of VBR connections. Moreover, the MAC mechanism should protect the bandwidth allocated to connections against bandwidth violations caused by traffic sources that exceed their expected traffic contracts.

1.5.2 Definition for a new PHY layer based on radio

The selection of radio channel as the transmission medium in the ATM WLAN, makes the second difference compared to ATM LAN. In the latter the transmission medium is highly reliable so the procedures handling transmission errors are minimum and confined in a single error correction-multiple error detection capability, and only for the header of the ATM cells. This has a straightforward impact on the overhead added by the physical layer to the transmitted packet (ATM cell). Opposite to fibre optics and cables (used in ATM LANs), the radio channel is a rather hostile transmission medium especially when the terminal moves, or the propagation environment changes because of movements. Finally, shadowing [ANDERS'95] could obstruct significantly the transmission and result to loss of communication link.

1.5.3 Possible implementation of Link Layer Control

The expected increase of the BER and PER in the radio channel could impose the definition of a Link Layer Control (LLC) layer, to introduce additional error handling procedures between PHY layer and ATM layer of the ATM WLAN. In chapter 3, possible solutions are discussed and their implications on the traffic management mechanisms and the PRM of the ATM WLAN. The introduction of DLC could alleviate the BER but also affect the traffic policing functions because it alters the inter-arrival sequence of received packets. Moreover, it adds on the overheads while it is not used all the time, because over the indoor radio channel, packets are received successfully most of the time [BASELINE].
1.5.4 Extension of C-plane protocols to support user mobility

ATM LAN does not support roaming of its users, so there is no provision for mobility management in its protocol family. Mobility is an advantage of wireless networks and is supported in ATM WLAN. Thus the protocols residing in ATM LAN have to be modified or extended to support mobility management issues like location update, connection re-routing, call set-up re-routing. In chapter 6 it is described how existing protocols of the ATM LAN PRM are modified to adapt to the mobility requirement within ATM WLAN. Furthermore in chapter 6 it is described how the MAC mechanism operates to support seamless hand-off.

1.5.5 Design objectives

From the above mentioned, it is concluded that there is the necessity for a new Physical Layer definition which:

- provides with channel transmission rate, BER and PER acceptable to support multi-media based on tables 1.1 and 1.2,
- minimises the overheads,
- provides uniform processing to all the packets that carry upper layer information like it happens in ATM LAN,
- provides information to upper layers to initiate hand-offs,
- has the capability of varying the modulation level and the channel coding rate to improve BER and follow the channel state transitions.

Moreover, a MAC Layer must be introduced in the PRM of ATM WLAN to:

- allocate bandwidth following the ATDM concept,
- provide with capacity for hand-off signalling and for the traffic of terminals executing hand-off,
- support asymmetric traffic since ATM WLAN should support distributive services like video on demand,
- preserve network capacity and bandwidth of established connections against bandwidth violations that occur by connections whose rate exceeds the expected one,
- enhance the functions of physical layer when channel estimation is based on the transmission of beacon signals or packets between the system entities.

Finally the upper layer protocols in the ATM WLAN PRM have to be specified to cope with:

- Network establishment
- Mobility Management
• Call set-up re-routing and in-call connection re-routing when hand-off occurs.

These protocols evolve from the C-plane and M-plane protocols in the PRM of ATM LAN as it will be described in chapter 6.

1.6 Putting it together

The study of the characteristics of ATM LANs set the design objectives for the ATM WLAN. Combined with the requirements imposed by the presence of radio as the transmission medium, it led to the specification of required functions, the identification of the layers to be included in the PRM of ATM WLAN, the requirements definition in every layer and finally, the set-up of a basic system design which was refined progressively, to include optimum solutions for the various operations that should take place in ATM WLAN. The details about how these operations take place will be described later in this document. However, Fig. 1.5 shows the initial system architecture to provide the reader with the baseline that was used when the existing ETSI standards were studied and modified or when decisions regarding the functions in the ATM WLAN were taken.

![Diagram of ATM WLAN architecture](image-url)

Fig. 1.5 Topology paradigm for private ATM WLAN
Compared to Fig. 1.2, Fig. 1.5 shows a similar system architecture in terms of topology and network entities. Of course, the small ATM switches at the access part of the ATM LAN have been replaced by ATM Base Stations (BSs) which however act like ATM switches in the ATM layer of the PRM of ATM WLAN. Moreover, backbone ATM switches in ATM WLAN have enhanced protocols to support mobility. This does not mean that the architecture of ATM switches has to be changed, just their protocols of the C-plane have to be extended. Moreover, ATM WLAN does not support only wireless terminals but integrates fixed and wireless terminals using the same backbone network.

Fig. 1.6 presents a rough PRM for ATM WLAN, where it is shown that DLC, MAC and ATM functions are integrated in a common layer called ATM-MAC. The UNI part of the ATM WLAN (between BS and terminals), called Mobile UNI (M-UNI) could be regarded as an ATM switch where transmitters are the input ports and receivers are the output ports. The bandwidth allocation mechanism should minimise the buffer overflows as well as the probability for packet blocking.
Chapter 1: Background

Fig. 1.7 M-UNI as ATM switch

1.7 MAC Baseline for ATM WLAN

Following the initial idea of the ATM WLAN concept, a baseline set of requirements was established regarding the MAC layer. The multiplexing protocol to be introduced as the multiple access scheme of ATM WLAN has to provide high bandwidth utilization and satisfy the delay requirements of different types of services. It should take advantage of the discontinuous data generation which characterizes voice and file transfers and enable the bandwidth allocation following the traffic fluctuation of VBR sources. Moreover, it should apply traffic monitoring operations that prevent congestion and enhance QoS provision.

1.7.1 Network Topology

Considering the wireless network topology, a star baseline configuration has been selected because of the following reasons:

I. ATM networks are point-to-point oriented considering physical connections.

II. There are some certain advantages of star based wireless systems, where a central element known as the Base Station (BS) controls the channel access, compared to peer-to-peer contention based wireless networks [BANTZ'94]:

A. there is no the hidden terminal phenomenon,
B. the required transmitted power by each terminal is less,
C. the bandwidth allocation can be easily achieved by the BS which has the knowledge of the traffic demands,
D. traffic policing functions can be applied enhancing the QoS provisioning as shown in chapter 4.
E. when interconnection to a fixed network is required then the functions of the access element can be supported by the BS,
F. there are advantages considering network security as the relevant functions are performed in the lower layers of the architecture,
G. frequency planning is easier,
H. the presence of the BS enables the accommodation of various traffic needs especially when different priorities have to be included and periodic traffic can be supported easier,
I. there is no power waste because of collisions and channel sensing.

1.7.2 Transmissions Defined in the baseline MAC

The structure of the air-interface for the proposed MAC protocol has to provide for some physical layer issues, such as equalization or synchronization and contain periodic message patterns which could be used for the monitoring of the channel quality, by both the base station and the wireless terminals. Moreover, the format of the ATM cell has to be extended or modified to provide for the required Physical layer operations. By reducing the length of VPI and VCI fields, a more powerful error control scheme can be supported for the ATM cell header over the radio.

Based on the baseline MAC design the following transmissions can occur in the uplink:

1. A packet is transmitted by a terminal in a non-contention mode, carrying user or signaling information. The VCI field determines the connection and the kind of transmitted information. The value 0 of the VCI field identifies the control channel in the system.
2. Terminals which are switched on, transmit registration messages in a contention based mode in order to declare their presence in the network and be assigned VPI values by the base station. The successful access of the radio channel by one of these terminals will result in the allocation of a control channel so that the required handshaking (for the registration) will take place.
3. Registered terminals without established connections access the uplink to request a control channel to initiate a call. Terminals in this case could be polled by the base station or content within specific time periods or frequency slots.
4. Terminals with established connections but inactive traffic sources send uplink messages (following contention or polling) to indicate the transition of their traffic sources to the active state. This operation enables the exploitation of discontinuous transmission and increases the multiplexing gain especially when highly asynchronous connections are multiplexed (e.g. connections carrying computer messages).
Chapter 1: Background

To support the above mentioned transmissions, terminals have to report their activity and issue messages like "transition to inactive", "increase bandwidth", "decrease bandwidth", "control channel request". Moreover, they could report their instantaneous bit rate to enable the Base station to schedule packet transmissions. The above mentioned messages could be carried in the header of the transmitted packets or in separate MAC control packets. Based on these messages, BS allocates capacity of the uplink channel to packets of active connections, to control channel requests and to registration requests, as was mentioned above. BS informs the registered terminals about the allocation of the uplink slots, transmitting appropriate broadcast packets in the downlink channel.

1.7.3 Channel Allocation Algorithm

To determine how to reallocate the uplink time slots, the base station executes the "Channel Allocation Algorithm" which assigns different priorities to different packets or types of services and uses scheduling methods borrowed from Operating Systems [TANENB'92], such as Round Robin scheduling. Furthermore, it could divide the set of established connections into channel capacity owners and channel capacity renters. The channel capacity owners acquire a requested or predefined portion of channel capacity as soon as possible after reporting their transition from inactive to active. The channel capacity renters access the channel if it is freed by the channel owners. This results to increased channel utilization. The terms channel capacity owners and renters were initially adopted in the design of MAC, but later they were replaced by the concept of packet access priorities.

1.8 Conclusion

The introduction of radio as the transmission medium of the ATM WLAN implies the definition of new transmission techniques in the physical layer, the addition of a multiple access scheme to be combined with ATM layer functions, and the specification of mobility management functions in the C-plane of ATM PRM. The resulting network architecture integrates fixed and wireless access by a uniform backbone network, capable to provide mobility management. The characteristics of ATM networks and protocols influenced the design of the baseline MAC considered for ATM WLAN. However, the study of existing standards further influenced the selection of the mechanism to be adopted as the multiple access scheme of ATM WLAN.
Chapter 2

ATM WLAN based on DECT and HIPERLAN

This chapter describes the basic characteristics of DECT and HIPERLAN, the two European candidates for indoor wireless communications. The aspects of these two standards which need modifications to support the concept of ATM WLAN are identified. Before ATM WLAN emerges as a market product, those wireless standards are expected to dominate the market of indoor wireless networks. It is valid then, to examine how they could be used to support the ATM WLAN concept, before proposing an optimum solution, since they offer well established and tested solutions for the operations related to transmission and access of the radio channel. Moreover, ATM technology could be used in the core network of a wireless system with a uniform transport and network layer, which integrates different radio technologies, and supports multimedia. The work presented in this chapter could be used for the integration of DECT and HIPERLAN radio technologies in such a system.

2.1 DECT overview

In chapter 1 a baseline of the ATM WLAN system architecture was given as well as the description of its PRM. That presented PRM applies at the M-UNI of the ATM WLAN, i.e. at the point where the user terminals are attached to the network. DECT with its proposed architecture could realise the M-UNI of a wireless system because it uses Cluster Controllers (CC) and Base Stations (BS) (called also Fixed Radio Parts- FRP) to provide
communication among user terminals, as shown in Fig.2.1. For DECT operation there is a frequency band allocated at 1880 - 1900 MHz divided into 10 carriers [DECT1]. The multiple access is TDMA/TDD with a frame containing 24 time slots and having duration of 10 msecs. In DECT a physical channel (bearer) is defined as a specific time slot on a specific carrier. DECT adopts Dynamic Channel Allocation Scheme (DCA) according to which a BS can use any time slot on any carrier as a bearer to a terminal with the constraint that there are no time slots overlapped in time and associated to the same FRP. DCA results to increased network capacity compared to Fixed Channel Allocation (FCA) adopted in GSM and DCS-1800 where a carrier with all its time slots is associated with a FRP.

Fig 2.1 Possible configuration for DECT system

2.1.1 DECT DLC

DECT specifications cover the four lower layers of the OSI architecture (Fig.2.2), when MAC and DLL are considered as two different layers. The Data Link Control Layer (DLC), the equivalent of the 3rd layer of the OSI standard, is divided in two independent parts [DECT3]:

![DECT System Diagram](image-url)
1. The DLC C-plane which deals with the control procedures of the DECT protocol stack and contains all of the internal DECT control information, supporting the signalling of network layer. It is expected that after applying the proposed modifications to the DECT standard, it will still use the same mechanisms for internal control and maintenance, so it would be preferable not to alter the DLC C-plane.

2. The DLC U-plane which deals with the user information and contains all the user information and user control.

U-plane and C-plane are independent processes. For their co-ordination the Lower Layer Management Entity (LLME) has been defined. It controls the routing of connections, their modification in terms of bandwidth, as well as the operations related to hand-off. LLME operations are defined in general, to allow for different implementations and this is something that was taken into account for the support of ATM WLAN by DECT technology.

The DLC U-plane contains a family of U-plane services and for each one of them a specific Service Access Point (SAP) has been defined. Each U-plane service can use certain combinations of the logical channels defined in the DLC of DECT, as well as certain transmission classes. A logical channel defines a connection in the DLC layer with certain service provision to the upper layer, while a transmission class defines the parameters of the service experienced by upper layers. Based on these conventions, an ATM-based service can be defined in the DLC. This new service can be accessed through its own SAP and support the requirements of the upper layers [UNI31] which in this case are ATM and AAL in the PRM of ATM LAN.

2.1.2 DECT MAC

In the MAC layer of DECT [DECT2], connections are defined between a wireless terminal and the Cluster Controller (CC). The MAC layer in DECT offers three classes of services which are accessed through different Service Access Points (SAPs). Among them, the Multi-Bearer-Control (MBC), a connection oriented service, is mainly used for establishing user connections and for carrying U-plane information. In DECT, MAC connections could belong in one of three types: basic, advanced and physical. Advanced connections have the privilege of being able to be modified in terms of allocated bandwidth. Moreover, connections in MAC
layer can be symmetric or asymmetric regarding the bandwidth allocated to the direction from terminal to BS and vice versa.

Furthermore, the MAC layer uses several virtual multiplexers, to map the DLC logical channels in the transmitted packets over the radio link. The transmitted packets in the MAC layer are composed by a field of constant length (A-field) which carries the DLC C-plane information and a by a field of variable length (B-field) which accommodates both DLC U-plane and DLC C-plane information. The format of the B-field depends on the error control strategy used in the MAC (protected or unprotected mode). The possible lengths of the B-field are defined in [DECT2]. In the protected mode, an ARQ-based strategy is used in the MAC, to deal with the transmission errors, resulting in connections of non fixed throughput at the upper layer. However, in order to have a fixed throughput for one connection the unprotected mode can be used, reducing the error detection/correction capabilities of the MAC layer.

In this work the main objective is to locate the ATM layer (and the layers above it) on the top of the DECT protocol stack in such a way, so that the result will be compatible with the PRM of ATM WLAN in Fig. 1.6.

2.2 Proposed modifications on DECT for support of ATM WLAN concept

It is expected that, in an ATM oriented DECT system, the operations related to the ATM layer and its upper layers are carried out according to the procedures governing the ATM WLAN. These procedures are described in subsequent chapters. However, the ATM layer should use a combination of DECT DLC and DECT MAC services to achieve peer-to-peer communication. In Fig. 2.3, the PRM that exists in the network entities forming the ATM WLAN based on DECT is presented. This PRM applies between the ATM wireless terminals and the CC of DECT. This happens because DECT DLC also is defined between the CC and a DECT terminal. A typical example of a DECT based ATM WLAN system is presented in Fig. 2.4(a).

![Fig. 2.3 PRM for ATM WLAN based on DECT](image)
The new ATM oriented service in DECT DLC, called LU-ATM is presented in Fig. 2.3. It resides below ATM layer and it preserves the order of transmission at the receiving end, delivers ATM cells with error free headers to the upper layer and provides a FEC-based error control method that is common for every kind of ATM connection, in order to reduce the BER. No retransmission of cells should be executed as it is expected that it will cause delay problems in time sensitive applications. Different instances of the LU-ATM service are used to support different ATM virtual connections of a terminal according to DECT specifications.

To avoid re-transmissions it is necessary to use unprotected transmission classes in the MAC layer [DECT2]. Although the MAC error control scheme is disabled, the throughput of the established connections is constant in terms of delivered MAC PDUs per sec, making the bandwidth allocation easier. Moreover the DLC provides the required functionality for error handling. It is proposed that $I_N$ _minimum_delay_service_ [DECT 2]
be used, where each U-type data-burst is equal to one \( I_N \) segment. The length of that segment depends on the used slot type. The MAC connection supporting the ATM connection should belong to the category advanced, because it should be able to be modified in terms of used bearers. The already existing procedures in the MAC layer can remain unaltered.

The PDU used in the DLC service for optimum ATM support should be compatible to the ATM cell regarding format and length. Moreover it should be compatible with the structures transmitted in the MAC layer [DECT2]. ATM cell is 53 bytes long where 5 bytes is the length of its header. In DECT there is no \( I_N \) segment with this length. However, a \( I_N \) segment 100 bytes long is supported and this length is proposed to be used, because it is the only length bigger than that of an ATM cell. Such a \( I_N \) segment will carry an ATM cell and 376 redundancy bits. These bits can be used:

- for the ATM cell header error control functions, which may be enhanced compared to those of wired networks, because of the hostile radio transmission environment.
- for the protection of the ATM cell payload in order to compensate for the radio channel impairments and reduce the offered BER to the upper layers.

The structure of the air interface for the DECT based ATM WLAN is presented in Fig. 2.4(b).

Having specified the operations of the DLC and MAC layers, it is appropriate to define the interface between the layers of the ATM based DECT protocol stack. It is assumed that the interface between the MAC layer and the proposed DLC is the same as the existing DECT DLC - MAC interface.

The primitives exchanged between ATM and lower layers in [UNI31] are PHY-UNITDATA.req and PHY-UNITDATA.ind, and they carry one ATM cell. In this case the physical medium is always available, so no initial procedures must take place in order to establish a physical connection. In the case of DECT, the DLC maintains a virtual permanent connectivity between a base station and a user terminal. This permanent connectivity is maintained without wasting network resources, mainly in terms of bandwidth. DECT DLC has the option of suspending and resuming connections [DECT3]. This gives the opportunity to all the user terminals which support ATM services and use the proposed, in this work, modification, to initiate an ATM connection using the standard DECT procedures. The first connection to be established will support the signalling channel of ATM layer. While a connection is idle it should be suspended. The same scheme can be used in the opposite direction. When it is required to carry user traffic a connection is resumed. When it is required to establish a call or exchange control information the signalling connection is resumed too. It is obvious that on top of the two primitives (PHY-UNITDATA.req and PHY-
UNITDATA.ind) which carry the ATM cells, there is a need for more primitives between the ATM layer and the DECT DLC. These primitives are provided by the LLME of DECT. For this reason, it is proposed to extend the LLME upwards, in order to locate the decisions related to connection modifications in the LLME at the ATM layer. Based on the knowledge about the traffic class of every established connection (obtained during call set-up), the LLME can adjust the time threshold that indicates a suspension of the DLC connection by initiating an ATM_connection_suspend.request to the lower layer LLME. Moreover, by counting the inter-arrival rate of ATM cells per Virtual Channel (VC), LLME can decide on the issue of an ATM_connection_modify.request to adapt the allocated bandwidth to a user terminal to the instantaneous requirements of its ATM traffic. Finally, for a suspended ATM connection, the presence of an ATM cell at the SAP between AAL and ATM will cause the issue of an ATM_connection_resume.request to the lower layer LLME in order to re-establish the DLC connectivity over the radio link.

In summary to support ATM over DECT it is proposed to:

- introduce a new DLC LU service that applies FEC on the transmitted ATM cells and preserves their transmission sequence,
- select a certain kind of DECT MAC connection (advanced, IN_min_delay) to support the above mentioned LU service,
- define at the interface between the DECT LLME and ATM layers primitives related to suspension and resumption of the DLC connections supporting the ATM traffic.

Following the above mentioned modifications, the ATM WLAN concept based on DECT can be established to offer statistical multiplexing in the air interface of DECT.

2.3 HIPERLAN overview

HIPERLAN stands for High Performance Radio Local Area Network. The standard was prepared by the Radio Equipment and Systems (RES-10) Group of the European Telecommunications Standards Institute (ETSI) and aims to cover the Physical and MAC layers of the OSI architecture. The main objectives of the standard are:

1. the easy establishment of wireless ad-hoc Local Area Networks (LANs), by using distributed topology and routing functions in the MAC layer, and
2. the transparent support of existing applications and protocols residing in layers above MAC layer, by defining a MAC Service Access Point (SAP) fully compatible to ISO 15802-1 specifications.
Therefore, HIPERLAN can provide a wireless platform that offers a connection-less mode of packet transmission and supports currently used protocols in the Data Link, Network and Transport layers of the OSI architecture.

The protocol reference model of HIPERLAN is presented in Fig.2.5. The Physical layer deals with the modulation and demodulation, channel coding, synchronisation and equalisation of the channel. The Channel Access Cycle (CAC) layer is related to the access of the radio link by the wireless terminals. MAC layer contains the control functions related to topology updating and route determination. Moreover, it assigns channel access priorities to packets based on their lifetime and on the priority assigned to them by upper layers.

The multiple access technique, used by HIPERLAN terminals to acquire the radio channel, is totally distributed, based on packet priority declaration and channel sensing.

2.3.1 The physical layer of HIPERLAN

The transmission range of the HIPERLAN transmitter is up to 50 meters enabling a channel bit rate (gross bit rate) around 23 Mbps. A specific region (5.15 GHz to 5.30 GHz) of the frequency spectrum is allocated by CEPT for HIPERLAN operation. The allocated frequency spectrum is divided in 5 frequency bands (each frequency band associated with a carrier). All the nodes which belong to the same HIPERLAN network should use the same unique carrier. Packets can be relayed to nodes residing out of the transmission range of source nodes, following the concept of forwarding which is controlled by the MAC layer of HIPERLAN.

The packets that contain upper layer information, or MAC control information consist of two parts, a low bit rate (1.5 Mbps) part which is transmitted using FSK modulation and a high rate (23.5 Mbps) part GMSK modulated (Fig. 2.6). The low rate part contains information which is duplicated in the high rate part and identifies the receiver of the packet, to fire only the equaliser of the appropriate receiver and not equalisers of irrelevant terminals. The high rate part consists of a well defined training sequence (450 bits long) and a number (1 up to 47) of blocks each having a length of 496 bits. Each one of these
blocks of 496 bits occurs after BCH(31,26) encoding of 16 segments of information with length of 26 bits and block interleaving.

<table>
<thead>
<tr>
<th>Low Bit Rate Part</th>
<th>Synchronisation and training sequence</th>
<th>High Bit Rate Part 1</th>
<th>High Bit Rate Part 2</th>
<th>• • •</th>
<th>High Bit Rate Part m</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>450 bits</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

496 bits resulted by BCH (31, 26) encoding
16 blocks with length equal to 26 bits
m: 1... 47

Fig. 2.6 Format of the transmitted packet in HIPERLAN

The Physical layer is also responsible to report an idle channel condition to the above layer. This happens by monitoring the received power and detecting whether it is below a well defined threshold for a period equal to 1700 high rate bit periods. The idle channel condition enables a HIPERLAN terminal to transmit the packet with highest channel access priority without following the contention based algorithm of MAC.

2.3.2 The Channel Access Control layer (CAC) of HIPERLAN

CAC layer deals with the decision to transmit a packet or not. It is based on a multiple access technique called Elimination Yield - Non pre-emptive Priority Multiple Access (EY_NPMA) which is used during the Channel Access Cycle. Transmitted packets are generally divided in unicast (one specific destination node is defined) and multicast (destination nodes of the packet are more than one) packets. A correct reception of a unicast packet is always followed by a transmission of an ACK packet (acknowledgement) by the destination node. This does not happen in the case of multicast packets. A Channel Access Cycle starts after the transmission of an ACK packet or after the end of the expected transmission of an ACK packet. It is assumed that all the terminals that have listened to, or expecting, the ACK packet and having packets to transmit, are synchronised and enter the channel access cycle. Every node contends for only one of its packets per channel access cycle, that with the highest access priority among the packets residing in the node. According to the EY_NPMA principles, the channel access cycle has four phases (Fig.2.7):

The first phase is priority resolution: Generally packets can have five different access priorities (0 to 4, 0 being the highest priority). The access priority of a packet is calculated at the transmitting node, based on the remaining lifetime and the user priority of the packet, as well as on the number of intermediate nodes (hops) that the packet must traverse before being delivered at the final destination. The number of hops is known and assigned by the MAC...
The second phase of the channel access cycle is called elimination: In this phase, up to twelve time slots (each one being 256 bits long) could be accessed (using a well defined bit sequence) by the nodes survived in the first phase of the channel access cycle. The transmission in these time slots is continuous and starts always in the first time slot. The duration of every node’s transmission is based on a probability of time slot access and is given by the binomial distribution. Every node transmits for the decided period and then listens to the channel to examine if there are any nodes trying to access the channel, too. Longer transmissions in this phase reject listening nodes from the channel access cycle. More than one nodes could survive in this phase whose duration is determined by the longest transmission (up to 12 x 256 bits).

The third phase of the channel access cycle is called yield: All the survived nodes decide to sense the channel for a period which is a multiple (1 up to 14) of a 64 high rate bits block. The node which transmits first, acquires the channel because it disables all the sensing nodes from the channel access cycle. However, in this phase also, more than one nodes can survive and this results to packet collision.

The fourth phase is the transmission of packets by nodes survived in the channel access cycle. If no collision happens (only one node survived finally), a unicast packet could be transmitted followed by an ACK packet indicating that the packet was received correctly.
or damaged by radio channel improperties. It should be noticed, that not all the nodes listening to the transmitted packet may be able to listen to the related ACK packet, because the destination node could have a different coverage area than the source node. To start a new channel access cycle, nodes should be able to be synchronised based on the end of the ACK transmission or at the end of the expected ACK transmission.

The mechanism used in HIPERLAN, offers a way to distinguish between delay sensitive and non-delay critical applications. However, as it will be shown in section 2.4 when delay sensitive services of the same kind are multiplexed in the same network, HIPERLAN cannot achieve the same capacity as centralised methods adopting adaptive bandwidth allocation.

2.3.3 The MAC layer in HIPERLAN

The MAC layer in HIPERLAN is involved in the following procedures:

1. Network establishment, addition of a node in a network and removal of a node from a network. These procedures are carried out by the LookUp function of HIPERLAN MAC.
2. Topology updates and packet routing determination as well as packet relay (forwarding), controlled by the Routing Information Exchange function and the User Data Transfer function.
3. Power conservation by declaring periods in which the receiver of a node is active and can listen to transmitted packets. The related control functions reside in the Power Conservation function of the HIPERLAN MAC.
4. Calculation of the channel access priority of packets to be transmitted, based on number of hops and remaining packet lifetime. This function is carried out in the HMPDU (HIPERLAN MAC Protocol Data Unit) Transfer function.

Moreover, HIPERLAN MAC deals with some other functions like encryption and assignment of alias names to network addresses. More details can be found in the document describing the standard [HIPERLAN'95].

Generally, packets submitted to the MAC for transmission are assigned with one out of two user priority levels (0 for high and 1 for low). The remaining lifetime (RML) of every packet can be calculated based on its total lifetime and on the transit delay it has experienced. The concept of Normalised Residual packet lifetime (NRML) is introduced and computed by dividing packet’s RML with the remaining number of hops, which is contained in the packet assuming that the source node has enough information to determine the route of the packet. Table 2.1 shows how the mapping between user priority and channel access priority takes
place based on the NRML of every packet. The node presents in the next channel access cycle the packet with the highest channel access priority among its packets.

**Table 2.1. Definition of channel access priority**

<table>
<thead>
<tr>
<th>NRML</th>
<th>User Priority</th>
</tr>
</thead>
<tbody>
<tr>
<td>0... 10 msecs</td>
<td>0</td>
</tr>
<tr>
<td>10 msecs ... 20 msecs</td>
<td>1</td>
</tr>
<tr>
<td>20 msecs ... 40 msecs</td>
<td>2</td>
</tr>
<tr>
<td>40 msecs ... 80 msecs</td>
<td>3</td>
</tr>
<tr>
<td>more than 80 msecs</td>
<td>4</td>
</tr>
</tbody>
</table>

The three first procedures of the MAC, described above, use multicast packets in order to flood corresponding information in the network. The calculation of the channel access priority of packets happens in every node independently and without the knowledge of the total traffic in the network. HIPERLAN does not have the ability to perform traffic policing functions because of its distributed mode of operation. However, such functions are expected to be used in future networks with multimedia capabilities to prevent congestion [ONV'94].

### 2.4 Modifications to HIPERLAN for ATM support

From the discussion in the previous paragraphs and in paragraph 1.6, it is obvious that there are many differences between ATM WLAN and HIPERLAN PRMs and the corresponding network topologies. The main problem arises from the randomness observed in the location of the nodes in a HIPERLAN network and the non-uniform topology patterns that are created.

In HIPERLAN there is no call establishment phase before the information transfer takes place, because it is assumed that it has happened in upper layers. So there is no knowledge about the total traffic available in the network and no call admission functions. This means that overload conditions in the network cannot be avoided because traffic policing functions are not available and quality of service cannot be satisfied even if the multiple access throughput scheme is high. Even if traffic control functions are assumed in the layers above MAC, HIPERLAN could not still support optimally the QoS of different services because the MAC layer deals with packet priorities (based on ageing) but not with service priorities. For example, if video transmissions and file transfers are multiplexed in the
network, although initially video packets will have higher priority over data packets, gradually the latter will acquire (by the MAC algorithm) high priorities and contend for the channel on equal terms with video packets.

Moreover, the multiple access mechanism in HIPERLAN cannot provide bandwidth guarantees which result to QoS preservation. This means that the MAC layer in HIPERLAN cannot disable sources from competing for channel resource when their traffic emission exceeds the values that were negotiated during call set-up, i.e. there are no traffic policing functions. However, in chapters 3 and 4 it is shown that traffic policing and bandwidth preservation are elementary functions in the multiple access mechanism of ATM WLAN. In Fig. 2.8 it is shown that the multiple access scheme of HIPERLAN does not provide with bandwidth guarantees for connections that conform to their negotiated traffic limits. VBR video connections based on [MAGL'88] are assumed, at 2 Mbps average bit rate. Fig. 2.8 presents the dropping probability ($P_d$) of conforming and non-conforming connections when the maximum capacity of the system (6 users at 2 Mbps average rate) is achieved, and for different amounts of bandwidth violation (in Mbps) by non-conforming connections. It is shown that both conforming and non-conforming connections experience similar dropping result, i.e. the network cannot protect the bandwidth of the conforming connections. This happens because the multiple access scheme of HIPERLAN tries to share equally the total bandwidth without taking into account any traffic contract. $P_d$ should be stable for conforming connections. Thus, there is no way to combat bandwidth violation occurring as a result of the short term variations of traffic. However, ATM networks incorporate traffic policing functions that deal with exactly this problem [TM'96], and such mechanisms are expected to reside in ATM WLAN, as they have a straightforward impact on the offered Quality of Service. Nevertheless, HIPERLAN offers the access mechanism described in 2.3.2 that could be modified to support transmission of ATM traffic over the radio link, if certain modifications
apply to the existing standard. This effort is in accordance with ETSI view which through RES-10 group foresees ATM WLAN as an extended or upgraded version of HIPERLAN.

2.4.1. Network Configuration for HIPERLAN based ATM WLAN

To achieve ATM WLAN functionality using HIPERLAN it is proposed that the standard be modified in order to adapt a centralised topology according to the ATM concept (Fig. 2.9(a)). Based on the requirement for dividing the functions of ATM into user part and network part we assume a system architecture consisting of a fixed part and a wireless part. The fixed part is a traditional ATM network, while the wireless part extends the ATM concept over the radio based Physical layer of HIPERLAN [HIPERLAN'95].

![Diagram](image)

**Fig. 2.9** (a) Scenario for HIPERLAN based ATM WLAN configuration  
(b) Structure of air interface

The wireless part consists of wireless terminals executing the user side of the protocol for ATM over HIPERLAN, and Base Stations (BSs) executing the network part of the corresponding protocol. The interface between BSs and backbone ATM switches could be based on PNNI [PNNI'96]. Within the proposed system architecture, BSs perform a superset of HIPERLAN forwarder [HIPERLAN'95] functions, i.e. in addition to packet switching they perform call admission control, bandwidth allocation and ATM address registration. BSs communicate through the fixed network and according to the ATM concept. The terminals in the wireless part of HIPERLAN based ATM act like HIPERLAN non-forwarder nodes, i.e. they expect a forwarder to relay their packets towards the final destination.
Chapter 2: ATM WLAN based on DECT and HIPERLAN

Within the proposed scenario many terminals are logically attached to only one BS whose coverage area defines a cell. A BS communicates with all its adjacent terminals using a carrier different than the ones used by neighbouring BSs. Because there are three main carriers always available [HIPERLAN'95] for HIPERLAN operation, a frequency re-use factor of three could be adapted within the ATM WLAN based on HIPERLAN. In HIPERLAN based ATM WLAN, the BS itself is the network part of UNI and acts as an access ATM switch with one wireless port.

2.4.2. MAC scheme for ATM WLAN operation

To support adaptive bandwidth allocation and traffic policing in the wireless part of the HIPERLAN based ATM WLAN, certain modifications apply to the CAC and MAC layers [HIPERLAN'95] of HIPERLAN. The objective is to define a super-frame structure for the transmissions on the air-interface in every cell (Fig. 2.9(b)). This super-frame consists of a contention based period where all the HIPERLAN specified transmissions can be supported, and a contention free period, where packets carrying ATM traffic are transmitted. The duration of the super-frame is fixed while the boundary between its periods is variable, based on the terminals population and total ATM traffic within a cell. To support the second period of the super-frame the HIPERLAN MAC and CAC have to be slightly modified and their interface has to be extended. Considering CAC after the applied modifications, it transmits a packet in one of the three following cases (where the first two are defined in the existing HIPERLAN standard):

1. The channel has been found free, a HC_FREE.indication has been given to MAC and a HC-UNIT-DATA.request has been issued at the Service Access Point between MAC and CAC,

2. A new Channel Access Cycle has been initiated, a HC_SYNC.indication has been signalled to MAC and a HC-UNIT-DATA.request has been issued at the Service Access Point between MAC and CAC. In this case CAC will compete for the channel entering a new channel access cycle.

3. A new time slot has begun, a HC SLOT.indication (time_slot_number) has been signalled to the MAC to indicate the beginning of this time slot and a HC-UNIT-DATA.request has been issued at the Service Access Point between MAC and CAC to request the transmission of a HIPERLAN packet that carries an ATM cell in its payload. All the packets transmitted this way are uni-cast packets and they are acknowledged according to the CAC procedures. The end of the acknowledgement transmission defines the end of the time slot.
The third case is the modification in the CAC and MAC layers of HIPERLAN to support the super-frame structure, while the addition of HC_SLOT.indication at the SAP of CAC is required, to indicate to the MAC the occurrence of a new time slot. For that reason, the Super-frame Delineation Function is added in the CAC layer. At the beginning of the contention free period of a super-frame, the BS transmits a Packet Schedule multi-cast packet to allocate the time slots to active ATM connections. The reception of this multi-cast packet forces CAC to enter the Super-frame Delineation Function and indicate the beginning of all the time slots to the MAC. Moreover, the Packet Scheduling Function (PCF) is added as an extension to HMPDU Transfer function in the MAC layer. PCF selects a packet from the ATM connection that is scheduled for the current time slot and delivers it to HIPERLAN CAC for transmission, if it receives an HC_SLOT.indication. As already stated the scheduling information is contained in the Packet Schedule multi-cast packet which is delivered to PCF.

Moreover, the existence of the Super-frame Delineation Function results to the reduction of the overheads added by HIPERLAN MAC within the contention free period. In this period, only packets carrying ATM cells are transmitted on the air-interface, so all the transmitted packets have the same size and format. Then, all the parts of the HIPERLAN MAC PDUs that are related to variable length and format, are obsolete in the contention free period. The same assumption could be made for the CAC layer. If the existing MAC, CAC and PHY layer overheads are added to transmit ATM cells, the transmission efficiency is 18%. Contentions occurring because of the channel access scheme of HIPERLAN will reduce further the bandwidth efficiency of the system. To increase it, when ATM traffic is considered, a variable length packet format is associated to the contention based period of the super-frame and a fixed packet format is used within the contention free period. In that case the Service Data Unit (SDU) delivered at the PHY layer of HIPERLAN during the contention free period, contains an ATM cell. The objective is to deliver to PHY layer, an SDU of maximum length equal to 416 bits because PHY accepts SDUs whose length is a multiple of 416 bits. After the elimination of unnecessary overheads of MAC and CAC, the overheads added by PHY [HIPERLAN'95] bring the transmission efficiency at 22%, achieving a 4% improvement compared to standard HIPERLAN packet format. The unnecessary overheads consist of the fields related to MAC source and destination addresses which are redundant because all the packets in the contention free period are exchanged between a terminal and the BS.
Moreover, the multiple access scheme adopted in the contention free period is expected to further improve the bandwidth efficiency compared to the standard HIPERLAN channel access scheme. Fig. 2.10 shows the format of the transmitted packets and the Protocol Data Units (PDUs) at the different layers in the Protocol Reference Model of the HIPERLAN based ATM WLAN. It is shown that the Low Bit Rate part (LBR) of the transmitted packet carries the VPI/VCI, CLP/PTI, HEC values of the ATM cell. Moreover, the High Bit Rate part (HBR) carries the ATM payload and the Checksum calculated by the CAC layer. This is the format for the corresponding PDUs in the contention free period. In the contention period of the super-frame, the format of the PDUs is defined in the existing HIPERLAN standard.

To avoid the violation of the contention free period, at the beginning of every super-frame, BS transmits a multi-cast packet indicating the duration of the contention based period. The Super-frame Delineation Function in CAC, checks the duration of a packet transmission and compares it with the time remaining to the end of the contention period. If the transmission will be extended into the contention free period, it is abandoned and an HC-STATUS.indication(TRANSFER_UNSUCCESSFUL) is delivered at the MAC layer.

The contention based period is used to support HIPERLAN Lookup function, Routing Information Exchange function, and Power Conservation function [HIPERLAN'95]. Moreover, it is used for transmissions of packets carrying UBR ATM traffic [TM'96] (connections with no delay or bandwidth requirements), or ATM C-plane traffic (messages related to call control), as well as for the ATM address registration procedure between the BS and a terminal joining the system.
Following the HIPERLAN LookUp and Routing Information Exchange functions, a terminal joining the system can learn about the existence of a cell and its corresponding BS. Having learnt the HIPERLAN MAC address of the BS, the terminal uses it as the MAC destination address of all its packets transmitted in the contention period of the super frame. Moreover, the BS learns through the Routing Information Exchange function about the MAC address of the terminals in its cell. Then, each pair of HIPERLAN MAC addresses (where one of them is the BS MAC address) defines a logical port between the terminal and the BS. To that logical port a specific VPI value can be assigned to identify all the packets between the BS and a certain terminal. This issue minimises the overhead added currently by both MAC and CAC layers. Furthermore, the traffic generated by control functions of HIPERLAN (Routing Information Exchange function, Power Conservation function) is minimised because terminals need to acquire only the information about the BS and not about other terminals (in existing HIPERLAN specifications topology information is flooded in the network).

According to the above proposed modifications the HIPERLAN PRM is extended to that presented in Fig. 2.11, to support ATM traffic. This PRM applies at the User-to-Network Interface (UNI) between terminals and a BS. Adapting the presented PRM, full compatibility is achieved, between HIPERLAN based ATM WLAN terminals and terminals in fixed ATM or B-ISDN.

2.4.3 Bandwidth allocation in HIPERLAN based ATM WLAN

The protocol used for the access of the radio channel should be flexible enough to allow bandwidth sharing among terminals according to their instantaneous capacity requirements. Moreover, it should be able to provide bandwidth guarantees and apply traffic policing functions to established connections. The Packet Scheduling Function, introduced in HIPERLAN MAC, deals with this problem. In the terminal side and for every established ATM connection, it keeps a separate virtual buffer, and based on the occupancy of this
buffer; it reports the current bit rate of the corresponding connection to the Packet Scheduling Function of the BS.

The calculation of the current connection rate is based on the ATM cell inter-arrival process within a certain period. To report the latest calculated rate of a connection, the terminal sends to the BS a Resource Management ATM cell with the VPI/VCI values identifying that connection and the value of the calculated bit rate. The Resource Management ATM cell is transmitted in a time slot allocated to the corresponding virtual connection or in the contention based is transmitted in a time slot allocated to the corresponding virtual connection or in the contention based period of the super-frame. The Packet Scheduling Function residing in the BS, uses the information in the Resource Management ATM cells to regulate the bandwidth allocation in the contention free period of the super-frame. The algorithm that is executed to decide on the bandwidth allocation could be the one of the algorithms presented in sections 3.1.4 or 4.3.

Moreover, the algorithm presented in [APOST'96] was extended so that, the traffic policing functions of ATM layer assign priorities to bandwidth requests, based on the negotiated traffic contracts. These contracts are defined during call set-up according to ATM standards. The above mentioned assigned priorities control the scheduling of packets in the contention free period of the superframe. Following this technique the bandwidth requirements of individual sources are monitored and bandwidth violations are avoided. Thus, QoS can be guaranteed.

For the presented simulation results, a super-frame lasting 1.215 msec, and consisting of a contention free period with 15 time slots was assumed. The first time slot of the contention free period is used by the Base station to schedule terminals transmissions in the rest of time slots. Figs. 2.12 and 2.13 present the packet delay and packet dropping probability when ATM connections are supported in both standard HIPERLAN and HIPERLAN based ATM WLAN. The VBR traffic sources have an average rate of 1 Mbps and are modelled based on [MAGL'88]. It is obvious that in the case of standard HIPERLAN, the high transmission overheads combined with the contention based multiple access result in poor system performance and user capacity. When, the presented modifications are applied there is a significant improvement on user capacity. Assuming 2% as the maximum acceptable packet dropping probability the capacity of standard HIPERLAN when it encapsulates ATM is only 1 user (1 Mbps on average) while the modified HIPERLAN can support 4 users (4 Mbps on average). Moreover, Fig. 2.14 shows that the packet dropping probability for conforming sources is kept constant in the case of HIPERLAN based ATM WLAN when the system operates at the maximum user capacity and irrespectively of the amount of excessive traffic offered by non-conforming functions.
In summary to exploit the HIPERLAN standard for support of ATM WLAN concept it is needed to:

- define a star topology were a Base Station schedules the packet transmissions by terminals,
- extend the HIPERLAN Protocol Reference Model by including the Packet Scheduling Function dealing with the signalling for the traffic scheduling,
- introduce in the HIPERLAN Protocol Reference Model, the Super Frame Delineation Function that specifies the boundaries of the contention free period in the air-interface. This period is dedicated to ATM traffic.

2.5 Conclusion

In this chapter, the main functions of DECT and HIPERLAN were described, and two proposals were presented for incorporating ATM features, in these two European standards. Then, ATM traffic could be supported over the radio multiple techniques described in DECT and HIPERLAN specifications.

DECT offers a highly centralised topology and protocol to support mobility by incorporating both intra-cell and inter-cell hand-off. It adopts Dynamic Channel Allocation (DCA) to reduce the probability of call dropping during hand-off (because channels used by the moving terminal could be passed from the old to the new RFP). Moreover, the co-channel
interference is reduced, because the selection of a channel is based on the co-channel interference metric and not only on a fixed channel allocation algorithm like the one existing in GSM. Because of these reasons, DCA results in higher user capacity compared to fixed channel allocation [JUSTIN'93].

However, it is obvious that DECT contains highly complicated layers and protocols that increase significantly the transmission overhead. Being mainly a synchronous system, it moves away from the ATM WLAN concept. The latter aims at a uniform processing of user information in the bottom layers of its PRM and introduces adaptive bandwidth allocation in its multiple access mechanism. DECT discriminates among different services and control information by providing different combinations of DLC and MAC services and packets. Moreover, a big part of DECT specifications is dominated by C-plane functions which means that the protocol overhead is significant compared to the U-plane functions. So, even if DECT transmission rate was higher than the currently supported, its radio access topology would not be the optimum solution for ATM WLAN operation.

Moreover, HIPERLAN introduces a simpler PRM and provides for adaptive bandwidth allocation in its multiple access mechanism by applying different priorities to packets with different lifetime. However, access delay is not the only metric to decide on the priority of a packet in ATM networks and accordingly in ATM WLAN. Furthermore, it was proven by simulation, that the protocol overhead to transmit an ATM packet is excessive. However, by applying the presented modifications to support the objectives of the ATM WLAN multiple access scheme, it was shown that the performance of a HIPERLAN based ATM WLAN improves. This shows that the CAC algorithm used in HIPERLAN is insufficient for ATM traffic. Furthermore, the training sequence in the transmitted packets is very long compared to an ATM cell. A new way of sensing and equalising the channel must be incorporated to improve the transmission bandwidth efficiency.
Chapter 3

New ATM WLAN

In this chapter, the issues that drove the design decisions of the ATM WLAN MAC are discussed and the operations that take place in its different layers are presented. Moreover, a comparison in terms of user capacity takes place between HIPERLAN multiple access scheme and a Dynamic TDMA based protocol that uses the packet priority mechanism found in HIPERLAN. This protocol could realise the multiple access technique in ATM WLAN. Furthermore, it is shown how the mechanisms for ad-hoc networking residing in HIPERLAN can be combined with the optimum solutions proposed for ATM WLAN to obtain a dual mode protocol stack. The latter enables wireless terminals to form ad-hoc networks when necessary, and operate in a cellular architecture to provide the full range of ATM services to their users.

3.1 Selection of MAC scheme

In chapter 1, it was stated that it is required to add the MAC layer in the PRM of the fixed ATM LAN in order to migrate to ATM WLAN. In the latter, the characteristics of the services [ONV'94] require a multiple access method with bandwidth allocation adaptive to traffic variations. Moreover, the delay sensitive applications require a minimum access delay. Moreover, the used MAC should guarantee the relative delay experienced by connections which define a composite call. From the network point of view, MAC should preserve capacity efficiency, by allocating bandwidth only to active sources. There are several
multiple access techniques satisfying these criteria [APOST’95], [GOODMAN’89], [RAYCH’92]. All of them, however, require the existence of a base station which schedules the transmissions of the network terminals. This requirement discounts the establishment of ad-hoc networks but nevertheless some users, as well as some equipment providers could require the capability of ad-hoc networking. However, ad-hoc networks usually support simple applications like file transfers and not demanding delay sensitive services. The objective of the ATM WLAN is the full support of enhanced services provided also by B-ISDN, so its design should follow this objective.

3.1.1 Candidate Multiple Access techniques for ATM WLAN

The objective of the multiple access mechanism is to co-ordinate multiple nodes when they share the same transmission medium, in order to minimise the probability of collided transmission and maximise the performance in terms of access delay and channel throughput. Today, there are different alternatives based on wired LAN protocols (Ethernet, FDDI, Token Ring), wireless LAN standards (like those presented in chapter 2), as well as Personal Communication Systems (PCS). Generally, there are contention free access methods such as FDMA, TDMA, CDMA and polling, as well as contention based techniques like Aloha, CSMA, CSMA/CD [TANENB’88], CSMA/CA [CHEN’94] and EY-NPMA [HIPERLAN’95]. There are also combinations of the above schemes like PRMA [GOODMAN’89], Multiservice Dynamic Reservation TDMA [RAYCH’92] and MASCARA [PASSAS’97], the latter used for the MAC of WAND [IPCOM’96].

3.1.1.1 CSMA based mechanisms

CSMA based protocols are characterised by low access delays when the traffic load in the network is low too. However, when the traffic in the network increases the delay increases exponentially [CHEN’94] and obtains non upper bounded values. This drives the network to instability and decreases the channel throughput drastically. Because there is no channel arbitration mechanism, there is no provision for bandwidth preservation even for optimised contention based mechanisms like EY-NPMA. In chapter 2 it was shown that no bandwidth guarantees could be obtained using EY-NPMA.

Another problem that arises in CSMA based protocols is that of hidden...
terminal [WASH'94]. In Fig. 3.1 it is shown that when node ‘A’ transmits, node ‘C’ residing out of node’s ‘A’ coverage range could assume the channel idle and start transmitting, causing packet collision at node ‘B’. Of course, this is not a problem if the packets transmitted by nodes ‘A’ and ‘C’ are destined to nodes E and F respectively. HIPERLAN takes advantage of this to increase the system throughput by acknowledging all the packets corresponding to single receiver. Moreover, IEEE 802.11, in an attempt to solve the hidden terminal problem, adds on CSMA/CA [IEEE] technique a handshaking described as follows:

In Fig. 3.1, if node ‘A’ wishes to send a packet to node ‘B’, it follows the CSMA/CA algorithm to transmit a short message called Request To Send (RTS). Successful reception of this message by all the nodes in node’s ‘A’ coverage range disable them from transmitting. Moreover it causes node ‘B’ to respond transmitting another short message called Clear To Send (CTS). Successful reception of this message by all nodes in node’s ‘B’ coverage area will stop them from transmitting and keep the channel clear for the transmission of the packet from ‘A’ to ‘B’. However, the RTS-CTS mechanism adapted in IEEE 802.11 alleviates without solving the hidden terminal phenomenon, because RTS and CTS are not received successfully all the time. Moreover they add on the packet transmission overheads.

3.1.1.2 FDMA

The frequency band allocated to a system is partitioned into narrower frequency bands (channels), and one or more could be allocated to a node for the whole duration of its call. This scheme enables channel sharing in time (many nodes transmitting simultaneously at different channels), but it does not satisfy the requirement for dynamic bandwidth allocation and cannot support efficiently VBR traffic. If the allocation of bandwidth takes place based on peak rate of the source it causes waste of channel capacity.

3.1.1.3 Fixed TDMA

The similar effect occur in Fixed TDMA, the dual method of FDMA in time domain. In TDMA case nodes transmit periodically on a given carrier. The instants of transmission are fixed and defined during call set-up. No statistical multiplexing gain can be achieved in both FDMA and Fixed TDMA. However, compared to contention based methods, the access delay obtained in FDMA or TDMA is fixed and upper bounded. In more sophisticated versions of TDMA (usually called Advance TDMA/ATDMA or Dynamic TDMA/D-TDMA), the frame is divided in request sub-slots and message slots. Then on demand bandwidth allocation can be realised.
3.1.1.4 CDMA

To establish CDMA over Spread spectrum signals, different sequences of digital pulses (chips) per bit must be assigned to different pairs of transmitters and receivers, provided that for a given pair the same chip sequence (code) corresponds to every information bit. Also, the code sequences assigned to different channels must be mutually orthogonal. Following this convention, a channel is defined by both a frequency band and a code.

In the case of CDMA, the signal to interference ratio is determined by the ratio of the desired signal power to the sum of the power of all the other signals which are background noise, external interference, other cell interference and other user interference. At the receiver, the undesired signals appear as additive interference. The level of it depends on the number of users. CDMA has the advantage of accommodating a large number of users if they transmit messages for short period [PROAK'95]. The number of users is given by the following equation:

\[ N = 1 + \frac{W}{R} \left( \frac{\frac{E_b}{n}}{S} \right) \]

where \( W \) is the bandwidth of the spreaded signal and \( R \) is the bit rate.

Using CDMA all the users up to \( N \) can simultaneously use the whole available bandwidth. But, it is shown from the above equation that the number of users is proportional to the spreading factor \( W/R \). The required system bandwidth should be increased compared to the channel gross bit rate by a factor equal to spreading factor (usually in the order of 100) and this makes CDMA prohibitive for bit rates higher than 1 Mbps.

Moreover, the above equation is valid when all the signals at the receiver arrive with equal power. For this reason, power control must be applied, otherwise the power of the received signals will depend on the distance between transmitter and receiver. However, in the case of CDMA, the transmitters of undesired signals close to the receiver interfere strongly with the desired signal and prevent proper communication leading to reduction of system capacity [PAHL'88].

3.1.2 Advantages of a centralised approach

The medium access technique and the topology of the network entities are strongly related. With contention based protocols, the decision for transmission is taken individually by every node, so the resulting topology is peer-to-peer. On the other hand when polling, or TDMA, or CDMA based protocols are used, it is mandatory to define an entity responsible for channel
assignment to terminals. This entity is the BS and the resulting topology is star with the BS acting as the central element scheduling all the transmissions in its coverage area (cell). The BS, not only schedules the transmissions but also relays the packets towards its coverage area or the fixed network where it is attached. Thus, the existence of the BS eliminates the hidden terminal phenomenon.

The installation of several BSs with partially overlapped coverage areas allows the definition of clusters of cells. To avoid co-channel interference, neighbouring cells use different carriers, so an FDM technique is applied among the BSs with the same carrier being used by two BSs separated by a well calculated parameter called re-use distance. Because BSs are fixed their coverage areas are deterministic or predictable and network planning or coverage area planning is easier compared to peer-to-peer networks. Furthermore, there are more benefits resulting from the presence of the BS in star based topologies:

- The transmission range of a terminal corresponds to the radius of a cell for a star based network, while in a peer-to-peer network the transmission range of every node should be equal to the diameter of the network. This results to lower power consumption for the star topology.

- Power consumption is also achieved because of transmission scheduling. Generally power of a terminal is consumed by both its transmitter and its receiver. In a packet scheduling scenario, power is not consumed by the receiver to sense the carrier. Moreover, terminals do not have to be involved in unsuccessful packet transmissions because of collision and enter the sleep mode until the instant that has been indicated by the BS for packet transmission or reception. In HIPERLAN, in order to introduce a power conservation function a comparatively high overhead has been included in the whole protocol architecture [HIPERLAN'95].

- BSs belong to the fixed part of the network, so they do not have power consumption problems. It is easier then to execute more efficient multiple access algorithms based on the global knowledge of traffic. Complicated channel access mechanisms are not recommended for portable terminals and there is no straightforward way for knowledge of the total traffic load. This has an impact on the capability of preserving bandwidth for hand-off and so on the capability for mobility support.

- Moreover, the capability of applying more sophisticated multiple access methods in star based networks, enable them to accommodate more than one and complex services with variable traffic patterns. This is a crucial issue considering the expectations from the ATM WLAN.
• The presence of the BS makes easy the deployment of control functions for security, mobility management, configuration and network management.

• If the wireless network must be interconnected with a wired infrastructure, the presence of the BS that are reachable by all terminals is an advantage. Then, the inter-networking functions can be integrated with the traffic controlling functions of the BS and guarantee access to the wired network for every terminal. Moreover, traffic from the access point to the wireless terminal will be a big part of the whole network traffic and the ability of a BS to tailor its multiple access scheme to the new traffic needs is quite advantageous.

Moreover, there are certain characteristics of ATM technology that favour the selection of a centralised solution regarding the topology and multiple access scheme of ATM WLAN:

• Point-to-point physical connections among network entities, with physically separated forward and backward directions of transmission. This results to centralised topology where terminals are connected to ATM concentrators or access ATM switches, to form the access part of the network. ATM concentrators are connected to high capacity ATM switches which form the backbone part of the network. The physical connectivity has a straightforward impact on the protocol architecture of ATM. The entities forming a network are divided into a group forming the user part (terminals) and a group forming the network part (switches). The signalling related to call control always takes place between an ATM terminal and an ATM switch or between ATM switches. There is no case to have signalling for call control between two ATM terminals. End-to-end signalling between terminals, exists only at the AAL layer of the ATM Protocol Reference Model (PRM) [UNI31] and in its upper layers, and only after a connection has been established. This issue influences the design of the ATM WLAN and the modifications applied to DECT and HIPERLAN.

• The existence of a call set-up phase, between a terminal and the network, for every connection in the system (whatever the connection lifetime is) where a traffic contract [TM'96] for the connection is defined as well as the route that will be followed by all the packets in that connection unless a fault condition occurs. From this perspective ATM networks are circuit switched. However, because the information is carried by packets with fixed size and format, it can be concluded that ATM networks combine circuit and packet switching.

• The deployment of dynamic (or on demand) allocation of bandwidth, based on the instantaneous traffic requirement of each connection, in contrast to ISDN where fixed bandwidth allocation takes place. Moreover, traffic monitoring functions are applied per
virtual circuit or per virtual path to prevent network congestion and bandwidth violation by connections that do not conform to the agreed (during call set-up) traffic contract.

Based on the above considerations, it was decided that a centralised solution should be pursued regarding the multiple access scheme of ATM WLAN. An initial design was based on Adaptive TDMA with Frequency Division Duplex.

### 3.1.3 D-TDMA/FDD based MAC

The initially adopted solution is a TDMA based multiple access scheme in which, transmissions from terminals to the BS occur in the uplink frequency band while the BS broadcasts every transmitted packet in the downlink frequency band. The uplink frame is divided in time slots and in every time slot one ATM cell is transmitted, followed by a guard period which is determined by the size of the coverage area of the BS (cell). The source or destination terminal is identified by the VPI field value in the ATM cell header.

In this study, a delay spread of 50 nsecs has been assumed [STEEL'92] giving a maximum channel symbol rate of 10 Mspis based on the formula:

\[
R_s < \frac{1}{2D}
\]

where \( R_s \) is the channel symbol rate and \( D \) is the delay spread.

If equalization is used then even higher channel rates could be achieved at the expense of more complicated receiver, power consumption and usually packet overhead [RAYCH'92].

The selection of the error control strategy and the synchronization scheme in the physical layer was out of the scope of this study. However, it was taken into account that the physical layer of the ATM WLAN must provide a uniform error control platform to all the connections, irrespectively of the service they support. Considering synchronization, the ATM cell header has been modified as shown in Fig. 3.2, to provide an elementary synchronization sequence. Moreover, by reducing the length of Virtual Path Identifier and Virtual Circuit Identifier fields, a more powerful error control scheme for the ATM cell header is supported.

In the uplink and during a time slot period, one of the following transmissions can occur (Fig. 3.3):
1. A packet is transmitted by a terminal in a non-contention mode, carrying user or signaling information. The VCI field determines the connection and the kind of transmitted information. The value 0 of the VCI field identifies the control channel in the system.

2. Terminals which are switched on, transmit registration messages in a contention based mode in order to declare their presence in the network and be assigned VPI values by the base station. The time slot in that case is subdivided into a number of "reservation request subslots" and every terminal to be registered selects one of these subslots and transmits in a contention mode its registration message. The successful access of the radio channel by one of these terminals will result in the allocation of a control channel so that the required handshaking, for the registration, will take place.

3. The time slot is divided into subslots which are used by registered terminals without established connections to indicate that they would like to access the control channel. These subslots are preassigned to terminals by the BS in the previous frame duration, so the transmissions of this kind happen in a contention free mode (polling).

4. The time slot is divided in "request back subslots". Each "request back subslot" is assigned, to an established connection to indicate a source transition from inactive to active. It is also followed by a guard period.

In the GFC field of the packets transmitted in the uplink, the following messages are carried, related to the MAC functions:

1. "Transition to inactive": time slots allocated to the corresponding connection should be reassigned to other, active connections until the current source becomes active again.

2. "Increase bandwidth": the source requests more network capacity.

3. "Decrease bandwidth": the source Quality-of-Service (QOS) could be satisfied by less allocated capacity.

4. "Control Channel Request": an already active terminal uses the transmitted packet of the corresponding connection in order to request bandwidth allocation for a control channel, so that it can initiate another call or exchange management information with the base station.
Based on the messages carried in the GFC field of the packet headers in the uplink, and in the "request back subslots", the base station reassigns the time slots of the uplink frame to active connections, to periods of control channel requests and to periods of registration, as was mentioned above. The base station informs the registered terminals about the allocation of the uplink slots, transmitting appropriate broadcast packets in the downlink. The allocation of "request back subslots" is fixed for the whole duration of a call.

In order to determine how to reallocate the uplink time slots, the base station executes the "Allocation Algorithm" which assigns different priorities to different types of services and uses scheduling methods borrowed from Operating Systems [TANENB’92], such as Round Robin scheduling. Furthermore, it divides the set of established connections into time slot owners and time slot renters. Different queues are assigned to owners and renters corresponding to different channel access priorities (high and low). The time slot owners acquire their time slots as soon as possible after having reported transition from inactive to active. The time slot renters are just taken into account by the Allocation algorithm when an available time slot has been found.

To achieve synchronization among different connections which have common end points and belong to the same application, it is required that these connections be accepted in the network as owners. In these case they are always assigned with their required bandwidth and synchronization, in terms of satisfied skew, is achieved.

The above described access mechanism introduces some delays to sources when transitions from inactive to active occur because of the delay between the transition report and the base station reaction, but for delay sensitive services these delays are always under the limits and in the case of owners they are equal to, or less than an uplink frame duration. Moreover, the bandwidth efficiency is preserved because time slots are used mainly to satisfy the currently active traffic sources. The existence of the "request back subslots" in the uplink wastes a small part of the available capacity at the MAC but an adaptive algorithm controlling the number of these subslots in uplink time slots will result to better bandwidth efficiency. Moreover, the usage of these subslots improves the performance of the multiplexing scheme and compensates for the above mentioned bandwidth waste.

3.1.3.1 Simulation Results

To prove that, when the above described solution is adopted, the delay requirements of the delay sensitive services are met, simulations were carried out using the BONeS simulation package. A network consisting of a base station and four terminals capable to overload the uplink was designed where each terminal could establish six connections simultaneously. In three of these connections the wireless terminal was transmitting data generated by simulated
voice, video and disc sources. In the other three connections the terminal was receiving data from the corresponding sources of another terminal.

The simulation model used for the video sources is based on [MAGL'88]. Every 33 msecs (picture frame inter-arrival period) a new data rate is calculated to emulate the VBR nature of MPEG coders. The voice source model is based on [BRADY'69] and is a Markov model with two states VS0 and VS1. The duration of VS0 is given by an exponential distribution with an average value of 650 msecs and during that state there is no source activity (silence). The duration of VS1 is derived by an exponential distribution with an average value of 350 msecs and is characterised by packet transmission at a constant bit rate of 32 or 64 kbps. Finally, the sources modelling the file transfers are defined by a two state Markov model where state DS0 has a duration given by an exponential distribution with a mean value of 14 msecs and no transfer takes place to emulate the seek time of a hard disk drive, while state DS1 lasts for 204.8 μsecs and is characterised by the transmission of 512 bytes (1 disk sector) of information at 20 Mbps (Fig. 3.4).

During the simulation, it was assumed that there are no transmission errors in order to measure the multiple access delay. The transmission errors should affect the performance of the presented scheme when they occur in the downlink packets which carry the information related to the allocation of uplink time slots to terminals.

In Fig. 3.5 the delays that were experienced by the different classes of packets under different patterns of traffic are presented. In simulation I there is no video source active. In simulation II, video sources become active and access the uplink frame after the middle of simulation period. Finally, in simulation III, the video sources are active from the beginning of the simulation causing the network to be intensely loaded.
The delays experienced by voice and video packets are always upper bounded and in the same range while the delays experienced by data packets vary according to the total active traffic in the network. The access delay for the data packets is non upper bounded and is affected mainly by the presence of video connections in the network.

It is expected that introducing a Call Admission Control which regulates the mix of the types of the accepted connections, even the delays experienced by file transfer connections will be reduced. In this case the blocking or call queuing probability will increase for the same call inter-arrival rate.

The simulation results showed that the delays experienced by time critical services are acceptable and upper bounded. The absence of a Call Admission Algorithm showed that low priority connections experience non upper bounded delays. The existence of such an algorithm should certify the more efficient bandwidth utilization and fairness among different types of services. Moreover, if the distinction of the established connections is based on their conformance to their negotiated traffic contract and not on the their delay requirements, the bandwidth allocation mechanism in the ATM WLAN should always guarantee a portion of the system capacity to non-delay critical connections. This approach was followed in the mechanism for the bandwidth allocation presented in chapter 4.

Moreover, to evaluate the efficiency increase compared to peer-to-peer solutions, it was decided to compare the distributed multiple access algorithm used in HIPERLAN with a D-TDMA mechanism that follows the same algorithm with HIPERLAN to assign access priorities to terminals. Then the difference in performance is due to the difference in the overheads required for the peer-to-peer and the centralised solutions.

Fig. 3.5 Simulation Results
3.1.4 D-TDMA/TDD mechanism

Based on the mechanism used in the HIPERLAN MAC for assigning channel access priorities to packets, a packet scheduling algorithm was defined, which is a modification of the one presented in the previous section, so that the priorities of the connections are defined during call set-up, as well as during the call, based on the residual lifetime of the corresponding packets or bursts. The defined scheduling algorithm could be used in the MAC layer of the ATM WLAN or be introduced in the MAC layer of HIPERLAN to support the ATM WLAN concept. The mechanism assumes the existence of BSs servicing wireless terminals. Two queues residing in the Base Station (BS) were defined in 3.1.3 [APOST'95], corresponding to different connection classes, and different priorities were assigned to them. In the D-TDMA/TDD scheme the multiple queues have been replaced by a unique queue and an algorithm which queues and sorts burst transmission requests, based on the assumed lifetime of the packets and on the property of the connection to be “owner” (guaranteed bandwidth on demand) of time slots. Terminals are still able to initiate fast bandwidth requests, using dedicated subslots at the end of a TDMA frame or the GFC (Generic Flow Control) field in the header of the transmitted packets. The base station informs about the slot allocation using broadcast packets. Terminals and base station transmit and receive at the same frequency band, and the transmitted packets have an ATM cell-like structure. The transition from FDD to TDD enables the definition of a dual mode PRM (as explained in section 3.2) and increases the bandwidth utilisation when calls with asymmetric traffic are initiated.

The overhead of the D-TDMA algorithm is fixed per TDMA frame and consists of the guard periods between time slots, the transmissions of the broadcast packets (every 15 time slots) and the subslots used to ensure bandwidth requests. It is calculated that for a frame with 6 msecs duration the maximum achievable channel throughput is 91.8 % while for a frame with duration 1 msec it becomes 89.04 %. It is expected that the flow control functions and the Call Admission Control will keep the actual channel throughput close to its maximum values while preserving the QoS for different services.
3.1.5 *Comparison between EY_NPMA and D-TDMA/TDD.*

To evaluate the performance of the EY-NPMA, used in HIPERLAN channel access mechanism, a simulation model in BONeS was designed, able to model several aspects of the technique under investigation.

In the model, the HMPDU transfer function in every node calculates the channel access priority of all the packets residing in that node whenever a channel access cycle starts or whenever the channel is assumed free. The packet with the highest access priority is considered for transmission. All terminals with packets follow the channel access cycle as it is described in [HIPERLAN'95].

Variable Bit Rate (VBR) services modelled based on [MAGL'88], with average bit rate of 2 Mbps and maximum lifetime 12 msecs per packet are assumed in both HIPERLAN and ATM WLAN. Packets experiencing delay longer than their lifetime are dropped. In the case of D-TDMA a frame with duration of 6 msecs was assumed and the transmitted packets were 424 bits long. In HIPERLAN, bits generated in one picture-frame duration (around 33 msecs) form one or more packets for transmission with lengths up to 23312 bits. So, in D-TDMA the sources produce constant length packets and bursts with different lengths, while in HIPERLAN they produce packets with different lengths.

Fig. 3.7 shows the average access delays, the packet dropping rate and the channel utilization. The channel bit rate is 20 Mbps. HIPERLAN does not contain any call admission control functions and assumes that the upper layers should deal with this problem. In Fig. 3.7, one can see HIPERLAN performance for populations up to 14 terminals. On the other hand, the Call
Admission Control functions in ATM WLAN, prevent the establishment of more than 8 connections simultaneously.

Assuming that a packet dropping rate around 2% can be tolerated, the maximum capacity of the D-TDMA is 8 VBR users (totally 16 Mbps on average) and for EY-NPMA is 6 VBR users (12 Mbps on average). Although the performance of EY-NPMA looks better for light traffic loads in the network, D-TDMA takes over in high traffic intensity situations. Considering delays, in the case of EY-NPMA, a packet could be transmitted after a channel access cycle is finished while in D-TDMA an indication for burst generation can be delivered at the base station after waiting for one frame period. This is why the delay experienced by packets in D-TDMA is relatively high even at low traffic loads. The adoption of shorter frames will decrease the reaction time of the base station to bandwidth requests and could improve the delay metrics. Moreover, it is worthwhile to mention that the capacity of standard HIPERLAN (EY-NPMA) is higher than that of the modified HIPERLAN for ATM presented in 2.4 (12 Mbps versus 4 Mbps in Fig. 2.12). This happens because more information bits are included in the standard HIPERLAN payload and thus the transmission efficiency is higher. In 2.4.2 it was shown that the transmission efficiency of modified HIPERLAN was 22%. However for standard HIPERLAN the transmission efficiency could reach 94%.

So, it is proven that a centralised solution for the multiple access scheme would increase the capacity of the ATM WLAN. Such a multiple access scheme requires the existence of a Base Station (BS) to control the transmissions of ATM WLAN terminals. The existence of the BS will bring more benefits like:

- Simplicity in terms of signalling and processing related to the network operation. The required signalling involves only the BS and the terminal and there is no need to exchange topology information among all the terminals like it happens in HIPERLAN.
- Adaptive bandwidth provision to connections (based on ATDM\(^7\)) and support of the QoS of the same types of applications as those existing in the fixed ATM LAN. This can happen by specifying the kind of connection according to [TM'96] during the call set-up phase, and introducing certain multiple access mechanisms for different connection kinds.
- Access of the network resources by user terminals with the least possible overhead, since the processing related to the packet scheduling occurs in the BS.
- Minimum reconfiguration overhead whenever a node joins or abandons the network. This happens because the entities involved are only the BS and the terminal that joins or

\(^7\) Asynchronous Time Division Multiplexing: the basic multiplexing concept in ATM oriented networks
abandons the network. Moreover, hand-off operations are easy to be defined and deployed within a centralised topology.

- Low power consumption by the wireless terminals, since they can schedule reception based on a fixed frame structure defined by the BS, or power on for transmission based on the packet scheduling information provided by the BS, too.

- Guaranteed allocated bandwidth to coexisting networks with different owners. This happens by allocating different frequency bands to different BS. Then, the deterministic coverage area of the BSs results to easy frequency planning. Moreover, the expected traffic in certain places could drive the frequency allocation and the placement of BSs.

![Diagram of ATM WLAN network configuration](image)

**Fig. 3.8 Network configuration for ATM WLAN based on fixed infrastructure**

So the study of various multiple access schemes and their implications on ATM traffic, and the comparison between the centralised TDMA scheme and the peer-to-peer EYNPMA mechanism justifies the selection of a star based solution for the topology of ATM WLAN, with a centrally controlled multiple access mechanism in the wireless access parts of it. Then, Fig. 3.8 presents the possible network configurations for an ATM WLAN that integrates wireless and wired access within the same backbone infrastructure (ATM switches). In this case ATM switches are enhanced with the mechanisms described in chapter 6 to support mobility of terminals.

Considering the multiple access scheme, Fig. 3.8 shows that it can be implemented under different network configurations depended on the customer needs. Fig. 3.8.(a) could apply in a residential or small business environment where every independent house or small enterprise deploys an ATM WLAN system which contains the wireless terminals and the BS responsible for the multiple access and the switching of traffic within its coverage area.
Moreover, this BS could be attached to a public ATM switch for provision of public services (e.g. telephony, videophone, internet access, video on demand etc.).

Fig. 3.8.(b) could apply in big business environments where a big amount of traffic has to be switched among the wireless access parts. Moreover, this configuration could be adopted for deployment of multimedia micro-cellular networks. Under this configuration, the multiple access mechanism is controlled by a processor within a port of the ATM switch. BTS acts only as a repeater, extending the radio transmission in the wire and vice versa. For large scale deployment the configuration in Fig. 3.8.(b) will result to reduced cost, compared to the configuration in Fig. 3.8.(a).

3.1.6 Fully Wireless ATM WLAN

In the case of several wireless standards [HIPERLAN’95], [IEEE] wireless terminals give the opportunity to establish ad-hoc networks (i.e. without requiring fixed network infrastructure like the one presented in Fig. 3.8). This is an advantage of HIPERLAN compared to DECT where communication between terminals is not possible without the FRP and the CC. In the case of an ad-hoc network it is expected that the services to be supported will be confined to single file transfers. Ad-hoc networks are established between users within small range so services like telephony or videophone will be obsolete. Moreover, services based on fixed infrastructure (e.g. video on demand) are out of the ad-hoc content.

As was presented above, ATM WLAN aims to achieve integration in two dimensions. First service integration is introduced over a uniform transport mechanism and according to the ATM concept. Second, the introduced service integration has to be supported for both fixed and mobile terminals (environment integration). One step forward, is to achieve service support for both ad-hoc networking and infrastructure based networks.

Furthermore, there could be the case where the full range of services must be supported within a region, but no wires can be used (e.g. listed house). Then the wired infrastructure is not available, ad-hoc networking is not adequate for multimedia and one BS cannot provide the required user capacity. Then, the backbone network has to be provided by wireless means (radio links).

Fig. 3.9 shows the two cases of a fully wireless ATM WLAN system. In Fig. 3.9.(a), an ad-hoc network is established among five ATM WLAN terminals. The applications in these terminals are not aware of the absence of the required BS. By exploiting the topology and routing functions of HIPERLAN, and introducing a multiple UNI controller within every terminal, a transparent virtual circuit can be set-up between any two terminals in the ad-hoc network. This solution is described in 3.2.7. Moreover, Fig.3.9.(b) shows that network
Chapter 3: New ATM WLAN

The PRM in effect at the above defined interfaces has to take into consideration the different environments (ad-hoc, fully wireless, integration of UNI and M-UNI by the same M-NNI). The objective is to define a global PRM for M-UNI or M-NNI with a super-set of component functions. Certain components in the different layers will be activated for different environments. For example, the UNI controller that defines the interfaces among terminals in an ad-hoc network is redundant when the terminal is serviced by a BS. Since the operating environment of the network and the multiple access mechanism are highly related, the components of the PRM that correspond to different operating environments should reside below ATM layer which is common for every environment.

In Fig. 1.6 the whole PRM for the ATM WLAN was presented based on the definition of B-ISDN PRM. Fig. 3.10 presents a refinement of the lower two layers of that PRM, namely ATM-MAC and PHY layers.

3.2.1 Radio Transmission Specific Functions.

These functions deal with RF transmission and reception, channel sensing, modulation and demodulation, line and channel coding, as well as packet synchronisation.

More advanced topics in this layer are related with the effort to increase the bit rate of the channel. This could be done using several approaches like multilevel modulation, adaptive equalisation and Multi-carrier Modulation (MCM) [SHEL'95]. The payload of the transmitted packet should be one ATM cell (possibly with a small overhead added by the MAC sublayer), because according to the ATM concept, the ATM cell is the elementary structure transferred in the ATM layer peer-to-peer communication and delivered by the ATM to the lower layers. Furthermore, the overhead added by the Physical layer should not be substantial compared to the ATM cell length (53 bytes). For example, this not the case in the HIPERLAN where there is a training sequence carried in every physical layer packet and being 450 bits long, i.e. longer than an ATM cell. If one ATM cell-like packet was transmitted the efficiency would be less than 50%.
Although it is a trend to research in the area of equalisation to achieve high channel rates, it is believed that other techniques like Multi-carrier Modulation or single user CDMA could be appropriate for the physical layer of the ATM WLAN. These methods do not require long training sequences and enable the existence of a common transceiver structure for every wireless environment.

Another issue concerns the error control method used in the physical layer (channel coding). The same physical layer is expected to support different kinds of services with different delay and BER requirements. The coding scheme should not impose delays that degrade the QoS. Although the error control method should be optimised according to the radio channel characteristics it should also take into account the service quality limitations. In some cases interleaving is used to alleviate the BER of the channel. In ATM WLAN interleaving involving more than one cells should not be used, at least in the physical layer, because of two reasons:

1. It increases the delay experienced by the ATM cells of the connection, contributing to the QoS degradation, because it reduces the delay to be allowed in other layers of the protocol stack and at the intermediate nodes of the network.
2. As it is expected that a packet with one ATM cell in its payload will be transmitted in the lowest layer, the process of interleaving will require a sequence of ATM cells to be delivered to the physical layer, to achieve a reasonable interleaving depth. The physical layer should accept one ATM-cell-like packet by the MAC layer with every data request, so the contradiction with the previous requirement is obvious.

It is expected that a concatenated code is more appropriate for the indoor channel which is characterised by burst errors.

### 3.2.2 Suitable Channel Access Cycle.

To support a hybrid channel access scheme for the ATM WLAN, different approaches have been considered. Initially, it was proposed that the MAC layer mechanism be based on the CDMA technique. In that case, when terminals are co-ordinated by a base station, a CDMA/TDMA approach could be followed. Terminals use one or more codes to transmit data in time slots allocated to them by the base station. TDD or FDM could be used to separate the uplink and downlink channels. Terminals do not have to keep information about the topology of the network. All the traffic is carried through the base stations. However, when terminals are gathered to form ad-hoc networks a CDMA/CSMA method could be used. Following this method, every terminal is assigned one receiving code or even more than one receiving codes. Whenever a node has a packet to transmit to one of its neighbours, it
selects one of the receiving codes of that neighbour and acquires the channel. Simultaneous transmissions using a specific code will cause multi-user interference to the corresponding receiver. Then, a CSMA technique could be used, to enable or disable terminals’ transmissions based on the level of that interference. Nodes should have the capability to denote free periods for every receiving code. They should have also the capability to inform when they accept reservation requests to allocate periods of time to a receiving code. To do so, terminals should be synchronised to have the same reference of time. The information about the frequency and the duration of the reservation periods can be obtained by the ATM layer and according to the traffic description of the established connections. In ad-hoc networking, terminals should acquire knowledge of the network topology exchanging messages between each other in a similar way to HIPERLAN or IEEE 802.11. Another serious problem in the case of CDMA/CSMA is that of the power control. The transmitted power by the terminals is totally uncontrolled. If a receiver uses one receiving code, the rest of the codes, used by other terminals to receive information, cause multi-user interface. For these reasons, the CDMA solution was dropped and it was decided to combine the HIPERLAN mechanism for channel access for ad-hoc networking and an optimum solution for ATM WLAN when the required infrastructure is present.

Thus the Channel Access Cycle sub-layer of the ATM WLAN PRM is used when ad-hoc networking takes place. It is based on the EY-NPMA concept that was described in 2.3.2. It is inherited from the HIPERLAN PRM to support a totally random multiple access mechanism. Every transmitted packet contains the MAC address of the next terminal to receive it. This terminal is in the path to the packet final destination. The successful reception of the packet by the indicated terminal is followed by an acknowledgement to confirm the successful access of the radio channel.

3.2.3 Super-frame Delineation Function

This function deals with the synchronisation to the air interface defined by a BS. This air interface is defined in detail in chapter 4. The detection of the packets broadcasted by the BS happens in this sub-layer. Moreover, the Super-frame delineation function reports the beginning of every new time slot to the upper sub-layer.

3.2.4 Packet Scheduling Function

The Packet Scheduling Function is the heart of the multiple access scheme in ATM WLAN and is applied for a system with BSs (fixed backbone infrastructure). Chapter 4 will describe in detail the adopted mechanism. Roughly, the terminal declares current bandwidth requirements for every established connection. Then, the BS predicts the packet inter-arrival
process based on these bandwidth requests, and calculates the multiple access priorities of the generated packets based on the traffic contract that was negotiated during call set-up. Moreover, the BS broadcasts the packet scheduling information in its coverage area. The declaration of bandwidth requests happens whenever the packet inter-arrival process of a connection changes.

According to figure 3.11, the channel has a framing structure. The length of the frame can vary according to the traffic which must be transferred between the base station and the wireless terminals. The base station can indicate the beginning of the next frame. Every frame begins with a sequence of short time slots allocated to calls for their whole duration. They are used by sources as sockets to indicate active state to the base station. The rest of the frame contains time slots that can accommodate the transmission of a physical layer packet of the ATM WLAN. They can be accessed by the base station or a wireless terminal. The base station sends periodically a slot allocation packet which indicates which terminals will transmit in the following time slots. This slot allocation packet could indicate also the beginning of the next frame.

3.2.5 Packet Transfer Function

In an ad-hoc network, there is no packet scheduling information provided by the BS. Moreover, a terminal has no knowledge about the established connections of other terminals. However, there must be a mechanism to sort the packets within a terminal according to their multiple access priorities. Packet Transfer Function provides this mechanism. It is similar to the MAC HMPDU Transfer Function defined in HIPERLAN, but it uses different criteria to set the packet access priorities. The calculation of the multiple access priorities is based on the traffic contract during call set-up. Then, a GCRA [TM’96] algorithm is used per connection to assign the priority of every packet. Moreover, the Packet Transfer Function has to rely packets towards their final destination.
3.2.6 Routing Information Exchange

The objective of this sub-layer is to advertise the MAC address of the terminal and its neighbours participating in an ad-hoc network, as well as to collect the MAC addresses of all terminals participating in this network. Moreover, it calculates the route to any terminal in the ad-hoc network. This information is passed to the Packet Transfer Function which adds the MAC address of the next hop to be followed by a transmitted packet.

3.2.7 Multiple UNI Controller

This component of the ATM WLAN PRM comes into effect when add-hoc networking takes place. For every other terminal that is detected in the current ad-hoc network, a different UNI has to be defined and associated with a different VPI value. Within the ATM-MAC layer, that UNI is specified by the MAC address of the corresponding terminal. This MAC address has been advertised as described in the previous paragraph, following a procedure similar to HIPERLAN LookUp function. Moreover, this MAC address is associated with the ATM address of the corresponding terminal. During Call Set-Up, the VPI value of the new connection is assigned based on the ATM address of the other communicating party. This VPI value is used by the Multiple UNI controller to define the MAC destination address in the transmitted packet.

![Fig. 3.12 Frame structure of the channel between switch and MUX or between switches](image)

3.2.8 Air Interface for wireless M-NNI

When the ATM switches are interconnected using radio links a MAC scheme has to be defined for the M-NNI. Fig. 3.12 shows the air interface applying between two ATM switches. According to Fig.3.12, the channel between two switches or one switching and one multiplexing terminal has a framing structure and is divided between the two elements on a TDD basis and in two equal (in terms of number of time slots) parts. Every part (or direction) contains a number of short slots which are used to carry fast requests for bandwidth
allocation. The rest is divided in time slots carrying physical layer packets in one direction. After the short slots the switch controlling one direction sends a control cell which contains also the number of the slots which will be used in that direction. The rest of the slots in that direction could be used by the other switch, i.e. a change in the direction could happen for some slots during one frame period.

3.2.9 ATM layer

ATM layer generally controls the way the MAC layer assigns the channel bandwidth to the terminals of the ATM WLAN. It executes Call Admission Control to adjust the long-term mix of traffic in the network. Call Admission Control is executed to all the intermediate nodes which belong to the communication path. Traffic policing functions disable the violation of the QoS. These functions take place mainly at the points where traffic multiplexing is performed, to enhance the network stability and prevent the degradation of QoS. Instantaneous traffic requirements are scheduled to define the short-term mix of traffic. These instantaneous channel requests are carried along the connection path to allocate the required bandwidth for the transferred packets. Moreover, ATM at the intermediate nodes of a fully wireless network, performs switching of the transmitted packets towards their final destination. Fig. 3.13 shows a proposed schematic diagram of the ATM layer related functions.

![Fig. 3.13 MAC-ATM layer](image-url)
The ATM WLAN supports several types of services with different traffic characteristics and different time constraints. These different kinds of services experience different priorities when they content for the channel capacity. However all the kinds of services use the same parameters to describe their traffic patterns. Currently, peak rate, average rate and burstiness are considered as the parameters for traffic characterisation. The messages that are used by the ATM layer (and possibly by the MAC layer) and are related to the bandwidth allocation should be the same for all the services. However, the internal resources, such as queues, could be different to satisfy the QoS.

ATM layer deals with these different types of traffic in a similar way. At the points of the network where traffic multiplexing is performed, ATM layer controls the bandwidth of every connection separately and tries to allocate bandwidth to connections only when it is required. At the points of the network where switching is performed (i.e. in the backbone network), the channel bandwidth is divided among the nodes based on the occupancy of their buffers.

### 3.3 ATM Adaptation Layer

AAL is the end-to-end transport protocol among the communicating ends of a call. There are different types of AALs according to the supported kinds of services. The already standardised AALs should be used in the ATM WLAN, enhanced by the service-specific error control functions, to compensate for the increased BER expected in the radio channel. Furthermore, AAL could deal with the synchronisation problems of the multiconnection calls. There are two kinds of multiconnection calls which can exist independently or in a combined mode:

1. Multiparty calls where different connections are associated with different communicating ends. These ends could be located in different places, so the propagation delays could be different.

2. Calls composed by parallel connections with the same source and destination. In this case the relative delays of the transmitted packets are of great interest. The upper layers could deal with this problem and most of the multimedia programs do. But they can tolerate a certain relative delay, so AAL should guarantee this allowed margin of delay.

However, if additional functions are considered in the AAL of the ATM WLAN a problem arises, that of the interoperability between an ATM WLAN and other ATM based networks. So an internetworking unit would be necessary to transform WLAN AAL packets to other AAL packets, but the fixed ATM LAN and the ATM WLAN would not be compatible at the ATM layer any more. Thus, in order to preserve the compatibility between
the two networks, it is appropriate to deal with the virtual connection synchronisation problem in the multiple access mechanism of ATM WLAN. Moreover, the alleviation of the increased bit error rate that is expected in the radio channel should be dealt in a layer below ATM.

AAL layer instances are initiated by the upper call management plane to provide reliable end-to-end transmission. The operations carried out are:
- maintenance of buffers to assemble and disassemble the SDUs,
- delivery of SDUs to upper layers and provision of cell payload to ATM layer,
- error control functions and related peer-to-peer communication,
- indication to the lower layer of the involved traffic source (idle, active, next burst ready).

The above mentioned functions are contained in the standardised specifications of the AAL.

3.4 Upper layers, Management layers

The services of the AAL layer are not available only for transfer of user data, but also for the transmission of messages among the management and maintenance functions of the network. These messages will be carried in ATM cells transmitted using ATDM by the MAC layer.

The introduction of ATM in the WLAN should result to simpler frame structures compared to already existing specifications (like DECT). Functions related to network addressing (allocation of ATM addresses and possibly VPI/VCI values to terminals) could happen according to the existing standards. The enhancement in the case of the WLAN is the addition of the hand-off function.

Based on Fig. 3.10, in the case of ad-hoc networking the Routing Information Exchange Function is used to carry out the registration, routing and location update procedures. However, when an optimum cellular like ATM WLAN is deployed, the existing signalling protocols of the ATM protocol suite could be exploited to support address registration, location management and in call-handoff. A proposed solution is presented in chapter 6.

Thus, the upper layers in the ATM WLAN protocol reference model deal with the following aspects:

I. Registration of terminals to a network
   A. access procedures
   B. registration messages (peer-to-peer communication)
   C. address allocation
   D. VPI/VCI values allocation
II. Network updating and mobility management
   A. updating messages (peer-to-peer communication)
   B. updating functions (hand-off)
   C. mobility management (topology data bases)
   D. routing determination based on the topology modifications

III. Call management
   A. establishment of permanent connections
   B. set-up and release of temporary connections
   C. VPI/VCI values allocation
   D. initiation of required lower layer instances and Service Access Points

3.5 Conclusion

From the study of several candidate multiple access schemes, it was concluded that contention based techniques waste bandwidth because of retransmissions and the overheads to execute the contention based algorithms. Moreover, they are characterised by non-upper bounded delays. Furthermore, CDMA techniques would require an extremely wide frequency spectrum to support the bit rates of ATM WLAN. Additionally, fixed TDMA or FDMA are not adequate to provide bandwidth allocation on demand. An adaptive TDMA technique seems to be the optimum solution especially when the benefits of a centralised approach are considered against peer-to-peer network architectures. However, since ATM WLAN objective is to support ad-hoc networking too, a superset of PRM functions was defined. Based on the mode of operation, certain functions are enabled. The mechanisms in ad-hoc case are inherited from HIPERLAN while the optimum centralised approach favours a reservation based adaptive TDMA/TDD scheme.
Chapter 4

Multiple Access Scheme with Traffic Policing for ATM WLAN

In this chapter, a TDMA based mechanism is presented, which provides bandwidth guarantees to virtual connections, and enables the QoS assurance when the information sources respect a traffic contract negotiated during the call set-up phase. Moreover, the proposed mechanism prevents the violation of the bandwidth assigned to a connection by sources which do not conform to their traffic contracts. The proposed mechanism in the MAC layer of the ATM WLAN can guarantee a minimum available bandwidth of a connection during the whole duration of the call. Furthermore, it has the capability to increase the bandwidth allocated temporary or permanently to a connection, if there is no violation of other connections' bandwidth. The proposed scheme offers a maximum packet access delay below the maximum affordable packet delay, for compliant virtual connections.

4.1 Introduction

The ATM WLAN should support all the possible services defined in ATM networks. Contention based methods deal quite well with asynchronous (bursty) traffic and allow adaptive bandwidth allocation (transmission rate follows source rate) under light traffic. This happens because every node controls its own transmissions. However, as presented in chapter 2, there are no bandwidth guarantees, i.e. violations of bandwidth allocated to connections could occur, and traffic with constant rate characteristics cannot be
supported efficiently. In the mechanism proposed in this chapter, the submitted traffic is assigned with access priorities based on the traffic contract negotiated during the call set-up phase. Moreover, bandwidth is allocated to terminals by the base station. Then, bandwidth of connections could be guaranteed and preserved, while congestion in the network could be avoided. Moreover, different applications are characterised by different traffic patterns (bursty, continuous, CBR, VBR) which can be declared during call set-up. Then, the allocation of bandwidth takes place according to the traffic patterns of the sources, to ensure the QoS of the relevant applications.

4.2 Traffic Management in ATM Networks

Traffic Management in ATM networks includes all the functions used to prevent, or react to congestion in the network. A subset of the traffic management functions are the traffic policing mechanisms (or Usage Parameter Control) that protect the network resources from malicious or unintentional misbehaviour which could affect the QoS of the established connections.

4.2.1 Service Categories in ATM Networks

According to [TM'96], ATM networks support five service categories (CBR, rt-VBR, nrt-VBR, UBR, ABR). These service categories relate traffic characteristics and QoS requirements to network behaviour. Whenever a new connection has to be established, its service category has to be defined. Moreover, a traffic contract between the traffic source and the network part performing traffic policing has to be agreed. The traffic contract consists of a traffic specification and a QoS commitment.

The traffic specification contains the values for the parameters characterising a connection of a specific service class. The parameters defined in ATM Traffic Management Specification are Peak Cell Rate (PCR), Sustainable Cell Rate (SCR), Minimum Cell Rate (MCR), Maximum Burst Size (MBS), Cell Delay Variation Tolerance (CDVT). Table 4.1 shows which traffic parameters are specified for every service category.

The QoS commitment contains the values for the QoS parameters associated to the service category of the established connection. The following QoS parameters are defined in ATM: Peak-to-peak Cell Delay Variation, Maximum Cell Transfer Delay, Cell Loss Ratio, Cell Error Ratio, Severely Errored Cell Block Ratio, Cell Mis-insertion Rate. The former three parameters are negotiated while the rest three are not. The QoS commitment means that the QoS parameters will be satisfied for the ATM cells that are emitted according to the
traffic specification. The traffic policing functions monitor the ATM cell stream of a connection and identify those cells possibly violating the agreed traffic specification.

<table>
<thead>
<tr>
<th>Traffic Parameters</th>
<th>CBR</th>
<th>rt-VBR</th>
<th>nrt-VBR</th>
<th>ABR</th>
<th>UBR</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCR</td>
<td>specified</td>
<td>specified</td>
<td>specified</td>
<td>specified</td>
<td>specified</td>
</tr>
<tr>
<td>SCR</td>
<td>unspecified</td>
<td>specified</td>
<td>specified</td>
<td>unspecified</td>
<td>unspecified</td>
</tr>
<tr>
<td>MCR</td>
<td>unspecified</td>
<td>specified</td>
<td>specified</td>
<td>specified</td>
<td>unspecified</td>
</tr>
<tr>
<td>CDVT</td>
<td>specified</td>
<td>specified</td>
<td>specified</td>
<td>unspecified</td>
<td>unspecified</td>
</tr>
</tbody>
</table>

**Table 4.1 ATM Layer Service categorisation based on definition of traffic parameters**

4.2.1.1 CBR

CBR service is intended to support real-time applications with tightly constrained delay and delay variation. However, it could be used by other applications too. As presented in Table 4.1, only the PCR parameter is specified. All the ATM cells transmitted at a rate up to the specified PCR should be considered conforming and experience the negotiated QoS. The source can emit ATM cells at the PCR at any time and for any duration.

4.2.1.2 rt-VBR

Like CBR, the real time VBR service category is intended for applications with tightly constrained delay and delay variation. However, in this case, if traffic is emitted at PCR for a long period the traffic conformance is violated. Then QoS cannot be guaranteed. Unlike CBR SCR and MBS are additional traffic parameters that must be defined to provide the traffic specification for a rt-VBR connection.

4.2.1.3 nrt-VBR

The non real time VBR service category is intended for applications without hard delay requirements. Unlike rt-VBR, where CTD and CDVT are specified as QoS requirements, in the case of nrt-VBR service there are no guarantees for delay constraints. Only the expected CLR is specified. However, the same set of parameters as in rt-VBR are used for the traffic specification.

4.2.1.4 UBR

Like nrt-VBR, the UBR service category is used for non real time applications. In UBR case, not even the CLR is specified. Moreover, no specification of any traffic parameter has to be provided.
4.2.1.5 ABR

ABR service category is intended for applications without tightly constraint delay requirements. It is the only service category where the traffic specification can change during the call and the QoS parameters keep on being satisfied. This is achieved through a flow control mechanism feeding information in the reverse direction, towards the traffic source. The latter is expected to use the feedback information to adjust its transmission rate and experience low CLR. PCR and MCR are the parameters required for the traffic specification.

4.2.2 Traffic Policing functions in ATM Networks

The traffic specification of a connection is provided for two reasons. First to provide the network with the necessary information to decide if the new call can be accepted or not. Based on the traffic specifications of the new call and the already established calls, the network predicts if it will be able to satisfy the QoS of all the above mentioned calls if it accepts the new call. Second the traffic specification parameters are fed in the mechanisms that monitor the traffic activity of a connection and report bandwidth violations. The latter occur as excessive transmission of ATM cells above the PCR, or above the PCR for a duration longer than a well defined period. The above mentioned monitoring and reporting mechanisms are realised by the Generic Cell Rate Algorithm (GCRA). Moreover, they reside in the network part of the UNI.

4.2.2.1 Generic Cell Rate Algorithm

GCRA is an algorithm that applies in a stream of transmitted ATM cells and characterises them as conforming or non-conforming (with respect to a certain traffic contract). The characterisation is based on the difference between the ATM cell actual and expected arrival times. The later are calculated from the traffic specification parameters.

GCRA is defined as a virtual scheduling algorithm or a Leaky Bucket Algorithm. The first ATM cell of a connection initialises the GCRA characterising all the subsequent ATM cells of this connection. Fig. 4.1 [TM'96] presents the two versions of the GCRA. Note that more than one instances of GCRA with different attributes could exist per connection. For example one GCRA examines conformance to the declared PCR and another GCRA checks ATM cell emission in respect to SCR. Then an ATM cell is conforming if all the GCRA instances characterise it as conforming. The state of the GCRA is updated only by conforming ATM cells (i.e. the expected time of the next ATM cell is set only when a conforming ATM cell is found). Moreover in every connection there are two different ATM cell flows identified by the value of the CLP bit in the header of the ATM cell. Since it is very hard that all the ATM cells of a connection conform to its traffic contract, the network
should be able to provide the required QoS for some non-conforming ATM cells. For this reason there are two GCRAs for every connection corresponding to CLP = 0 and CLP 0+1. Then an ATM cell with CLP = 0 must conform to both instances of the GCRA while an ATM cell with CLP = 1 must conform only to the GCRA applied to CLP = 0+1 flow.

**Fig. 4.1** Two equivalent versions of the GCRA

4.3 Traffic Policing in the MAC of ATM WLAN

In chapter 2 it was shown that the multiple access scheme of HIPERLAN is not adequate to prevent bandwidth violations in the air interface. HIPERLAN’s channel access cycle tries to provide a fair share of the radio channel to all the terminals with pending packets, and takes into account the delay requirements of the individual connections. However, bandwidth violations still occur mainly because of two reasons:

1. The high priority connections could emit traffic at high rates for long periods. Such traffic could correspond to non-conforming traffic in ATM. In the case of HIPERLAN, this traffic steals radio resources from complying traffic. This is presented in Fig. 2.8,
where a gradual increase on the packet dropping is observed for both complying and non-complying connections as the bandwidth violation increases.

2. In HIPERLAN, channel access priorities depend on the access delay experienced by individual packets. Thus connection priorities are not considered. However, as low priority packets are delayed, their access priority increases and, eventually, becomes higher than that of delay sensitive packets. Then violations of high priority connections occur.

It becomes apparent that the multiple access scheme of ATM WLAN should take into account the traffic specifications of individual connections, and assign radio resources based on these specifications and not based on the relative significance (or delay requirements) among established connections.

4.3.1 QoS metrics in the ATM WLAN

The QoS metrics taken into account are:

- **Access Delay**: at the air interface of the ATM WLAN, the objective for the packet access delay to be experienced within a connection has to be defined. The specified access delay should allow for the satisfaction of the end-to-end delay objectives of the supported services.

- **Relative Access delay variation** defined by the statistics of the time interval between the transmission of two subsequent packets of the same connection, and the statistics of the packet generation process.

- **Packet dropping rate** of a connection, calculated over the whole duration of calls or simulations as the ratio of dropped packets over generated packets.

- **Length of dropped burst** defined by the statistics of the number of packets (or bits) dropped between two subsequent transmissions.

4.3.2 Objectives of the proposed multiple access mechanism

The objectives of the multiple access layer, regarding the QoS of connections are:

- The transmission of submitted packets within the negotiated limits of access delay. End-to-end packet delay consists of the following components.
  1. **Access delay** in the calling party’s access network.
  2. **Queuing delay** in the backbone network.
3. Possibly, access delay in the called party's access network. Maximum access delays are defined per virtual connection during call set-up. Generated packets having experienced access delays longer than the negotiated are dropped.

- The preservation of bandwidth assigned to connections in the access network. It is assumed that a part of the total bandwidth is assigned per connection. When VBR sources transmit at a rate higher than their allocated bandwidth, they steal bandwidth allocated to other connections. It is expected that the backbone ATM network contains the mechanisms reported in section 4.2 to prevent such violations. Similar mechanisms should exist in the access network too.

- The assignment of bandwidth only to connections with pending packets or with active sources. This is another way to preserve QoS because available bandwidth allocated to a silent source can be given to support the excess bandwidth of an active connection.

The meeting of the above mentioned objectives will result in the satisfaction of the defined QoS requests.

4.3.3 Traffic patterns of connections

In this study the following categories of connections were identified based on the way that information is submitted by the corresponding sources:

- CBR connections, assumed active for the whole duration of the call. The management of these connections is rather easy. A constant portion of the total bandwidth, equal to the bit rate of the connection, is allocated for the whole duration of the call. In order to preserve the total bandwidth and enable several terminals to access the channel, the call duration could be specified.

- On-off connections, characterised by two states. An idle state when no transmission occurs and an active state when information is transmitted at a peak rate. An adaptive multiple access scheme could take advantage of the idle periods of connections and multiplex sources of the same type with total peak bandwidth more than the bandwidth of the network. The duration of active and idle states is given usually by exponential distributions and this provides a mechanism to adjust the multiplexing gain. For example, when a voice source is characterised by one idle state with mean duration 650 msecs and an active state with mean duration 350 msecs and bit rate of 32 kbps, it is expected that, on average, 3 sources could be multiplexed on a 32 kbps channel. The multiple access scheme could disable active connections to use the channel when they have exceeded a
negotiated period in the active state, or when they have exceeded a negotiated ratio of active period over silence period.

- VBR connections, transmitting at a range of bit rates. Although they can be assigned an average bit rate (calculated as generated bits over call period) it is expected that rarely they transmit at this rate. However, most of the time the transmission rate is close to the average rate. A useful information is the period of time they transmit at a constant rate. For example, when video is transmitted at a picture frame rate of 30 frames per second and the bit rate changes per image frame, every 33 msecs there is a change in the bit rate of the video source. In that case a bandwidth re-negotiation could take place every 33 msecs. Then, bandwidth required for a longer period of time can be calculated in an initial short period and packet scheduling can become easier. For example in a TDMA scheme with frame duration equal to 6 msecs, during a TDMA frame the current transmission rate can be calculated which will be valid for 5 more TDMA frames.

4.3.4 Exploiting silence and low bit rate periods of connections

When sources enter the silent period, the bandwidth assigned to them could be re-allocated to connections with pending packets. The silent sources should have a way to indicate their transmission to active in order to be granted the required bandwidth. This indication should be fast enough, to enable the fast allocation of bandwidth, although the mechanism used should preserve the total bandwidth in the network.

Let’s assume that voice sources, characterised by silences and talkspurts, are multiplexed in the network, and that the voice-source bit rate during talkspurt is 32 Kbps, while the affordable access delay is 20 msecs. Moreover, let’s assume that the transmitted packet has a payload of 384 bits (ATM cell like) and convolutional code (2,1) is applied to the generated voice information. Then, during talkspurts, one packet is delivered at the MAC layer every 6 msecs. If the packet experiences delay more than 20 msecs it is dropped. When a voice source becomes active at $t_0$, it should be granted with bandwidth to transmit its first packet, not before $(t_0+6)$ msecs and not later than $(t_0+26)$ msecs.

In PRMA [GOODMAN’89], transmissions take place in time slots which are divided in traffic time slots (assigned to active connections by the base station) and in reservation time slots which occur periodically (in well defined time intervals). All the sources entering into their active state, try to transmit the first packet of the active state in these reservation time slots. So, in PRMA it is guaranteed that if a voice source becomes active at $t_0$ it will contend for bandwidth after $(t_0+6)$ msecs. If the transmission is successful, then the base station allocates traffic channels to the active source, otherwise subsequent
attempts for transmission in the reservation slots take place. So, in PRMA it is not guaranteed that the voice source, becoming active, will transmit the first packet within $(t_0+6)$ msecs. This could lead to packet dropping because of excessive delay and because failure to transmit the first packet of the active state could influence the rest of the packets in that state. Bandwidth loss occurs as a result of the contention based (slotted ALOHA) multiple access of the reservation slots. The maximum achievable throughput is 40% (same sources in the network) or 50% (non-uniform sources) [TANENB'88]. However whenever a transmission in a reservation slot is successful, one information packet is transmitted and thus network bandwidth is preserved. Assuming that 2 slots are used for reservation every 6 msecs, the bandwidth assigned for reservation slots is 128 kbps (every slot carries 384 information bits). The minimum wasted bandwidth in that case ranges from 64 kbps to 76.8 kbps. Moreover, even if the bandwidth loss is affordable, the access delay experienced due to the contention based access of reservation slots could be prohibitive for multimedia services.

However, polling could be used as another mechanism to indicate the transition of a source from idle to active. A base station allocates periodically traffic slots to idle sources. Bandwidth is lost because of the lack of transmissions. The time interval between two traffic slots allocated to the same idle source could be based on the delay requirements of the source or on the statistics describing the source transitions. Assuming that one traffic slot is assigned every 20 msecs to an idle source (to compensate with maximum affordable delay), the wasted bandwidth is 19.2 kbps per idle source. Then 3 or 4 sources should be simultaneously idle to have the same wasted bandwidth as in PRMA.

Moreover, short time slots (called subslots) could be used to indicate transition to active or the existence of the first packet of an active source. Based on the information carried in these subslots, the base station could assign traffic slots to sources or it could keep excluding idle sources from the bandwidth allocation. Assuming 8 bits per subslot and 16 msecs between two subsequent subslots of the same idle source, the bandwidth assigned for subslots is 500 bps per idle source. Then, 128 sources could be simultaneously multiplexed and idle to have the same wasted bandwidth as in PRMA. Different kinds of services have different delay requirements, so the polling interval could be different for different connections. This could lead to a multi-frame structure on the air interface which consists of different cycles corresponding to connections with different polling requirements. A class of connections could be polled (using subslots) every 1 msec (polling channel: 8 kbps per connection), another class every 6 msecs (1.3 kbps) etc. Following this method we separate reservation and information transfer and we avoid contention in the traffic slots which are allocated to connections by the base station following packet scheduling mechanisms.
Chapter 4: Multiple Access Scheme with Traffic Policing for ATM WLAN

The ATM WLAN will support not only temporary virtual connections that will exist for the duration of the corresponding call, but also semi-permanent connections which could exist on a constant basis (i.e. the connection between an X-terminal and an X-server). When semi-permanent virtual connections are established it is expected that they will be highly asynchronous, so it will not be efficient to allocate reservation sub-slots to them or apply polling to them. It is better to provide a contention based mechanism for such connections. Moreover, it is more likely to have successful transmission when a lot of sort sub-slots are available for access than when there are few and longer time slots. To indicate pending bursts, it is proposed to introduce a number of sub-slots that will be accessed according to the slotted ALOHA principles. Occasionally, these sub-slots could be uniquely allocated to connections for polling.

4.3.5 Communication between terminals in the same cell

To preserve bandwidth of the network, the mobile terminals should be able to listen to each other’s transmissions (when they reside to each other’s coverage areas), to avoid the need of packet overlaying by the base station. Let’s assume a TDMA frame of 6 msecs where in every time slot, 384 information bits are carried. The bit rate per slot is then 64 kbps and this is the network bandwidth needed to transmit a packet between two terminals with overlapping coverage areas. For a packet to be transmitted between two nodes, two time slots are requested or a channel of 128 kbps, if there is packet overlay by the base station. It is assumed that bandwidth is still assigned by the base station. However, when communicating terminals are outside each other’s coverage area, packet overlay by the base station should take place. A means of estimating quality of reception among terminals should be provided, to decide if direct communication will be followed, or the base station will relay the packets of a connection. Moreover, when the quality of the radio link between two wireless terminals deteriorates (because of shadowing or mobility) a kind of “in-cell” hand-off should be executed and communication through the base station should be established.

Let’s assume that the VPI field in the transmitted packets identify the path between the wireless terminal and the base station (i.e. specifies the terminal) and that the VCI field specifies the virtual connection used by an application. Then, during call set-up the communicating terminals could learn each other’s VPI and VCI values. Moreover let’s assume that a value 0 in the GFC field specifies the link between a terminal and the base station while a value 1 defines the direct path between two terminals. In that case the transmitting terminal uses a value of 1 in the GFC and the VPI/VCI values associated with the receiving node.
However, the above described scenario is more suitable for portable but not highly mobile terminals. More problems occur in the case of packet multi-casting since some of the receivers would acquire the multicasted packets without overlaying and some others would not. Then there is problem scheduling the multicast transmissions.

4.3.6 Re-negotiation of allocated bandwidth

In order to multiplex various kinds of traffic sources in the ATM WLAN and preserve a high utilization on the air interface, it is appropriate to introduce the concept of in-call bandwidth negotiation. During call set-up, a minimum and maximum bandwidth for each service need to be defined based on the QoS characteristics of that service. If the minimum bandwidth cannot be guaranteed, the call is dropped or blocked otherwise it is accepted and allocated the required minimum. The excess bandwidth can be allocated to the connection when there is available bandwidth. Moreover, as the traffic builds up, the bandwidth of some of the connections could be reduced again to the required minimum. For no delay sensitive applications the bandwidth re-negotiation could be based on the ABR traffic model. The information required at the MAC could be extracted from the ATM flow control cells defined in the ABR traffic class. The proposed MAC could support future delay sensitive applications based on ABR traffic class too.

4.3.7 Mechanism for QoS assurance.

During call set-up, it is decided if the new connection is highly asynchronous or it can be characterised by states associated with constant bit rates. Bursts and source rates are indicated by wireless terminals, and assigned priorities according to the conformance of generated traffic to the contract that was negotiated during call setup. Moreover, the actual rate associated with a state of a source can be decomposed in a set of rates, each one assigned with a different priority.

For highly asynchronous connections, every burst is assigned with an access priority based on the length of the burst, the time of the last occurrence of a burst in that connection, and the length of the previous burst. Moreover, different parts of the burst could be assigned with different access priorities.

For connections defined by states transmitting at different constant rates, a new rate is reported using a sub-slot or a specific field in the header of the transmitted packets. This rate is interpreted in a set of pairs each pair having two components: time interval for the indication of next packet and access priority associated with that indication. So, the reported rate is decomposed in a set of rates, each one with different access priority.
Four levels of priority are defined in the MAC layer of the ATM WLAN:

1. Priority 0, being the highest, corresponds to guaranteed bandwidth of connections and is given to packets which will be definitely transmitted at the guaranteed rate.

2. Priority 1, assigned to packets carrying urgent control messages.

3. Priority 2, given to packets which do not violate the traffic contract but still correspond to bit rate higher than the guaranteed bandwidth of the connection. The combination of priorities 0 and 1 will force VBR services to approach the guaranteed bit rate under high traffic load in the network, while under light traffic conditions, it will provide with low access delays.

4. Priority 3, given to packets violating traffic contracts. When the network is overloaded, priority 3 packets are dropped and the network is not driven in a congestion region. However, if the network can support the excess rate it could transmit priority 3 packets.

The priority assigned to the packets corresponding to different rates could be given following two ways:

- Statically during call-setup when a set of relations of the type \{rate - priority\} is provided. Then, the behaviour of the source is not taken into account during the call. For example, always priority 0 is assigned to a rate up to 10 Mbps, priority 2 is assigned to an excess rate (from 10 Mbps) up to 15 Mbps, and priority 3 is assigned to all the excess rates above 15 Mbps.

- Dynamically based on an agreed traffic contract and on the conformance of the source to that contract. Then the base station has to monitor the traffic behaviour of every source and to relate the current state to the past history of the traffic. For example, it can be agreed that priority 0 is assigned always to rates up to 10 Mbps but, to assign priority 2 to excess rates of 15 Mbps the source should not transmit at this rate within a predefmed period.

The assignment of access priorities to bursts or parts of bursts, as well as to rates describing the actual states of the sources is implemented as follows: For every asynchronous source there is a timer to indicate the time instant when a number of tokens equal to the length of a conforming burst, will be generated. Whenever a new set of tokens is generated, tokens that were possibly left during the previous cycle are now being dropped. For connections with constantly fluctuating rate, state report instants are defined. At these instants a source has to report the current bit rate. At the base station, the reported rate $R$ is

---

8 The proposed mechanism could be regarded as a modification to the leaky bucket scheme.
used to define the instants when a timer will indicate packet generation at the source. If \( R_G \) is the priority 0 rate and \( R_H \) is the priority 2 rate then and \( R = R_H + R_X \) where \( R_X \) is the priority 3 rate. The values \( R_G \) and \( R_H \) are used to create token with access priorities 0 and 2, according to the following algorithm.

report of new state for connection occurs;
<calculate \( R_G \) and \( R_H \) values>;
<define activation time for timer 1 based on \( R_G \)>
<define activation time for timer 2 based on \( R_H \)>

\( \text{timer 1 active} : \text{diluting\_tokens} ++; \)

\( \text{timer 2 active} ; \)

\( \text{if } (\text{diluting\_tokens} > 0) \text{ then} \)
   \( \text{number\_of\_Priority0\_tokens} ++; \)
   \( \text{diluting\_tokens} --; \)

\( \text{else} \)
   \( \text{number\_of\_Priority2\_tokens} ++; \)

According to the above algorithm, priority 2 tokens are generated based on \( R_H \) and some of them are characterised as priority 0 tokens based on rate \( R_G \).

The following algorithm is shown how access priorities are assigned to packets of a virtual connection. As it was mentioned, the generation of packets is estimated based on the rate \( R \) and the instant of the state report.

new packet occurs;

\( \text{if } (\text{number\_of\_Priority0\_tokens} > 0) \text{ then} \)
   \( \text{access\_priority\_for\_new\_packet} = 0; \)
   \( \text{number\_of\_Priority0\_tokens} --; \)

\( \text{else if } (\text{number\_of\_Priority2\_tokens} > 0) \text{ then} \)
   \( \text{access\_priority\_for\_new\_packet} = 2; \)
   \( \text{number\_of\_Priority2\_tokens} --; \)

\( \text{else} \)
   \( \text{access\_priority\_for\_new\_packet} = 3; \)
Chapter 4: Multiple Access Scheme with Traffic Policing for ATM WLAN

The output of the mechanism for QoS assurance is a sorted list of active virtual connections to which time slots should be allocated in a FCFS policy. The input of the mechanism consists of the fast bandwidth requests contained in the above mentioned subslots, the knowledge of the activity of some sources (indicated by appropriately set timers), the property of connections to have guaranteed bandwidth or not, and the traffic contracts negotiated during call setup. Bursts with equal priorities are sorted based on the access delay they have experienced. This access delay is estimated by the base station based on the instance when the indication of the burst occurred.

The proposed mechanism exploits the traffic behaviour of the connections and is able to support future applications with different traffic characteristics. For connections that consist of CBR periods, the mechanism enables the declaration of such a period and the allocation of the requested bandwidth if this is available or guaranteed. For highly asynchronous services, the mechanism enables the indication of a burst waiting to be transmitted.

The total guaranteed bandwidth cannot be higher than the total bandwidth of the network, to cover the case were the whole guaranteed bandwidth is requested simultaneously by the connections. Moreover, most of the times, a portion of the total bandwidth should not be guaranteed to connections. This portion of network capacity could be used to absorb excessive transmission rates of the sources. It could be characterised as guard bandwidth. The size of this bandwidth could be calculated based on the variance that is exhibited by the traffic sources.

4.3.8 Air Interface to contain the QoS assurance mechanism.

The air interface consists of a sequence of frames. Every frame contains traffic slots where the base station and the wireless terminals transmit ATM-cell like packets, as well as sub-slots accessed by the wireless terminals. Wireless terminals use the sub-slots to indicate the existence of a burst, or the bit rate of a source expressed in packets per frame. The sub-slots are accessed following the slotted ALOHA method but periodically some of them are assigned to specific connections for polling. For this reason, the base station indicates which sub-slots are used for polling and which ones can be used for contention, by broadcasting a packet with the relevant information.

The sub-slots accessed on a slotted ALOHA basis are followed by a packet transmitted by the base station to acknowledge successful transmissions in the sub-slots and indicate collisions that possibly occurred.

Traffic slots multiplex traffic of both uplink and downlink. Within every frame,
every 15 time slots the base station broadcasts a packet specifying which connections will transmit packets in the following 14 time slots. The rest of the time slots are accessed by the base station or the wireless terminals to transmit packets.

![Diagram showing the air interface of ATM WLAN](image)

**Fig. 4.2** Air interface of ATM WLAN

It is assumed that all the signalling messages and the user information are carried in the transmitted packets. The sub-slots carry only information related to the MAC layer.

### 4.3.8.1 Description of the ATM WLAN model

To evaluate the performance of the above mentioned TDMA based multiple access scheme, a BONeS simulation model was used. The scenario initially considered, consists of one base station and a number of multimedia terminals and simulates the operations taking place at the air interface of Fig 4.2.

Every terminal contains three client applications and three server applications. A client can request the transmission of a stream of specific media (audio, video, file transfer) by another terminal. A server accepts requests by clients and initiates the transmission of such a stream associated with a specific type of media. Totally, a wireless terminal can transmit over up to three connections simultaneously. The modelling of the VBR video source is based on [MAGL’88]. Moreover, a CBR video generator can be used in our model. Voice source is modelled according to [BRADY’69]. To model the file transfer, a finite state machine was selected with two states. During the first state, whose duration is given by an exponential distribution with mean 14 msecs, no information is generated to emulate the
movement of a disk head between subsequent sectors of a file. The second state emulates the transmission of a sector 512 bytes long at a rate of 40 Mbps.

The base station contains the functions related to the:

- establishing and releasing of calls and connections
- slot allocation to virtual connections and bandwidth preservation/guarantee.
- registration of terminals in the base station coverage area.

It is assumed that the transmitted packets have the format presented in Fig. 4.3(a). This format corresponds to the currently identified requirements of the multiple access method. It should be modified, to cover all the introduced aspects in the ATM WLAN and reflect realistic requirements imposed by the physical layer. SYNC is used to estimate the boundaries of the transmitted packet and possibly to acquire knowledge of the radio channel. GFC is used as follows:

- bit indicates the logical link. GFC 0 = 0 identifies a transmission between the base station and one or more terminals while GFC 0 = 1 identifies a transmission from one terminal to another.
- bit indicates transition to silence. GFC 1 = 1 means that the source of the corresponding virtual connection enters an idle state
- bit indicates request of an ATM control channel. GFC 2 = 1 means that the virtual connection or the corresponding terminal have to exchange signalling information to handle exceptional case such as hand-off or disconnect because of an error.
Chapter 4: Multiple Access Scheme with Traffic Policing for ATM WLAN

- bit indicates transition in the behaviour of the source corresponding to the connection and the request of a subslot where a state report could be delivered to the base station. This could happen when the transmission rate of the source changes or when a new burst is generated by the source before the previous one was transmitted by the terminal.

Compared to the ATM cell the transmitted packet in the ATM WLAN devotes one byte less for VCI field, because it is expected that a wireless terminal will not have more than 256 connections simultaneously established. The free byte is used for the SYNC field.

The header of the information packets do not contain any field related to the current rate of the information source because it is expected that source state transitions and operations related to such transitions will be rare compared to the packet transmission rate (i.e. between two transitions a lot of packets are transmitted). As a result a lot of packets will carry no information in such a field and this will increase the wasted bandwidth.

The sub-slots have the format presented in Fig. 4.3(a). The presented format implies that only registered terminals and established connections access the sub-slots, because VPI values are assigned to terminals whose registrations are accepted and VCI values are assigned during the establishment of connections. However the same sub-slots could be used by terminals which have not joined the network and wish to indicate their presence and connect to a base station.

4.3.8.2 Simulation Results

The capability of the proposed algorithm to apply traffic policing on connections, is proven by simulation of a network with Variable Bit Rate (VBR) video sources, modelled based on [MAGL'88]. The following approach was followed: Initially the capacity of a coverage area (in terms of number of sources) was estimated by calculating the packet dropping probability when all the sources transmit at the same average bit rate of 2 Mbps (Fig. 4.4) and conform to that average. Assuming 2% as an acceptable packet dropping probability, the capacity is 8 VBR sources (16 Mbps on average). Then, the packet dropping probability is calculated when 5 sources conform to the average rate (2 Mbps VBR) while the rest of them transmit at a higher average rate (2.5 to 4 Mbps) to simulate non-conforming sources (Fig. 4.5). From Fig. 4.5, it is obvious that the packet dropping probability of the conforming sources remains unaffected by the presence of the non-conforming connections. Moreover, the latter experience high packet dropping rate because the network prevents the allocation of radio capacity to them, in order to preserve the bandwidth and the QoS of the conforming connections. However, when no traffic policing applies conforming connections experience increased packet dropping which reduces the user capacity of the system.
Chapter 4: Multiple Access Scheme with Traffic Policing for ATM WLAN

4.4 D-TDMA performance over the fading channel

The performance metrics presented in the previous section and in chapter 3 did not include the effect of BER or PER on the transmissions. However, the transmissions in the air-interface of the ATM WLAN are not error free. Then, it is appropriate to apply a packet retransmission scheme to deal with packet errors. However, re-transmissions will decrease...
the available bandwidth on the upper layers and the QoS of the supported services will be affected. System capacity could be affected too.

In Fig. 4.6 the optimised air interface for the ATM WLAN is presented. It is based on the air-interface presented in section 4.3. In this extended version, there is provision for re-transmissions of packets with errors and for channel monitoring. The latter is used to initiate hand-offs as described in chapter 6 and possibly to adopt the coding rate at the physical layer of the ATM WLAN.

The main difference compared to the air-interface presented in section 4.3, is the existence of sub-slots following the traffic slots between two broadcast packets by the Base Station. These sub-slots are called ACK sub-slots to differentiate from the sub-slots in the beginning of every frame. ACK sub-slots carry the acknowledgements from MTs to BS and vice versa. The acknowledgement mechanism follows a reply on success policy: a transmission in an ACK sub-slot means correct reception of a packet, while silence means reception of a packet with errors or packet loss. There is a one-to-one mapping between the traffic slots and the ACK sub-slots to avoid scheduling of ACK sub-slot transmissions.

![Optimised Air interface for ATM WLAN](image)

**Fig. 4.6 Optimised Air interface for ATM WLAN**

to cope with packet re-transmissions

The mechanism with the ACK sub-slots and re-transmissions is introduced to compensate for variations that could happen within one frame duration. In that case, additional bandwidth should be provided by the network to support re-transmissions. However, it is expected that this bandwidth will represent only a small portion of the total
rate characterising a connection. According to [BASELINE] 10% of the bandwidth is required for re-transmissions. To provide the additional bandwidth for the ACK sub-slot transmissions the transmission rate in the Physical layer has to be increased accordingly.

Moreover, the packet re-transmissions that may occur should not affect the traffic policing functions that are employed in the ATM layer and in the MAC layer of the ATM WLAN. This could happen as the packet delay variation that occurs in a stream of successfully transmitted and re-transmitted packets could violate the traffic contract of a connection. However, the mechanism presented in section 4.3, takes into account the lifetime of the packets within a connection. This lifetime could not only represent the maximum access delay that packets can experience, but also the maximum time they have to wait for, without violating the traffic contract. One solution is to re-transmit as many packets as possible within the frame they are scheduled to be transmitted and drop the rest. However, for UBR traffic class re-transmissions of corrupted packets could be scheduled for next frames. Whatever the solution is, additional bandwidth is required to carry the re-transmissions and the BS has to take into account this bandwidth requirement when it schedules the terminal transmissions.

Based on the above mentioned considerations, the simulation model of the ATM WLAN air interface was extended to introduce errors in the transmissions and support packet and sub-slot re-transmissions. The effects on the packet dropping probability and on the experienced packet access delay are of interest. Fig. 4.7 shows the performance of the system when it is loaded to its maximum capacity and for different BERs. A code with 10 error detection capability within a transmitted packet is assumed. The effect on the packet dropping probability becomes significant for BERs worse than $10^{-4}$. 

---

Fig 4.7 Performance of the ATM WLAN MAC scheme versus introduced BER
As it is presented in Fig. 4.7, the total Packet Loss Rate (PLR) on the air interface of ATM WLAN is affected by the packets that must be retransmitted and do not find the required bandwidth for retransmission. Moreover, the higher mean access delay experienced by packets for BER equal to $10^{-4}$ is explained by the fact that more packets will go through successfully in the retransmission period compared to the case of BER $10^{-3}$, contributing to the slight increase of the mean access delay. Furthermore, Fig. 4.8 shows the effect on BER on the channel throughput considering slot utilization.

4.5 Conclusion

In this chapter, it was shown how the concept of traffic management defined for fixed ATM networks is introduced in ATM WLAN. A traffic policing mechanism was introduced in the MAC layer of the ATM WLAN for assigning bandwidth to individual virtual connections, based on their conformance to their traffic contracts. The proposed mechanism has the ability to preserve the bandwidth of a conforming connection, and to assure the QoS of the corresponding application. An air interface structure suitable for the above mentioned mechanism was introduced so that the bandwidth allocation follows the traffic variations in every virtual connection, and the channel capacity is assigned only to active connections so that the system efficiency is maximised.

![Traffic Slot Utilisation for the ATM WLAN](Fig 4.8 Traffic Slot Utilisation for the ATM WLAN MAC Scheme versus BER)
Chapter 5

Support of TCP over ATM WLAN

In this chapter it is described how the Transport Control Protocol (TCP) performs over ATM WLAN. TCP deploys algorithms to avoid and react to congestion in networks. However, it does not provide with solutions to prevent congestion before it occurs in the ATM layer of ATM networks. Then, ATM buffer overflows occur that result to TCP throughput degradation, and TCP aware mechanisms have to be defined to overcome the problem of low throughput at the TCP level. It is shown that the same algorithms can be applied in ATM WLAN as well, to improve the throughput of the TCP layer. Moreover when TCP resides on top of wireless protocols, packet losses are interpreted as congestion. Then the congestion avoidance mechanisms of TCP are activated resulting in a degraded throughput. It is shown that by adopting the cell re-transmission policy presented in chapter 4, a high TCP throughput can be achieved.

5.1 Introduction

Before native ATM applications become widespread, ATM systems must support Internet based applications. This happens, mainly, by supporting TCP/IP over ATM Adaptation Layer 5 (AAL5) in the Protocol Reference Model (PRM) of ATM networks. However, several problems arise when TCP runs over ATM or networks with wireless links. In a real system TCP offers compatibility between terminals residing in different networks, since it is the end-to-end transport protocol dominating in Internet. Since TCP deals mainly with congestion problems it does not performs efficiently when packets are dropped because of channel errors in wireless networks. Moreover, when used over ATM, TCP performance
deteriorates if cells are dropped due to buffer overflows in ATM switches, unless Early Packet Discard (EPD) [ROMA'95] is adopted.

The above mentioned problems must be taken into account within the multiple access mechanism of ATM WLAN in order to improve the performance of TCP. Like any other ATM oriented network, ATM WLAN aims to support any service that can be offered through Internet, so it has to support TCP in an optimum way.

5.2 TCP Principles

Transport Control Protocol (TCP) [STEV'94] provides for reliable stream transport. It is the second (after IP) most important protocol in the TCP/IP family. It provides:

- a reliable stream of bytes,
- guarantees that all bytes are delivered uncorrupted and that they are delivered in sequence,
- virtual circuits for connection oriented communication

The byte stream is buffered and packetised typically in 4 Kbyte segments for sending in IP datagrams. A "Push" operation can be used to force immediate delivery if needed. Full duplex (two-way simultaneous communication) is supported. Reliability and in-sequence delivery is obtained with the sliding window principle. Piggybacking is used. TCP allows several application programs to have active connections simultaneously (multiplexing).

The sliding window protocol of TCP solves the flow control problem by letting the acknowledgements (ACKs) to contain information that lets the sender modify the size of its sending window. The sequence numbers are based on octet numbers (byte numbers) instead of frame numbers i.e. the sequence numbers are octet numbers in the total stream, not segment numbers. The acknowledgements are cumulative: all octets up to the acknowledged octet have arrived correctly. Each ACK contains the amount of available buffer space in the receiver (receiver window size). If this size is 0 the sender stops completely. Note that this mechanism only handles flow control, i.e. the overload problem at the receiver. Other mechanisms are needed to handle congestion control, i.e. avoiding that intermediate routers get congested.

5.2.1 Measurement of Round Trip Time

Fundamental to TCP's timeout and retransmission is the measurement of the round-trip time (RTT). As the RTT changes, TCP should modify its timeouts. Moreover, it is important not only to track the RTT but also estimate its variance.
It was [STEV'94] found that old versions of TCP could not keep up with wide fluctuations in RTT, which leads to more retransmissions when the network is already loaded. So the tracking of the RTT variance was proposed, and formulas that compute the timeout interval based on the mean and variance of RTT were given.

The timeout time is set adaptively: The current RTT is estimated with exponential averaging:

\[ \text{RTT} = a \times \text{Old_RTT} + (1-a) \times \text{New_rtt_sample} \]

Alpha has a recommended value of 0.9. Then the timeout interval is calculated by the following algorithm which also estimates the variance of the RTT:

\[
\begin{align*}
\text{DIFF} &= \text{New_rtt_sample} - \text{Old_RTT} \\
\text{RTT} &= \text{Old_RTT} + d \times \text{DIFF} \\
\text{DEV} &= \text{Old_DEV} + d \times (\text{abs(DIFF)} - \text{Old_DEV})
\end{align*}
\]

d is usually an inverted power of two, e.g., 0.125. DEV is the estimated standard deviation of the RTT, and the timeout is set to:

\[ \text{Timeout} = \text{DEV} \times \text{DEV} \times \text{RTT} \]

since the variance is the square of the standard deviation.

5.2.2 Karn's Algorithm

However, there is still one problem when getting an RTT sample: When retransmissions have occurred it is not possible to determine if the acknowledgement comes from the first or second try. If the first is always assumed, the estimate will be too large, which decreases efficiency since the sender waits too long. If the second is always assumed, the estimate will become too short and unnecessary retransmissions will occur. The solution is Karn's algorithm. When retransmissions occur, the RTT estimate is not updated. Moreover, when timeouts occur "timer backoff" is done:

\[ \text{New_timeout} = g \times \text{Timeout}, \ g \text{ is typically 2} \]

5.2.3 Delayed ACKs

The goal is to send less than one ACK per received segment. Rather than sending an ACK immediately, TCP waits ~200ms hoping that there will be data in the reverse direction thus enabling a piggybacked ACK.
5.3 TCP Congestion Control Scheme

TCP congestion control scheme consists of the "Slow start" and "Congestion Avoidance" phases. During Slow Start whenever the sender receives an acknowledgement for a transmitted packet, it doubles its window. When the size of the sender window exceeds a certain threshold, the sender enters the Congestion Avoidance phase and increases its window by the inverse of its current size. The sender follows the timeout policy to retransmit unacknowledged packets. A timeout is an indication for congestion in the network and the sender enters again the Slow Start phase. Then the link could stay idle because of a false assumption for congestion.

5.3.1 Slow Start

Old TCP versions would start a connection with the sender injecting multiple segments into the network, up to the window size advertised by the receiver. Then, problems can arise if there are routers and slower links between the sender and the receiver. Some intermediate router must queue the packets, and it's possible for that router to run out of space. This approach can reduce the throughput of a TCP connection drastically.

The algorithm to avoid this is called slow start. It operates by observing that the rate at which new packets should be injected into the network is the rate at which the acknowledgements are returned by the other end. Slow start adds another window to the sender's TCP: the congestion window, called "cwnd". When a new connection is established with a host on another network, the congestion window is initialised to one segment (i.e., the segment size announced by the other end, or the default, typically 512 bytes). Each time an ACK is received, the congestion window is increased by one segment. The sender can transmit up to the minimum of the congestion window and the advertised window. The sender starts by transmitting one segment and waiting for its ACK. When that ACK is received, the congestion window is incremented from one to two, and two segments can be sent. When each of those two segments is acknowledged, the congestion window is increased to four. At some point the capacity of the internet can be reached, and an intermediate router will start discarding packets. This tells the sender that its congestion window has got too large.

5.3.2 Congestion Avoidance

Congestion can occur when data arrive through a big pipe (e.g. a fast LAN) and are sent out through a smaller pipe (e.g. a slower WAN). Congestion can also occur when multiple input streams arrive at a router whose output capacity is less than the sum of the inputs.
Congestion avoidance is a way to deal with lost packets. The assumption of the algorithm is that packet loss caused by damage is very small (much less than 1%), therefore the loss of a packet signals congestion somewhere in the network between the source and destination. There are two indications of packet loss: a timeout occurring and the receipt of duplicate ACKs. When congestion occurs TCP must slow down its transmission rate of packets into the network, and then invoke slow start. Slow start continues until TCP is halfway to where it was when congestion occurred, and then congestion avoidance takes over.

Congestion avoidance and slow start require that two variables be maintained for each connection: a congestion window, `cwnd', and a slow start threshold size, `ssthresh'. In Slow start `cwnd' begins at one segment, and is incremented by one segment every time an ACK is received. Congestion avoidance dictates that `cwnd' be incremented by (segsize*segsize/cwnd) each time an ACK is received, where `segsize' is the segment size and `cwnd' is maintained in bytes. This is a linear growth of `cwnd', compared to slow start's exponential growth.

5.3.3 Fast Retransmit

TCP may generate an immediate acknowledgement (a duplicate ACK) when an out-of-order segment is received. This duplicate ACK should not be delayed. The purpose of this duplicate ACK is to let the other end know that a segment was received out of order, and to tell it what sequence number is expected. Since TCP does not know whether a duplicate ACK is caused by a lost segment or just a reordering of segments, it waits for a small number of duplicate ACKs to be received. It is assumed that if there is just a reordering of the segments, there will be only one or two duplicate ACKs before the reordered segment is processed, which will then generate a new ACK. If three or more duplicate ACKs are received in a row, it is a strong indication that a segment has been lost. TCP then performs a retransmission of what appears to be the missing segment, without waiting for a retransmission timer to expire.

5.3.4 Fast Recovery

After fast retransmit sends what appears to be the missing segment, congestion avoidance, but not slow start is performed. This is the fast recovery algorithm. It is an improvement that allows high throughput under moderate congestion, especially for large windows. The reason for not performing slow start in this case is that the receipt of the duplicate ACKs tells TCP that more than just a packet has been lost. Since the receiver can only generate the duplicate ACK when another segment is received, that segment has left the network and is in the receiver's buffer. That is, there is still data flowing between the two ends, and TCP does
not want to reduce the flow abruptly by going into slow start. The fast retransmit and fast recovery algorithms are usually implemented together.

5.4 Performance issues of TCP over ATM

Several publications deal with the problems arising when running TCP over fixed ATM networks [LAKS'97], [ROMA'95], [GOYAL'97]. TCP segments are segmented in ATM cells by AAL5. When an ATM cell is lost because of buffer overflows in the ATM layer, a whole TCP segment has to be retransmitted in the TCP layer. This increases the traffic load of the ATM layer and decreases the throughput at the TCP layer simultaneously. The published work reports mainly experiment results of TCP over ATM.

In [ROMA'95], the absence of congestion control mechanisms defined in ATM is assumed. ATM cell loss because of congestion (buffer overflows) causes throughput degradation, because TCP segments with lost ATM cells have to be retransmitted. More bandwidth is wasted because of the transmission of their non dropped (useless) ATM cells. Partial Packet Discard (PPD) algorithm is proposed as a solution but found non optimal. It drops the rest of the higher layer packet when one of its cells is dropped. However, The Early Packet Discard (EPD) algorithm provides optimal throughput. As soon as a buffer occupancy exceeds a certain threshold, whole upper layer packets are discarded. However, this results to unfair sharing of bandwidth among UBR connections. Moreover, [ROMA’95] investigates experimentally the performance of EPD and PPD with respect to the size of the switch buffers, TCP windows, TCP segments. To improve the TCP throughput the following are to be avoided: smaller switch buffers, large TCP windows, large TCP packet sizes.

In [GOYAL'97], experiment results of TCP traffic over ATM connections are discussed. To achieve the maximum throughput, no TCP segment loss must be achieved. Thus, the buffer size of the ATM switches must be equal to the sum of the receiver window sizes of all the TCP connections. If no specific algorithms are deployed, there is problem of unfairness experienced by TCP connections over UBR traffic class. He claims that EPD alleviates the problem of poor TCP throughput but does not improve fairness. Selective packet drop per virtual circuit must be deployed to improve fairness.

In [LAKS'97] after the analysis of two versions of TCP, namely Tahoe and Reno, the following are concluded:
1. When the TCP sender enters the slow start phase of the TCP mechanism, and the size of the buffer is smaller than 1/3 of the bandwidth-delay product, buffer overflows occur leading to a second slow start phase and further throughput decrease.
2. TCP is highly vulnerable to random losses and inadequate to multiplex different kinds of services in the same connection. Fortunately ATM networks provide their own mechanisms for multiplexing different kinds of services. Then, the fluctuations of the bandwidth available for TCP connections could lead to random losses and deterioration of network utilization. However if there is adequate buffering, TCP could provide with satisfactory throughput figures.

3. Connections with higher propagation delays experience unfairness compared to those with lower propagation delays. TCP connections should have reserved bandwidth throughout the network. In that case ABR traffic class should be used to support TCP connections in an ATM network.

4. Loss because of congestion is the only way to provide feedback in TCP so congestion cannot be predicted.

5. TCP Behaviour over Wireless Links

In [BALA'96] different mechanisms deployed to support and improve TCP performance over wireless links are compared. The main problem is that TCP deals only with congestion and does not take into account packet loss because of increased BER. Dealing with the latter by deploying congestion avoidance algorithms results to highly degraded throughput.

Generally, there are two approaches to improve TCP in lossy environments. The first approach hides any non-congestion related losses from the TCP sender. It requires no changes to existing sender implementations. The intuition behind this approach is that since the problem is local, it should be solved locally, and that the transport layer need not be aware of the characteristics of the individual links. Protocols that adopt this approach attempt to make the lossy link appear as a higher quality link with a reduced effective bandwidth. As a result, most of the losses seen by the TCP sender are caused by congestion. The second class of techniques attempts to make the sender aware of the existence of wireless hops and realise that some packet losses are not due to congestion. The sender can then avoid invoking congestion control algorithms when non-congestion related losses occur. Finally, it is possible for a wireless-aware transport protocol to coexist with link-layer schemes to achieve good performance.

Moreover, the existing schemes are classified into three basic groups, based on their fundamental philosophy: end-to-end proposals, split-connection proposals and link-layer proposals. The end-to-end protocols attempt to make the TCP sender handle losses through the use of two techniques. First, they use some form of selective acknowledgements (SACKs) to allow the sender to recover from multiple packet losses in a window with-out resorting to a
coarse timeout. Second, they attempt to have the sender distinguishing between congestion and other forms of losses using an Explicit Loss Notification (ELN) mechanism.

Furthermore, split-connection approaches completely hide the wireless link from the sender by terminating the TCP connection at the base station. Such schemes use a separate reliable connection between the base station and the destination host. The second connection can use techniques such as negative or selective acknowledgements, rather than just standard TCP, to perform well over the wireless link.

The third class of protocols, link-layer solutions, lie between the other two classes. These protocols attempt to hide link-related losses from the TCP sender by using local retransmissions and perhaps forward error correction over the wireless link. The local retransmissions use techniques that are tuned to the characteristics of the wireless link to provide a significant increase in performance. The main advantage of employing a link layer protocol for loss recovery is that it fits naturally into the layered structure of network protocols. The link-layer protocol operates independently of higher layer protocols and does not maintain any per connection state. The main concern about link layer protocols is the possibility of adverse effect on certain transport layer protocols such as TCP.

5.5.1 *Indirect TCP*

This was one of the early protocols to use the split-connection approach. It involves splitting each TCP connection between a sender and receiver into two separate connections at the base station: one TCP connection between the sender and the base station, and the other between the base station and the receiver.

However, the choice of TCP over the wireless link results in several performance problems. Since TCP is not well-tuned for the lossy link, the TCP sender of the wireless connection often times out, causing the original sender to stall. In addition, every packet has to go through TCP protocol processing twice at the base station (as compared to zero times for a non-split connection approach). Another disadvantage of this approach is that the end-to-end semantics of TCP acknowledgements is violated, since acknowledgements to packets can now reach the source even before the packets actually reach the mobile host. Also, since this protocol maintains a significant amount of state at the base station per TCP connection, handoff procedures tend to be complicated and slow.

5.5.2 *The Snoop Protocol*

The snoop protocol introduces a module, called the snoop agent, at the base station. The agent monitors every packet that passes through the TCP connection in both directions and
maintains a cache of TCP segments sent across the link that have not yet been acknowledged by the receiver. A packet loss is detected by the arrival of a small number of duplicate acknowledgements from the receiver or by a local timeout. The snoop agent retransmits the lost packet if it has it cached and suppresses the duplicate acknowledgements. Considering the above mentioned classification of protocols, the snoop protocol is a link layer protocol that takes advantage of the knowledge of the higher layer transport protocol (TCP). The main advantage of this approach is that it suppresses duplicate acknowledgements for TCP segments lost and retransmitted locally, thereby avoiding unnecessary fast re-transmissions and congestion control invocations by the sender. The per-connection state maintained by the snoop agent at the base station is soft, and is not essential for correctness. Like other link layer solutions, the snoop approach could also suffer from not being able to completely shield the sender from wireless losses.

5.5.3 Selective Acknowledgements

Since standard TCP uses a cumulative acknowledgement scheme, it often does not provide the sender with sufficient information to recover quickly from multiple packet losses within a single transmission window. However, TCP enhanced with selective acknowledgements performs better than standard TCP in such situations. A reliable link-layer protocol that uses knowledge of TCP (LL-TCP-AWARE) to shield the sender from duplicate acknowledgements arising from wireless losses gives a 10-30% higher throughput than one (LL) that operates independently of TCP and does not attempt in-order delivery of packets. Also, the former avoids redundant retransmissions by both the sender and the base station, resulting in a higher throughput. The selective acknowledgement scheme is quite effective in dealing with a high packet loss rate when employed over the wireless hop or by a sender in a LAN environment and especially when packet losses occur in bursts.

5.6 Simulation of TCP over ATM WLAN

Fig. 5.1 presents the network scenario modelled and the Protocol Reference Model that resides in the terminals. Two ATM WLAN segments are defined. Terminals in one segment transmit to terminals to the other segment and vice versa. Thus all the ATM traffic has to go through the two access points and the presented backbone ATM switch.

The operations that take place in the layers of the Protocol Reference Model presented in Fig. 5.1 are as follows.
5.6.1 Functions supported in the Simulation Model

5.6.1.1 Application layer

Different traffic generators can be considered in the application layer, such as video traffic source based on [MAGL'88], Poisson traffic generator and constant bit rate traffic source. For the simulation a Poisson traffic generator was considered with an average inter-arrival time of 40 ms and a fixed packet length of 4095 bits. Thus, the average bit rate of the traffic source is around 100 Kbps. Furthermore, the transmission rate in the air interface is fixed at the 1Mbps. Although this rate does not correspond to the adopted transmission rate of ATM WLAN (20 - 25 Mbps) it allows faster simulation execution without violating the qualitative study of TCP performance over ATM WLAN.

5.6.1.2 TCP layer

TCP layer, implements the Slow Start Algorithm and Congestion Avoidance Algorithm as presented in section 5.3, as well as Fast Retransmit and Fast Recovery. When a duplicate acknowledgement is received, the implied TCP segment is retransmitted out of order by the TCP transmitter. This happens since all the ATM cells, and consequently the TCP segments of the same connection, will follow the same route. Moreover the timeout interval for every transmitted packet is calculated dynamically according to the algorithms presented in sections 5.2.1 and 5.2.2. The receiver delivers to the Application layer the TCP segments received in order and stores those that are received out of order. Moreover, the implemented model supports variable maximum sliding window size, maximum TCP segment length and packet timeout period.

5.6.1.3 ATM Adaptation Layer

AAL5 is modelled for this part. AAL performs segmentation of TCP segments at the transmitting side and re-assembly at the receiving side. Moreover it checks for dropped cells
within an AAL frame. At the receiving side, AAL delivers to the upper layer (TCP) AAL frames without lost cells. AAL does not execute error detection or correction.

5.6.1.4 ATM

In the simulation model this layer has significance at the ATM switches. It performs switching of cells towards their final destination. EPD is implemented in the ATM layer of the switches. Moreover, the ATM connections supporting TCP belong to the UBR traffic class.

5.6.1.5 WATM MAC

This layer performs buffering of the ATM cells and traffic scheduling at the air-interface of the ATM WLAN. No bandwidth guarantees apply for the TCP connections multiplexed in the air-interface of ATM WLAN. Moreover, a Cell Loss Rate (CLR) is modelled to simulate cell losses because of corrupted transmissions. In this layer ATM cell retransmission can be selected at the beginning of the simulation. EPD algorithm is deployed in the WATM MAC and is expected to increase the TCP throughput when cells are dropped because of buffer overflows.

5.6.1.6 Tx/Rx Radio

This layer is not modelled, only a user defined CLR can be introduced.

5.6.2 Simulation Results

To examine the performance of TCP over ATM WLAN, a simulation program that corresponds to 20 secs was executed. Acknowledgements are transmitted with highest priority and are not affected by the channel errors. Thus no ACKs are lost and their propagation delay is eligible. An infinite maximum tolerable access delay for ATM cells was assumed to make the buffers of access points overflow.

Two cases were considered. The first one examines the TCP behaviour as the traffic load in the network increases while the second investigates the TCP performance when transmission errors are introduced (in terms of Cell Loss Rate-CLR).

5.6.2.1 TCP vs. Traffic Load

In the first case no CLR is assumed and the behaviour of the TCP connections is examined for different traffic loads (expressed as number of enabled transmitters) and when EPD is active or not. Figs. 5.2 to 5.5 correspond to this case.

Two main observations occur from this case. The first is that when EPD comes into effect, there is a certain improvement on TCP performance in terms of throughput and end-to-end delays. This is shown in Fig. 5.2 where the generated, transmitted and received TCP
segments are presented, and in Fig. 5.3 where the TCP throughput is drawn, averaged over the period of 20 secs. This performance improvement is experienced because cells of corrupted TCP segments are dropped freeing bandwidth for uncorrupted TCP segments.

Fig. 5.2 TCP segment profile

The second observation is that there is a limit on the traffic that can be accepted in order to achieve maximum throughput. As shown in Fig. 5.3 the maximum throughput is achieved for 7 users while for 8, 9 and 10 there is a significant throughput decrease. The same phenomenon has been reported in [ROMA’95] where the TCP throughput decreases as the number of TCP connections increases. This shows that TCP cannot prevent congestion in the air-interface and that the traffic monitoring mechanisms proposed in chapter 4 are necessary. Fig. 5.4 draws the TCP offered traffic averaged over the simulation period. There it is shown that the total offered TCP traffic increases as the number of sources increase up to the congestion limit. This reflects also the increase of the total ATM throughput presented in Fig. 5.5.

Fig. 5.3 TCP layer throughput

Fig 5.4 TCP offered traffic

Fig 5.5 ATM layer throughput
5.6.2.2 TCP vs. Cell Loss Ratio and Retransmission

In the second case the total traffic load in the air-interface is kept constant (7 enabled traffic sources corresponding to the maximum achievable throughput) but the Cell Loss Rate (CLR) is varied (0.0, 0.001, 0.005, 0.01, 0.015, 0.02, 0.05, 0.1). Moreover, a cell retransmission scheme in the WATM MAC is adopted to examine the potential improvement of the system performance. Figs 5.6 to 5.8 represent this case. Fig. 5.6 shows the throughput degradation in the TCP level for the above mentioned CLR values, when no cell retransmissions take place. Moreover, in the same figure it is shown how re-transmissions improve the throughput performance of TCP. For low CLR and adopting re-transmissions, the throughput stays close to the maximum achievable. Fig. 5.7 shows the corresponding offered TCP traffic. For the no retransmission schemes the congestion avoidance algorithm is invoked decreasing dramatically the offered traffic in the network. Moreover, Fig. 5.8 shows the effect of TCP behaviour and retransmission policy over the ATM traffic. Since retransmission enables TCP to keep a high throughput the traffic offer in ATM layer is kept high compared to the no retransmission case.
5.7 Conclusion

To examine the performance of TCP over ATM WLAN a simulation model was built supporting the main TCP functions and modelling the air-interface of the ATM WLAN. The aim of the simulations was to show an improvement of TCP performance in terms of throughput when EPD and cell retransmission are adopted in the ATM/MAC layer of ATM WLAN. Indeed the introduction of cell retransmission brings the TCP performance close to the maximum achievable. Moreover, the adoption of EPD improves TCP performance by freeing bandwidth that would be wasted by cells which belong to corrupted TCP segments. Moreover, it was found that there is a maximum TCP traffic load that can be sustained by ATM WLAN in order to keep maximum TCP utilization. Further increase of TCP traffic results to decrease of the aggregate TCP throughput.
Chapter 6

Mobility Management in ATM WLAN based on Standardised Protocol Specifications

This chapter shows how a seamless hand-off procedure can be established over the ATM WLAN air-interface to cope with two cases. First, the hand-off to a different carrier within the same coverage area because the channel quality or the available bandwidth of the current carrier are not adequate to satisfy the QoS of established connections. Second, the hand-off between two different Base Stations (inter-cell hand-off) as a mobile crosses the borders of two coverage areas.

Moreover, it is shown how Private/Public User Network Interface (P-UNI), and Private Network Network Interface (P-NNI) protocols can be exploited, to support mobility in an ATM based network, with wireless access parts and mobile users. The proposed modifications deal with the mobility management functions. The new protocols can integrate different radio access technologies and be used in different radio environments, as they involve the upper layers of ATM Protocol Reference Model. So, they can also be supported on top of the proposed multiple access scheme for ATM WLAN. The standardised address registration mechanism at P-UNI is extended for registration of mobile users, and support of hand-off: P-NNI routing protocols are modified to enable the location update for roaming users and the in-call re-routing of virtual connections. This is achieved by keeping the PNNI topology and routing mechanisms of the fixed network infrastructure (ATM switches), and by introducing the concept of Location Management Information Bases (LMIB) and LMIB Servers for the mobility management operations. The functionality and location of LMIB Servers are defined. Finally in this chapter, extensions to call control signalling in P-NNI are proposed to support the in-call connection re-routing.
Chapter 6: Mobility Management in ATM WLAN
based on Standardised Protocol Specifications

6.1 Hand-off procedures in ATM WLAN air-interface

As presented in chapter 4, the dynamic TDMA multiple scheme is defined over one carrier. This means that a Mobile Terminal (MT) can use only one carrier to operate. However, it is desirable to deploy more than one carriers within a certain coverage area, to increase the user capacity. Then an entity per carrier must exist to schedule the transmissions based on that particular carrier. However, one entity within the coverage area could deal with the addressing aspects of terminals and the interconnection to the backbone network irrespectively of the used carrier by every terminal. The following elements should be considered in the wireless part of the ATM WLAN:

- **Access Points (APs)** which are network entities controlling the MTs’ transmissions in different carriers. For every carrier there is a corresponding AP which contains all the functions required for the multiple access of that carrier by MTs.

- **Base stations** which can be viewed as the termination points of the private ATM network, or the equipment performing the network part functions of the private UNI. A BS could have more than one access points i.e. use more than one carriers. In that case different carriers could be used to cover different areas around a BS or provide with higher user capacity over a certain area. While a MT communicates with a BS it keeps the same private ATM address even if it switches among different APs residing in the same BS (i.e. different carriers of the BS). The maximum bandwidth available to a MT is defined by the bandwidth of one carrier.

- **Mobile Terminals (MTs)** which are capable to initiate more than one call at a time and establish more than one connection simultaneously (for one call or more than one calls). It has been assumed that each MT is associated with one Virtual Path Identifier (VPI) value which is unique per Base Station (BS) and identifies the signalling channel between the BS and the specific MT.

Based on the definition of the entities in the assumed wireless ATM network and their corresponding functions two cases of hand-off exist: Intra-cell hand-off where a MT changes the carrier it uses because of channel quality or bandwidth shortage, and inter-cell hand-off because it moves to a new Base Station coverage area.

6.1.1 **Intracell hand-off**

As mentioned before, changing the carrier for communication to the same BS is a kind of intra-cell hand-off and can be initiated by the BS or the MT. It happens when the quality of reception by either BS or MT deteriorates in the carrier currently used. Moreover it could happen because the traffic demand on the current carrier by the MT to handoff or by other
MTs increases. The communication between BS and MT should be continued on a different carrier and through a new AP.

Generally speaking, the MT can monitor the quality of the downlink by:
- receiving the broadcast packets transmitted by the APs of the BS. These are the packets carrying the information related to slot and subslot allocation, BS identity and possibly neighbouring BSs identities,
- receiving the packets of the established connections in the downlink and examining the corresponding BER or Packet Error Rate (PER).

Accordingly the BS can monitor the quality of the uplink by:
- receiving the packets of the established connections in the uplink
- receiving the F5 and F4 flow OAM\(^9\) packets [UNI31] (carrying information related to connectivity and congestion in the network) transmitted periodically by idle MTs or by MTs with active calls but idle connections,

According to [UNI31], in the OAM packets there is provision for 16 different messages but only three messages are specified. So we define new messages to support the hand-off procedures (Fig. 6.1).

<table>
<thead>
<tr>
<th>ATM cell header</th>
<th>OAM Cell Type</th>
<th>Function type</th>
<th>Function specific fields</th>
<th>CRC-10</th>
</tr>
</thead>
<tbody>
<tr>
<td>VPI identifies MT</td>
<td>1111: Intracell Hand-off</td>
<td>1001: Hand-off Request</td>
<td>e.g. when Hand-off Request a list with the available carriers and the corresponding free bandwidth</td>
<td></td>
</tr>
<tr>
<td>0110: Intercell Hand-off</td>
<td>1000: Hand-off Indicate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0000: Carrier Indicate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0110: Carrier Allocate</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 6.1 Introduced F5 OAM packets to assist the hand-off procedures in the Wireless ATM Network

In the case of BS, loss of the physical connection to a MT means initialisation of an AP switch procedure. If the BS contains more than one APs, it assumes that it will listen from the MT within T1 msecs. For this reason the BS expects to get a message containing the VPI associated to the "lost" MT. To do so, for T1 msecs the BS assigns one time subslot in every one of its carriers to the lost MT to enable it to indicate its existence. If the BS has only one AP (controls only one carrier) or if it does not hear from the MT within T1, it assumes that the MT executes hand-off with another BS and that it will hear from the MT

\(^9\) Operation Administration and Maintenance
within T2 msecs. If it does not happen, the BS assumes the MT disconnected from the network and releases its connections.

If the MT looses (for some reason) the connectivity to AP and consequently to BS, it should scan all the carriers of the BS and access successfully the subslots of a BS’s carrier (allocated to it by the BS) to indicate its existence and the carrier of its preference within T1 msecs. It is assumed that there is a bilateral agreement on the carrier sequence where time slot allocation is performed by the BS and scanning is performed by the MT. If within T1 msecs MT cannot access a carrier of the current BS it should start looking for the carriers of another BS i.e. execute inter-cell hand-off. The old BS should be notified within (T2+T1) msecs after the connectivity to the MT was lost.

In the case where the BS decides that a MT should switch to another carrier and still can communicate with that MT, the BS instructs the MT to start a procedure similar to the one presented above. This notification is given in a new UNI F5 flow packet, called Intra-cell Hand-off Indicate packet from the BS to the MT and contains the carriers that can stand the traffic related to the MT. If no carrier can accommodate that traffic the BS could instruct the MT to hand-off to another BS or to terminate some of its connections while informing it about the available bandwidth in the carriers (i.e. provide the MT with a list of pairs of the type \((\text{carrier}, \text{available bandwidth})\)). Assuming that every pair occupies two octets of the transmitted packet, every Intra-cell Hand-off Indicate packet can accommodate information for 22 carriers.

Accordingly when the MT is notified by the BS to change carrier and keeps its connectivity to the current AP, it starts scanning the carriers of the current BS contained in the BS’s notification. The scanning happens in the periods when MT does not have to communicate with the current AP (and for this reason AP could suspend the connections of the MT for one frame duration). BS will allocate time slots to MT in subsequent frames of the initial carrier and in those slots the MT will indicate in ascending order of reception quality the carriers where it could switch. To do so, MT uses a new F5 OAM packet called Carrier Indicate packet which contains the MT’s carriers of preference. BS will compare the available bandwidth in every carrier to the required bandwidth for the MT’s connections and it will ask the MT, using a Carrier Allocate packet, to switch to a new carrier. If no carrier can accommodate MT’s connections some of them should be terminated or the MT should hand-over to another BS.
If the MT detects excessive degradation on the current carrier, but it still retains its communication with the current AP then it is the MT which notifies the BS for the carrier switch using the Intra-cell Hand-off Request message. The BS replies using the Intra-cell Hand-off Indicate packet with the list of its carriers and the available bandwidth per carrier and the procedure mentioned before is followed. Fig. 6.2 shows the message exchange for MT and BT initiated intra-cell hand-off.

6.1.2 Inter-cell hand-off

Considering inter-cell hand-off, the BS or the MT could initiate the appropriate procedures because the uplink or the downlink respectively does not provide the expected quality. In this paragraph we describe the procedures taking place to decide the carrier that will accommodate the traffic of the MT executing hand-off. These procedures involve the MT, BS_old and BS_new. The procedures related to location management and connection rerouting involve entities in the fixed part of the private ATM network and are discussed in next sections.

Hand-off between two base stations happens when the AP-switch procedure is failed, i.e. there is no carrier of the current BS with acceptable quality or enough bandwidth. The MT has to scan the carriers of the neighbouring BSs. The identifiers of these carriers could have been given to MT during its registration in the coverage area of the BS or could be
given to the MT with the notification for carrier switch. Among these carriers the MT will select one as the favourite carrier to hand-off. In order to notify the new AP and the new BS for its existence, and acquire bandwidth for the transmission of the required messages to the new AP, the MT must access successfully (after contention) at least one of the sub-slots at the beginning of every frame on the AP's carrier. A successful access of a sub-slot will result to the transmission of the Inter-cell Hand-off Request by the MT.

Because BSs are not synchronised, there could be a time shift between the beginning of the frame at one carrier and the beginning of the frame at another. MT should be able to switch between the two carriers the one to be abandoned and the one to be joined. For that reason MT can indicate to the two APs, involved in the hand-off, in what part of their corresponding TDMA frames exchange of packets should take place between the corresponding AP and the MT executing hand-over. This way, there is guarantee that the MT can execute seamless hand-over. A straightforward solution would be for the MT to request a certain number of time slots in a certain part of the frame of the new AP for signalling, and use the rest of the time for communication with the abandoned AP. For that reason, the MT after contending successfully in the randomly accessed subslots in the new carrier is provided with time slots which it uses to transmit the F5 OAM Inter-cell Hand-off Request packet to indicate to the new AP the range of the time slots that could be used by the MT to transmit or receive the hand-off related messages to/from the new AP.

The BS to be joined sends to the MT a Inter-cell Hand-off Indicate with the list of its carriers and their available bandwidth. The MT responds then with the Carrier Indicate packet which indicates the carriers of preference of the MT. The BS then responds with the Carrier Allocate packet indicating the AP that will serve the MT. Note that inter-cell hand-off is always MT initiated. If the hand-off is totally rejected, the connections of the MT will be dropped and the MT will be de-registered from the network, assuming that it looses the connection to the previous BS, too. However, if the MT is accepted it follows the address registration procedure described in section 6.4. Moreover, it can indicate the successful inter-cell hand-off to
the old BS. The latter will then receive by the backbone network the address of BS_new. This action will trigger the rerouting of the MT's connections and the release of all the resources related to the MT at BS_old. The release could be slightly delayed to allow the transmission of packets in the downlink (BS_old to MT).

If the MT looses the connectivity to BS_old before initiating hand-off to BS_new, then its connections are not terminated if the connection re-routing message arrives to BS_old within T1+T2 msecs after loss of the connection. This happens because, MT could establish communication with BS_new and the re-routing of the connections is based on information carried by the backbone network. The deployed mechanism is presented in next sections. Fig. 6.3 presents the required signalling for inter-cell hand-off.

6.2 Initiative for Mobility Management in ATM WLAN.

Mobility in Wireless ATM networks is a research area of great interest, and significant related work has already been presented [ACAMPORA'94], [YUAN'96], [AKYOL'96], [ACAMPORA'96], [WALKE'96]. It originates from the requirement to combine the support of multimedia communications and the provision of mobility. Classical examples are the Mobile Broadband System (MBS) [WALKE'96] and the Universal Mobile Telecommunication System (UMTS) [BUIT'95]. Unlike ATM WLAN, these systems aim to cover large geographical areas and support users highly mobile. ATM WLAN will cover smaller areas but its users could be highly mobile too (i.e. provision of multimedia within a university campus).

Furthermore, support of mobility in ATM, completes the work that deals with the definition of advanced multiple access schemes for optimised Wireless ATM networks (WATM-nets) [CHINI'96], [RAYCH'92], [MAHM'96], [APOST'95]. Until recently, the design of ATM networks did not take into account the mobility factor. So, the existing protocols for ATM deal mainly with the varying traffic characteristics of fixed connections but not with the implications imposed by user mobility. For these reasons, the existing ATM protocols have to be enhanced to incorporate mobility management functionality.

The work presented in this thesis differs from that described in [ACAMPORA'94], [YUAN'96], [AKYOL'96], [ACAMPORA'96], in the sense that, it presents a new proposal for a signalling protocol for mobility management that is based on existing ATM protocol specifications. Moreover, since the proposed protocol could be applied in any environment, it could be used combined with the proposals in [ACAMPORA'94], [AKYOL'96], [SCWARTZ'96], to improve the network performance during hand-off. Moreover, compared
Chapter 6: Mobility Management in ATM WLAN based on Standardised Protocol Specifications

Furthermore, it is believed that the backbone network of a WATM-net will have a mesh and not a deterministic topology as required in [ACAMPORA’94], [ACAMPORA’96], [AKYOL’96] and [WALKE’96]. Then, the mechanisms of Private Network-to-Network Interface specification (PNNI) [PNNI’96] can be exploited, not only because they are topology independent, but also, because they support communication of topology information, minimise the traffic for topology update, and calculate an optimum route for a given destination.

However, the exploitation of PNNI for support of mobility management is proposed within ATM Forum [LIVLIST]. However, that proposal does not integrate both location management and network hand-off within the same signalling framework. Moreover, it results to more signalling exchange compared to the work presented in this thesis.

6.3 Protocol Reference Model for WATM-NET

Fig. 6.4 presents a possible network configuration for a WATMnet, and the Protocol Reference Model (PRM) applying at different parts of the network. The Radio Access Layer functions were presented in chapters 3 and 4. As shown, Mobile Terminals (MTs) communicate, through Base Stations (BSs), with ATM switches that form the backbone part of WATM-net. BSs act like modems which extend the transmission from the wire to the radio and vice versa. Before any transfer of information from the layers above Radio Access Layer (RAL) begins, MT receives a RAL number that identifies uniquely the MT in a given BS coverage area. The reception of this number is accompanied by the reception of all the required information for ciphering and data encryption in the RAL layer, so that all the subsequent transmissions carrying user profiles and identities are protected against malicious interceptions. This is an advancement compared to GSM system where User Identity is transmitted un-ciphered [GSM0902]. Moreover, the ATM switch port number that is connected to a BS, and the RAL number of a MT attached to that BS, form the Physical Interface Index of the RAL connection to the MT, in the access ATM switch.

In an effort to relate WATM and B-ISDN terminology, the interface between a terminal and the access ATM switch has been called Mobile User Network Interface (M-UNI) while the interface between switches has been called Mobile Network to Network Interface (M-NNI). As the MT moves among BSs connected to the same ATM switch, only its Physical Interface Index changes and the rest of the network is not affected. Connection
re-routing happens only in the access ATM switch and the MT keeps its current WATM address.

Fig. 6.4 Protocol Reference Model for WATM-net

The above mentioned interfaces are based on the existing ATM specifications [UNI31],[UNI4],[PNNI'96], and contain the required modifications to support mobility management functions. The rest of the chapter explains how P-UNI management functions, and P-NNI management and call control signalling are used, or modified, to cope with user mobility. The procedures covered in the following sections are:

1. Address registration, i.e. the allocation of an ATM address to a terminal attached at a given M-UNI and the association between the terminal address and the user WATM identity.

2. Location update, i.e. the concept of tracking a user or a terminal location as it moves and changes the M-UNI, so that new calls can be routed to the roaming user automatically through an optimum route.

3. Virtual connection re-routing (during hand-off) as communicating parties change their location.
Although provision for user or terminal authentication is taken into account this function is not treated in detail within this work. Section 7.2 outlines further research needed in order to complete the work presented in this chapter.

6.4 Address Registration at the WATM-UNI

The address registration procedure at the M-UNI is based on the procedure standardised for B-ISDN and described in [UNI31]. For that reason, the ATM addressing scheme in [UNI31] is assumed in WATM-net. Then, WATM addresses consist of a Network Prefix provided by the network and a User Part provided by the MT. Moreover, a Mobile User is expected to be identified by a globally unique WATM User Identifier (WATM-UI). Then, a MT has always two WATM addresses:

1. The MT current WATM address associated with the current location of the MT. The Network Prefix is provided by the access ATM switch, while the User Part is the WATM-UI of the mobile user. MT current WATM address is used in Call Set-Up messages and in messages related to connection re-routing.

2. The MT permanent WATM address based on the location that is considered as the home location of the user. The User Part is the WATM-UI of the corresponding mobile user. The Network Prefix combined with a default value for the User Part, specifies the WATM address of a Location Server equivalent to GSM HLR [STEEL'92]. This Location Server is described in section 6.6.2. MT permanent WATM address is used to identify the mobile user/terminal when location update and address registration take place.

For address registration, a client-server mechanism across the M-UNI is defined where a terminal acts like a client requesting a WATM address to be registered for itself. The client-server mechanism is based on Simple Network Management Protocol (SNMP) [WASH'94] (Fig. 6.5). SNMP messages are passed to Signalling ATM Adaptation Layer [UNI4] where the required error control functions take place, and SNMP messages are segmented into ATM cells. These cells, carrying SNMP messages, are transferred in a Virtual Circuit (VC) with a well defined Virtual Circuit Identifier (VCI). To support the address registration procedures, two Management Information Base (MIB) Tables are defined. The first is called Network Prefix Table and contains network prefixes in effect at the M-UNI. The second is called Address Table and contains the registered ATM addresses at the corresponding M-UNI. The Network Prefix Table resides in the MT and is modified by SNMP messages issued by the network side of the M-UNI. The inverse stands for the Address Table. For this reason, in every side of the M-UNI there is a UNI Management Entity (UME) to process the received SNMP messages, manipulate the corresponding
current WATM address takes place. Furthermore, the re-routing of the MT active connections is triggered, as it is described in section 6.7.

In Figs. 6.6 and 6.7, it is shown that all the required information is transferred after the ciphering information is provided to the MT. So all identities information is transferred protected. The message exchange described above, is independent of the Radio Access Layer algorithms which are used for the multiplexing of traffic on the air interface. The only requirement is the allocation of the RAL number to a MT that accesses the radio channel in a BS coverage area for the first time. Thus, a uniform signalling protocol is defined for hand-off and mobility management and for the different radio access technologies that are proposed for WATM networks.
6.5 Hierarchical Routing in PNNI Domain

WATM-net, consisting of backbone and access ATM switches, can be organised according to PNNI specification [PNNI’96]. As stated previously, it is realistic to assume a mesh topology for the ATM switches of a wide area ATM network. Then, PNNI specification can be exploited since it defines a protocol for distributing topology or routing information, and a signalling protocol for call set-up and release. To define the M-NNI, which applies between any two ATM switches of WATM-net, PNNI signalling has to be extended in order to support location update, in-call connection re-routing, and routing of the Call Set-up message to a mobile terminal or user. In this section, the logical organisation of an ATM network is described, based on PNNI specification. In section 6.6, it is shown how WATM-net can be organised based on PNNI principles. There, the required modifications are presented to lead to the complete definition of M-NNI.

6.5.1 Logical Organisation of ATM Network based on PNNI.

According to PNNI, a switching system is defined as a set of one or more ATM switches that are represented as a single switch (called also Logical Node or LN) for the purpose of routing. To implement efficient routing in PNNI, a logical hierarchy is defined, where in its bottom layer, real ATM switches form groups considered as individual LNs in the next upper layer of the hierarchy. Moreover, LNs form groups to be treated as LNs of even higher levels.
For each group of LNs or switches (called Peer Group), one of its nodes, called Peer Group Leader (PGL), carries out all the required operations that represent the Peer Group as a single LN. Although a PGL could be elected periodically among the nodes of the group, there is provision in PNNI for defining the Peer Group leadership by network configuration [PNNI'96]. While at the bottom layer of the P-NNI hierarchy switches are interconnected using physical links, LNs of upper layers are interconnected by logical links. PNNI specifications describe how physical links between real switches of the bottom layer form logical links between LNs of upper layers. Neighbouring LNs exchange PNNI Topology State Packets (PTSPs) which contain topology and link state information. Thus, the information required for the routing is flooded in the network at the different layers and is used to update the topology databases of the switches. An example of hierarchical organisation of the PNNI routing domain is shown in Fig. 6.8.

Fig. 6.8 Hierarchical organisation of the P-NNI routing domain and address reachability (in italics the advertised address summaries)

6.5.2 Address summarisation and advertisement in PNNI routing domain.

ATM addresses are used in such a way that a specific sub-part of the address refers to a specific layer of the PNNI domain. An address summarisation and advertising method is used to flood the address reachability information [PNNI'96] in the whole network. According to this method, switching systems and LNs advertise the set of the ATM addresses they...
represent by broadcasting a summary address, which is the common prefix of the ATM addresses they represent. Similarly, LNs of upper layers advertise address summaries to their neighbour LNs (same layer), and to the lower layer LNs that compose them. The higher the layer of the PNNI domain, the shorter the address summaries and possibly the volume of the address advertisement. Thus, if the virtual organisation of the network follows the topological distribution of ATM switches, routing information among separated ATM switches becomes minimum and the same stands for the associated traffic. Fig. 6.8 shows an example of hierarchical network organisation and address summarisation according to PNNI specifications. The address value that is passed from LN A to LN B is A. This value is further passed from LN B to LNs B1 and B2 and consequently to all the real ATM switches in Peer Group B. Then all the nodes in Peer Group B know that they have to route their packets to LN A through LN B. However, if for some reason there is an ATM address that does not fit in any address summary, then this address must be advertised un-summarised. This is a problem when MT permanent WATM address is considered. This address must be used as an index in the location servers that hold entries about the associated user or MT. However, this address cannot be summarised by access switches or LNs. In section 6.7, it is described how the MT permanent WATM address is used in the network. However, the above mentioned address advertisement and summarisation scheme covers the MT current WATM addresses and, as presented in section 6.7.2, it enhances the Call Set-Up and in-call re-routing procedures.

6.5.3 Routing of the Call Set-Up message in PNNI domain

Based on the above mentioned hierarchical organisation, every switch does not need to know about the states of all the physical links in the network, or the accurate ATM addresses accessible by every real ATM switch. However, every ATM switch knows all the ATM addresses that are accessible by a LN of an upper layer, because the latter has advertised their address summary downwards. Matching the called MT address to the longest possible advertised address summary, the access ATM switch of the calling MT knows the LN of the lowest possible layer that must receive the Call Set-Up message in the route to the called MT.

Moreover, every LN knows the detailed route in its Peer Group that links to another LN in the same Peer Group or to an LN of an upper layer. The Call Set-Up message contains the current ATM address of the called MT. In the following paragraph routing of the Call Set-up message is presented, assuming that the called MT current WATM address is known. The mechanism for learning this address is presented in section 6.7.2.
Based on the above conventions, a Call Set-Up message containing the called MT current WATM address, is routed as follows [PNNI'96]: A coarse route is included in the Call Set-Up message by the access ATM switch servicing the calling MT. This route contains the upper layers to be crossed by the Set-Up message in order to reach a higher layer LN which logically stands “above” the access ATM switch of the called MT. This LN is the intermediate destination of the Set-Up message. The calculated route is characterised as hierarchically complete source route. It is described as a set of lists called Designated Transit Lists (DTLs). Every DTL describes the LNs of a specific hierarchical layer that must be traversed by the Call Set-Up message to its intermediate destination. When the Set-Up message reaches its intermediate destination a new set of DTLs will be calculated to forward the Set-Up message downwards in the PNNI hierarchy to its final destination (the called MT).

Whenever a LN (let’s say LN1) receives the Call Set-Up message it checks for the next LN to be reached (let’s say LN2). Since a LN is itself a Peer group of lower layer LNs, an optimum route has to be followed within LN1 towards LN2. Following a repetitive procedure in every upper layer LN, the projection of the calculated routes on the bottom layer of the PNNI hierarchy will give the optimum route from the calling MT to the called MT. As it is mentioned in [PNNI'94]: The process of traversing and creating lowest level DTLs is similar to driving across a continent. One needs a high level map to determine what countries to go through and which border crossings to take. At each border, one can discard the country map he/she was following and pick up a new detailed map of the country just entered. An example of Call Set-Up is given in [PNNI'96].

6.5.4 Crank-back during Call Set-Up

During the call set-up procedure, and as the Call Set-Up message propagates from the source to the destination node, certain LNs are defined as crank-back addresses. At these points a re-routing of the call can happen to go around a part of the network that cannot accept the new call because of traffic overload or faulty links. Usually as crank-back points are defined the LNs where the call set-up message enters a new Peer Group. Whenever a LN cannot forward further the Call Set-Up message, it sends backwards a Call Release message [PNNI'96]. This message is sent backwards until it reaches a crank-back address where the re-routing of the call will take place. Note that a call re-routing can occur only during call set-up and not during a call in progress. In section 6.7.3, the crank-back mechanism is extended for in-call connection re-routing that happens because a MT executes hand-off between different access ATM switches.
6.6 WATM-NNI Functionality for Mobility Management

6.6.1 Definition of WATM-NNI routing domain.

Based on the above mentioned hierarchical organisation of PNNI routing domain, a proposal for the WATM-NNI routing domain could be given to, minimise the traffic for mobility management in the network, guarantee service provision to mobile users, and assure security in terms of network resource utilization and user billing.

The whole set of access and backbone ATM switches that will span over different countries will form the WATM-NNI routing domain. Every logical layer of the latter will correspond to a different kind of physical region. For example, the lowest layer could be the Street NNI layer containing different Peer Groups, each one consisting of the ATM switches servicing a street. Moreover, the next upper layer could be the Neighbourhood layer where Peer Groups are composed by LNs where each node is a street layer Peer Group. Similarly and going upwards, City, County, Region/State, Operator, Country layers can be defined where Country layer is the most top layer. Fig. 6.9 shows an example of WATM-NNI routing domain for the three lowest layers. Peer Group Leadership can be configured by Network Administration. Moreover, division of switches into Peer Groups can take place according to expected traffic at certain regions.

Minimisation of traffic for mobility management is achieved by confining it within the area where user mobility takes place. Guarantee of service provision and security are assured by the ability of the network to communicate User profiles, authenticate the MT and update its location.

6.6.2 Extension of PNNI routing/topology functions for Location Update

As already described in 6.5.2, there is a mechanism in PNNI routing domain for communicating address reachability. This mechanism takes advantage of the address summarisation to minimise the topology information flooded in the network. However, if addresses have to be advertised un-summarised this mechanism does not minimise the related traffic.

Assuming that a MT uses always its MT permanent WATM address, and following the address advertisement mechanism, whenever a MT crosses the border of a Peer Group two sets of messages have to be transferred to all the ATM switches within these Peer Groups. The first set contains the messages informing about the removal of the MT from the previous ATM switch. The second set contains the messages that inform about the MT attachment to the next ATM switch. Although every ATM switch can directly calculate the
hierarchically complete source route to the terminal after the reception of these two sets of
messages, it is believed that unnecessary traffic is created. So in the case of MTs, the address
summarisation benefit characterising PNNI is lost.

![Diagram of WATM-NNI routing domain up to City Level]

Fig. 6.9 Example of WATM-NNI routing domain up to City Level

For example in Fig. 6.9, when a terminal is detached from node 'A.2.2' and is
attached to 'A.1.3', a message from 'A.1.3' and a message from 'A.2.2' will have to be
received by all the switches in Peer Groups 'A1' and 'A2' (i.e. 9 nodes are involved).
Moreover, if a terminal moves from 'A.2.2' to 'B.1.1', all the switches in Peer Groups 'A1',
'A2', 'B1', 'B2' and 'C' will be involved (i.e. 17 nodes will be involved). It is obvious that
the location update signalling is not confined in the area where mobility occurs.

The following solution is used to reduce traffic for location update and integrate
signalling for in-call connection re-routing: A new MIB is defined in every Peer Group of
every level in WATM-NNI routing domain. It is called Location Management Information
Base (LMB), and it contains for every MT a record with the following information elements:

1. The MT permanent WATM address.

2. Authentication information for every MT registered within the Peer Group or within a
   Logical Node in the Peer Group.
3. The address of the Logical Node in the Peer Group that has to be reached towards the current MT location. For a Street Level Peer Group, the Logical Node is a real ATM switch.

If there is no information about a MT in a given LMIB, then an upper level LMIB is expected to contain such information. However, if there is not such LMIB then the Global LMIB, described later in this section, has to be interrogated about the location of the MT.

LMIB is realised in a switch or a LN called LMIB Server. LMIB server could reside in the Peer Group Leader (PGL). A different Semi-Permanent Virtual Connection (SPVC) is defined for each Logical Node of the Peer Group and the PGL to exchange messages related to mobility. So in Fig. 6.9, in Peer Group ‘A1’ there are 3 SPVCs for mobility management between ‘A.1.4’ and ‘A.1.1’, ‘A.1.2’ and ‘A.1.3’ respectively. Moreover, there is a SPVC from LN ‘A1’ to LN ‘A2’ since ‘A2’ is the LMIB server of Peer Group ‘A’. This SPVC is realised in the link connecting switches ‘A.1.4’ and ‘A.2.5’. Accordingly, there is an SPVC between node ‘A.2.5’ and ‘C’ to realise the SPVC for mobility between LN ‘A’ and LN ‘C’. In this case LN ‘C’ is the PGL for the City area (City Level).

The information that is stored in the LMD3 of a street level Peer Group per MT is: (MT permanent WATM address, valid, address of advertising access ATM switch or LN). Moreover the access switch has assigned a MT current WATM address to a MT as described in section II. The information stored per MT in upper level LMD3s is (MT permanent WATM address, valid, address of advertising LN). The address of advertising LN could identify a lower layer or upper layer node. Note that, since ‘A.2.5’ is Peer Group Leader of Peer Group ‘A2’ and Logical Node ‘A2’ (represented by ‘A.2.5”’) is Peer Group Leader of Peer Group ‘A’, ‘A.2.5’ stores information related to Peer Group ‘A2’ and Peer Group ‘A’ (i.e LMIBs related to 2 virtual levels of the WATM-NNI routing domain). In that case, there is no need to keep any information referencing LN ‘A2’, since this information is already available in the LMIB corresponding to the lower level.

Based on the definition of LMIBs and LMIB servers, WATM address advertisement does not take place among all the nodes of the network but only among LMIB servers. This results to fast location update and reduced signalling. If a MT moves from ‘A.2.2’ to ‘A.1.3’ then nodes ‘A.1.4’ and ‘A.2.5’ will be involved on top of ‘A.1.3’ and ‘A.2.2’, as LMIB servers of Peer Groups ‘A1’ and ‘A2’ respectively (i.e. 4 nodes are involved compared to 9 in the above mentioned example). If the MT moves from ‘A.2.2’ to ‘B.1.1’, then nodes ‘A.2.5’, ‘C’ and ‘B.1.4’ will be involved as LMIB servers on top of ‘A.2.2’ and ‘B.1.1’ (i.e. 5 nodes are involved compared to 17 in the previous example).
Based on the above mentioned conventions a MT always uses the MT permanent WATM address for Mobility Management procedures. This address is provided by the user of the MT, by inserting in the terminal a Subscriber Identification Module card, like it happens in GSM network with SIM card [GSM0217]. Moreover, the MT permanent WATM address identifies the location of the LMIB server that has to be interrogated regarding the location of a user, if no other LMIB has knowledge about that user location. A MT permanent WATM address, contains a Network Prefix that, combined with a fixed and well known value for the address User Part, addresses a LMIB Server in an Operator Level of WATM-NNI routing domain (i.e. it specifies country and operator). That LMIB server, called Global LMIB Server, acts like GSM HLR for the WATM operator subscribers and like GSM VLR [STEEL’92] for the WATM users that are serviced by this operator but are subscribers of another WATM operator. The User Part of the MT permanent WATM address contains the WATM-UI that has been assigned by the WATM operator to its subscriber.

So, compared to PNNI specifications, the concept of MIB has been extended to that of LMIB for storage of information related to Mobile users, and the concept of LMIB Server has been added in order to minimise the traffic related to Location Update. Moreover, the address advertisement and summarisation mechanism has been changed for mobile MTs, since reachability of MT current WATM addresses is not flooded anymore to every switch in the WATM network. Only reachability of MT permanent WATM addresses is exchanged among LMIB servers. However, PNNI mechanisms stay unaffected for the backbone infrastructure of WATM-net.

As it was mentioned in 6.5.1, there is an Information Exchange mechanism in PNNI (based on PTSPs) that floods topology information in the nodes of the PPNI routing domain. The same mechanism is used for transferring the Mobility Management Information among the LMIB Servers of the WATM-NNI routing domain. Moreover, the reception of such information by a LMIB Server could initiate the in-call connection re-routing for MTs executing hand-off.

6.7 Mobility Management Procedures in WATM-NNI

6.7.1 Location Update

Based on the definition of WATM-NNI routing domain, a Location Update procedure has to take place whenever a MT is switched on, or it changes the access ATM switch of attachment to WATM-net. In both cases, and as mentioned in section 6.4, there is an access ATM switch listening to MT for first time, and assigning to MT a RAL number to initiate the address
registration procedure. Fig. 6.10 shows the message exchange for Location Update of a MT that moves from node 'A.2.2' to node 'B.1.1'(in Fig. 6.9).

Generally, after receiving the MT permanent WATM address and the authentication information, the access switch will issue a Reachable_MT_PTSP and transmit it to the LMIB Server of its Peer Group. The Reachable_MT_PTSP contains the MT permanent WATM address, and the required authentication information provided by the MT. The access ATM switch will then wait for the response to that PTSP in order to register its Network Prefix in the Network Prefix Table of the MT, and initiate the registration of the MT current WATM address. Whenever a LMIB Server receives a Reachable_MT_PTSP, it checks if there is a valid entry for the carried MT permanent WATM address in its LMIB. If there is no valid entry, an upper LMIB Server must be interrogated, and this procedure is repeated until an LMIB server in the Country Layer is reached. If the MT is not registered within the country then its Global LMIB server has to be interrogated. Global LMIB Server is referenced by the Network Prefix of the MT permanent WATM address. So eventually, there will be an LMIB containing a valid entry for the MT permanent WATM address under consideration, which at the worst case will be its Global LMIB. The first LMIB Server that will find a valid entry for the MT permanent WATM address, will try to Authenticate the terminal, and return the feedback to the LMIB Server that issued the last Reachable_MT_PTSP. The feedback is contained in the packet acknowledging that Reachable_MT_PTSP. An inverse direction is
followed, and all the LMIB servers in the path from the authenticating LMIB Server to the access ATM switch must set a valid entry for that MT permanent WATM address. Moreover, the authenticating LMIB Server notifies all the LMIB Servers that had previously valid information for the MT about the new advertising LMIB Server. Section 6.7.3, describes how the location update procedure initiates the in-call re-routing for the active connections of the moving MT.

6.7.2 Address Resolution and Call Set-Up to Mobile User

Whenever a Call Set-Up message is issued by a MT, the WATM backbone network has to forward this message to the current location of the called MT. Moreover, as the Call Set-Up propagates towards its final destination, the corresponding virtual connection is established, by allocating the required resources at the intermediate nodes in the path to the called MT.

As it was described in section 6.5.3, the Call Set-Up message passes through Peer Groups of different layers. Within every Peer Group, an available optimum route is calculated always. So the total route from source to destination is an optimum route (out of the more than one that could exist), and is provided by PNNI mechanisms.

Considering mobile users, the main objective of the Call Set-Up transfer mechanism is to find the access ATM switch that services the called MT, learn its current WATM address (address resolution phase) and use it in the Call Set-Up message. Then, the Call Set-Up procedure described in 6.5.3 can be followed (Call Set-Up phase). So, a calling terminal provides the MT permanent WATM address of a called MT. In 6.6.2, it was described how the user WATM-UI is associated to the MT permanent WATM address. The access ATM switch of the calling terminal inserts in the Call Set-Up message the MT current WATM address of the called user's MT.

The address resolution phase involves LMIB Servers in two different paths, and is triggered as soon as the access ATM switch of the calling MT receives the Call Set-Up message. One path contains the LMIB Servers from the Peer Group of the calling MT access switch to the first LMIB Server that contains a valid entry for the called MT permanent WATM address. In the worst case this LMIB Sever is the called MT LMIB Server. The second path contains all the LMIB servers to the Peer Group of the called MT access switch.

The access ATM switch of the calling MT sends a Resolve_Address_PTSP(permanent WATM address value) to the LMIB Server of its Peer Group and expects as response, the current WATM address of the called MT. Similarly to the procedure for Location Update, every LMIB Server checks its LMIB to find the next LMIB
server with valid information about the called MT. If there is a valid entry in the LMIB for the called MT, then the Resolve_Address_PTSP is forwarded to next LMIB server that is indicated in the valid LMIB entry. Otherwise, it is sent in the LMIB server of the upper layer in WATM-NNI domain.

If the Resolve_Address_PTSP reaches the Country layer of the WATM-NNI routing domain of the calling terminal, it means that the Global LMIB Server of the MT must be interrogated. This Global LMIB Server is identified by the Network Prefix of the MT permanent WATM address. Then the address resolution procedure will go through the Global LMIB Server of the called MT.

At the end, the access ATM switch of the called MT will be found and declare the MT current WATM address. Then, the called MT current WATM address value will be returned via all the interrogated LMIB servers to the access ATM switch of the calling terminal, to start the Call Set-Up phase as described in section 6.5.3. The SPVCs defined for communication of Location Update messages are used for the transfer of Resolve_Address_PTSPs and their responses.

The introduction of the address resolution phase is required to keep a Call Set-Up phase compatible with the existing PNNI specifications. To avoid its overhead when the called terminal is a fixed one, specific WATM address Network Prefixes could differentiate between mobile and fixed terminals. An access ATM switch receiving a Call Set-Up message from a MT can decide based on the prefix of the called MT WATM address, if the call is destined to a fixed or mobile terminal. Then, the access ATM switch can follow or skip the address resolution phase.

6.7.3 Network Hand-off, In call re-routing

Hand-off operation involves two access ATM switches. The abandoned switch is referred hereafter as Old Access Switch, while the joined one is referred as New Access Switch. As mentioned in section 6.4, the hand-off operation and the requirement for in-call connection re-routing is indicated by the MT to the Old Access Switch, during the procedure of Address Registration (Fig. 6.7). Moreover, if the connectivity to a MT is lost by the Old Access Switch, the latter could set the status of the corresponding Physical Interface Index to hand-off for a short period before releasing the connections of the lost MT. This will enable the network to inform the Old Access Switch about the new location of the MT if it executes hand-off.

In section 6.7.1, it was mentioned that the first LMIB Server in the path from New Access Switch to the Global LMIB Server of the MT, with a valid entry for the moving MT,
will indicate to all LMIB Servers in the path to the Old Access Switch the new attachment point of the MT. This information will reach the Old Access Switch that will ask for the re-routing of all the established connections of the MT (if any). LMIB Servers are not involved in the Connection re-routing.

For every established call, the Old access Switch will issue a Call Release packet which contains one more field, compared to the message format standardised in PNNI [PNNI'96]. The new field is called New Destination Address. This field contains the New Access Switch Address. Moreover, the field “Cause” in the Call Release message gets the value NumberChanged. A LN then, must be defined as the optimum point for the re-routing of a connection ‘C’ to New Access Switch. This LN is found as following.

Every LN ‘X’ that receives the Call Release packet, compares the address of New Access Switch to the known address summaries (for the fixed part of the network). These have been advertised to it using the standard PNNI address advertisement mechanism. Then all the possible Routes (not only the optimum) to New Access Switch can be calculated based on PNNI mechanism.

If at least one of these calculated Routes contains the hop from LN ‘X’ to next LN ‘Y’ within the route of connection ‘C’, LN ‘X’ forwards the Call Release message to LN ‘Y’. Otherwise LN ‘X’ is the node to re-route connection ‘C’.

LN ‘X’ will issue a Call Re-route Message with destination address the New Access Switch address. Call Re-route is a new Message added in PNNI signalling. In WATM-net, a portion of backbone resources (link capacity, buffer space) should be allocated for in-call re-routed connections, to minimise the call dropping probability. Call Re-route message requests for that portion of reserved resources. Moreover, it carries a copy of the call profile at the Old Access Switch to re-establish the WATM-UNI environment between the MT and the New Access Switch. After issuing the Call Re-route message, LN ‘X’ switches traffic in connection ‘C’ towards the New Access Switch. It is assumed that, the Call Re-route message will not undergo a crankback procedure, since there will be reserved bandwidth for in-call connection re-routing.

Moreover, LN ‘X’ issues a Call Release Complete Message as a response to Call Release message. Then all the nodes in the previous path of connection ‘C’ (towards Old Access Switch), will release the resources allocated for ‘C’. When Old Access Switch Receives the Call Release Complete message, it will send the SetRequestResponse(InterfaceIndex) (Fig. 6.7) to MT and release all the resources allocated
for it. Then, MT will register a new MT current WATM address in the New Access Switch and support traffic of connection 'C' through the new route.

It must be emphasised that, the above described in-call re-routing mechanism gives an optimum route from the calling MT to the new location of called MT. This route is calculated according to PNNI specification.

6.7.4 Example of in-call connection re-routing

With Fig. 6.9 as a reference, let's assume the case where a MT1 initially attached to 'A.1.1', communicates with a MT2 attached to 'A.2.2'. Moreover, let 'A.1.1' - 'A.1.4' - 'A.2.5' - 'A.2.4' - 'A.2.2' be the route of the corresponding connection 'C'.

When MT2 moves towards 'A.1.3', 'A.2.2' becomes the Old Access Switch and 'A.1.3' the New Access Switch. MT2 still uses 'A.2.2' to carry traffic of connection 'C' while it is in address registration phase with 'A.1.3'. Let's assume that authentication has been carried out by 'A.2.5' (PGL of 'A'). According to Fig. 6.7, MT2 has received the Network Prefix from 'A.1.3' and has issued the hand-off flag to 'A.2.2'. When MT2 receives the corresponding acknowledgement by 'A.2.2', it will stop dealing with 'A.2.2' and will register a MT current WATM address in 'A.1.3'. As soon as it receives the acknowledgement by 'A.1.3', it will start carrying traffic of 'C' using 'A.1.3'.

The Location Update procedure will invoke LMIB Servers 'A.1.4' and 'A.2.5', since the latter contains all the required information for authenticating MT2. LMIB Server 'A.2.5' will change the address of advertising LN to 'A.1.4' in the valid entry for MT2. LMIB Server 'A.1.4' will set a valid entry for MT2 with 'A.1.3' as address of advertising LN. Moreover, LMIB Server 'A.2.5' will ask 'A.2.2' to de-register MT2.

'A.2.2' sends the Call Release message to 'A.2.4'. There, it is found that all the available Routes to 'A.1.3' are 'A.2.4' - 'A.1.4' - 'A.1.3' and 'A.2.4' - 'A.2.5' - 'A.1.4' - 'A.1.3'. Since the second route contains the hop 'A.2.4' - 'A.2.5' of connection 'C', the Call Release Message is forwarded to 'A.2.5'. Similarly, it reaches 'A.1.4' which finds out that it is the node to re-route the connection. Still, MT2 receives traffic in connection 'C' through the initial route 'A.1.1' - 'A.1.4' - 'A.2.5' - 'A.2.4' - 'A.2.2'.

'A.1.4' takes two actions: first, it issues a Call Release Complete message to 'A.2.5' and stop switching traffic of C towards 'A.2.5'; second, it issues a Call Re-route message to 'A.1.3', followed by the traffic of 'C'.

The Call Release Complete message will arrive eventually to 'A.2.2' and it will be the last information element of connection 'C' transmitted in the path 'A.1.4' - 'A.2.5' -
‘A.2.4’ - ‘A.2.2’. The reception of this message will cause ‘A.2.2’ to acknowledge the hand-off procedure by sending to MT2 the SetRequestResponse(InterfaceIndex) (Fig. 6.7). Then, MT2 will register a MT current WATM address in ‘A.1.3’ and use this New Access Switch for the traffic in connection ‘C’.

If MT2 had more than one active connections, the reception of the last Call Release Complete message would trigger the acknowledgement of hand-off by ‘A.2.2’. If the physical connectivity to ‘A.2.2’ is lost by MT2, then the latter will proceed to the registration of a MT
current WATM address within a specified period after the SetRequest (InterfaceIndex.status, "hand-off") (Fig. 6.7) has been issued to ‘A.2.2’.

‘A.1.3’ will find in the Call Re-route message, all the required information to establish the corresponding call profile at the network part of WATM-UNI. Moreover, ‘A.1.3’ will buffer the cells of connection, if required, until MT2 registers a new MT current WATM address.

6.8 Conclusion

B-ISDN is expected to integrate over a uniform transfer platform (ATM layer of its protocol reference model) different applications with different characteristics in terms of population of communicating users, number of active connections, traffic profiles and quality of service requirements. However, the initial specifications of protocols related to B-ISDN, do not take into account the requirement for mobility.

In the proposed scenario for hand-off in ATM WLAN, the signalling required to decide the next carrier that will accommodate the traffic of the MT was introduced. An extension of OAM flows to carry the required messages was proposed, and eight new F5 OAM packets were introduced with the format presented in Fig. 6.1. Moreover, it was shown that the carrier selection, is not based only on the quality of reception but also on the bandwidth requirements of the MT and the bandwidth availability in the candidate carriers. This is an innovative concept compared to single call mobile networks (like GSM), imposed by the capability of ATM networks to support multiple calls over the same physical connections.

Furthermore, considering mobility management in the core network, the proposed modifications presented in this chapter, apply to the management functions of the P-UNI, where the standardised address registration mechanism is extended to support authentication and registration of a mobile terminal, and indicate hand-off between different access ATM switches. Moreover, the hierarchical organisation of an ATM network according to PNNI specifications, as well as, the mechanisms for topology update and routing are kept to define the WATM-NNI routing domain (the set of fixed network infrastructure). Furthermore, PNNI was extended to cope with location update, Call Set-Up to mobile user, in-call connection re-routing when hand-off takes place. For that reason, the concept of LMIAs and LMI Servers at different hierarchical layers of WATM-NNI routing domain was introduced, to provide the required knowledge regarding the location of mobile users. The signalling among LMI Servers is an extension of the Routing Specification and uses similar Protocol Data Units. It deals with location update of a terminal and is used to resolve the current WATM address of
a mobile terminal whenever a Call Set-Up is issued to its user. Moreover, it was shown how the Location Update process could indicate a hand-off to an access ATM switch.

The resulting signalling protocol could be used to integrate different multiple access technologies in different environments (pico-cellular, micro/macro-cellular, satellite) by providing a common ATM backbone network with mobility support, and a common signalling layer (UME) at the access part (UNI). Then, the selection of a multiple access scheme becomes irrelevant.

Compared to [YUAN'96], the proposed protocol introduces fewer messages in the backbone network (3 compared to 8) and less complicated signalling to support mobility management. Unlike the proposal in [LIVLIST] it provides for in-call connection re-routing. Moreover, it produces less messages for location update. This happens because in [LIVLIST] the location update information has to reach all the ATM switches in the vicinity of the MT while the proposed protocol defines signalling only among a sub-set of the switches (the LMIB Servers). Furthermore, unlike [YUAN'96], the generic nature of the presented signalling can be combined with different multiple access schemes or even solutions for improvement of hand-off performance [AKYOL'96].
Chapter 7

Conclusion and Further Research

This chapter discusses the extent to which the objectives have been met and proposes further research issues that need to be addressed.

7.1 Conclusion

ATM has been introduced as the technology to support statistical multiplexing of multiple types of services in the same network. Compared to traditional network solutions, it aims to improve the delivered QoS in terms of end-to-end delays and offered capacity. Furthermore, the introduction of wireless technology in computer networks offers low cost of installation and relocation, as well as user and terminal mobility. For these reasons ATM WLAN emerges as a key technology in the area of wireless multimedia communications and this is proven by the research taking place in this area.

The design solutions for ATM WLAN must follow the ATM concept and be according to the characteristics of ATM networks. Moreover, to achieve compatibility between fixed and wireless terminals the ATM standardised protocols must be incorporated in ATM WLAN, too. The objective of the work presented in this thesis was to define the system architecture for ATM WLAN and propose its Protocol Reference Model. To do so, new components have to be added in the PRM of ATM networks, or existing ones need to be modified. The introduction of radio as transmission medium in ATM WLAN implies the
Chapter 7: Conclusion and Further Research

definition of a multiple access scheme in the PRM of ATM WLAN. In order to decide on such a technique three factors must be taken into account:

1. the mechanisms for traffic handling that characterise ATM networks
2. the adopted system architecture of ATM WLAN
3. the mechanisms for multiple access adopted in existing wireless standards like HIPERLAN and DECT.

Moreover, to support mobility, the signalling related to hand-off execution, location update and in-call connection re-routing must be introduced in ATM PRM. The study of existing protocols at the User-to-Network Interface and Network-to-Network Interface showed that it is possible to exploit and extend them to support the above mentioned functions related to mobility.

There are certain characteristics identifying ATM technology and driving the design of ATM WLAN:

- Point-to-point physical connections among network entities, with physically separated forward and backward directions of transmission. The entities forming the ATM network, are divided into user part (terminals) and network part (ATM switches). Always, the signalling related to call control, takes place between an ATM terminal and an ATM switch, or between ATM switches. End-to-end signalling between terminals, exists only above the ATM Adaptation Layer (AAL) of the Broadband Integrated Services Digital Network (B-ISDN) PRM only after a virtual connection has been established.

- The call set-up phase, between a terminal and the network, for every virtual connection in the system. Then, a traffic contract for that connection is defined, as well as the route that will be followed by all the packets unless a fault condition occurs. From this perspective ATM networks use the ‘circuit switching’ concept. However, ATM networks also make use of the ‘packet switching’ scheme, because the information is carried by packets with fixed size and format (ATM cells).

- Dynamic (or on demand) bandwidth allocation, based on the instantaneous traffic requirement of each connection. Traffic monitoring functions are also applied per virtual circuit or per virtual path to prevent network congestion and bandwidth violation by connections that do not conform to the agreed traffic contract.

Based on the above mentioned characteristics, and the requirements imposed by the selection of radio as the transmission medium, the design objectives for the ATM WLAN were set:
Chapter 7: Conclusion and Further Research

- A system architecture that integrates wired and wireless transmission media (fixed and mobile access), within the same infrastructure. The deployment of wireless and mobile access should follow the ATM characteristic of dividing the protocol functions into user and network part.

- Definition of Multiple Access Control (MAC) that supports dynamic bandwidth allocation with traffic policing functions to prevent congestion of the air-interface. Furthermore, this protocol should provide for hand-off execution, enhance the operations of the PHY (by providing information for channel monitoring periodically), and apply error control to decrease further the BER achieved by PHY layer operations.

- Modifications on the signalling protocols of UNI and PNNI to support mobility management and handoff.

However, the design decisions were influenced by the study of existing standards. The two European standards for indoor wireless communications, namely DECT and HIPERLAN, were selected. Moreover, it was examined how these standards can be adapted to support ATM traffic.

In DECT a new Data Link Control (DLC) User Plane (U-plane) service was defined that deals with ATM cell reordering and channel coding. Moreover, the interface between the Lower Layer Management Entity (LLME) and the ATM layer was specified to provide for connection suspension and resumption based on the instantaneous traffic in the ATM connection.

In HIPERLAN, the PRM and multiple access scheme were extended to adopt to a centralised topology where a base station or access point schedules the transmissions of wireless terminals. During the HIPERLAN study it was examined if a TDMA based multiple access scheme would outperform HIPERLAN's contention based channel access technique. It was concluded that the TDMA solution offers higher capacity, stability for high traffic loads while HIPERLAN's technique offers lower delays for low traffic loads. However, it was proven that in HIPERLAN there is no QoS provision and that connection bandwidth violations could occur. Moreover, the Low Bit rate part of the transmitted packet, the long training sequence for channel estimation and the acknowledgement that follows every unicast packet contribute significantly to the reduction of the transmission efficiency. Thus, the modifications of the multiple access scheme were followed by an efficient design of the transmitted Protocol Data Units for HIPERLAN based ATM WLAN.

The characteristics of ATM networks and the advantages of centralised topologies for wireless networks compared to peer-to-peer schemes drove the decision for an ATM
WLAN system with a star topology. Adaptive or Dynamic TDMA was selected as the multiple access scheme because CDMA requires very wide bandwidth to spread the high frequency (because of high bit rate) signal that carries ATM traffic, and static TDMA or FDMA cannot provide bandwidth allocation on demand. Moreover CSMA based techniques are characterised by high overheads for ATM cell transmission and non upper bounded delays. The initial ATDMA scheme proved that it is possible to multiplex in the radio different types of services with different bandwidth requirements and traffic patterns.

Moreover, influenced by the capability of HIPERLAN to form ad-hoc networks, a dual mode PRM for ATM WLAN was proposed to enable the establishment of ad-hoc ATM WLANs. The HIPERLAN channel access scheme applies in ad-hoc ATM WLAN. However the transport capabilities of the latter will be reduced compared to the star based ATM WLAN configuration.

The air-interface of ATM WLAN was designed in a way to achieve multiplexing of different ATM traffic classes, maximisation of efficiency by allocating bandwidth only to active connections, and bandwidth allocation on demand. For this reason, short time slots (sub-slots) are defined in the air-interface and accessed by terminals to send MAC control messages to the BS. These messages are transmitted in polling or slotted ALOHA mode and inform the BS about terminals that perform hand-off or registration, sources being activated, or sources changing their bit rates. These messages are used by the BS to assign packet access priorities and schedule all the packet transmissions (both uplink and downlink).

Moreover, traffic policing functions characterise ATM networks and their concept was adopted in the multiple access control of ATM WLAN, too. Then ATM connections receive guaranteed bandwidth while they conform to their traffic contract, and consequently they can satisfy the QoS of the corresponding applications. Furthermore, congestion on the air-interface is avoided.

In order to support Internet applications ATM WLAN has to support TCP in the most possible efficient way. For this reason the solutions adopted in fixed ATM networks to increase TCP throughput are also incorporated in ATM WLAN. Moreover, the cell retransmission scheme adopted in the ATDMA based air-interface enhances TCP operation and contributes to TCP throughput increase.

The capability of ATM terminals to initiate more than one connections simultaneously differentiates radio hand-off in ATM WLAN from that in existing cellular systems. In a network configuration where a BS controls more than one carriers, intra-cell hand-off could happen because of bandwidth limitation in the carrier already used by a terminal. Still user mobility necessitates for inter-cell hand-off. The OAM functions defined
in ATM were extended in ATM WLAN to support the signalling required for both cases of hand-off.

Moreover, the signalling protocol defined at the User-to-Network Interface (UNI) was extended to provide for address registration of mobile terminals and initiate mobility management functions and connection re-routing in the core ATM network. To do so, the standard signalling protocol of ATM networks (P-NNI) was extended. It was shown how the address advertising mechanism and the routing organisation of P-NNI can be exploited to support network hand-off and location update by introducing the concept of Location Management Information Base (LMIB) and LMIB Server.

7.2 Further Research Issues

Significant research has to be carried out in the area of the Physical layer for ATM WLAN. Orthogonal Frequency Division Multiplexing (OFDM), seems to be a promising technique for ATM WLAN since it is characterised by spectral efficiency, reduced computational complexity and results to shorter transmission overheads compared to equalised single carrier solutions. The Physical layer solution has to exploit the ATM WLAN air-interface structure to obtain the required channel estimation and adjust the number of carriers and the code rate in order to achieve low BERs. Furthermore, the performance metrics of the proposed ATDMA technique over OFDM have to be measured.

Moreover, it was shown that although the introduction of EPD improves the throughput performance of TCP, still it cannot avoid congestion in the air interface of ATM WLAN and the total throughput decreases as the total offered traffic increases. However, by combining TCP with the traffic policing functions defined in ATM WLAN MAC congestion should be avoided. Then it has to be examined what the improvement of TCP performance is, and if EPD is still required.

Furthermore, considering mobility management, it was mentioned that during the in-call re-routing procedure, the User Terminal uses the Old Access Switch to carry WATM traffic, until the Call Release Complete message reaches the Old Access Switch. Then, if the Call Re-route message arrives to New Access Switch earlier, the re-routed traffic has to be buffered in the New Access Switch. This could affect the QoS of the re-routed connections and it must be further investigated. By letting a User Terminal to switch to New Access Switch before the Call Release Complete message reaches the Old Access Switch, there is no need for buffering, but a portion of the traffic going through Old Access Switch could be lost. Although the transport layer (AAL) of the WATM network should deal with the lost portion
of traffic, loss or retransmission of packets could affect the QoS and must be measured or calculated.

Furthermore, the in-call re-routing of a connection assumes that bandwidth is available for that reason. However, user mobility affects the Call Admission Control in an area covered by multiple Base Stations, and must be considered when allocating resources of both access and backbone part of the network.
References


151
[GSM0902] ETSI/European Telecommunication Standards Institute, *GSM Specification 09.02*.


[IEEE] IEEE 802.11 specs.


