Medium Access Control Protocols for Multimedia Traffic in Wireless Networks

Ian F. Akyildiz and Janise McNair, Georgia Institute of Technology
Loren Carrasco Martorell and Ramon Puigjaner, Universitat de les Illes Balears
Yelena Yesha, University of Maryland at Baltimore County

Abstract

This article presents a survey on medium access control protocols for multimedia traffic in wireless networks. A basic overview of MAC protocol concepts is presented, and a framework is developed on which to base qualitative comparisons. The MAC protocols covered in this article include third-generation TDMA and CDMA schemes intended for use in a single-hop wireless system. The operation of each protocol is explained, and its advantages and disadvantages are presented. Finally, a qualitative comparative outline of the discussed protocols is provided, based on multimedia traffic requirements.

One of the driving forces of the next generation of wireless communication and computing networks is the promise of high-speed multimedia services. Third-generation systems, such as the International Mobile Telecommunication System (IMT-2000) network (formerly known as the Future Public Land Mobile Telecommunication System, FPLMTS) and the Universal Mobile Telecommunication System (UMTS), promise to provide multimedia services to mobile and fixed users via wireless access to the global telecommunications infrastructure [1, 2]. The IMT-2000 is a universal, multifunction, globally compatible digital mobile radio system that plans to integrate all traffic types and all wireless systems under a common set of formats. The standards for IMT-2000 are being developed by the International Telecommunication Union (ITU) standards organization for global communications [1]. UMTS is a similar global wireless solution, and is being standardized by the European Telecommunications Standards Institute (ETSI) [2]. Among the requirements for third-generation systems is the ability to support multimedia traffic. The various classes of traffic can be distinguished as suggested by the Asynchronous Transfer Mode (ATM) Forum [3]:

- Constant bit rate (CBR) traffic (digital voice and video)
- Real-time variable bit rate (rt-VBR) traffic (compressed voice and video)
- Non-real-time variable bit rate (nrt-VBR) traffic (data)
- Available bit rate (ABR) traffic (non-time-critical data)
- Unspecified bit rate (UBR) traffic (file transfer, system backup, e-mail)

In a wireless system consisting of a number of mobile terminals that transmit traffic of any type on a shared medium to a centralized base station, a procedure must be invoked to distribute packet transmission among all users. This procedure is known as a medium access control (MAC) protocol. MAC protocols are often classified according to their method of resource sharing, as well as their multiple access technology [4]. The resource sharing methods include *dedicated assignment*, *random access*, and *demand-based assignment*. Dedicated channels assign each user a predetermined and fixed allocation of resources, regardless of the user's need to transmit. Dedicated assignment schemes are appropriate for continuous traffic, but can be wasteful for bursty traffic. On the other hand, random access channels allow all users to contend for the channel by transmitting as soon as packets are available to send. Random access is suitable for bursty data traffic, but is not desirable for delay-sensitive traffic. Demand-based assignment schemes assign resources according to requests, or reservations, submitted by users. Once the requests are transmitted (using either dedicated or random access channels) and processed, users can be assigned resources according to the results. Demand-based channels are useful for variable-rate traffic and the hybrid conditions of multimedia traffic. However, the additional overhead and delay caused by the reservation process can degrade performance.

In addition to the resource sharing method, the *multiple* (or *multi*) *access* scheme of a MAC protocol establishes a method of dividing the resources into accessible sections. Three accepted methods for resource division are frequency-division multiple access (FDMA), time-division multiple access

The work of Ian F. Akyildiz and Janise McNair is supported by Department of Defense under grant number MDA 904-97-C-1105-0003. The work of Loren Carrasco Martorell is supported by the Comisio'n Interministerial de Ciencia y Tecnologi'a under Grant (TIC 98-0263), Spain, and the Comunitat Aut'onoma de les Illes Balears, Spain.

(TDMA), and code-division multiple access (CDMA). FDMA schemes divide the resource into portions of spectrum, referred to as *channels*. TDMA schemes divide the resource into time slots. Finally, CDMA divides the resource into a collection of codes through which assigned users can coexist on the same channel. First-generation mobile systems, such as the Advanced Mobile Phone System (AMPS), used FDMA schemes to support analog communication [5]. For second-generation systems, the most used multi-access schemes have been TDMA and CDMA. North American Interim Standard 54 (IS-54) and the European standard Global System for Mobile Communications (GSM) each support TDMA, while the North American standards IS-95 and IS-136, the European standard Universal Terrestrial Radio Access (UTRA), and the Japanese standard Personal Digital Cellular (PDC) support CDMA [6]. In recent years there has been considerable debate on the issue of whether TDMA or CDMA is the best candidate for standardization for third-generation networks and beyond [4, 7–10].

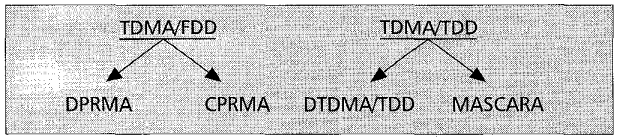
This survey examines both of the leading technologies for third-generation wireless communication, TDMA and CDMA. Other surveys have explored the earlier development of MAC protocols, and focus on either TDMA or CDMA protocol development, but have not considered comparisons from both categories of protocols. In [4] the author emphasizes the importance of the nature of the traffic to be transported in the selection of the resource sharing method. In addition, the author demonstrates CDMA equivalence with a spread-spectrum ALOHA channel when a common code is assigned to all users. In [5] the author shows the evolution of the various MAC protocols (FDMA, TDMA, and CDMA) from the first generation to the second, but does not examine individual protocols and does not discuss the transport of multimedia traffic. In [11] the authors present proposed MAC protocols for wireless ATM systems but restrict the discussion to TDMA techniques. A protocol comparison is provided based on complexity, offered quality of service (QoS), overhead, and random access technique, but the survey contains protocols that support only speech and data traffic, and it does not consider priority assignment techniques.

Although the properties of voice and data traffic can be used to create more resource availability for MAC protocols [12], the addition of CBR traffic and rt-VBR traffic adds complexity to the system and requires a technique to assign priority to various users and/or traffic types. In addition, the use of contention and retransmission of packets becomes a more important issue for the wireless channel since the resources are severely limited compared to wireline networks. The corresponding delays will have a greater impact. To serve the requirements of multimedia traffic a MAC protocol must:

- Provide simultaneous support for the wide variety of traffic types mentioned above
- Support traffic that requires delay and jitter bounds
- Assign bandwidth resources in an efficient manner (e.g., on demand)
- Support both fair and prioritized access to resources

We focus on protocols that have specifically considered the transport of multimedia traffic with QoS constraints and prioritization. Many MAC protocols have been developed to support speech and data traffic and can be found in [12–18]. Other MAC protocols that have considered multimedia traffic can be found in [9, 19–22], but these did not include either QoS or priority considerations. Investigations of MAC protocols for ad hoc networks can be found in [23–25].

In this survey, we present proposals for third-generation TDMA and CDMA MAC protocols for wireless networks carrying multimedia traffic. We have selected protocols that have been designed with multimedia traffic in mind. The procedures consider various multimedia traffic classes with varying bit rates, as well as QoS issues such as bit error rates (BERs), priority access, and delay requirements. In the next section we explore TDMA MAC protocols which are well able to support bursty traffic sources and asymmetric communication applications, such as Internet downloads. Then we discuss CDMA MAC protocols which support wideband communication with increased transmission options for multimedia traffic. Finally, we provide a qualitative comparative outline of the protocols according to suggested multimedia-based criteria.



■ Figure 1. *TDMA MAC protocols.*

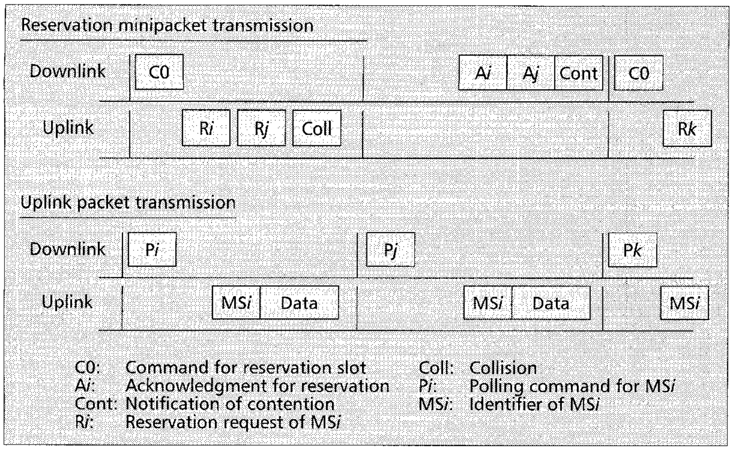
Time-Division Multiple Access MAC Protocols

As mentioned previously, TDMA is a method of resource division that divides the spectrum into time slots. We have further categorized the TDMA protocols according to the duplexing technique. Frequency-division duplex (FDD) provides two carrier frequencies, while time-division duplex (TDD) provides only one carrier frequency. In FDD the uplink frequency carries traffic from the remote terminal to the base station, while the downlink frequency carries traffic from the base station to the terminal. FDD allows the possibility of almost immediate feedback from the base station, enabling the remote terminal to find out quickly if its contending reservation request was unsuccessful and a retransmission is necessary. Thus, FDD impacts the delay encountered by user traffic as well as the resource availability of the wireless channel. Figure 1 shows the TDMA MAC protocols under their duplex methods.

TDMA with FDD: Dynamic Packet Reservation Multiple Access — Dynamic Packet Reservation Multiple Access (DPRMA) is a MAC protocol developed for wireless ATM networks [26]. The concept is based on the classical PRMA protocol [5, 12]. PRMA supported voice and data and enabled more users to be supported than time slots by using the silent periods characteristic in voice traffic to serve intermittent data traffic. Like PRMA, DPRMA attempts to take advantage of the bursty nature of multimedia traffic. It is a demand-based assignment scheme that uses slotted ALOHA for reservation contention periods.

DPRMA time slots are assigned to users according to the amount of bandwidth required. The user submits an initial rate request, or a change in rate request, by setting the appropriate reservation request (RR) bit in the header of the uplink slot. All users may submit reservations, but real-time requests have a higher priority than non-real-time traffic. The results of the contention period are transmitted downlink via several reservation acknowledgment (RA) bits in the header of downlink messages. After a contention period, the base station allocates as much of the user's requested rate as possible. If only a partial allocation can be assigned to the user, the remaining request is kept by the base station in order to accommodate the full rate request when capacity becomes available. The successful user monitors the slot reservation (SR) bits in the message header of the downlink channel to determine in which slots it may send packets.

For time-dependent traffic, user requests are filled using as many of the available empty slots as necessary. When additional slots are needed the slots occupied by data traffic are preempted,



■ Figure 2. *Data and signaling channels of the CPRMA protocol.*

and the data users are placed in a queue to await further service when bandwidth becomes available. For real-time users, packets are dropped if the guaranteed rate cannot be met. Data traffic is lost only when the buffer for preempted packets overflows.

The primary feature of the DPRMA protocol is dynamic assignment of slots according to requested bandwidth as well as the time-dependent nature of the traffic. The authors show that DPRMA performs well in a system with voice, videoconferencing, and data users present [26]. This protocol was chosen for its simplicity of bandwidth assignment and its simple approach to priority access for different traffic classes. However, DPRMA's use of full-sized request slots for contention periods can waste the limited wireless bandwidth, and thereby degrade overall performance.

Centralized PRMA — The CPRMA MAC protocol was designed for a microcellular environment [27]. The goal is to grant transmissions at each slot according to the terminal with the most urgent need to transmit. Like DPRMA, CPRMA is a demand access scheme with contention-based reservation periods that expands the concepts of PRMA to multimedia traffic. Unlike DPRMA, CPRMA attempts prompt retransmission of corrupted packets. In CPRMA, the base station again plays a central role in scheduling packet transmissions for remote terminals.

As shown in Fig. 2, RRs are transmitted in the available slots of the uplink channel. Since a request contains a limited amount of information, such as terminal identifier, time spent by the first packet in the buffer, and a cyclic redundancy check (CRC) field, requests can be sent within a minipacket. A successful reservation is acknowledged by the base station. Then, when the base station issues a polling command with the remote terminal identifier, the terminal can transmit packets on the assigned reserved slots. If a collision occurs, a cycle for collision resolution is initiated and continues until all of the collided minipackets have successfully been transmitted.

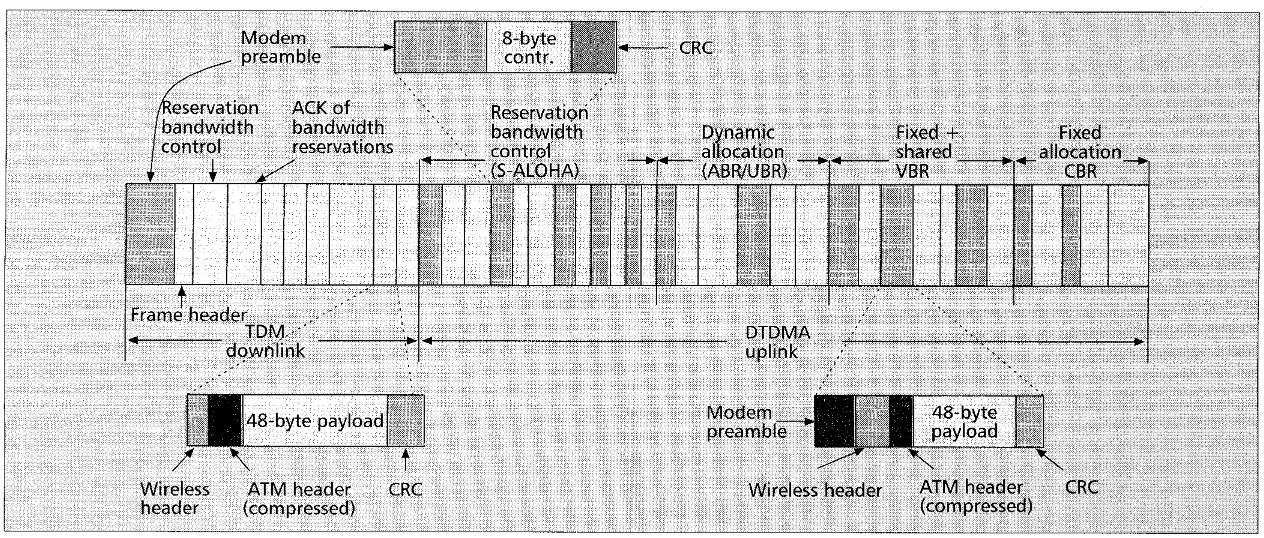
Multimedia traffic is accommodated in this protocol via the polling process. The polling sequence for the reserved terminal is generated by a scheduling algorithm, whose purpose is to provide QoS to all users according to the various loss, delay, and bandwidth requirements, and to assign slots according to appropriate priority constraints.

The CPRMA protocol [27] was selected to illustrate a more complex approach to priority assignment than the DPRMA approach [26], and to show an alternative to full slot contention. In CPRMA slots are dynamically assigned, and the scheduling algorithm provides a more accurate statistical combination of traffic. Delay requirements, as well as loss and bandwidth, can be served according to individual user requests, rather than real-time versus non-real-time traffic classes.

Although the FDD techniques provide a faster method to determine if retransmission is necessary, TDD techniques also have advantages to benefit multimedia traffic. TDD systems have only one frequency carrier. However, TDD schemes can allow asymmetric traffic to be transmitted to/from the remote terminal from/to the base station, which is appropriate for Web browsing or Internet downloads.

TDMA with TDD

Dynamic TDMA with TDD — DTDMA/TDD was designed for the wireless ATM network (WATMnet), a prototype microcellular WATM network capable of providing integrated multimedia communication service to remote terminals [28]. The MAC protocol combines all three resource sharing methods:



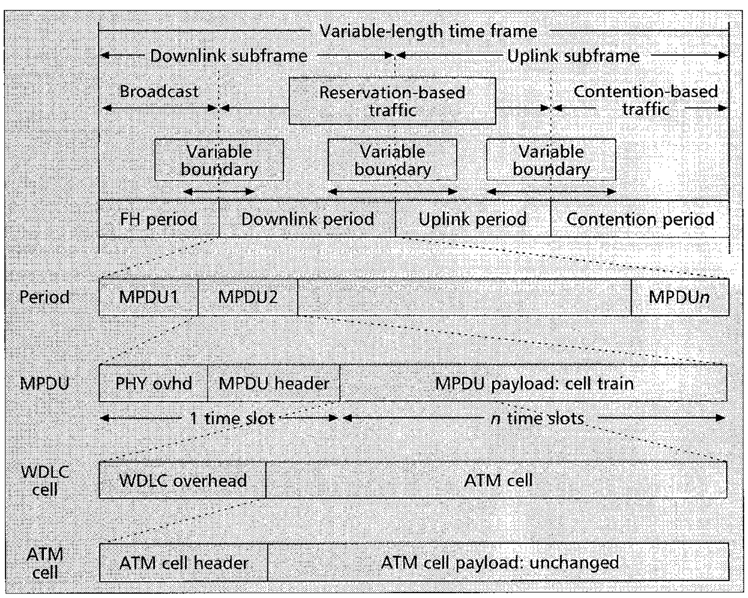
■ Figure 3. *The DTDMA/TDD frame format.*

dedicated, random, and demand assignment. Instead of a strictly dynamic allocation of VBR traffic, DTDMA/TDD provides both fixed and shared allocation of VBR. Dynamic allocation is used for ABR and UBR traffic; Fig. 3 shows the DTDMA/TDD frame format.

Users send transmission requests to the base station in the dedicated reservation slots using slotted ALOHA random access. The requests are then processed, resulting in a schedule table based on the QoS parameters of user traffic. The base station then broadcasts slot allocation and acknowledges successful reservations. For CBR traffic, slot allocation is performed once during call establishment. A fixed allocation of slots is assigned according to user requests. When CBR slots are no longer available, arriving CBR calls are blocked. VBR slots are assigned based on a Usage Parameter Control (UPC) statistical multiplexing algorithm. Like CBR, VBR slots have fixed allocation, but unused slots are shared with other traffic classes. Arriving VBR calls are also blocked when VBR slots are not available. Finally, for ABR/UBR traffic, slot allocation is performed on a burst-by-burst basis via dynamic reservation of ABR/UBR slots and unused CBR and VBR slots. ABR/UBR calls are always accepted and inserted with contiguous allocation where possible.

The DTDMA/TDD protocol was chosen as another increase in complexity for transmission scheduling according to traffic class. Since slots are apportioned for CBR, VBR, and ABR/UBR categories (circuit mode for CBR/VBR and dynamically for ABR/UBR), the channel is not dominated by the most demanding user. As in CPRMA, minislots are used for reservation in order to preserve bandwidth.

Mobile Access Scheme Based on Contention and Reservation for ATM — The MASCARA protocol was proposed as part of the Wireless ATM Network Demonstrator (WAND) project being developed with the support of the European Community [29, 30]. It is a demand assignment scheme with contention-based reservations. The uplink subframe is divided into a contention period to transmit reservation requests or some control information, and an uplink period for uplink data traffic. As shown in Fig. 4, each period within a frame has a variable length depending on the instantaneous traffic to be carried over the wireless channel.



■ Figure 4. *The MASCARA time frame structure.*

If a remote terminal has cells to transmit, it sends a reservation request, which is either piggybacked in the MASCARA protocol data units (MPDUs) the remote terminal sends in the uplink period, or contained in special *control MPDUs* sent for that purpose in the contention period. At the end of a frame, the base station schedules the transmissions of the next frame, according to reservation requests sent by the remote terminals, the arriving cells for each downlink connection, and the traffic characteristics and QoS requirements of all connections. In the frame header of the downlink, the base station broadcasts information which contains a descriptor of the current time frame (including the lengths of each period), the results of the contention procedures from the previous frame, and the position of the slot allocated to each downlink and uplink connection. To minimize physical-layer overhead, MASCARA uses the concept of a *cell train*, which is a sequence (1–n) of ATM cells belonging to one remote terminal and having a common header. The length of the overhead plus that of the MPDU header is equal to one time slot, which is defined as the length of an ATM cell (53 bytes).

The master scheduler uses an algorithm called Priority Regulated Allocation Delay-Oriented Scheduling (PRADOS) to schedule transmissions over the radio interface. The PRADOS algorithm is based on priority classes and the delay constraints of each active connection [31]. PRADOS assigns priorities for each connection according to its service class, ranging from 1 to 5; in order of increasing priority, the classes are UBR(1), ABR(2), nrt-VBR(3), rt-VBR(4), and CBR(5). The PRADOS algorithm combines priorities with a leaky bucket traffic regulator. The regulator uses a token pool that is introduced for each connection. Tokens are generated at a fixed rate equal to the mean ATM cell rate of each VC. The size of the pool is equal to the maximum number of ATM cells that can be transmitted with a rate greater than the declared mean. Starting at priority 5 and ending with priority 2, the scheduler satisfies requests for connections as long as tokens and slots are available. For every slot allocated to a connection, a token is removed from the corresponding pool. PRADOS also schedules transmissions as close to the deadline of each ATM cell as possible in order to maximize the fraction of ATM cells transmitted before their deadline.

The MASCARA protocol provides variable capacity to terminals in multiples of slots via the cell train concept. However, the variable-length frame creates difficulties in assigning capacity to CBR packets. Like DTDMA/TDD, a complex scheduling algorithm is employed to manage QoS requirements of user traffic. Both slotted ALOHA and a stack algorithm are being investigated for the contention procedure [30].

CDMA MAC Protocols

CDMA is an access technique for spread spectrum where all remote terminals in a cell transmit using the whole spectrum of the channel simultaneously. Spread-spectrum techniques may use frequency hopping or direct sequence. Currently, CDMA technology presents new challenges to multimedia MAC protocols, since they must not only efficiently share resources, but also guarantee the stability of the system (i.e., control the interference level in peak bursty periods). Moreover, for different types of traffic coexisting on the same channel, or if packet switching is used, channel users can be affected

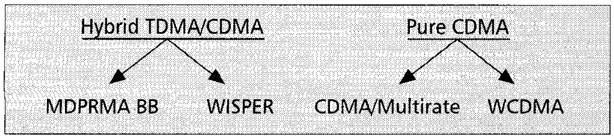


Figure 5. *CDMA wireless multimedia MAC protocols.*

by fluctuating interference levels. Thus, the interference level experienced by each type of user may be different and must be considered, and global estimations must be calculated to accept high-bit-rate users.

In 1998 ETSI decided to use wideband CDMA (WCDMA) in the paired band (FDD) and hybrid TDMA/CDMA in the unpaired band (TDD) [32]. The development of third-generation systems include the evolution of GSM and UTRA to the TDMA/CDMA and WCDMA standards. For this reason, we classify the wireless multimedia CDMA MAC protocols as shown in Fig. 5: hybrid TDMA/CDMA protocols and pure CDMA protocols. The hybrid protocols apply to both the TDMA and CDMA categories, but are discussed here in relation to CDMA.

Hybrid TDMA/CDMA MAC Protocols

Hybrid TDMA/CDMA schemes benefit from both the capacity of TDMA schemes to handle high-bit-rate packet-switched services and the flexibility of CDMA techniques that allow smooth coexistence of different types of traffic.

Multidimensional PRMA with Prioritized Bayesian Broadcast (MDPRMA BB) [33] is a derivation of the classic PRMA protocol [12] adapted for hybrid schemes and multimedia traffic, while the Wireless Multimedia Access Control Protocol with Bit Error Rate Scheduling (WISPER) protocol [34] uses a more deterministic approach where the base station schedules transmissions according to BER requirements of the different traffic classes.

Multidimensional PRMA with Prioritized Bayesian Broadcast — The MDPRMA BB MAC protocol is a proposed strategy for UMTS for the uplink channel of UTRA [33]. This protocol is suitable for any hybrid FDMA/TDMA or TDMA/CDMA scheme. However, in [33] the focus is restricted to the UMTS TDMA/CDMA TDD access selected by ETSI to be used in the UMTS unpaired bands. MDPRMA BB is a derivation of the classic PRMA protocol adapted to hybrid access schemes and multimedia traffic. As in PRMA, time slots of fixed length are grouped into frames, but MDPRMA BB further divides each time slot into subslots using up to eight spreading codes, as shown in Fig. 6. These subslots are either available for contention (C subslot) or reserved for the information transfer of a particular terminal (I subslot). A mobile that holds no reservation but is admitted to the system will contend for a reservation using C subslots. The mobile will contend in a certain C subslot with a specified probability which is service-class- and time-slot-dependent.

Probabilities for each type of service and each time slot of the next uplink frame are broadcast by the base station in the downlink frame. The probabilities depend on the estimated number of *backlogged* terminals (terminals that have a packet to transmit but no reservation) currently in the system, and are calculated according to protocol stability and best delay-throughput performance. Moreover, these probabilities are connected to a load-based access control to ensure control of the interference level of the CDMA component and the stability of the system. The downlink broadcast will also include the acknowledgments for successful requests in the previous frame. Usually, implicit reservation is used, that is, once the user receives the acknowledgment it will start transmitting in the C subslot used to sent the request. From that moment this C subslot becomes an I subslot and will be reserved for that user. The reservation will last until the end of the spurt in case of voice users, or for a certain number of frames (F) in case of data services. By controlling the contention access (transmission probabilities) and allocation (number of frames) for data services, the protocol is able to track delay requirements and dropping probabilities for different services.

One of the main drawbacks of this protocol is that it is not able to support high-bit-rate data services or real-time services such as video which will multi-subslot allocations in the same frame. Another problem is that it does not take into account the problems derived by the CDMA component; it allows different services (with different BER requirements) to share the same time slot. Thus, the capacity of the slots will be variable and limited by the most demanding service (with respect to BER).

Wireless Multimedia Access Control Protocol with Bit Error Rate Scheduling — The WISPER protocol was developed to take advantage of the power control characteristics of the IS-95 standard [34]. This protocol was designed so that the BER of the transmission channel is maintained below a given specification. It schedules packet transmission according to the BER requirements of several different traffic classes. In this proto-

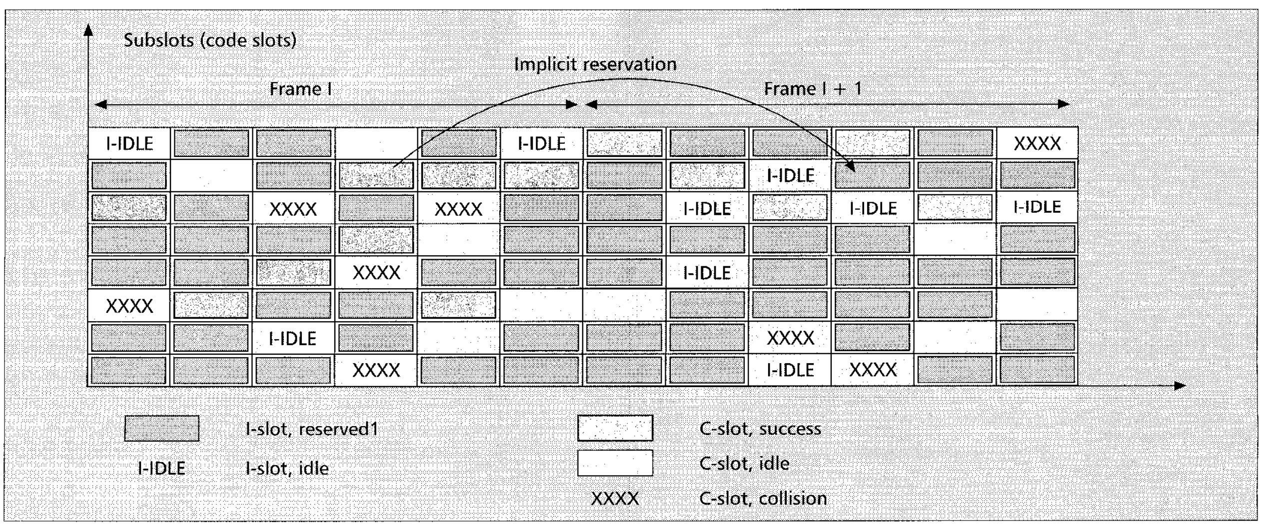


Figure 6. *Slots and frames in MD PRMA BB.*

col, the total available bandwidth is divided into two bands, one for the uplink and one for the downlink. For both bands time is divided into frames, and each frame is divided into slots. The length of a frame is chosen to coincide with the packet arrival rate of the most abundant traffic class (usually voice). Figure 7a shows the relative timing of the upstream and downstream frames.

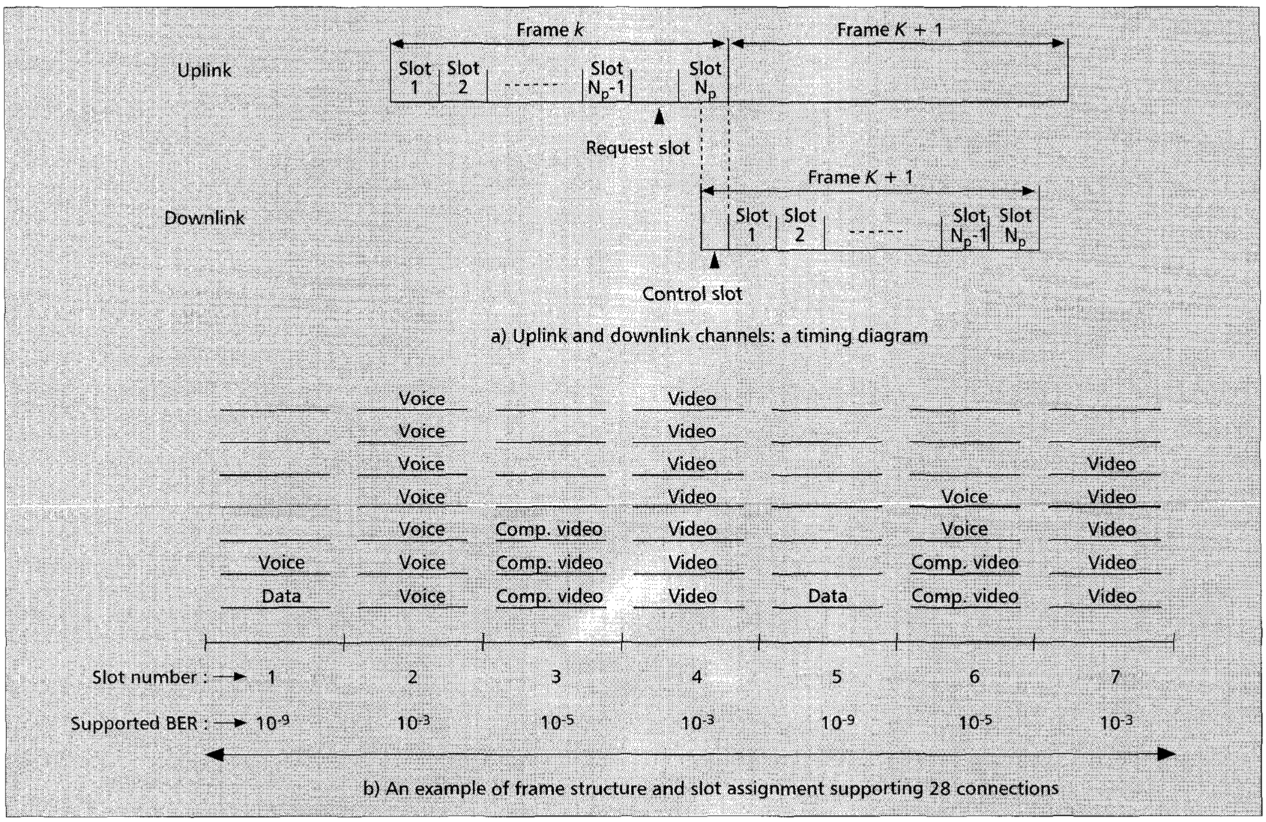
For the uplink, each frame is divided into packet slots and one request slot. Each packet slot can carry any class of traffic. The request slot can be used for two purposes: to place admission requests by new remote terminals that want to be admitted to the wireless network, and to place transmission requests by remote terminals currently registered in the wireless network. Whenever a remote terminal has new packets ready for transmission, it must send a transmission request to the base station, indicating the number of packets in the new batch as well as the corresponding timeout value of the packets. A remote terminal sends the transmission request by either using the request slot or piggybacking the request in a previously transmitted data packet. The latter method is used whenever possible in order to reduce contention in the request slot. In either case, requests are transmitted using an assigned primary pseudo-noise code. Once a request has been received, a data structure is used by the base station to keep track of the batch associated with the request. The data structure contains information such as the remote terminal that owns the batch, the packet's timeout value, and the number of packets in the batch. This information is kept until the packets in the batch have been received successfully, or until they timeout and are discarded. When a base station responds to a request for packet transmission, it specifies the slot(s) and the corresponding number of packets that can be transmitted in the next frame.

WISPER designates slots that can support certain BERs, and schedules packet transmissions in these slots in such a way that the wireless bandwidth can be used efficiently. To use the available bandwidth in an efficient manner, packets that have either equal or almost the same maximum BER specifications are transmitted in the same slot. In other words, the packet with the most stringent BER specification determines the maximum number of packets which can be transmitted simultaneously in that specific slot. Figure 7b shows an example of frame structure and slot assignment supporting 28 connections. The transmission order of the packets is determined by a novel scheduler. For each new frame, the packet scheduler prioritizes packet transmissions and accommodates the higher-priority packets in the frame so that throughput is maximized. The transmission order is determined according to packets' timeout values and the number of packets ready for transmission at each remote terminal.

WISPER maintains different BERs according to the specifications of the packets being transmitted. This results in an important increase in total system capacity over those schemes that use a single BER threshold. The throughput is maximized by ordering packet transmissions according to traffic classes and by scheduling the packet transmissions at the remote terminal's maximum possible transmission rate. In addition, packet losses are minimized by a packet prioritization scheme that determines packet transmission order by considering remaining times before packet timeout occurs.

Pure CDMA MAC Protocols

Current research efforts in strictly CDMA-oriented MAC protocols concern the complex WCDMA physical layers where many transmissions options are possible. The first protocol



■ Figure 7. *A WISPER timing diagram and frame structure.*

described in this section is based on prioritized queuing and the assignment of transmission probabilities to remote terminals [35]. The second protocol was specifically designed for the WCDMA physical layer adopted by Europe and Japan [36].

A MAC Protocol for a Cellular Packet CDMA Carrying Multirate CDMA — This protocol is designed for a cellular direct sequence/CDMA (DS/CDMA) network carrying multiple traffic types [35]. It is a packet-oriented MAC protocol which has prioritized queuing for different types of traffic. Each packet comprises a synchronization header, an ATM cell, and an error correction control parity trailer. It is assumed that an ideal feedback channel exists for the transmission acknowledgment. The users are classified into types according to their traffic rate. Where traffic of the same rate has different priorities, different traffic types can be created even for the same rate. As traffic arrives from the source it is buffered in a finite-length buffer for each traffic type. If there is a packet in the queue, the user attempts transmission at the beginning of the next slot. The user can assume three states: idle, active, and blocked, based on the state of the buffer and the success of the previous transmission. If there is no packet queued in the buffer, the user assumes the idle state.

An active or blocked user may further assume a substate based on how many packets are queued in the transmission buffer. If an idle user's information source generates a packet(s), the packet(s) is(are) queued in the transmission buffer and the user assumes the active state. An active user attempts to transmit the head-of-the-queue packet with probability P, which may differ in value for different users in order to ensure priority treatment for different queues. A lower P corresponds to a higher priority. If the transmission succeeds, the user attempts to transmit the next packet in the queue if it is not empty, and the user's state remains unchanged; if the successful transmission emptied the queue and no new packets have arrived, the user assumes the idle state. If the transmission failed, the user assumes the blocked state. An active user in the blocked state attempts a retransmission with probability P. If the retransmission fails, the user remains in the blocked state. Otherwise, the user assumes the idle or the active state depending on whether the buffer has been emptied or not.

This protocol can handle a mixture of traffic types and rates, including relatively high traffic rates. It does not impose any limits on the number of traffic types carried. An advantage of the protocol is its simplicity of implementation. It can also ensure a considerable multiplexing gain as bandwidth increases. However, the use of information packets in contention periods not only greatly increases delays due to contention resolution, but increases the likelihood of packet loss. This can adversely affect the delivery of throughput-dependent traffic (e.g., file transfers) that are sensitive to loss of information.

Wideband CDMA MAC Protocol for Real-Time and Non-Real-Time Services — The WCDMA MAC protocol for real-time and non-real-time services [36] is a proposal to be used on top of the WCDMA physical layer adopted by the Japanese Association of Radio Industries and Business (ARIB) and ETSI for their IMT-2000 standard systems [32] (ETSI has adopted WCDMA only in the paired band).

The physical layer is capable of simultaneously transferring data from different services with a single remote terminal. Data streams originated from different services are multiplexed at the physical layer. These services can have different QoS requirements in terms of BERs. These different requirements can be met at the physical layer by applying different types of coding techniques for different types of services. Moreover, it is possible for the remote terminals to transmit with variable bit rates out of a discrete set of possible bit rates. This can be done if the terminal has a dedicated bidirectional channel (DCH) at his disposal, each DCH having a dedicated code. The DCH has an associated control channel on which power control bits and rate information bits are transmitted. Power control is used to mitigate the effect of fast fading, and the rate information bits indicate the actual rate of the dedicated channel. In the uplink, the DCH rate is modified through changes in the spreading factor. It is possible to change the rate every 10 ms.

For packet data services, there are two basic methods to transmit data. First, short packets can be transmitted in an ALOHA basis using the random access channel (RACH), a channel common to all terminals in the cell, used for issuing transmission requests as well. This mode can be used to transmit short and infrequent packets with no delay and minimum overhead. In the case of larger packets, the terminal will request a DCH on the RACH (i.e., a dedicated code with fast-power-controlled transmission). Since the length of the packet is large, the overhead caused by the reservation mechanism will then be negligible. The request will include the type of service and eventually the packet length. After evaluating if the required resources are available, the network will answer the request indicating a set of possible transmission formats (TFs). If the load is low, the system will also indicate the specific TF and the time the user can start transmitting. In heavily loaded situations, first only the set of TF formats will be sent to the user. Then the user has to issue another transmission request to receive the specific TF to be used. While the terminal is transmitting, the network will decrease or increase its transmission rate depending on the network load. Once the transmission is finished, the terminal will maintain the link for a certain time. If a new packet is generated during that time, the mobile can start transmitting immediately with the prescribed TF. However, if the generated packet is very large, the terminal will first have to issue a request on the DCH (piggybacked request) asking for permission. Lastly, if no more packets arrive during the holding time, the link will be lost, but the terminal will keep the TFs for a certain period in order to facilitate future transmissions.

Real-time services have an allocation procedure that is very similar to the data case. Once the remote terminal has data to transmit, it issues a request on the RACH. The network will then answer with a set of TF formats, and the terminal can start to transmit immediately using any TF out of its set. This degree of freedom at the terminal will allow for variable-rate transmission. If the network load is high, the network can limit to a subset the initial set of TFs that the terminal can use for its transmission. For multiplexed services or multiple services transmitted from the same terminal, if the terminal wants to transmit, the network will assign a set of TF for each service. The terminal will be able to use any TF for real-time services, while the network will assign a specific TF for data services. An additional constraint will be put on this kind of multiservice remote terminal in terms of a maximum power/rate threshold.

The WCDMA protocol deals with a complex and very flexible physical layer where many transmission options are possible. The control is performed through a demand assignment protocol (except for very short packets) relying on a complex resource management performed at the base station and possibly in the network. In fact, the stability of the protocol depends on higher layer-functionalities such as call admission control (CAC) and congestion control (CC). For example, when the cell is highly loaded and current users suffer a link degradation due to excessive interference, new calls will be blocked (CAC), and CC will order the MAC to decrease the transmission rates (lower profile TFs) of the terminals already connected to the base station. Unfortunately, these high-layer functionalities are still not completely defined. The protocol is lacking a method for the net-

MAC protocol	Slot/code assignment	Access contention	QoS support	Priority access	Complexity
		TDMA protocols			
TDMA-FDD					
DPRMA [26]	According to BW needs	Full-sized request slots	Based on required BW	Real-time over non-real-time	Low
CPRMA [27]	According to BW requests and availability	Minisized request slots	Loss, delay, and traffic requirements	Scheduling of polling sequence	Low
TDMA-TDD					
DTDMA/ [28]	CBR,VBR - circuit mode VBR,ABR – dynamic	Minisized request slots	Based on UPC values	Fixed no. of slots vs. variable	High
MASCARA [29, 30]	According to BW needs and priority class	Piggybacked request slots	Required BW and delay	Priority classes	High
		CDMA (and hybrid) protocols			
TD/CDMA					
MDPRMA BB [33]	According to traffic class and required traffic rate	Full-sized slots	Traffic rate and delay	Different trans. probabilities	High
WISPER [34]	According to required BER and traffic class	Piggybacked request slots	Required BER and delay	Prioritized pkt. transmission	High
Pure CDMA					
Multirate [35]	According to traffic class and required traffic rate	Information packets	Traffic rate and delay	Diff. trans. probs. & pkt. queue size	Medium
WCDMA [36]	According to load, traffic class, and rate	Request packets	Required BER and delay	Different trans. format specified	High

■ Table 1. *A qualitative comparison of MAC protocols for multimedia traffic in wireless networks.*

work to evaluate if a new user could be admitted to a cell, as well as the procedure followed by the congestion control functionality. Another characteristic that deserves further evaluation is the extensive use of an ALOHA access in the RACH for short packet transmission, and several types of transmission request messages. In highly loaded conditions (or if the remote terminals make uncontrolled use of short packet transmissions), the performance of the protocol could be severely degraded, not only because of excessive packet collisions that imply delays in accessing the channel and loss of short data packets, but also because of the increase in the interference level produced by many non-fast-power-controlled packet transmissions.

An Overview of MAC Protocols

In order to explore the ability of MAC protocols to support the criteria outlined in the introduction, we must explore the following questions:

- What channel access method is being used?
- What are the slot/code designations within each frame? Are frame lengths variable or fixed?
- How are resource assignment decisions made?
- How are different traffic types effectively integrated and QoS constraints managed?
- For terminals contending for the same resources, how are unsuccessful transmissions resolved?

The channel access method can greatly affect the delays experienced by user traffic. For example, many of the protocols employed demand-based slot (or code) assignment with a random access reservation scheme. Collisions that occur during the reservation period must be resolved and successful packets acknowledged in such a manner as to quickly remove administrative traffic and thereby achieve more efficient utilization of available bandwidth. Likewise, the use of minislots and piggybacking is useful to minimize the resources consumed by overhead and to provide more throughput of information. Most infrastructured networks assume a centralized access technique, taking advantage of the base station as a gateway between the mobile terminals and the network. However, slot assignment at the base station must incorporate knowledge about the required BERs, permitted delays, and agreed-upon throughput in order to implement a QoS policy. At the same time, the algorithms used to manage user requirements must be balanced against the corresponding complexity. In Table 1 we qualitatively compare the multimedia MAC protocols according to a set of multimedia-based issues: bandwidth, or slot, assignment; contention slots; QoS support; priority access support; and complexity.

Conclusion

In this article we present a qualitative overview of MAC protocols for multimedia traffic in wireless networks. We outline the design objectives for providing service to various traffic types and develop an appropriate set of criteria for multimedia support with QoS. We describe the operation of several contemporary MAC protocols intended for both TDMA and CDMA systems, including those intended for third-generation systems.

In summary, most current protocols employ demand assignment with some form of contention reservation period. Since wireless networks cannot take advantage of the wireline technique of collision detection, delays are incurred by each terminal awaiting an acknowledgment to determine if its transmission was successful, and then by terminals that must execute some retransmission strategy. Future research must focus on developing methods to reduce contention and the delay in retransmitting unsuccessful packets. Slot/code assignment provides an opportunity to maximize utilization of bandwidth according to the traffic and BER requirements of the transmitted packets. Priority techniques increase the complexity of the schemes, but are necessary for the realization of QoS constraints of varying types of multimedia traffic. In addition, MAC protocols designed for IMT-2000 systems will have to deal with complex and very flexible physical layers, where many transmission options will be available. Successful MAC protocols will find a balance between the complexity of service guarantees for multiple service classes, the ability to use available resources in an efficient manner, and the ability to react quickly to failed transmissions.

References

[1] R. Pandya *et al.*, "IMT-2000 standards: Network aspects," *IEEE Pers. Commun.*, Aug. 1997, pp. 20–29.

[2] A. Samukic, "UMTS universal mobile telecommunications system: Development of standard for the third generation," *IEEE Trans. Vehic. Tech.*, vol. 47, Nov. 1998, pp. 1099–1104.

[3] ATM Forum, "Traffic Management Specification, Version 4.0," Apr. 1996.

[4] N. Abramson, "Multiple access in wireless digital networks," *Proc. IEEE*, vol. 82, Sept. 1994, pp. 1360–70.

[5] R. Steele, "The evolution of personal communications," *IEEE Pers. Commun.*, vol. 1, 2nd qtr. 1994, pp. 6–11.

[6] E. Dahlman *et al.*, "UMTS/IMT-2000 based on wideband CDMA," *IEEE Commun. Mag.*, vol. 36, Sept. 1998, pp. 70–80.

[7] P. Jung, P. W. Baier, and A. Sreil, "Advantages of cdma and spread spectrum techniques over fdma and tdma in cellular mobile applications," *IEEE Trans. Vehic. Tech.*, vol. 42, Aug. 1993, pp. 357–63.

[8] R. Meidan, "To spread or not to spread, this is question," *IEEE VTC*, vol. 1, 1994, pp. 56–59.

[9] D. Raychaudhuri and N. D. Wilson, "ATM-based transport architecture for multiservices wireless personal communication networks," *IEEE JSAC*, vol. 12, Oct. 1994, pp. 1401–14.

[10] R. R. Roy, "Network constraints in multimedia conferencing and the role of ATM networks," *AT&T Tech. J.*, vol. 73, July/Aug. 1994, pp. 97–108.

[11] J. Sanchez, R. Martinez, and M. Marcellin, "A survey of MAC protocols proposed for wireless ATM," *IEEE Network*, Nov./Dec. 1997, pp. 52–62.

[12] D. J. Goodman and S. X. Wei, "Efficiency of packet reservation multiple access," *IEEE Trans. Vehic. Tech.*, vol. 40, Feb. 1991, pp. 170–76.

[13] N. Amitay, "Resource auction multiple access (RAMA): Efficient method for fast resource assignment in decentralized wireless PCS," *Elect. Lett.*, vol. 28, Apr. 1992, pp. 799–801.

[14] J. De Vile, "A reservation based multiple access scheme for a future universal mobile telecommunications system," *Proc. 7th IEE Conf. Mobile and Pers. Commun.*, Brighton, U.K., Dec. 1993, pp. 210–15.

[15] M. J. Karol, Z. Liu, and K. Eng, "An efficient demandassignment multiple access protocol for wireless packet (ATM) networks," *ACM/Baltzer J. Wireless Networks*, vol. 1, Oct. 1995, pp. 267–79.

[16] F. D. Priscoli, "Medium access control for the median system," *Proc. ACTS Mobile Summit '96*, Granada, Spain, Nov. 1996, pp. 1–8.

[17] P. Smulders and C. Blondia, "A MAC protocol for ATM-based indoor radio network," rep. for European Cooperation in the Field of Scientific and Technical Research (EUCO-COST), COST 231, TD (94) 055, Apr. 1994.

[18] L. Tan and Q. Zhang, "A reservation random-access protocol for voice/data integrated spread-spectrum multiple-access systems," *IEEE JSAC*, vol. 14, Dec. 1996, pp. 1717–27.

[19] G. Anastasi, D. Grillo, and L. Lenzini, "An access protocol speech/data/video integration in TDMA-based advanced mobile system," *IEEE JSAC*, vol. 15, Oct. 1997, pp. 1498–1510.

[20] X. Qui, V. Li, and J. H. Ju, "A multiple access scheme for multimedia traffic in wireless ATM," *J. Mobile Networks and Apps.*, vol. 1, Dec. 1996, pp. 259–72.

[21] N. D. Wilson *et al.*, "Packet CDMA versus dynamic TDMA for multiple access in an integrated voice/data PCN," *IEEE JSAC*, vol. 11, Aug. 1993, pp. 870–84.

[22] Z. Zhang and Y.-J. Liu, "Performance analysis of multiple access protocols for CDMA cellular and personal communication services," *IEEE INFOCOM '93*, vol. 3, 1993, pp. 1214–21.

[23] C. Zhu and M. Corson, "A fivephase reservation protocol (FPRP) for mobile ad hoc networks," *Proc. IEEE INFOCOM '98*, San Francisco, CA, Apr. 1998, pp. 322–31.

[24] A. Muir and J. J. Garcia-Luna-Aceves, "A channel access protocol for multihop wireless networks with multiple channels," *Proc. ICC '98*, Atlanta, GA, June 1998.

[25] S. Singh and C. S. Raghavendra, "PAMAS—power aware multi-access protocol with signaling for ad-hoc networks," *ACM Comp. Commun. Rev.*, vol. 28, July 1998.

[26] D. A. Dyson and Z. J. Haas, "The dynamic packet reservation multiple access scheme for multimedia traffic," *ACM/Baltzer J. Mobile Networks & Apps.*, 1999.

[27] G. Bianchi *et al.*, "CPRMA: A centralized packet multiple access for local wireless communications," *IEEE Trans. Vehic. Tech.*, vol. 46, May 1997.

[28] D. Raychaudhuri *et al.*, "WATMnet: A prototype wireless atm system for multimedia personal communication," *IEEE JSAC*, vol. 15, Jan. 1997, pp. 83–95.

[29] F. Bauchot *et al.*, "MASCARA, a MAC protocol for wireless ATM," *Proc. ACTS Mobile Summit '96*, Granada, Spain, Nov. 1996, pp. 647–51.

[30] N. Passas *et al.*, "Quality of service-oriented medium access control for wireless ATM networks," *IEEE Commun. Mag.*, vol. 35, Nov. 1997, pp. 42–50.

[31] N. Passas *et al.*, "MAC protocol and traffic scheduling for wireless ATM networks," *ACM Mobile Networks and Appls.*, vol. 3, Sept. 1998, pp. 275–92.

[32] T. Ojanpera and R. Prasad, "An overview of air interface multiple access for IMT-2000/UMTS," *IEEE Commun. Mag.*, vol. 36, Sept. 1998, pp. 70–79.

[33] A. Brand and A. Aghvami, "Multidimensional PRMA with prioritized Bayesian broadcast—a MAC strategy for multiservice traffic over UMTS," *IEEE Trans. Vehic. Tech.*, vol. 47, Nov. 1998, pp. 1148–61.

[34] I. F. Akyildiz, D. A. Levine, and I. Joe, "A slotted CDMA protocol with BER scheduling for wireless multimedia networks," *IEEE/ACM Trans. Net.*, vol. 7, no. 2, Apr. 1999, pp. 146–59.

[35] R. Pichna and Q. Wang, "A medium-access control protocol for a cellular packet CDMA carrying multirate traffic," *IEEE JSAC*, vol. 14, Dec. 1996, pp. 1728–36.

[36] R. Roobol *et al.*, "A proposal for an RLC/MAC protocol for wideband CDMA capable of handling real time and non real time services," *IEEE VTC*, Ottawa, Canada, May 1998.

Biographies

IAN F. AKYILDIZ [F] (ian@ee.gatech.edu) received his B.S., M.S., and Ph.D. degrees in computer engineering from the University of Erlangen-Nuernberg, Germany, in 1978, 1981, and 1984, respectively. Currently he is a professor with the School of Electrical and Computer Engineering, Georgia Institute of Technology, where he serves as the elected chair of the Telecommunications Area, and director of the Broadband and Wireless Networking Laboratory. He has held visiting professorships at several universities. He received the Don Federico Santa Maria Medal for his services to the Universidad of Federico Santa Maria, Chile. He served as a National Lecturer for ACM from 1989 until 1998 and received the ACM Outstanding Distinguished Lecturer Award for 1994.

JANISE MCNAIR [M] (mcnair@ee.gatech.edu) received her B.S. and M.S. degrees from the University of Texas at Austin in 1991 and 1994, respectively. Currently she is a Ph.D. candidate in the School of Electrical and Computer Engineering at the Georgia Institute of Technology, Atlanta, and a research assistant in the Broadband and Wireless Networking Laboratory. She was a National Science Foundation Fellow from 1991 to1994. Her research interests include wireless multimedia networks, mobility management, and satellite networks.

LOREN CARRASCO MARTORELL (loren@ipc4.uib.es) received her Diploma in telecommunications engineering from Barcelona University, Spain, in 1991. In January 1992 she joined the Wireless Communication Department of the national research center of the Spanish PTT, Telef6nica, where she was involved in several European RACE (CODIT, MONET) and Eurescom projects until 1995. In 1997 she entered as an associate professor at the University of the Balearic Islands, Spain. Her current research interests involve multiple access and wireless networks for the support of multimedia services.

RAMON PUIGJANER [M] (putxi@ps.uib.es) received his Industrial Engineer degree from the Polytechnic University of Catalonia, Barcelona, Spain, in 1964, his Master's degree in aeronautical sciences from the Ecole Nationale Suprieure de l'Aronautique de Paris, Franc), his Ph.D.. from the Polytechnic University of Catalonia in 1972, and his license in informatics from the Polytechnic University of Madrid, Spain. From 1966 to 1987 he taught and researched automatic control, computer architecture, and computer performance evaluation at the Polytechnic University of Catalonia. From 1970 to 1987 he held several positions in industry at UNIVAC (after, SPERRY and finally, UNISYS), where he was in charge of computer performance measuring and modeling for tuning and sizing in Spain. Since 1987 he has been a professor of computer architecture and technology in the Department of Computer Science of the University of the Balearic Islands, Palma de Mallorca, Spain. His current research interests are the performance evaluation of computer systems and computer networks and the diffusion of these techniques in the industrial milieu mainly in the field of real-time systems. He was involved in ESPRIT II project 5409, COMPLEMENT and ESPRIT IV project 22453, HELIOS, and was reviewer of ESPRIT II project IMSE, ESPRIT III project PYTHAGORAS, and ESPRIT project MERCURY. Currently he is participating in ESPRIT IV project 23242, SUCSEDE, and reviewing ESPRIT IV project BISANTE.

YELENA YESHA [SM] (yelena@cesdis.gsfc.nasa.gov) received her B.Sc. degree in computer science from York University, Toronto, Canada, in 1984, and Ph.D. degrees in computer and information science from The Ohio State University in 1986 and 1989, respectively. Since 1989 she has been with the Department of Computer Science and Electrical Engineering at the University of Maryland Baltimore County, where she is presently a professor. In addition, since December 1994 she has served as director of the Center of Excellence in Space Data and Information Sciences at NASA. During 1994, she was director of the Center for Applied Information Technology at the National Institute of Standards and Technology. She is presently serving as a member of the U.S. delegation to the G7 program on the global marketplace for small and medium enterprises. She is presently consulting for IBM's team developing the leading software products for global electronic commerce. Her research interests are in the areas of distributed databases, distributed systems, electronic commerce, and trusted information systems.