Wireless ATM

Nikos Passas and Lazaros Merakos

Department of Informatics & Telecommunications
University of Athens
Panepistimiopolis
15784 Athens, Greece
Tel: +30 1 7275327
Fax: +30 1 7275601
Email: {passas,merakos}@di.uoa.gr

1. Introduction

Broadband and mobile communications are presently the two major drivers in the telecommunications industry. Asynchronous Transfer Mode (ATM) is considered the most suitable transport technique for the Broadband Integrated Services Digital Network (B-ISDN), due to its ability to flexibly support a wide range of services with Quality-of-Service (QoS) guarantees. These services are categorized in five classes according to their traffic generation rate pattern: Constant Bit Rate (CBR), real-time Variable Bit Rate (rt-VBR), non-real-time Variable Bit Rate (nrt-VBR), Available Bit Rate (ABR) and Unspecified Bit Rate (UBR). On the other hand, wireless communications are enjoying a large growth in the last decade. Wireless Local Area Networks (LANs) in particular, are becoming popular for indoor data communications because of their tetherless feature and increasing transmission speed. The combination of wireless communications and ATM, referred to as *"wireless ATM"*, aims at providing freedom of mobility with service advantages and QoS guarantees.

Wireless ATM is mainly considered for wireless access to a fixed ATM network; in this sense, it is mostly applicable to wireless LANs. A typical wireless ATM network (Figure 1) includes the following main components:

- *Mobile Terminals* (MTs), the end user equipment, which are basically ATM terminals with a radio adapter card for the air interface,
- *Access Points* (APs), the base stations of the cellular environment, which the MTs access to connect to the rest of the network,
- an *ATM Switch* (SW), to support interconnection with the rest of the ATM network, and
- a *Control Station* (CS), attached to the ATM switch, containing mobility specific software, to support mobility related operations, such as handover[1], which are not supported by the ATM switch.

In many proposals, the CS is considered integrated with the ATM switch in one network module, referred to as *"Switch Work Station"* (SWS). Even though this is the most common architecture, other schemes are possible. For example, APs could be equipped with switching and buffering capabilities, as proposed in [1]. This, in principle, could expedite mobility and call control operations, but could also increase the overall cost of the system significantly, since the APs need to be more complicated, implementing the full signaling ATM stack.

The main challenge for wireless ATM is to harmonize the development of broadband wireless systems with B-ISDN/ATM, and offer similar advanced multimedia, multiservice features for the support of time sensitive voice communications, LAN data traffic, video, and desktop multimedia applications to the wireless user. A sensible quality degradation is unavoidable, due to the reduced bandwidth of the wireless channel and the presentation capabilities of the MTs, but the network should be able to guarantee a minimum acceptable quality. Towards this direction, there are several problems to be faced, mainly because of the incompatibilities of the ATM protocol and the wireless channel. First of all, ATM was originally designed for reliable, point-to-point optical fiber links. On the contrary, the wireless channel is a multiple access channel that suffers from high, time varying, bit error rates, mainly due to fading and interference. This leads to the need for advanced multiple access control and error control mechanisms, for the efficient and reliable sharing of the scarce available bandwidth of the wireless channel, among different kinds of connections. Second, ATM was also designed for large bandwidth environments, following a bandwidth consuming policy to attain simplicity and fast switching of data packets. This leads to a packet header (ATM cell header, in the ATM terminology), which consumes approximately 10% of the available bandwidth (5 out of 53 bytes). For gigabit-per-second optical fibers used in B-ISDN, this is not considered a drawback, compared to fast switching and packet delivery. But for a wireless channel of tens of megabits per second this can be vital for the overall performance. As shown later in this article, the usual practice is to perform header compression, to reduce overhead as much as possible. Finally, ATM signaling enhancements are definitely an important subject for wireless ATM, mainly for mobility. To support it, several additional functions and signaling need to be added in traditional ATM, for registration, location update, handover, etc. Particularly for handover, the comparatively high transmission speed,

[1] Mobility issues will be explained in detail later in this article

combined with the requirements of some real-time applications (e.g., videoconference), ask for fast and efficient handover techniques. All these mobility-related functions are usually implemented in the CS shown in Figure 1 to leave the conventional ATM switches intact. Mobility issues are discussed in detail later in this article. Additionally, some standard call control procedures of fixed ATM need also to be enhanced to cover the particularities of the wireless channel. Especially connection setup requires advanced call admission control algorithms that consider the instabilities of the wireless channel.

The rest of the article is organized in two main sections. Section 2 describes the basic issues and solutions for the medium access control, concluding with the most important standards. Some important protocols are discussed, and their effectiveness in servicing ATM traffic is analyzed. In section 3, the required signaling enhancements for call and mobility control are discussed. The section starts with a basic signaling architecture, and continues with connection setup (especially call admission control), and handover. Finally, section 4 contains our conclusions.

2. Medium Access Control (MAC)

2.1. MAC Protocol Structure

In wireless ATM networks, an advanced MAC protocol is required, able to provide adequate support to the traffic classes defined by ATM standards, together with efficient use of the scarce radio bandwidth. Additionally, this protocol should be adaptive to frequent variations of channel quality.

MAC protocols can be grouped, in general, into five classes [2]: i) fixed assignment, ii) random access, iii) centrally controlled demand assignment, iv) demand assignment with distributed control, and v) adaptive strategies. Fixed assignment techniques permanently reserve one constant capacity subchannel for each connection for its whole duration and they perform very well with constant bit rate connections, both in terms of service quality and channel efficiency. However, their performance decreases dramatically when they are asked to support many infrequent users with variable rate connections. In such cases, random access protocols usually perform better. A typical example of such a protocol is ALOHA, which permits users to transmit at will; whenever a collision occurs, collided packets are retransmitted after some random delay. It is well known that, although ALOHA-type protocols are easy to implement and attain minimum delays under light load, they suffer from long delays and instability under heavy traffic load. Enhancements of ALOHA include collision resolution techniques that increase the maximum achievable stable throughput. Centrally controlled demand assignment protocols reserve a variable portion of bandwidth for each connection, adjustable to its needs. Unlike random access techniques, these protocols operate in two phases: reservation and transmission. In the reservation phase, the user requests from the system the portion of bandwidth required for its transmission needs, and the system responds by reserving the bandwidth and informing the user, while in the second phase the actual transmission takes place. Demand assignment protocols are usually complex, but are also stable and perform well under a wide range of conditions, although the reservation phase results in time and bandwidth consumption. With distributed control, the users themselves schedule their transmissions, based on broadcast information. Finally, adaptive schemes combine elements from the above techniques, and aim at supporting many different types of traffic [3].

Concerning the multiple access technique, the proposed protocols for the radio interface of wireless ATM networks are in general based on Frequency Division Multiple Access (FDMA), or Code Division Multiple Access (CDMA), or Time Division Multiple Access (TDMA), or combinations of these techniques. The scarcity of available frequencies, and the requirement for dynamic bandwidth allocation, especially for variable bit rate connections make the use of FDMA inefficient. On the other hand, CDMA limits the peak bit rate of a connection to a relatively low value, which is a problem for broadband applications (>2Mbps). Accordingly, most of the proposed protocols use an adaptive TDMA scheme, due to its ability to flexibly accommodate a connection’s bit rate needs, by allocating a variable number of time slots, depending on current traffic conditions.

Beyond this general choice of a TDMA-based scheme, the MAC protocols proposed in the literature differ in the technique used to build the required adaptivity in the TDMA scheme. The three main techniques used, alone or in combinations, are *contention*, *reservation*, and *polling*.

Contention-based random access protocols are simple and require minimal scheduling. An example is the slotted ALOHA with exponential back-off protocol presented in [4]. Functionality that can be omitted from the MAC layer, such as handover and wireless call admission control, is pushed to the upper layers. These protocols, attain good delay performance under light traffic, and fit well with the statistical multiplexing philosophy of ATM. Nevertheless, their performance is questionable under heavy traffic conditions, or when multiple traffic classes must be supported with guaranteed QoS.

Another group of protocols uses reservation techniques, mainly through reservation/allocation cycles, to dynamically allocate the available bandwidth to connections, based on their current needs and traffic load. A well-designed representative protocol of this group can be found in [5]. It is a TDMA Time Division Duplex (TDD) protocol, where time is divided in constant length frames and every frame is subdivided in a request subframe and a data subframe. The request subframe is accessed by MTs, through a simple slotted-ALOHA protocol, in order to declare their transmission needs, while the data subframe is used for user data transmission. The allocation of data slots is performed by the AP, based on a scheduling algorithm, and the MTs are informed through broadcast messages. This kind of protocols are more complex and introduce some extra delays, due to the required reservation phase, but on the other hand, they are stable under a wide range of traffic loads and can guarantee a predictable quality of service, which is very important in wireless ATM networks. Their performance depends to a large extent on the scheduling mechanism used for the allocation of the available bandwidth. A number of scheduling algorithms has been proposed in the literature, which try to separate real-time and non-real-time connections. For example, a minimum bandwidth can be allocated to non-real-time connections, while real-time connections are served as soon as possible. In [6], a delay-oriented scheduling algorithm, referred to as PRADOS, is proposed to meet the requirements of the various traffic classes defined by the ATM architecture. In order for PRADOS to maximize the fraction of ATM cells that are transmitted before their deadlines, each ATM cell is initially scheduled for transmission as close to its deadline as possible. After that, a packetization process ensures that no time slots will be left empty.

A third group of protocols uses adaptive polling to distribute bandwidth among connections (e.g., [7]). A slot is given periodically to each connection, without request, based on its expected traffic. Compared to reservation-based protocols, these protocols are simpler, since there is no reservation phase, but their performance depends on the algorithm that determines the polling period for each connection. If the polling period is shorter than needed, then such protocols might suffer from low utilization, since many slots will be empty. On the other hand, if the polling period is longer than needed, they result in increased delays and poor QoS. The problem becomes more difficult for variable-bit-rate bursty connections. Several proposals suggest an adaptive algorithm to decide on the polling period of each connection, based on total traffic load, expected traffic for each connection, and required QoS [7].

Finally, to improve performance, a combination of the above schemes is possible; for example, a protocol that is based mainly on reservation, but has also a random-access part for urgent traffic. A typical representative of this category is MASCARA [8]. The multiple access technique used in MASCARA for uplink (from the MTs to the AP of their cell) and downlink (from the AP to its MTs) is based on TDMA/TDD, where a time slot is equal to the time required to transmit an ATM cell. The MASCARA time frame is divided into a DOWN period for downlink data traffic, an UP period for uplink data traffic, and an uplink CONTENTION period used for MASCARA control information. Each of the three periods has a variable length, depending on the traffic to be carried on the wireless channel. The AP schedules the transmission of its uplink and downlink traffic and allocates bandwidth dynamically, based on traffic characteristics and QoS requirements, as well as the current bandwidth needs of all connections. The current needs of an uplink connection from a specific MT are sent to the AP through MT “reservation requests”, which are either piggybacked in the data MPDUs the MT sends in the UP period, or contained in special “control MPDUs” sent for that purpose in the CONTENTION period. Protocols belonging to the same category can be found in [5, 9].

To minimize overhead added by the ATM header, header compression techniques can be used. A straightforward solution is the replacement of the 3 bytes long VPI/VCI, used for addressing in ATM, with a shorter MAC specific identifier (MAC_ID), whose length is at most 1 byte, depending on the environment. The MAC_ID is used only for wireless channel transmission, and after this it is replaced with the original VPI/VCI.

2.2. Error Control

In wireless ATM, fulfilling the strict QoS requirements of ATM over an unreliable wireless channel is a challenging problem, and error control is very important. The error control mechanisms used can be thought of as belonging to a sublayer of the MAC layer (usually the upper part), referred to as Wireless Data Link Control (WDLC) sublayer. WDLC is responsible for recovering from occasional quality degradations of the wireless channel, and for providing an interface to the ATM layer in terms of frame format and required QoS.

Error control techniques, in general, can be divided in two main categories, namely Automatic Repeat reQuest (ARQ) and Forward Error Correction (FEC). In ARQ techniques, the receiver detects the erroneously received data and requests retransmission from the transmitter. Since retransmissions imply increased delays, ARQ is efficient for non-real-time data. ARQ techniques are conceptually simple and provide high system reliability at the expense of some extra delay and bandwidth consumption due to retransmissions. FEC, on the other hand, is efficient for real-time data. A number of bits is added in every transmitted data unit, using a predetermined error-correction code, which allows the receiver to detect and correct errors up to a predetermined number per data unit, without requesting any additional information from the transmitter. It is clear that FEC techniques are fast at the expense of lower bandwidth utilization due to the transmission of additional bits.

In wireless ATM, where both real-time and non-real-time data must be supported, a hybrid scheme combining ARQ and FEC is usually used. According to this, for real-time connections (e.g., CBR, rt-VBR) FEC bits are included in the header of every MAC data unit, to allow the receiver (AP or MT) to correct most of the errors. For non-real-time connections (e.g., nrt-VBR), no extra bits are included, and the AP (MT in the downlink) requests from the MT (AP in the downlink) the retransmission of erroneously transmitted MAC data units.

2.3. MAC Standards

Currently, the MAC technology for wireless ATM is served mainly by two standards, both based on TDMA. The 802.11 standard [10], developed by the IEEE 802 LAN standards organization, and the HIgh PErformance Radio LAN Type 2 (HIPERLAN/2) [11], defined by the European Telecommunications Standards Institute (ETSI) RES-10 Group. Although both standards were mainly designed for conventional LAN traffic, they can definitely serve as a medium for passing ATM traffic, with the proper QoS guarantees. Here we focus more on HIPERLAN/2 because it provides more flexibility for ATM traffic. IEEE 802.11 operates at 2.4GHz and considers data traffic up to 2Mbps. The medium can alternate between a contention mode, known as the *contention period* (CP), and a contention-free mode, based on polling, known as the *contention-free period* (CFP). 802.11 supports three different kinds of frames: management, control, and data. A management frame is used for MT association/de-association, timing, synchronization, and authentication/de-authentication. A control frame is used for handshaking and positive acknowledgements during a CP, and to end a CFP. Finally, a data frame is used for transmission of data during a CP or CFP. On the horizon there is the need for higher data rates, for applications requiring wireless connectivity at 10Mbps and higher. This will allow 802.11 to match the data rates of most wired LANs. There is no current definition of the characteristics for the higher data rate signal. However, for many of the options available to achieve it, there is a clear upgrade path for maintaining interoperability with 2 Mbps systems, while providing higher data rates as well.

HIPERLAN/2 systems on the other hand, operate at the 5.2 GHz unlicensed band and attain transmission rates ranging from 6 Mbits/sec to 54 Mbits/sec (with a typical value being 25 Mbits/sec).

In that sense, it serves better the desired transmission speed for ATM applications. The MAC protocol of HIPERLAN/2 is based on a TDMA/TDD scheme. Time is divided in MAC frames, which are further divided in time slots. Time slots are allocated to the connections dynamically and adaptively depending on the current needs of each connection. Slot allocation is performed by a MAC scheduler which takes into account QoS requirements of each connection. A MAC scheduling algorithm has not yet been specified by the HIPERLAN/2 standards. An efficient algorithm that will be able to meet the requirements of different connections should be developed. The duration of each MAC frame is fixed to 2ms. Each frame comprises transport channels for broadcast control, frame control, access control, downlink and uplink data transmission and random access. All data between the AP and the MTs is transmitted in the dedicated time slots, except for the random access channel where contention for the same time slot is allowed. The length of the broadcast control field is fixed, while the length of the other field may vary according to the current traffic needs.

HIPERLAN/2 error control entity supports 3 different modes of operation: Acknowledged mode, Repetition Mode, and Unacknowledged mode. **Acknowledged mode** provides for reliable transmissions using retransmissions to compensate for the poor link quality. The retransmissions are based on acknowledgements from the receiver. The ARQ protocol that is used is Selective Repeat (SR) allowing various transmission window sizes to be used depending on the requirements of each connection. In order to support QoS for delay critical applications (e.g., voice, real-time video), error control may also utilize a discard mechanism for discarding data units that have exceeded their lifetime. **Repetition mode** provides for reliable transmission by repeating data units. In repetition mode, the transmitter transmits new data units consecutively, and is allowed to make arbitrary repetitions of each data unit. No feedback is provided by the receiver. Finally, **Unacknowledged mode** provides for unreliable, low latency transmissions. In unacknowledged mode, data flows only from the transmitter to the receiver. No ARQ retransmission control or discard messages are supported.

From the above short description of the two standards, it is clear that HIPERLAN/2 provides more alternatives to better satisfy the requirements of different ATM connections. Nevertheless, we should notice that, mainly due to increased complexity, HIPERLAN/2 products are not yet available in the market, while there is a wide range of 802.11 equipment from a number of vendors.

3. Signaling Enhancements

Terminal mobility in wireless ATM requires a number of additional operations not supported in fixed ATM networks. These operations include:

Registration/De-registration: When a MT is switched on, it needs to inform the network and be accepted by it to be able to send/receive calls. This operation is called registration. An important part of registration is authentication, where the MT is recognized as authentic and it is permitted to continue registering. The opposite operation to registration, when the MT is switched off, is called de-registration, and informs the network that the MT is no longer available.

Location update: When a MT has no active connections, it is practically untraceable by the network. So a passive operation is required, in which the system periodically records the current location of the MT in some database that it maintains, in order to be able to forward an incoming connection, when a new connection setup request arrives.

Handover (also referred to as "handoff"): It is the operation that allows a connection in progress to continue as the MT changes channels in the same cell, or moves between cells. In a multi-channel system, handovers within the same cell, where the connections are transferred to new radio channels, are referred to as *"intra-cell handovers"*. The case where the MT's connections are transferred to an adjacent cell is referred to as an *"inter-cell handover"*. One of the key issues in wireless ATM is maintaining the QoS of different connections during a handover.

Connection setup: Standard protocols for connection setup in fixed ATM networks assume that the terminal's address implicitly identifies its attachment point to the network. However, this is not the case in wireless ATM. Thus, the ATM connection setup protocols must be augmented to dynamically resolve a MT endpoint location. Additionally, connection admission control (CAC), as part of the connection setup process, is much more difficult in wireless ATM. This is because the wireless channel quality varies in time, due to temporary interference or fading, so the available resources are not fixed. A proposal for wireless ATM CAC can be found in [12].

Registration/de-registration and location update solutions are more or less generic and do not have extra requirements in a wireless ATM environment. Below we elaborate on connection setup and handover, and analyze their requirements and constraints, starting with the signaling architecture.

3.1. Signaling Architecture

Current trends in designing the access network (AN) part of fixed B-ISDN aim at concentrating the traffic of a number of different User Network Interfaces (UNIs) and routing this traffic to the appropriate Service Node (SN) through a broadband V interface (referred to as VB). The main objective in AN design is to provide cost effective implementations without degrading the agreed QoS, while achieving high utilization of network resources. This is reflected in both the reduction of the AN physical equipment and in the limitations imposed on the AN functionality, such as the inability to interpret the full ATM layer control information and signaling. The use of only low-level operations in AN forces the establishment of several internal mechanisms that are used to unambiguously identify the connection an ATM packet belongs to, and to convey only those connection parameters that are absolutely necessary for traffic handling.

In this framework, a fast control protocol running over a universal VB interface can be introduced [13], which serves a number of AN internal functions while preserving the highest possible degree of transparency at the SN. The protocol is based on the Local exchange Access network Interaction Protocol (LAIP), which was developed to accommodate the SN-to-AN communication requirements, as identified in the early study and design of the dynamic $VB_{5.2}$ interface, i.e., the interface between the fixed ATM AN and the SN. In the relevant standardization bodies, the presence of such a protocol has been firmly decided and has been given the name Broadband Bearer Channel Control Protocol (B-BCCP). The services of the $VB_{5.2}$ control protocol enable the dynamic AN operation by conveying the necessary connection-related parameters required for dynamic resource allocation, traffic policing, and routing in the AN, as well as information on the status of the AN before a new connection is accepted by the SN.

The signaling access architecture for wireless ATM considered here is an extension of the broadband V interface, where an enhanced version of the $VB_{5.2}$ control protocol is used to enable the dynamic operation of the AN and to serve the AN internal functions. It is assumed that a mobility-enhanced version of the existing B-ISDN UNI Call Control (CC) signaling is employed to provide the basic call control function and to support the handover-related functions. In addition, pure ATM signaling access techniques, based on Metasignaling, are adopted for the unique identification and control of signaling channels. These features allow us to minimize the changes required to the signaling infrastructure used in the wired network, and, in this respect, they can guarantee the integration of the wireless ATM access system with fixed B-ISDN. However, when striving for full integration, the mobile-specific requirements imposed by the radio access part need to be taken into account.

In today's wired ATM environment, the user-network interface is a fixed port that remains stationary throughout the lifetime of a connection. The current B-ISDN UNI protocol stack uses a single protocol over fixed point-to-point or point-to-multipoint interfaces. On the other hand, in wireless ATM, mobility causes the user access point to the wired network to change constantly, and the mobile terminal connections must be transferred from access point to access point, through a handover process. The support of the handover functionality assumes that the fixed network of the access part has the capability to dynamically set-up and release bearer connections during the call. A well-accepted methodology to support these features is the call and bearer separation at the UNI. The use of the extended $VB_{5.2}$ interface control protocol for wireless ATM access systems serves for the set-up and reconfiguration of fixed bearer connections of the same call, supporting in this way the call and bearer control separation in the AN part.

Based on the above terminology, the following types of signaling interaction for the communication of peer entities can be identified [14]:

• ***Mobile Call Control Signaling (MCCS)***: This includes an enhanced B-ISDN call control signaling protocol (denoted as Q.2931*), based on the ITU recommendation Q.2931, for the set-up, modification, and release of calls between the MT and the CS. The enhancements required in the current signaling standards are related to the support of the handover function (e.g., inclusion of handover-specific messages).

• ***Mobility Management Signaling (MMS)***: This is responsible for the MT registration/authentication and tracking procedures.

• ***Bearer Channel Control Signaling (BCCS)***: This serves for providing the traffic parameters to the AP, and handles the establishment, modification/reconfiguration, and release of fixed ATM connections between the AP and the CS.

• ***Radio Channel Control Signaling (RCCS)***: This deals with low-level signaling related to the radio interface consisting of messages between the MT and the AP (MAC and physical layer specific messages).

At the user plane, the MT has a typical ATM protocol stack on top of a radio specific physical layer and a MAC layer. The AP acts as a simple interworking unit that extracts the encapsulated ATM cells from the MAC frame, and forwards them to the CS through a proper ATM virtual connection. The MAC functionality realized at the AP is based on a MAC scheduler, which, on the basis of the ATM connection characteristics declared at connection set-up and current transmission requests, allocates the radio bandwidth according to the declared QoS requirements and service type of each connection. As already mentioned, such a mechanism provides a degree of transparency to a subset of broadband/ATM services, and achieves efficient sharing of the scarce radio bandwidth among the mobile users. The CS realizes the typical B-ISDN protocol functionality of the U-plane.

3.2. Connection Setup

Connection setup procedures used in traditional ATM networks assume i) reliable gigabit links with fixed capacity, and ii) stationary users. Accordingly, CAC algorithms do not need to be constantly informed about the available resources and the users' attachment points. But this is not the case in wireless ATM. The wireless channel impairments and MAC layer overheads can result in lower bandwidth than the theoretically available, while a mobile users' attachment point with the network can change anytime. Below we describe typical connection setup scenarios in wireless ATM, focusing on the differences with fixed ATM.

When a MT initiates a new call, its signaling channel transparently conveys a standard connection SETUP_REQUEST signaling message to the CS. Upon receipt of this request, the CS identifies the calling MT and the called terminal, and contacts the location server to track the location of the calling MT and the called terminal (if it is mobile). An initial connection acceptance decision is made, based on the user service profile data and on the QoS requirements set by the MT.

In case the request is accepted, the AP of the calling MT should be notified by the CS (using BCCS) on the expected new traffic so that it can decide on the admission in the wireless channel, and allocate radio resources accordingly. To this end, the traffic parameters of the new connection, or at least a useful subset of them, should be communicated to the AP of the calling MT. This information gives the opportunity of exercising a policing functionality at the AP, implemented implicitly by its radio bandwidth allocator. It also protects the CS from the unlikely case where, although the CS expects availability of radio resources, these are exhausted due to additional overheads of the MAC layer, or a temporary reduction in radio link quality. The latter is useful in case the CAC of the CS does not take into account issues specific to the wireless access. Since the final CAC decision is taken at the CS, it is possible to implement a connection acceptance algorithm customized to the specific wireless access system. Traffic characteristics will appear at the AP together with the QoS requirements, declared as the class of service the specific connection will support. This gives the ability to the MAC to implement a set of priorities according to the connection an ATM cell belongs to. To be able to recognize the particular connection class, it is necessary to declare also the VPI/VCI values that will be used.

The task of the AP-CS communication and bearer channel establishment in the fixed access network is very important in this case. An ALLOC message is generated and forwarded to the AP, through

BCCS. The AP will reply with an ALLOC_COMPLETE or an ALLOC_REJECT message indicating whether it agrees or not with the CAC decision. The latter implies that the call is rejected at the AP. Upon receipt of an ALLOC_COMPLETE, the CS returns a CALL_PROCEEDING message to the calling MT and initiates the connection establishment procedures towards the core network (B-ISUP IAM message) if the called terminal is a fixed one.

In case the called terminal is another MT (i.e., intra-CS call), the call processing module of the CS forwards the setup request towards the AP of the called terminal, where similar functions to those described above take place. The signal exchanges for this case are shown in Figure 2. In the fixed-to-MT (incoming) connection set-up scenario, the CS receives an incoming SETUP_REQUEST message, identifies the called MT, tracks its location, draws an initial CAC decision, and asks the corresponding AP of the called MT [14].

In all cases, the ALLOC message transfers to the AP all the connection-related information required for the AP operation. This includes the bandwidth requested by the connection, the service class, the QoS parameter values, etc. An improvement, in the case where the requested bandwidth or the QoS cannot be supported by the radio part of the communication path, is for the AP to generate an ALLOC_MODIFY message indicating this situation and suggesting a QoS degradation needed for the connection to be accepted. This useful “fallback” mechanism intends to set-up connections with the highest available bandwidth. However, such a capability is useless if the standard ATM signaling does not support QoS negotiation to let the CS and the MT negotiate the new situation. In all scenarios, we have implicitly assumed that MTs remain stationary at connection set-up. If we assume that a MT may move during connection set-up, the set-up might not succeed. In this case, the new location of MT is determined and another set-up should be attempted following the same procedures. The calling or called party can initiate the release of a call. Upon receipt of a RELEASE message, the CS releases all the resources associated with that call and triggers the release towards the AP, the core network, or the MT.

3.3. Handover

Among mobile specific operations, handover is probably the most difficult to perform, due to the diversity of requirements of different kinds of connections, and the constraints imposed by the wireless channel. In any case, an unavoidable period of time is required, during which the end-to-end connection data path is incomplete. This means that some data might get lost or should be buffered for later delivery. The effect that this increase of losses or delays has on each application depends on the nature of the application and the duration of the disruption. Current proposed protocols for handover in wireless ATM may be grouped in four categories [15]:

Full connection rerouting: It is the simplest kind of handover, where the system establishes a completely new end-to-end route for each handover – as if it was a new connection. Clearly, this kind of handover is simple in terms of implementation, but can result in unacceptable delays and losses, depending on the distance between the two parties.

Route Augmentation: In this case, the original connection is extended with an additional hop to the MT's next location. For users with limited mobility, this solution can result in low delays and very limited or no losses, since no actual rerouting is performed. But if the MT starts changing cells more often, the additional extensions will result in a very long connection path, increasing delays and reducing network utilization.

Partial connection rerouting: This kind of handover attempts to perform a more efficient re-routing, by preserving as much of the old connection path as possible and re-routing the rest. The key issue here is to locate the nearest ATM network node that is common to both the old and the new data path. Then the common node will handle the tearing down of the old part and establishing the new, taking also care of the data that are on the way in the old part, when the switching is performed. Temporary buffering before switching or temporary rerouting after switching can be used to minimize losses. Partial connection rerouting is the most common handover type found in the literature.

Multicast connection rerouting: In this kind of handover, more than one connection paths are maintained at a time, although only one path is operational. When the MT moves to a new cell, data can immediately start flowing towards the new direction. This eliminates the need for establishing a new path during handover (partial or full) and leads to lower delays and losses. On the other hand, since the system cannot maintain a path for every cell a MT can move to, an intelligent algorithm is required to predict the MT's movement and pre-establish paths on the neighboring cells, while at the same time, paths that are no longer needed are canceled. An extension of this kind of handover could also permit the same data to flow in more that one data paths when the MT is at the threshold between two cells (this is also referred to as "macrodiversity"). This allows a MT with multiple receivers (antenna diversity) to get data from more than one APs, and keep only the correctly transmitted information, reducing in this way the bit error rate.

Another categorization in wireless ATM handover is based on who performs what. In general, a handover mechanism involves a continuous procedure of channel measurements, and starts with a handover request initiation. In that sense, there are three fundamentally different categories of handover mechanisms: *network-controlled*, *mobile-assisted*, and *mobile-controlled*. In network-controlled handover, the MT is completely passive. All measurements are performed by the network (basically the AP) and the handover request is initiated by the AP. This is a simple solution, which does not perform well in the case where the signal received by the AP is good, while the signal received by the MT is bad. This weakness is overcome by the mobile-assisted handover, where both the AP and the MT are measuring the strength of the received signal, however the handover request is initiated by the AP. The MT can only send its measurements to the BS in order for it to have a better picture of the situation. Finally, in mobile-controlled handover all measurements and handover requests are executed at the MT. If the handover request is executed via the "old" AP (the AP that the MT is leaving), we have a *"backward handover"*, and if it is executed via the "new" AP, we have a *"forward handover"*. Backward handovers are in general more seamless than forward, so the usual practice is for the MT to prefer backward handover, and, only if this is not possible (in case of an abrupt signal strength reduction), to perform forward. Mobile-controlled handover can operate either alone or in conjunction with network-controlled, or mobile-assisted handover.

No matter what handover algorithm is used, the main target should be to maintain the QoS of active connections, not only during, but also after handover in the new cell. During handover, temporary buffering can be used at the switching point to ensure delivery of loss sensitive data. For delay sensitive data that cannot be buffered, the only solution is to ensure simple and fast handover operation. On the other hand, maintaining QoS in the new cell is not always possible. In fixed ATM, if an efficient CAC algorithm decides that a connection can be accepted with the requested QoS, then the network can guarantee this QoS throughout the duration of the connection. The same cannot be said for a connection to a MT, which can be rerouted when the MT is handed over to a new cell. For example, if this new cell is overcrowded, there might not be enough resources to support the QoS of the connections of the newly arrived MT. In this case, the smallest possible number of the MT's connections should be rejected, to leave enough resources for the rest. This decision is usually taken by the CS, because it has a more global view of the system. A more advanced solution is to re-negotiate the QoS in the new cell in order to avoid connection rejection, as much as possible. In this case, the MT will be asked to reduce its requirements if it wants to maintain its connections in the new cell. Below we describe a simple but typical mobile-controlled handover procedure.

When the MT decides that a handover should be performed, it sends a HANDOVER_REQUEST message towards the CS, transparently via the old AP. This message contains identification of the MT, the call and the target AP. The MT may have multiple active connections at the same time, as multimedia applications are to be supported. If this is the case, during the request for handover the MT could also indicate the priorities of the different connections in case the new AP cannot accommodate all of them.

A fast control protocol between AP and CS, is required for the release/establishment of the old/new bearers in the fixed network part, and for performing possible QoS renegotiations during handover. Upon receipt of the HANDOVER_REQUEST, the CS identifies the MT, and initiates a state machine for the handover. Similar procedures to those described for connection set-up are performed between the CS and the new AP. In this way, the CS informs the target AP about the expected QoS and bandwidth requirements to allocate radio resources accordingly.

When the CS receives the response from the new AP (ALLOC_COMPLETE), it sends a HANDOVER_RESPONSE message to the MT to inform it about the handover results and possible QoS modifications, and reconfigures the ATM connections of the ATM switch towards the new AP. After receiving the HANDOVER_RESPONSE, the MT releases its radio connection with the old AP, and establishes a radio link with the new AP. Special ATM (and lower) layer cell relay functions take place at the MT and the CS to coordinate the switching of traffic, and to guarantee the transport of user data at an agreed QoS level in terms of cell loss, ordering, and delay. Finally, the CS updates the location server about the new location of the MT, and sends a RELEASE message to the old AP to notify it that the connection no longer exists and to de-allocate the corresponding radio resources.

The handover process described above is expected to be fast. In the unlikely case that a MT moves again before the handover is accomplished, handover is again attempted to the current destination AP, until it eventually succeeds. The forward handover scenario is similar to the backward one. The MT releases the old radio connection and communicates directly with the new AP. Since all signaling is passed through this new AP, a dynamic signaling channel allocation scheme is employed, in order for the MT to obtain a signaling channel for passing the messages to the CS.

4. Conclusions

The design of wireless ATM systems to offer ATM services to wireless users has attracted considerable attention during the past few years, and a large number of proposals exists in the literature dealing with specific design issues. The most important of these issues are the medium access control and the signaling enhancements.

Medium access control is much more demanding in wireless ATM than in traditional wireless networks, owing to the, often conflicting, requirements of the various ATM traffic types. The current trend is for flexible, TDMA/TDD protocols with variable time frame, enabled with a sophisticated traffic scheduling algorithm that adjusts the bandwidth given to a connection to its time-varying requirements, without violating the contract made with other active connections.

On the other hand, enhancements are required to standard ATM signaling, to cover issues like wireless call admission control and handover. Wireless call admission control is part of the overall call admission control process, handling available resources in the wireless link. Since the available wireless bandwidth is time varying, due to temporary deterioration of the radio signal, the procedure should always have up to date information on the current status of the radio link. Finally, handover is a completely new issue for wireless ATM signaling. Handover mechanisms should be fast and efficient, in order to minimize losses or delays, which could influence the QoS provided to the user. The usual practice is to introduce a special purpose Control Station, centrally located in the wireless ATM network, which handles all the extra signaling and implements wireless call admission control and handover mechanisms.

As a final comment we can say that the future trends in wireless communications tend to be towards wireless IP based systems with QoS provision (i.e., IPv6), rather than wireless ATM. Nevertheless, the issues and problems are more or less the same, so techniques and mechanisms developed for wireless ATM can, with proper adjustments, be used in wireless IP as well.

5. References

[1] M. Veeraraghavan, M.J. Karol, and K.Y. Eng, Mobility and connection management in a wireless ATM LAN, *IEEE Journal in Sel. Areas in Commun.*, vol. 15, no. 1, pp. 50-68, January 1997.

[2] F.A. Tobagi, *Multiaccess link control*, Computer Network Architectures and Protocols, P.E. Green, Jr., ed., New York: Plenum Press, 1982.

[3] E.Ayanoglu, K.Y. Eng, and M.J. Karol, Wireless ATM: limits, challenges, and proposals, *IEEE Pers. Commun. Magazine*, Vol. 3, No. 4, pp. 18-34, August 1996.

[4] J. Porter and A. Hopper, An ATM-based protocol for wireless LANs, *Olivetti Research Ltd. Tech. Rep. 94.2*, April 1994, available at ftp://ftp.cam-orl.co.uk/pub/docs/ORL/tr.94.2.ps.Z

[5] D. Raychaudhuri, L.J. French, R.J. Siracusa, S.K. Biswas, R. Yuan, P. Narasimhan, and C.A. Johnston, WATMnet: A prototype wireless ATM system for multimedia personal communication, *IEEE Journal on Sel. Areas in Commun.*, vol. 15, no. 1, pp. 83-95, January 1997.

[6] N.Passas and L.Merakos, Traffic scheduling in wireless ATM networks, *Proc. IEEE ATM '97 Workshop*, Lisbon, Portugal, May 1997.

[7] C.-S. Chang, K.-C. Chen, M.-Y. You, and J.-F. Chang, Guaranteed quality-of-service wireless access to ATM networks, *IEEE Journal on Sel. Areas in Commun.*, vol. 15, no. 1, January 1997.

[8] F.Bauchot, N.Passas, G.Marmigere, L.Merakos and S.Decrauza, MASCARA: A MAC protocol for wireless ATM, *Proc. International Conference on Telecommunications*, Melbourne, Australia, April 1997.

[9] F. D. Priscoli, Adaptive parameter computation in a PRMA, TDD based medium access control for ATM wireless networks, *Proc. IEEE GLOBECOM '96*, vol. 3, pp. 1779-1783, London, UK, November 1996.

[10] IEEE 802.11D5, Wireless LAN Medium Access Control (MAC) and PHysical Layer (PHY) specifications, July 1996.

[11] ETSI TR 101 683 (V1.1.1), Broad Radio Access Network (BRAN); HIgh Performance Radio Local Area Networks (HIPERLAN) Type 2; System overview, February 2000.

[12] O.T.W. Yu and V.C.M. Leung, Adaptive resource allocation for prioritized call admission over an ATM-based wireless PCN, *IEEE Journal on Sel. Areas in Commun.*, vol. 15, no. 7, pp. 1208-1225, September 1997.

[13] I.S. Venieris et al., Architectural and control aspects of the multi-host ATM subscriber loop, *Journal of Network and Systems Management*, 5 (1997), pp. 55-71.

[14] N. Loukas, N. Passas, L. Merakos and I. Venieris, A signaling architecture for wireless ATM access networks, ACM Wireless Networks Journal, vol. 6 (2000), pp. 145-159, November/December 2000.

[15] I.F. Akyildiz, J. McMair, J Ho, H. Uzunalioglu, and W. Wang, Mobility management in current and future communication networks, *IEEE Network*, vol. 12, no. 4, pp. 39-49, July/August 1998.

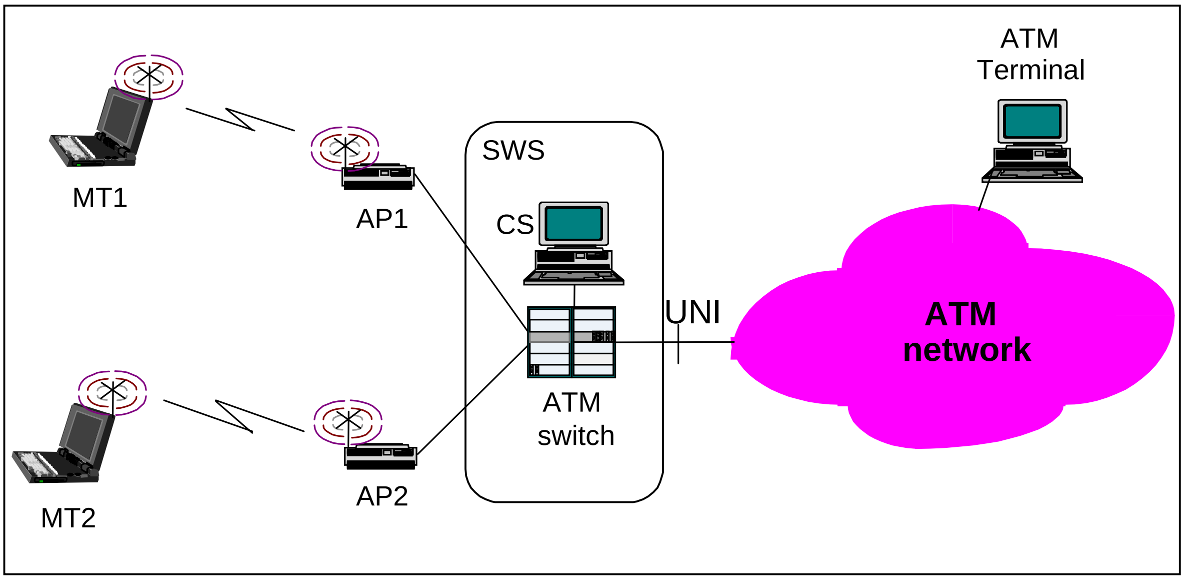

Figure 1: A typical wireless ATM network

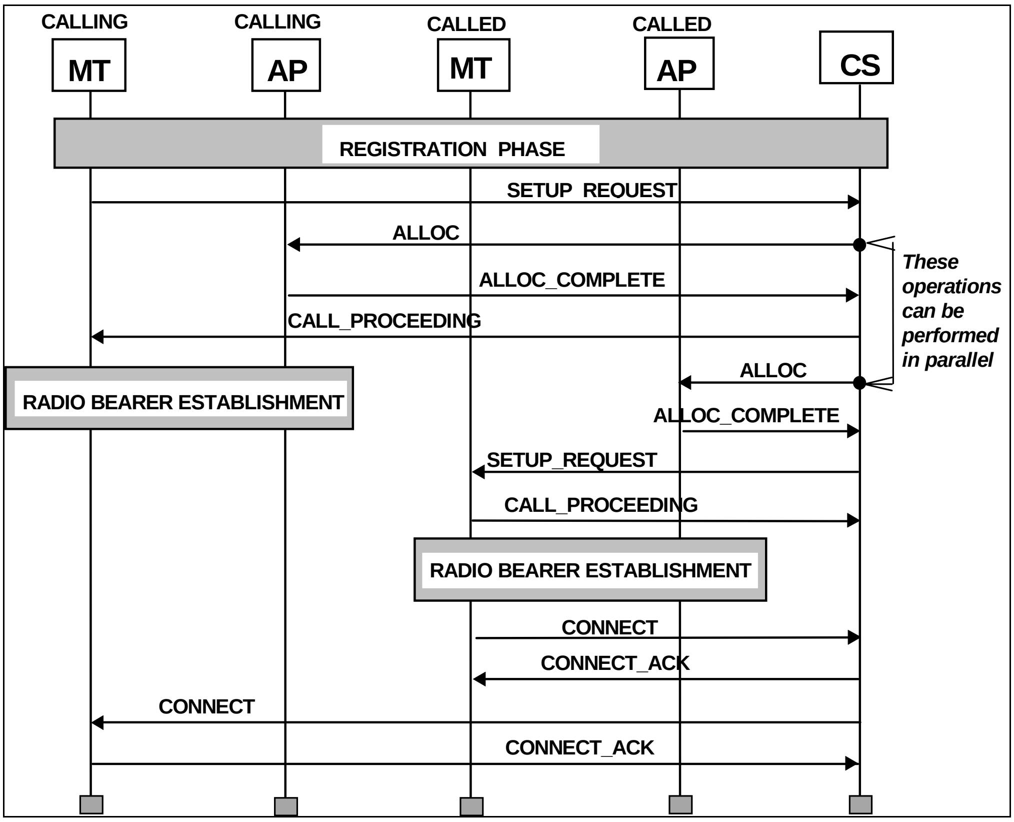

Figure 2: Connection setup procedure between two MTs