

Providing deterministic packet delays and packet losses in multimedia wireless networks[‡]

Mounir Hamdi*[†] and Felix K. L. Lee

Department of Computer Science
The Hong Kong University of Science and
Technology
Clear Water Bay, Kowloon
Hong Kong

Summary

‘Anytime, anywhere’ communication, information access and processing are much cherished in modern societies because of their ability to bring flexibility, freedom and increased efficiency to individuals and organizations. Wireless communications, by providing ubiquitous and tetherless network connectivity to mobile users, are therefore bound to play a major role in the advancement of our society. Although initial proposals and implementations of wireless communications are generally focused on near-term voice and electronic messaging applications, it is recognized that future wireless communications will have to evolve towards supporting a wider range of applications, including voice, video, data, images and connections to wired networks. This implies that future wireless networks must provide quality-of-service (QoS) guarantees to various multimedia applications in a wireless environment.

Typical traffic in multimedia applications can be classified as either Constant-Bit-Rate (CBR) traffic or Variable-Bit-Rate (VBR) traffic. In particular, scheduling the transmission of VBR multimedia traffic streams in a wireless environment is very challenging and is still an open problem. In general, there are two ways to guarantee the QoS of VBR multimedia streams, either deterministically or statistically. In particular, most connection admission control (CAC) algorithms and medium access control (MAC) protocols that have been proposed for multimedia wireless networks only provide statistical, or soft, QoS guarantees.

In this paper, we consider deterministic QoS guarantees in multimedia wireless networks. We propose a method for constructing a packet-dropping mechanism that is based on a mathematical framework that determines how many packets can be dropped while the required QoS can

*Correspondence to: Mounir Hamdi, Department of Computer Science, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong.

[†]E-mail: hamdi@cs.ust.hk

[‡]This work is supported in part by HKTTIT.

still be preserved. This is achieved by employing: (1) An accurate traffic characterization of the VBR multimedia traffic streams; (2) A traffic regulator that can provide bounded packet loss and (3) A traffic scheduler that can provide bounded packet delay. The combination of traffic characterization, regulation and scheduling can provide bounded loss and delay deterministically. This is a distinction from traditional deterministic QoS schemes in which a 0% packet loss are always assumed with deterministically bounding the delay.

We performed a set of performance evaluation experiments. The results will demonstrate that our proposed QoS guarantee schemes can significantly support more connections than a system, which does not allow any loss, at the same required QoS. Moreover, from our evaluation experiments, we found that the proposed algorithms are able to out-perform scheduling algorithms adopted in state-of-the-art wireless MAC protocols, for example Mobile Access Scheme Based on Contention and Reservation for ATM (MASCARA) when the worst-case traffic is being considered. Copyright © 2002 John Wiley & Sons, Ltd.

KEY WORDS

admission control
quality-of-service guarantees
multimedia
wireless networks

Published online: 11 July 2002

1. Introduction

With the rapid advances in wireless transmission technologies, there is a great interest in using high-speed wireless networks for the transmission of not only traditional discrete media such as text and still images, but also multimedia traffic such as real-time images, voice and video [1]. Multimedia applications are different from traditional applications in that they require QoS guarantees in terms of delay, delay variation and loss rate.

The traffic characteristics and real-time nature of these multimedia applications pose new challenges

to the design, implementation and management of future wireless networks [2]. To provide QoS to multimedia applications, a resource reservation scheme is needed to allocate wireless resources to the different connections. Since the characteristics of most multimedia traffic are in the category of VBR sources with considerable burstiness, designing an optimal or a good resource allocation scheme for these VBR sources becomes especially difficult. If the resources are reserved according to the average traffic rates, unacceptable high delays or unacceptable high packet losses may result when the sources transmit near its peak rate. On the other hand, if the resource reservations are based on the peak rates, the network may be underutilized most of the time [3].

In general, there are two ways to guarantee the QoS of the VBR connections, either deterministically or statistically. Statistical models normally would employ stochastic traffic models, including Markov modulated, self-similar, S-BIND, among others [4]. However, these models are either not powerful enough to capture the important burstiness and time correlation of realistic sources, or are too complex for practical implementation. Deterministic models would normally characterize their traffic by burst length, average rate and/or peak rate. Because of their pessimistic nature, the throughput and the channel utilization of deterministic models are relatively smaller than that of statistical models. The advantage of deterministic models, unlike statistical models, is that they can provide deterministic guarantees such that no connection would violate its desired QoS [3].

In this paper, we consider deterministic QoS guarantees in a wireless environment. We propose a method for constructing a packet-dropping mechanism that makes decisions as to when we need to drop packets while still not violating any QoS guarantees. This is achieved through an accurate traffic characterization of the VBR multimedia traffic streams. We also employ a traffic regulator that can provide bounded packet losses and a traffic scheduler that can provide bounded packet delays. Hence, their combination can provide bounded loss and delay deterministically. This is a distinction from traditional deterministic QoS schemes in which a 0% packet loss is always assumed with deterministically bounding the delay. By allowing certain bounded packet loss, it is possible to admit more connections to the wireless networks. Most multimedia connections, for instance video and audio connections, can accept a certain percentage of losses while still providing an acceptable QoS.

We employ a traffic regulator at the entrance position of the network for each connection to deterministically bound the packet loss while reducing the burstiness and the peak rate of the traffic. Then, the regulated traffic is passed to the scheduling algorithms, which can provide packet delay-bounds, to test whether such connections should be admitted. Since the traffic regulator can provide a bounded loss and the scheduler can provide a bounded delay guarantee, their combination can provide *both* bounded loss and bounded delay deterministically.

1.1. Background: Medium Access Control Protocols for Wireless Networks

Radio channels constitute a finite precious resource in wireless networks, and therefore they must be

efficiently shared by all associated mobile stations. As a result, a key technical issue related to the design of wireless networks is the selection of a suitable wireless channel sharing (access) technique [1,5]. Wireless channel access schemes could be based on circuit switching or packet switching. However, packet switching has been widely recognized as more flexible than circuit switching especially for multimedia communications [1]. Furthermore, wireless access schemes could be based on code division multiple access (CDMA), frequency division multiple access (FDMA), time division multiple access (TDMA) or combinations of these techniques. Because of the lack of available frequencies and the requirement for dynamic bandwidth allocation, especially for VBR connections, it makes 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 (> 2 Mb/s). Accordingly, most protocols in the area of high-speed wireless networks use an adaptive TDMA scheme, due to its ability and flexibility to accommodate a connection's bit rate needs by allocating more or fewer time slots depending on current traffic conditions [6]. As a result, we will focus on TDMA schemes in this paper.

In a TDMA wireless network, each mobile station, transmitting data, is allocated one or more time slots within a frame in the downstream (base to mobile users) and upstream (mobile users to station) directions [2]. In the downstream direction, the base station broadcasts to the active mobile stations in a Time Division Multiplex (TDM) format. In the upstream direction, each active mobile station transmits to the base station using well-defined wireless access protocols.

A plethora of wireless access protocols have been proposed and/or implemented for wireless personal communication systems [1,3]. They can be broadly classified into three categories:

- 1) *Random access protocols*: The mobile stations access time slots within a frame on demand and with no coordination between them. Thus, when more than one packet is transmitted at the same time, a collision occurs, and the information contained in all the transmitted packets is lost. As a result, a collision resolution mechanism must be devised to avoid high-packet delays or even unstable operation under high loads. The wireless access mechanisms under this category are known

as ALOHA-type protocols or CSMA/CD-type protocols [7,8]. These protocols attain good performance under light traffic, however, their performance is questionable under heavy-traffic conditions.

- 2) *Reservation-based protocols*: A mobile station that is preparing a packet transfer has to reserve a time slot through reservation/allocation cycles, to dynamically allocate the available bandwidth to connections based on their current needs and traffic loads. The allocation of data slots is performed by the base station on the basis of a scheduling algorithm, and the mobile terminals are informed through broadcast messages. These kind of protocols are more complex; but, on the other hand, they are stable under a wide range of traffic loads and can guarantee a predictable QoS. Numerous MAC protocols have been proposed, which fall under this category and include the Dynamic slot Assignment (DSA++) protocol [9], the Distributed Queuing Request Update Multiple Access (DQRUMA) [10] and the MASCARA [11].
- 3) *Polling-based protocols*: The third group of protocols uses adaptive polling to distribute bandwidth among connections. A slot is periodically given to each connection, without request, based on its expected traffic. Chang *et al.* [12] proposed a polling scheme with non-preemptive priority (PNP), where information exchange occurs in time to meet QoS of CBR sources and VBR sources. Mahmoud *et al.* [13] have also proposed an adaptive polling protocol.

Most of the above wireless access protocols are targeted towards voice and messaging-type data applications. Currently, new wireless access protocols aiming at providing services for multimedia applications are the object of strong research work [14]. These new protocols represent a good starting point for providing service for future multimedia traffic needs. In particular, they are able to differentiate between various traffic classes. However, these protocols are not designed to provide the QoS guarantees that are needed by multimedia applications. The QoS guarantees can be achieved only through the design of efficient scheduling and admission control algorithms tailored towards a given wireless access protocol.

1.2. Motivation

Wireless channels constitute a finite precious resource in a wireless network, and therefore they must be efficiently shared by all associated mobile stations. Thus, unlike wireline communications wherein an increasing demand for bandwidth can be met by deploying additional wire (or fiber) facilities, the available radio spectrum cannot arbitrarily be expanded. As a result, a key technical issue related to the design of wireless networks is the selection of a suitable wireless media MAC protocol [1,5].

In addition, many CAC algorithms and MAC protocols have been proposed for multimedia wireless networks. Nearly all of the proposed algorithms only provide statistical, or soft, QoS guarantees. These MAC protocols can only provide QoS guarantees to the connections on the general case. Even if a connection is admitted to the network, the network cannot fully (100%) guarantee that no QoS of any connections will be violated. The required or the desired QoS of some connections may still be violated in some cases.

On the other hand, there are some multimedia applications that need deterministic, or hard, QoS guarantees. In multimedia wireless networks, there are some proposed protocols that can provide deterministic QoS guarantees. Chang *et al.* [12] proposed a polling scheme to provide such deterministic guarantees. Liebeherr *et al.* [15] proposed a CAC mechanism capable of providing deterministic guarantees. Most of these protocols can only provide deterministic bound for delay with zero packet loss. To the best of our knowledge, there are no protocols or algorithms that can provide deterministic bounding for both packet delay and packet loss in multimedia applications.

In this paper, we propose a CAC algorithm and a MAC protocol that aim to provide such deterministic bounds for both packet loss and packet delay in a wireless environment. By allowing certain bounded packet loss for the connections, it is possible to admit more connections to the network and increase the network utilization. This is motivated by the fact that most multimedia connections, for instance video and audio connections, can accept a certain percentage of packet losses while still providing an acceptable QoS.

1.3. Organization of the Paper

In this paper, we present traffic characterization methods and packet-dropping schemes that yield deterministic bounds for both packet loss rate and packet delay

while maintaining an efficient utilization of a wireless network. In the introduction of this paper, we have provided the context and motivation for the problems we address in this paper. The remainder of this paper is organized as follows:

In Section 2, we discuss the necessary network support in order to achieve a bounded QoS and describe its three main components: traffic characterizations, packet-scheduling disciplines and the delay bound tests.

In Section 3, we review some related research by focusing on the deterministic traffic characterizations, their respective policing mechanisms and packet-scheduling disciplines that can provide deterministic QoS guarantees in a wireless environment.

In Section 4, we introduce various proposed methods for bounding the packet loss and packet delay deterministically and also prove the optimality of one of our proposed algorithms for a single class of connections.

In Section 5, we conduct performance evaluations for our proposed schemes. We compare our results with the well-known wireless MAC protocols, for example, MASCARA.

In Section 6, we conclude and summarize our contributions to the paper.

2. Network Support for a Bounded Quality-of-service

2.1. Introduction

A network with a bound-delay service requires a resource reservation scheme to allocate network resources for different connections. In the design of this reservation scheme, two components are crucial. The first one is the admission control scheme, which limits the number of connections admitted to the network in order to avoid any of the connections violating the desired QoS. The second one is the traffic-policing mechanism that monitors whether the traffic input rate of each individual connection matches its specified traffic rate. In this Section, we describe the three central components to the design of the admission control and the traffic-policing mechanisms: traffic characterization, packet-scheduling disciplines and delay-bound tests [16].

2.2. Traffic Characterization

In order to provide deterministic bounded QoS, the traffic characterizations must satisfy three

requirements. First, it must provide the worst-case description of the traffic source such that the network can determine the upper bound of the traffic arrivals. Second, it must be policeable as the traffic-policing mechanism is one of the essential components. Finally, it should describe the traffic accurately so that the admission control mechanisms do not overestimate the resources required by the connections.

Since a deterministic service provides worst-case guarantees, a traffic characterization must specify the worst-case traffic of a connection. We use A to denote the actual traffic of a connection, where $A[\tau, \tau + t]$ denotes the amount of traffic arrivals in the time interval $[\tau, \tau + t]$. The worst-case traffic characterization A is described by a traffic-constraint function A^* , which provides an upper bound on A . A traffic-constraint function A^* should satisfy two important properties, namely time-invariance and subadditivity [15,17]. Since a time-constraint function A^* bounds the maximum traffic over any time interval t , the delay-bound tests can be made independent of the starting time of a connection and it should satisfy the following equation for $\tau \geq 0$ and $t \geq 0$:

$$A^*(t) \geq A[\tau, \tau + t]$$

A traffic-constraint function A^* is subadditive if it satisfies the following inequality:

$$A^*(t_1) + A^*(t_2) \geq A^*(t_1 + t_2) \text{ for all } t_1, t_2 \geq 0$$

The subadditive nature of the traffic-constraint function allows the arrivals on a connection to attain the bound given by A^* . In other words, it is feasible that $A^*(t) = A[\tau, \tau + t]$ for any $t \geq 0$. Figure 1 further illustrates the relationship between A and A^* .

2.3. Packet-scheduling Disciplines

The packet scheduler is central in controlling the end-to-end delay of packets in a QoS network. Packets

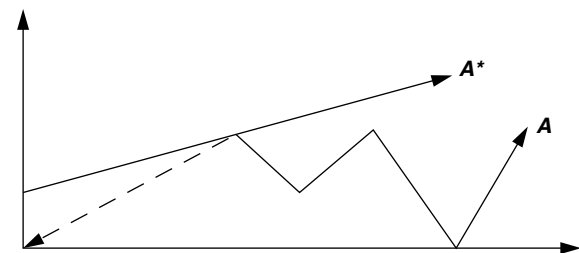


Fig. 1. Relationship between traffic A and traffic-constraint function A^* .

from different connections waiting to be transmitted are stored in the transmission queue of the link and the packet scheduler determines the order of the transmission. There are many varieties of packet-scheduling disciplines. We can classify them into two major categories, either work-conserving or non-work-conserving. The work-conserving packet scheduler is never idle if the queue is not empty, while the non-conserving-packet scheduler may be idle even if there are some queued packets.

A scheduling discipline may also be classified as preemptive or non-preemptive. A preemptive-scheduling discipline may suspend the transmission of one packet in order to transmit another, while a non-preemptive discipline does not interrupt the transmission of packets. In this paper, we consider packet schedulers that are both work-conserving and non-preemptive.

There are a large number of packet-scheduling disciplines that can provide bounds on delay, but not all of them use the network resources efficiently. The performance of a packet scheduler in providing bounded QoS can be determined by the degree to which it satisfies the following requirements [15]: efficiency, flexibility, complexity and analyzability.

2.4. Delay-bound Tests

The delay-bound test is central to the CAC. The decision of the admission control is mainly dependent on whether the delay-bound test is passed or not. If the delay-bound test fails after the network admits the new connection, the failure to pass the test means that some connections may violate the desired QoS; so the admission control will reject the admission of the new connection.

A delay-bound test depends heavily upon the choice of the packet scheduler and the traffic characterization model. Normally, the delay-bound test is designed specifically for each packet scheduler and uses the traffic-constraint functions as parameters. In addition, the properties of the packet scheduler and the accuracy of the traffic model are directly reflected in the delay bound test.

3. Related Work

3.1. Introduction

Although there is also a lot of related research on multimedia wireless MAC protocols and their CAC algorithms, nearly all of the proposed algorithms only

provide statistical QoS guarantees. As a result, these MAC protocols may not be able to satisfy connections that demand deterministic QoS guarantees.

In addition, there is also a lot of related research on both traffic characterization and packet scheduling, but much of this work cannot be directly applied to support traffic that requires a bounded QoS. For example, some approaches characterize the traffic sources using sophisticated stochastic models such as Markov-modulated [18], autoregressive [19,20], self-similar [2,21,16], TES [22], and S-BIND [23]. These approaches are not capable for providing worse-case bounds on traffic arrivals and, therefore, cannot be used as the traffic characterization for deterministic bounded connections. Furthermore, the design of a real-time traffic-policing mechanism for a stochastic traffic model is very difficult.

In this Section, we review some approaches, that are suitable for use in networks with bounded QoS.

3.2. Medium Access Control Protocols for Wireless Environments

As discussed previously, most multimedia wireless MAC protocols use adaptive TDMA schemes because of their flexibility to accommodate a connection's bit rate needs by allocating a variable number of time slots depending on current traffic conditions. In particular, most TDMA proposals use random access techniques such as Slotted ALOHA (or variants of this protocol) for the dialup process and reservation of slots for transmission from mobile terminals to the base station [9].

In this section, we review six wireless MAC protocols because they are regarded as state-of-art wireless access protocols in this area.

3.2.1. Packet reservation multiple access

There is plethora of MAC protocols that have been proposed for wireless networks. One of the most well-known and the most-widely studied protocol is the Packet Reservation Multiple Access (PRMA) [24]. PRMA is a Slotted ALOHA reservation-based TDMA protocol. But unlike Slotted ALOHA, the contentions for a time slot in PRMA occur only at the beginning of each talk-spurt of the conversation and unlike traditional TDMA, PRMA allows a voice user to reserve a slot only during each talk-spurt rather than during the whole conversation [22].

There are a lot of variants of PRMA that have been proposed. For example, [1] proposed and studied an

improved version of PRMA in which a minimum portion of the available channel capacity is dedicated to the reservation channel. Slotted ALOHA contentions occur in some slots that are further divided into mini-slots. As a result, the throughput performance under high-load conditions is improved [22].

3.2.2. Packet reservation multiple access with dynamic allocation

Packet Reservation Multiple Access with Dynamic Allocation (PRMA/DA) is another enhanced version of the PRMA protocol [25]. PRMA was originally designed for voice and data traffic only, while PRMA/DA considers CBR, VBR and data traffic [9]. In PRMA/DA, a dynamic allocation mechanism is proposed to estimate the number of contention slots needed. As a result, this mechanism help resolve the contention situation quickly while avoiding the waste of bandwidth in allocating too many slots as contention slots. One of the criticisms of PRMA/DA is that it does not use mini-slots for the access request [9].

The frame format of PRMA/DA is shown in Figure 2.

3.2.3. Distributed queuing request update multiple access

DQRUMA protocol was proposed by Karol *et al.* [10]. DQRUMA considers a time-slotted system with no frame reference, where the request access and packet transmission channels are formed on a slot-by-slot basis [9]. The uplink is divided into a series of mini-slots used for requesting access, each one followed by a slot for packet transmission. The downlink consists of a series of mini-slots for acknowledgement of request accesses, each followed by a slot for

packet transmission. After the base station receives a successful request from a mobile terminal, it immediately sends the corresponding acknowledgment in the appropriate downlink mini-slot.

The frame format of DQRUMA is shown in Figure 3.

3.2.4. Dynamic slot assignment

The DSA++ protocol was proposed by Petras *et al.* [26]. DSA++ is a modified version of the Dynamic Slot Assignment (DSA) and it uses a variable-length frame structure. The assignment of capacity is based on a priority calculation for each mobile terminal being served. The priority is determined according to a set of dynamic parameters, which include the number of waiting packets and their due dates. The advantage of this protocol is that it allows the base station to implement power control as the broadcast of the information about the next signalling period is in a single downlink burst.

The frame format of DSA++ is shown in Figure 4.

3.2.5. Mobile access scheme based on contention and reservation for ATM

One of the most well-respected wireless MAC protocols is MASCARA, which is proposed by Bauchot *et al.* [11] as the MAC protocol for the WATM Network Demonstrator (WAND) project being developed with the support of the European Community [9]. In this paper, we conducted performance evaluation of our proposed protocol and compared our results with MASCARA. An in-depth description of the MASCARA protocol and the evaluation results will be given later in the paper.

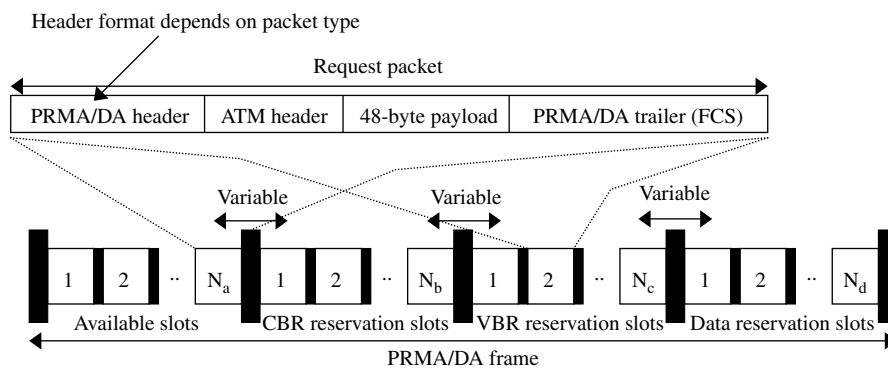


Fig. 2. Frame format of PRMA/DA.

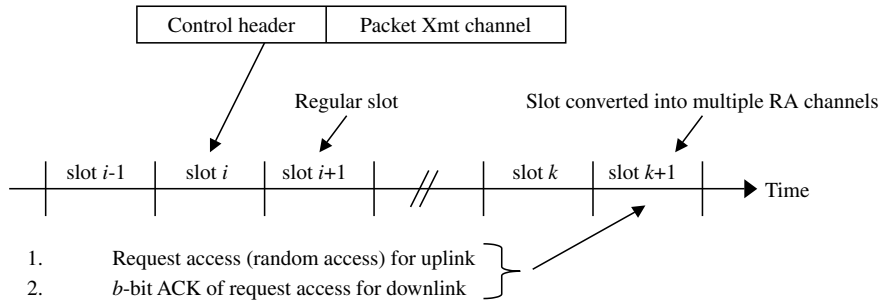


Fig. 3. Frame format of DQRUMA.

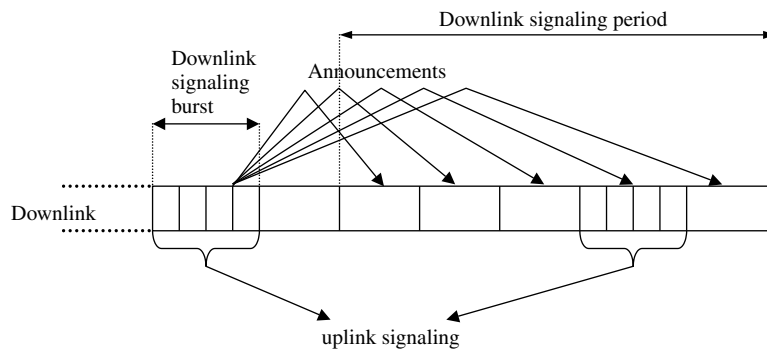


Fig. 4. Frame format of DSA++.

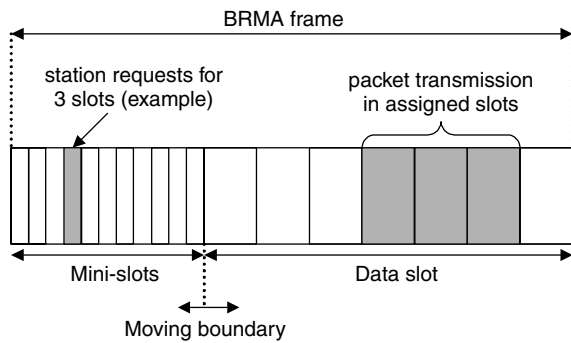


Fig. 5. Frame format of BRMA.

3.2.6. Bandwidth reservation multiple access

The Bandwidth Reservation Multiple Access (BRMA) is proposed by Z. Zhang *et al.* [22] to resolve users contention and assign bandwidth among multiple users trying to gain access to a common channel such as in mobile users contending for resources in a wireless local area network. In BRMA, the assignment of the minislots is deterministic; thus, both the request channels and the data channels are collision-free.

The frame format of BRMA is shown in Figure 5.

3.2.7. Discussion of tradeoffs

For reference, we summarize the characteristics of the protocols reviewed in Table I. Some of the entries are based on subjective judgments rather than formal analysis.

From the comparative results, it appears that the MAC protocols that use FDD can deal with the access contention procedures more quickly because they can send acknowledgment signals almost immediately after receiving the packets. However, TDD can be advantageous in some situations in which frequencies are scarce or the uplink and the downlink is unbalanced.

3.3. Traffic Characterization

In order to provide deterministic bounded QoS, traffic characterization must specify the worst-case traffic characteristic of a connection. In this section, we review six models: the peak-rate model [27], the (σ, ρ) -model [5], the $(\vec{\sigma}, \vec{\rho})$ -model [28], the (r, T) -model [29], the $(X_{\min}, X_{ave}, I, s^{\max})$ -model [30] and

Table I. Summary of wireless MAC protocols.

	PRMA	PRMA/DA	DQRUMA	DSA++	MASCARA	BRMA
Physical Layer Type	FDD	FDD	FDD	FDD	TDD	TDD
Frame Type	No Frame	Fixed	No Frame	Variable	Variable	Variable
QoS Support	voice	voice, video, data	VBR	CBR, VBR, ABR	CBR, rt-VBR, nrt-VBR, ABR, UBR	CBR, VBR, ABR
Control overhead	Low	Medium	Low	High	High	High
Random Access Technique	S-ALOHA	S-ALOHA	S-ALOHA, binary stock	Splitting algorithm (2 and 3)	Not defined yet	No
Call Admission Control	No	Yes	No	No	Yes	No

the D-BIND model [31]. We formulate the traffic-constraint function A^* for each of these traffic models and also discuss the traffic-policing mechanisms used for enforcement. We then evaluate and compare the trade-offs involved with each traffic model.

3.3.1. The peak-rate model

The peak-rate model [27] is the simplest and the most widely used of all traffic models [7]. This peak-rate model describes the traffic by two parameters: minimum interarrival time, X_{\min} and the maximum transmission time, s^{\max} . The traffic-constraint function is given by the following:

$$A^*(t) = \left(\left\lfloor \frac{t}{X_{\min}} \right\rfloor + 1 \right) s^{\max} \text{ for all } t \geq 0$$

The peak-rate model can specify CBR traffic very well but over-estimates the resource requirements for VBR sources.

3.3.2. The (σ, ρ) -model

The (σ, ρ) -model [5] describes its traffic by two parameters: the burstiness σ and an average traffic rate ρ . The traffic-constraint function is given by the following:

$$A^*(t) = \sigma + \rho t \text{ for all } t \geq 0$$

This model enforces a rate ρ while allowing some burstiness up to σ . The traffic of this model can be policed with the leaky-bucket policing mechanism.

3.3.3. The $(\vec{\sigma}, \vec{\rho})$ -model

This model is a generalization of the (σ, ρ) -model [28]. It describes its traffic by a set of (σ_i, ρ_i)

pairs. The amount of traffic admitted to the network is limited by each of the (σ_i, ρ_i) pairs. The traffic-constraint function is given by:

$$A_m^*(t) = \min\{\sigma_i + \rho_i t\} \text{ for all } t \geq 0 \text{ and } 1 \leq i \leq m$$

Note that when $m = 1$, this model is identical to the $(\vec{\sigma}, \vec{\rho})$ -model.

3.3.4. The (r, T) -model

The (r, T) -model [29] describes its traffic by two parameters: a rate parameter r and a framing interval T . Time is partitioned into frames of length T and the maximum traffic of a connection during any frame is limited to rT bits. Thus, this model enforces an average rate of r while allowing for moderate bursts. The traffic-constraint function is given by the following:

$$A^*(t) = \left(\left\lceil \frac{t}{T} \right\rceil + 1 \right) rT$$

The maximum burst that may enter the network is $2rT$. The traffic of this model can be policed with the jumping window policing mechanism.

3.3.5. The $(X_{\min}, X_{\text{ave}}, I, s^{\max})$ -model

The $(X_{\min}, X_{\text{ave}}, I, s^{\max})$ -model [30] describes its traffic by four parameters. X_{\min} is the minimum packet interarrival time, X_{ave} is the maximum average packet interarrival time over any time interval of length I , and s^{\max} is the maximum packet transmission time. The traffic-constraint function is given by the

following:

$$A^*(t) = \left\lfloor \frac{t}{I} \right\rfloor \frac{I s^{\max}}{X_{\text{ave}}} + \min \left(\left\{ \left\lfloor \left(\frac{t}{I} - \left\lfloor \frac{t}{I} \right\rfloor \right) \times \frac{I}{X_{\min}} \right\rfloor \frac{I}{X_{\text{ave}}} \right\} s^{\max} \right)$$

The traffic of this model can be policed with a combination of a cell-spacing mechanism and a moving window mechanism.

3.3.6. The D-BIND model

The D-BIND model is a general traffic model that uses a number of rate-interval pairs (R_i, I_i) [4]. The maximum rate over any interval of length I_i is restricted to R_i for all pairs i . The traffic-constraint function is given by:

$$A^*(t) = \frac{R_i I_i - R_{i-1} I_{i-1}}{I_i - I_{i-1}} + R_i I_i$$

for all $I_{i-1} \leq t \leq I_i$.

The traffic of this model can be policed with n moving window mechanisms, one for each (r, I) pair.

3.3.7. Discussion and trade-offs of the various traffic models

For reference, we summarize the relationship between the traffic models and their policing mechanisms in Table II.

A study by Reibman and Berger [31] and another by Rathgeb [32] evaluate the accuracy with which the various traffic models can characterize VBR video. Rathgeb shows how the parameters of each traffic-policing mechanism can be expressed in terms of parameters of the other mechanisms, enabling a direct

comparison of the various mechanisms. The examples presented indicate that the leaky bucket is superior to both the windowing mechanisms for describing VBR video since neither the jumping windowing nor the moving windowing are capable of capturing the short-term burstiness.

Wrege *et al.* [16] showed that the $(\vec{\sigma}, \vec{\rho})$ traffic models, which employ multiple leaky-bucket mechanisms, can accurately characterize VBR video.

Both Dittmann [33] and Rathgeb [32] showed that the moving window mechanism is significantly more difficult to implement than either the jumping window or leaky-bucket mechanisms.

For these reasons, the networking community has focused primarily on characterizations that can be policed by the leaky-bucket mechanisms, that is, the (σ, ρ) and $(\vec{\sigma}, \vec{\rho})$ traffic models.

3.4. Packet-scheduling Schemes and Delay-bound Tests

Packet-scheduling schemes for providing bounded QoS can be classified into two classes: delay-based disciplines and rate-based disciplines. The delay-based schedulers can provide maximum delay guarantees to connections. The admission control mechanisms for delay-based schedulers calculate delay guarantees for a connection based on the worst-case traffic characteristic of all connections and the properties of the packet scheduler. On the other hand, the rate-based schedulers can provide minimum throughput guarantees. They allocate a fraction of the available bandwidth to each connection and calculate delay guarantees on the basis of their traffic characterizations.

Table II. Summary of traffic models.

Traffic Model	Policing Mechanisms	Traffic-constraint Function
Peak-rate	Packet spacer	$A^*(t) = \left(\left\lfloor \frac{t}{X_{\min}} \right\rfloor + 1 \right) s^{\max}$
(σ, ρ)	Jumping window	$A^*(t) = \sigma + \rho t$
$(\vec{\sigma}, \vec{\rho})$	Leaky bucket	$A_m^*(t) = \min\{\sigma_i + \rho_i t\}$
(r, T)	Multiple leaky buckets	$A^*(t) = \left(\left\lfloor \frac{t}{T} \right\rfloor + 1 \right) rT$
$(X_{\min}, X_{\text{ave}}, I, s^{\max})$	Moving window and spacer	$A^*(t) = \left\lfloor \frac{t}{I} \right\rfloor \frac{I s^{\max}}{X_{\text{ave}}} + \min \left(\left\{ \left\lfloor \left(\frac{t}{I} - \left\lfloor \frac{t}{I} \right\rfloor \right) \times \frac{I}{X_{\min}} \right\rfloor \frac{I}{X_{\text{ave}}} \right\} s^{\max} \right)$
D-BIND	Multiple moving windows	$A^*(t) = \frac{R_i I_i - R_{i-1} I_{i-1}}{I_i - I_{i-1}} + R_i I_i$

3.4.1. Delay-based scheduling disciplines

In this section, we review two delay-based scheduling disciplines that have been previously considered for bounded QoS: Earliest-Deadline-First (EDF) [34] and Static-Priority (SP) [35]. We then review their corresponding delay-bound tests that can be employed by the admission control.

Earliest-deadline-first (EDF). An EDF scheduler assigns each arriving packet a time-stamp corresponding to its deadline. For instance, a packet from a connection j with a delay bound d_j that arrives at time t is assigned a deadline of $t + d_j$. The EDF scheduler always selects the packets with earliest deadline for transmission.

The EDF scheduler can provide optimal network utilization as the EDF scheduler can support any connections that can be supported by any other schedulers. The high achievable network utilization makes EDF an excellent candidate for bounded QoS.

Without loss of generality, connections are assumed to be ordered so that $i < j$ whenever $d_i < d_j$. Thus the schedulability condition for an EDF scheduler is given as follows for a set of connections C :

$$t \geq \sum_{j \in C} A_j^*(t - d_j) + \max s_k$$

Static-priority (SP). An SP-scheduler classifies the connections with the same delay bound into the same connection set. The connection sets are sorted in order such that $d_p < d_q$ for $p < q$. Thus, SP maintains a set of priority FIFO queues and the priority of a queue is high if its delay bound is small. A packet of a connection j in connection set C_p is inserted into FIFO p . At the beginning of a busy period, or after completing the transmission of a packet, the SP-scheduler always selects the first packet in the nonempty FIFO queue with the highest priority for transmission.

Because of its simplicity, it enables packet scheduling at very high data rates. The following necessary and sufficient schedulability condition for SP schedulers is presented:

$$t + \tau \geq \sum_{j \in C_p} A_j^*(t) + \sum_{q=1}^{p-1} \sum_{j \in C_q} A_j^*((t + \tau)^-) - s_p^{\min} + \max_{r > p} s_r$$

3.4.2. Rate-based scheduling disciplines

All rate-based scheduling disciplines emulate one of two systems: (1) a TDM system that divides time into fixed-sized frames that are in turn divided into time slots, allocating these slots to connections, or (2) a Generalized-Processor-Sharing (GPS) system that allocates a service share to each connection and provides service to each connection in proportion to its share.

Note that similar delay-bound tests in delay-based schedulers are not available in rate-based schedulers.

3.4.3. Discussion and trade-offs

With regard to a high achievable network utilization, EDF is the better choice than any of the rate-based disciplines for a number of reasons. First, the EDF is the optimal packet scheduler for a single network switch. Second, delay guarantees for rate-based packet schedulers are only available for a restricted set of traffic models. Finally, the delay guarantees for rate-based packet schedulers are not in the form of delay-bound tests that can be easily applied to admission control mechanisms but are rather computed as a maximum delay bound based on throughput guarantees.

4. Proposed Approach to Bound Packet Losses and Packet Delays Deterministically

4.1. Introduction

As was shown in the previous Section, there are many admission and scheduling algorithms that have been proposed to provide deterministic delay-bound guarantees while trying to provide efficient utilization of the network resources. Most of these schemes assume that no packets of any connection can be lost. However, in practical situations, most traffic can accept a certain percentage of packet losses while still providing acceptable QoS. For example, in an Motion Picture Experts Group (MPEG) video, B frames are relatively less important than the I frames and the P frames. If a certain percentage of B frames in an MPEG video stream is lost, the resulting video would only have minor effects and would still be considered in a good quality. Another example involves QoS for voice connections. If the average packet loss rate of a voice connection is below a certain percentage, the difference in voice quality is also minor, and normally, people cannot distinguish the difference.

Obviously, allowing some packet losses of the connections would possibly allow to admit more connections to the network. However, to the best of our knowledge, there are no schemes that can provide deterministic bounded delay and packet loss guarantees for QoS networks. Thus, in this paper, we aim to provide such deterministic bounds for both packet loss and packet delay.

4.2. The Proposed Algorithm

Different from the previous research efforts that do not allow any packet loss, our QoS methodology includes an additional scheme for dropping packets. The major task of this packet-dropping scheme is to decide which packets are dropped in order to maintain the desired QoS.

There are two ways to implement the packet-dropping scheme. One method is to place the packet-dropping scheme together with the packet-scheduling scheme. Actually, most of the algorithms that provide statistical bounded QoS employ this method. The advantage of this method is that packets only need to be dropped when the network cannot serve the packets, but the disadvantage is that the droppings cannot be deterministically estimated or calculated. Thus, this kind of method is not suitable for any algorithm that aims to provide deterministic bounded QoS.

The other method is to place the packet-dropping scheme before the packets pass into the network. The packet-dropping scheme acts as a filter for the traffic, which reduce the burstiness and the peak rate of the traffic as much as possible while maintaining the loss below the desired level. As the dropping is performed before the packets pass into the network, possibly that some unnecessary dropped packets are dropped too; but the advantage of this approach is that the loss rate can be deterministically bounded.

The idea of our proposed algorithm is generated from the second-mentioned packet-dropping scheme. The traffic of each connection is passed through a packet-dropping scheme before passing into the network and the resulting traffic is scheduled by the packet scheduler, which can provide zero packet loss and bounded delay guarantee. Since the packet-dropping scheme can provide a bounded loss and the packet scheduler can provide zero loss and bounded delay guarantee, their combination can provide bounded loss and delay deterministically.

Because packets are allowed to be lost, and the packet-dropping process reduces part of the burstiness

and the peak rate of the traffic, the overall throughput and the channel efficiency of the network can be increased, *which is the major reason behind our scheme*. We will discuss the performance of the proposed algorithm later.

4.2.1. The proposed wireless MAC protocol

The proposed wireless MAC protocol is motivated by the BRMA [22]. Since the assignment of the mini-slots in BRMA is deterministic, both the request channels and the data channels are contention-free. This characteristic is suitable for us to provide deterministic QoS guarantee. Similar to other TDMA systems, time is divided into time slots and time slots are further divided into mini-slots and data slots. Mini-slots are for sending requests from the terminals to the base station, while the data slots are for sending real data or packets to the base station.

In BRMA, each connection is assigned one mini-slot in each frame. As a result, each connection does not need to contend for the channel with other connections. But in our proposed protocol, we have to deal with VBR, CBR, ABR and UBR traffic and only CBR and VBR connections need to have deterministic QoS guarantees. So, in our protocol, only CBR and VBR connections will be assigned one mini-slot in each frame, while ABR and UBR will contend for the channel through random-access methods for example, Slotted ALOHA. In our proposed MAC protocol, we simply use Slotted ALOHA to reduce the complexity of the network. The structure of the proposed protocol is shown in Figure 6.

4.2.2. The proposed traffic characterization

The proposed traffic model is based on the (σ, ρ) -model but with an important modification. Two more parameters are added which are the maximum traffic rate, ρ_{\max} and the minimum traffic rate, ρ_{\min} .

In the (σ, ρ) -model, σ defines the burstiness of the traffic. In the worst case, the (σ, ρ) -model assumes that all the burstiness arrive at the same time, but in general, the burstiness do not come at the same time and it would normally come within a certain

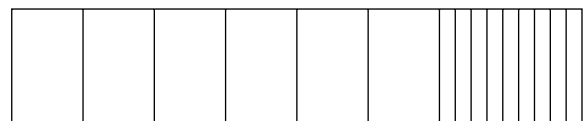


Fig. 6. Frame format of the proposed wireless MAC protocol.

duration. The objective of adding the maximum traffic rate ρ_{\max} is to characterize this fact. After adding the maximum traffic rate ρ_{\max} the traffic characterization model becomes similar to the special case of the (σ, ρ) -model with two pairs, (σ, ρ) and $(0, \rho_{\max})$.

In addition, most of the multimedia traffic is continuous traffic, for instance, video, audio and voice connections. During the connection, packets are sent continuously and thus the connections have a minimum traffic rate. To characterize this fact, we add a minimum traffic rate ρ_{\min} as a parameter to the proposed traffic characterization model.

As a result, the proposed traffic model has four parameters: burstiness σ , maximum traffic rate ρ_{\max} , minimum traffic rate ρ_{\min} and average traffic rate ρ . The traffic-constraint function is given by the following:

$$A^*(t) = \min(\sigma + \rho t, \rho_{\max}^*) \text{ for all } t \geq 0$$

4.2.3. The proposed packet-dropping mechanism

The task of the packet-dropping mechanism is to select the packets to be dropped for each connection. A good packet-dropping scheme for a bounded QoS should satisfy the following criteria: (1) reducing the burstiness of the traffic, (2) reducing the maximum traffic rate, (3) distributing the dropping of packets evenly and (4) conforming to the specified loss rate.

Because of the fact that the packet-dropping mechanism has to deal with real-time traffic and it is impossible to determine the duration of the traffic before the transmission is ended, the loss rate has to be kept at or below the desired level all the time.

THEOREM 4.1 *The optimal dropping of $L\%$ packets for the worst case of the $(\sigma, \rho, \rho_{\max}, \rho_{\min})$ traffic model is the same as the worst case of the $(\sigma', \rho', \rho'_{\max}, \rho'_{\min})$ traffic model where $\sigma' = \max(0, \sigma(1-Lf))$, $\rho' = \rho(1-L)$, $\rho'_{\max} = \min(\rho(1-L), (\rho_{\max}\rho)(1-Lf) + \rho(1-L))$, $\rho'_{\min} = \rho_{\min}$, and $f = \frac{1}{1 - (\rho_{\min}/\rho)}$.*

Proof. In the worst case, the traffic of the sources are similar to an extremal periodic on-off model, which is similar to the one represented in Figure 7.

An optimal packet-dropping algorithm should drop the same amount of packets in every burst. If more packets are dropped in one burst more than the other bursts, we can easily reduce the overall traffic

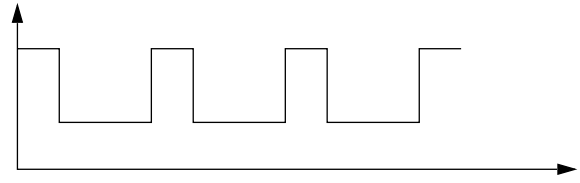


Fig. 7. The pattern of worst-case traffic.

burstiness by dropping one packet less in one burst and adding one more in the other burst. Since the packet dropping should be equal for all bursts, we can concentrate on one on-off cycle only as all cycles are the same.

Let: ρ_{\min} be the minimum traffic rate
 ρ_{\max} be the maximum traffic rate
 ρ be the average traffic rate
 σ be the maximum burstiness
 L be the allowable loss rate
 T_{on} be duration of the on cycle
 T_{off} be duration of the off cycle
 T be duration of one on-off cycle

According to Figure 7, we have:

$$T = T_{\text{on}} + T_{\text{off}}$$

According to the definition of maximum burstiness, $\sigma + \rho T_{\text{on}} \geq \rho_{\max} T_{\text{on}}$ and $2T_{\text{on}}\rho_{\max} + T_{\text{off}}\rho_{\min} \leq \rho(T + T_{\text{on}}) + \sigma$.

Since we are considering the worst case, we have by solving the equations:

$$T_{\text{on}} = \frac{\sigma}{\rho_{\max} - \rho}$$

and

$$T_{\text{off}} = \frac{\sigma}{\rho - \rho_{\min}}$$

To maintain the desired loss rate L , the maximum number of packets allowed to be dropped is $T\rho L$ in each on-off cycle.

$$\begin{aligned} &\text{maximum allowable losses} \\ &= \rho L \left(\frac{1}{\rho_{\max}/\rho - 1} + \frac{1}{1 - (\rho_{\min}/\rho)} \right) \end{aligned}$$

Without violating the desired loss rate, all allowable dropped packets can only be dropped while the traffic is in burst. We find that the resulting traffic can be characterized by the burstiness σ' , the average traffic rate ρ' , the maximum traffic rate ρ'_{\max} and the minimum traffic rate ρ'_{\min} .

As L percentage of the packets are dropped, obviously the average rate is also reduced by L . Therefore, we have:

$$\rho' = \rho(1 - L)$$

As packets are only dropped in the burst, the number of packet arrivals in the time interval $[0, T_{\text{on}}]$ is equal to the allowable loss plus the number of packet arrivals in the time interval $[0, T_{\text{on}}]$ for the new traffic characterization. Therefore, $\sigma + \rho T_{\text{on}} = \sigma' + \rho' T_{\text{on}} - \sigma L f$ and we have:

$$\sigma' = \max(0, \sigma(1 - Lf))$$

where

$$f = \frac{1}{1 - (\rho_{\min}/\rho)}$$

The length of the on and the off cycle of the resulting traffic is unchanged, so we can find the new value of ρ'_{\max} by $\sigma'/T_{\text{on}} + \rho'$. Thus, we have:

$$\rho'_{\max} = \min[\rho(1 - L), (\rho_{\max} - \rho)(1 - Lf) + \rho(1 - L)]$$

The minimum rate of the traffic remains unchanged, so we have:

$$\rho'_{\min} = \rho_{\min}$$

According to Theorem 4.1, the traffic of a $(\sigma, \rho, \rho_{\max}, \rho_{\min})$ -model with a loss rate L can be reduced to a $(\sigma', \rho', \rho'_{\max}, \rho'_{\min})$ -model in the worst-case scenario of the traffic arrival. It is obvious that the traffic of $(\sigma, \rho, \rho_{\max}, \rho_{\min})$ -model can be reduced to $(\sigma', \rho', \rho'_{\max}, \rho'_{\min})$ -model when the traffic arrival is not in the worst case.

As a result, we can use two leaky buckets to filter the traffic of the $(\sigma, \rho, \rho_{\max}, \rho_{\min})$ -model and reduce it to a $(\sigma', \rho', \rho'_{\max}, \rho'_{\min})$ -model. One of the leaky buckets initializes its credit counter to σ' and the its incrementing rate of the credit counter to ρ' . The other one initializes its credit counter to 1 and its incrementing rate of the credit counter to ρ'_{\max} . The leaky bucket allow the packets to go into the network when the credit counter is non-zero and drop the packets otherwise. The credit counter is decremented for each bit that enters the network and it is continuously incremented at the specified incrementing rate when its value is less than the initial value.

4.2.4. The packet-scheduling scheme

The aim of this paper is to provide bounded loss and delay and the design of the packet-dropping scheme

and the traffic characterization model is independent of the packet-scheduling scheme. Therefore, at the current stage, we use the optimal scheduler, EDF scheduler, as the packet-scheduling scheme.

4.2.5. The schedulability condition

Without loss of generality, we assume that connections are ordered such that $i < j$ whenever $d_i < d_j$. With the proposed traffic model and the packet-dropping mechanism, we have the schedulability condition as follows:

$$t \geq \sum_{i=1}^j n_i(\sigma'_i + \rho'_i(t - d_i)) + \max_{k < j} \max$$

for $d_i \leq t < d_j + 1, 1 \leq j < N$, and

$$t \geq \sum_{i=1}^j n_i[\sigma'_i + \rho'_i(t - d_i)] \quad \text{for } (d_N \leq t)$$

As long as the stability condition above is satisfied, for instance, $\sum_{j=1}^N \rho_j < 1$, we have:

$$d_j + \frac{\sigma'_j}{\rho'_{\max j} - \rho'_j} \geq \frac{\sum_{i=1}^j n_i(\sigma'_i - \rho'_i d_i) + \max_{k < j} \max}{C - \sum_{i=1}^j n_i \rho'_i}$$

4.3. Optimality for Single Type of Traffic Sources

As shown previously, the packet-dropping mechanism is optimal when there is only one connection. We now briefly argue for the optimality for the proposed algorithm when a number of single type of traffic are presented.

For the same type of sources, their desired packet delay, average traffic rate, maximum traffic rate and minimum traffic rate are the same. Therefore, the pattern of their worst-case traffic is identical. Thus, one of the optimal solutions should be that the dropping pattern of all connections be the same. In other words, we can only consider one of the connections as they are all identical. In particular, since our proposed method is optimal for a single connection; it also must be optimal for a set of connections that are in the same classes.

5. Performance Evaluation

5.1. Introduction

To evaluate the performance of our proposed MAC protocol and QoS guarantee mechanisms, we perform a series of evaluation tests in this Section. This Section is divided into two sections. In Section 5.2, we evaluate the performance of the protocol itself by varying different factors, for example traffic burstiness, average traffic rate, maximum traffic rate, minimum traffic rate and so on. The purpose of these evaluations is to see how these factors affect the performance of the proposed protocol. In Section 5.2, we evaluate the performance of the proposed protocol by comparing it with one of the existing MAC protocols, MASCARA.

5.2. Basic Evaluation

In this section, two sets of tests are conducted. The first test is for single type of traffic. We evaluate the performance of our proposal by varying different factors such as the burstiness σ , the peak rate ρ_{max} the minimum rate ρ_{min} and the desired loss rate L . In the second test, we evaluate the performance using multiple classes of traffic.

5.2.1. Single type of traffic

Unless otherwise specified, all of the results in this section are generated with the parameters shown in Table III.

First, we investigate the effect of the ratio of the minimum traffic rate to the average traffic rate on the channel utilization. As shown in Figure 8, we found that the channel utilization increases from 0.4 to 1 when the ratio of the minimum traffic rate to the average traffic rate increases from 0 to 1. A higher minimum traffic rate implies less variation of the

Table III. Traffic parameters for single type of traffic.

Channel Capacity	45 Mbps
Cell Size	53 bytes
ρ	0.15 Mbps
ρ_{max}	0.9 Mbps
ρ_{min}	0.09 Mbps
σ	100 cells
d	30 ms
L	5%

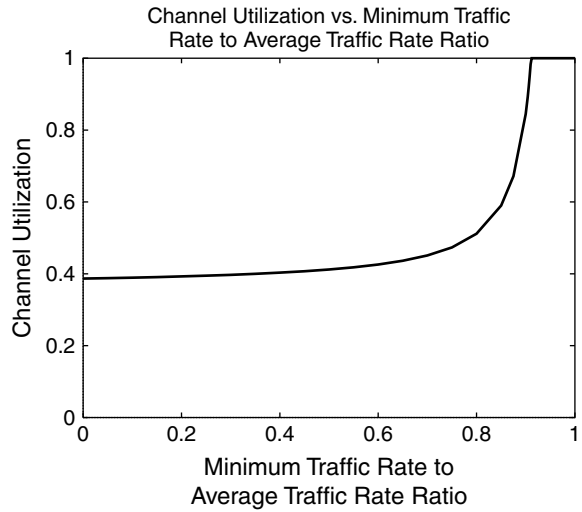


Fig. 8. Utilization vs ratio of minimum traffic rate to average traffic rate (ρ_{min}/ρ).

traffic; thus the scheduling of traffic is easier and can be more efficient.

Next, we investigate the effect of the ratio of the maximum traffic rate to the average traffic rate on the channel utilization. As shown in Figure 9, we found that the channel utilization decreases from 1 to 0.2 when this ratio increases from 0 to 20. The channel utilization drops sharply when the ratio is below 10, and it drops slowly when the ratio is above 10. In particular, a lower minimum traffic rate implies less variation of the traffic; thus the scheduling of this traffic is easier and can be more efficient.

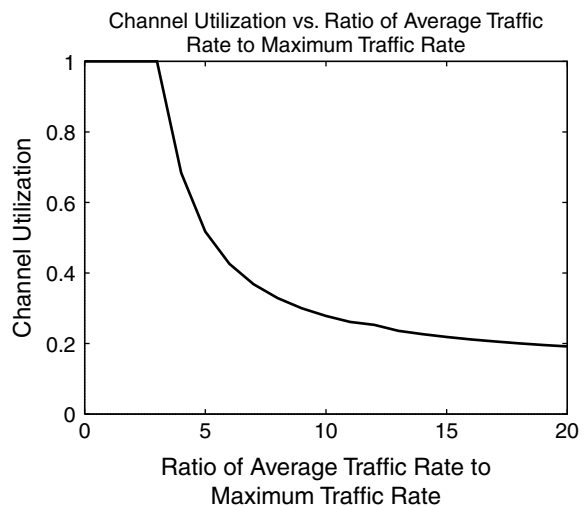


Fig. 9. Utilization vs ratio of maximum traffic rate to average traffic rate (ρ_{max}/ρ); the scheduling of the traffic is easier and can be more efficient.

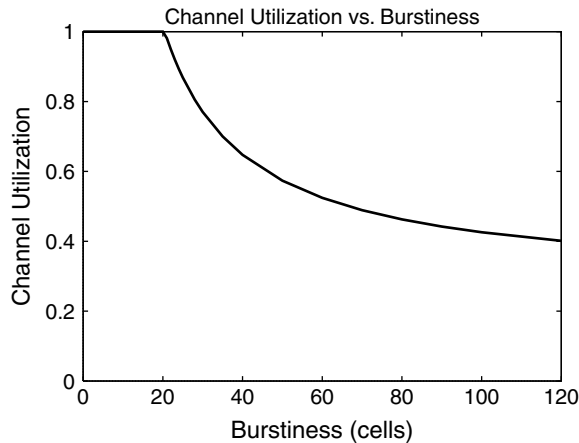


Fig. 10. Utilization vs burstiness of the traffic.

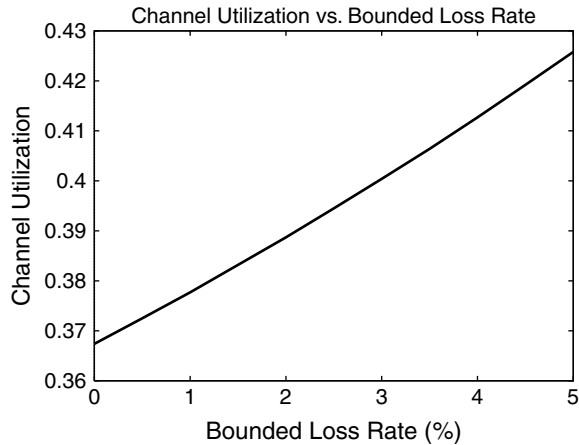


Fig. 11. Utilization vs desired.

We also studied the effect of the traffic burstiness on the channel utilization as shown in Figure 10. We found that the channel utilization decreases from 1 to 0.4 when the burstiness of the traffic increases from 0 to 120 cells. The result illustrates the fact that lower burstiness of the traffic implies less variation of the traffic; thus Figure 11 illustrates the effect of the packet loss rate on the channel utilization. The channel utilization increases from 0.368 to 0.425 when the desired bounded loss rate increases from 0% to 5%. By allowing 5% of packet loss for the connections, the channel utilization increases by 15%, thus 15% more connections can be admitted. According to the figure the channel utilization is directly proportional to desired bounded loss rate.

Table IV. Traffic parameters for multiple types of traffic.

Connection Class	1	2
ρ_i	0.15 Mbps	0.15 Mbps
σ_i	25 cells	250 cells
d_i	30 ms	50 ms
L_i	5%	5%

5.2.2. Multiple types of traffic

Now, we assume that we have two classes of traffic and the channel capacity is 45Mbps. The other traffic parameters are shown in Table IV.

According to Figure 12, the maximum number of class 1 connections that can be supported increases as the number of class 2 connections decreases, and vice versa. When the number of class 2 connections reduces to below a certain value, the maximum number of class 1 connections that can be supported stops to increase. The reason is that the resources saved by not serving class 2 connections are not enough to serve any more class 1 connections as the desired QoS, for instance delay, of class 1 connections is much higher than that of class 2 connections. The admission region increases as the ratio of minimum traffic rate to average traffic rate increases. This is due to the fact that the smaller ratio implies less variation of the traffic. When the ratio is 0.8, the area of the admission region of the proposed algorithm is around 40% larger than that of the algorithm which do not allow dropping.

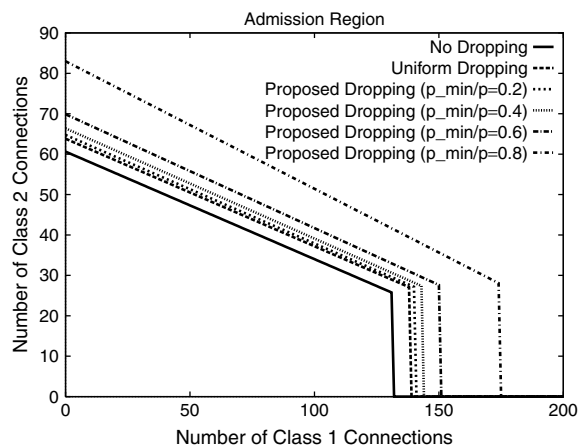


Fig. 12. Admission region for two different classes of traffic.

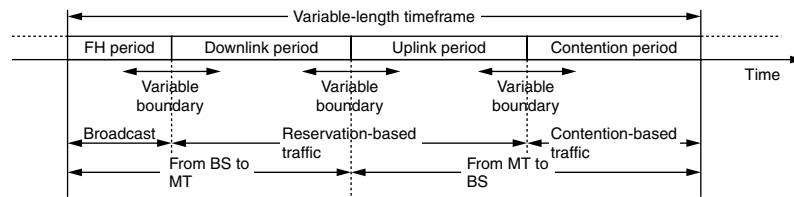


Fig. 13. Frame format of MASCARA.

5.3. Comparison with Other MAC Protocols

To further evaluate the performance of the proposed protocol, we have to compare it with some of the existing MAC protocols. Since there is no wireless MAC protocol that can provide deterministically bounded loss and delay, we are not able to directly compare our deterministic MAC with others. Instead, we compare our proposed protocol with a statistical protocol. Among the existing MAC protocols, we have selected the MASCARA protocol to perform the comparison since it is receiving the most attention [11].

5.3.1. Description of MASCARA

This MASCARA protocol was proposed by Bauchot *et al.* [11] as the MAC protocol for the WAND project being developed with the support of the European Community (EC). The frame format of MASCARA ++ is shown in Figure 13.

This protocol operates in a hierarchical mode by means of a master scheduler (MS) in the base station and a slave scheduler at each mobile terminal. The Down-link (DL) traffic is transmitted in TDM mode, while the Up-link (UL) packets are transmitted in a mix of reservation and contention modes. In MASCARA, time is divided into variable-length time frames, which are further sub-divided into time slots. The time-slot duration is equal to the time needed to transmit an Asynchronous Transfer Mode (ATM) cell payload plus the radio and MAC-specific headers. Slot allocation is performed dynamically to match current user needs and attain high-statistical multiplexing gain to provide the QoS required by the individual connections.

To handle the transmission process, MASCARA defines the concept of a “cell train”, which is a sequence of ATM packets belonging to one mobile terminal with a common header. The length of a time slot as well as the length of the MPDU header are defined as the length of an ATM packet. The concept of “cell train” provides variable capacity to mobile terminals in multiples of slots that have

the standard size of one ATM packet. The base station uses an algorithm called Priority Regulated Allocation Delay-Oriented Scheduling (PRADOS) to schedule transmissions over the radio interface. This algorithm is based on the priority class, the agreed characteristics and the delay constraints of each active connection. Passas *et al.* describe this algorithm in more detail [6].

5.3.2. Comparison results

As we need to compare a deterministic MAC protocol with a statistical MAC protocol, we have to perform the evaluation in two different cases, worst case and average case. The high gain from statistical scheduling and the tighter bounds of deterministic protocol would normally favor statistical MAC protocols over deterministic MAC protocols in the average case. On the other hand, in the worst case, the declared QoS guarantee of most connections will be violated using a statistical MAC protocol. As a result, we performed the evaluation using three metrics: average loss rate, percentage of connections violating declared QoS and admission region.

Average loss rate. In Figure 14, we found that our proposed protocol out-performs MASCARA in the worst-case traffic. When the number of users is higher than 56, the average loss rate in MASCARA has exceeded the declared loss rate of 5% in this case. In our proposed protocol, we can still keep the average loss rate below the declared loss rate until the number of users exceeds 60. (Note: we can only admit at most 60 connections as the admission control determines that we cannot provide hard guarantee to all connections if we admit more than 60 connections).

In Figure 15, we found that our proposed protocol under-performs MASCARA in the average case traffic. The average loss rate in MASCARA only exceeds the declared loss rate when the number of users exceeds 75. In our proposed algorithm, we only admit at most 60 connections as it is limited by our CAC.

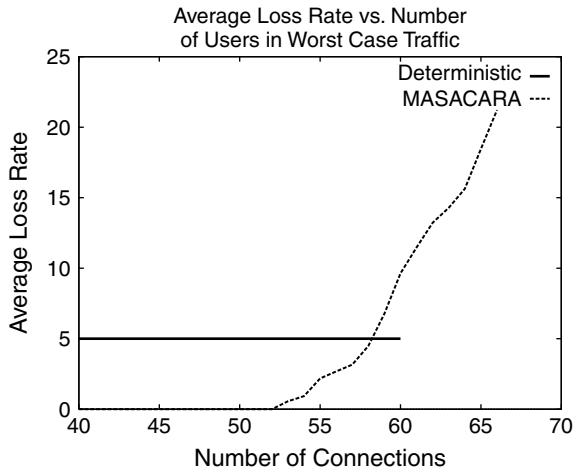


Fig. 14. Admission region for two different classes of traffic.

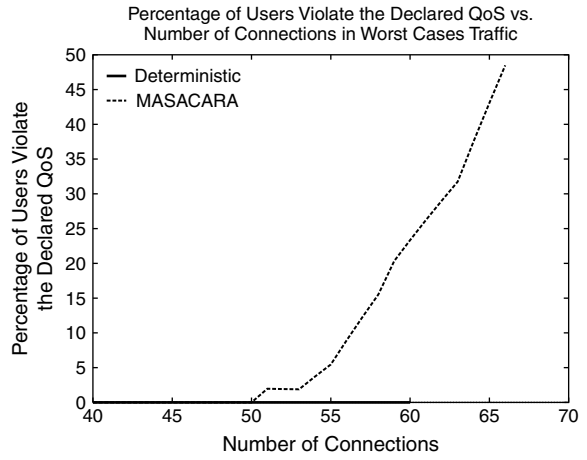


Fig. 16. Percentage of connections violate declared QoS vs number of users in worse-cases traffic.

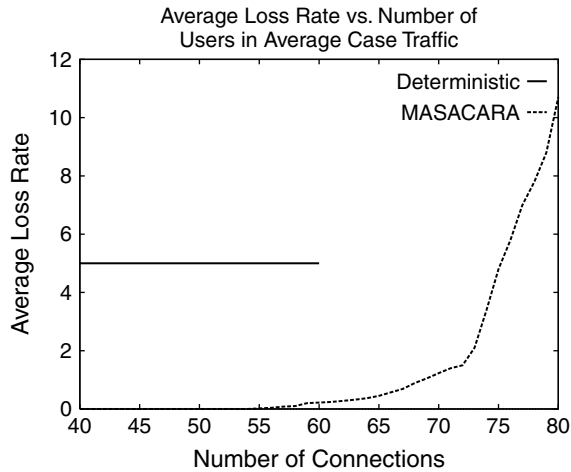


Fig. 15. Average loss rate vs number of users in worse-cases traffic.

Percentage of connections violating the declared QoS. Sometimes, the average loss rate is not a good performance metric. For example, suppose we get only one connection with zero loss while all others exceed the declared loss rate. The average loss rate may still be below the declared loss rate, but most connections receive a poorer than declared QoS. Thus, in this section, we use the percentage of connections violating the declared QoS as the metric.

In Figure 16, we found that some connections in MASCARA start to violate the QoS when the number of users increase to 50. The violating percentage increases as the number of users increase. In our proposed protocol, the percentage of connections that violate the declared QoS can be kept at zero when the

number of users is less than the calculated value from our CAC, for example 60. When the number of users is 60, there are around 23.3% of all connections that have violated the declared QoS using MASCARA.

In Figure 17, we can only find some connections violating the declared QoS when the number of users is more than 71 when using the average-case traffic.

Admission region. In Figure 18, we found that the admission region for our proposed protocol is much larger than that for MASCARA in the worst-case traffic. Using MASCARA, we can only admit 75 type 1 connections or 65 type 2 connections without having any connection violating the declared QoS. Using our proposed protocol, we can admit 100 type

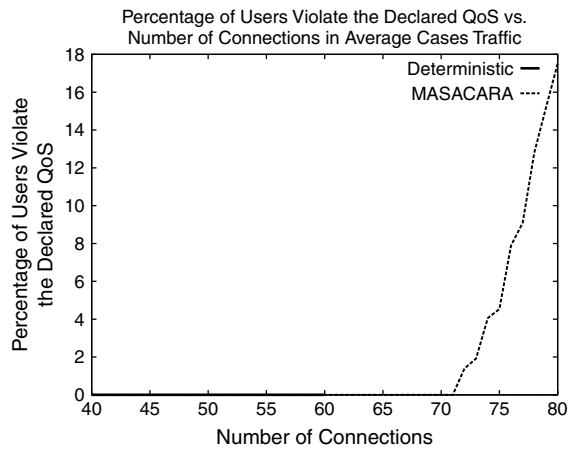


Fig. 17. Percentage of connections violate declared QoS vs number of users in average-cases traffic.

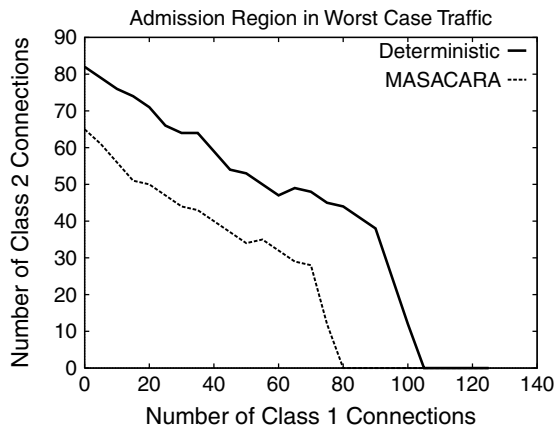


Fig. 18. Admission region for worst-case traffic.

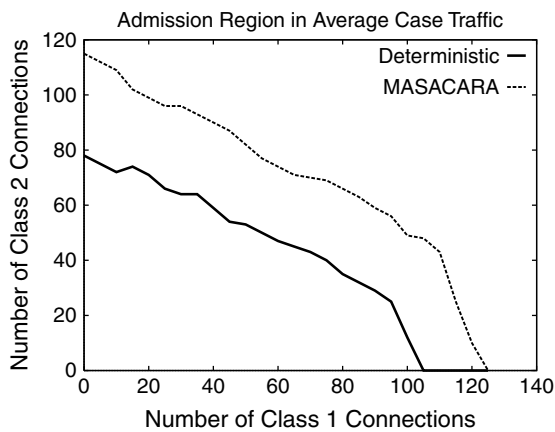


Fig. 19. Admission region for average-case traffic.

1 connections and 82 type 2 connections under the same condition.

In Figure 19, we found that the admission region for our proposed protocol is less than that for MASCARA in the average-case traffic. Using MASCARA, we can admit 120 type 1 connections or 115 type 2 connections without having any connection violating the declared QoS.

6. Conclusion

We proposed a *deterministic* QoS guarantees method for multimedia wireless networks, which is based on a mathematical framework that accurately characterizes multimedia traffic streams in conjunction with efficient scheduling and dropping algorithms. The uniqueness of this scheme is that it can provide bounded loss and bounded delay deterministically. By allowing certain bounded loss for the connections,

more connections can be admitted into the network. We have performed a set of performance-evaluation tests that demonstrated that our proposed algorithm can significantly support more connections than a system that does not allow any loss. Moreover, from the evaluation, we found that the proposed algorithm is able to out-perform well-known MAC protocol, for example, MASCARA, in worst-case traffic.

References

1. Acampora A. Wireless ATM: a perspective on issues and prospects. *IEEE Personal Communications* 1996; **3**: 8–17.
2. Adas A, Mukherjee A. On resource management and QoS guarantees for long range dependent traffic. In *Proc. IEEE Infocom*, April, 1995; pp. 779–787.
3. Anastasi G, Grillo D, Lenzini L. An access protocol for speech/data/video integration in TDMA-based advance mobile systems. *IEEE Journal on Selected Areas in Communications* 1997; **15**(8): 1498–1514.
4. Braden R, Clark D, Shenker S. Integrated services in the internet architecture: an overview. In *IETF RFC 1633*, July 1994.
5. Cruz RL. A calculus for network delay, part I: network elements in isolation. *IEEE Transactions on Information Theory* 1991; **37**(1): 114–131.
6. Passas N, Paskalis S, Vali D, Merakos L. Quality-of-service-oriented medium access control for wireless ATM networks. *IEEE Communications Magazine* 1997; **35**: 42–50.
7. Aras CM, Kurose JF, Reeves DS, Schulzrinne H. Real-time communication in packet-switched networks. *Proceedings of the IEEE* 1994; **30**(1): 122–139.
8. Eckberg AE. Meeting the challenge: congestion and flow control strategies for broadband information transport. In *Proc. IEEE GLOBECOM '89*, 1989; 49.3.1–49.3.5.
9. Sanchez J, Martinez R, Marcellin MW. A survey of mac protocols proposed for wireless ATM. In *IEEE Network*, November/December, 1991; pp. 52–62.
10. Karol MJ, Liu Z, Eng KY. Distributed-queueing request update multiple access (dquma) for wireless packet (ATM) networks. In *Proc. IEEE INFOCOM '95*, 1995; 1224–1231.
11. Bauchor F. Mascara, a mac protocol for wireless ATM. In *Proc. ACTS Mobile Summit '96*, November, 1996; pp. 17–22.
12. Chang CS, Chen KC, You MY, Chang JF. Guaranteed quality-of-service wireless access to ATM networks. *IEEE Journal on Selected Areas in Communications* 1997; **15**(1): 106–118.
13. Mahmoud AS, Falconer DD, Mahmoud SA. A multiple access scheme for wireless access to broadband ATM LAN based on polling and sectored antennas. *IEEE Journal on Selected Areas in Communications* 1996; **14**(4): 596–608.
14. Knightly EW. Traffic models and admission control for integrated services networks. *Ph.D. thesis, University of California, Berkeley*, May 1996.
15. Liebeherr J, Wrege DE, Ferrari D. Exact admission control in networks with bounded delay services. *IEEE/ACM Transactions on Networking* 1996; **4**: 885–901.
16. Wrege DE, Knightly EW, Zhang H, Liebeherr J. Deterministic delay bounds for VBR video in packet-switching networks; fundamental limits and practical tradeoffs. *IEEE/ACM Transactions on Networking* 1996; **4**(3): 352–362.
17. Crowcroft J, Wang Z, Smith A, Adams J. A rough comparison for the IETF and ATM service models. *IEEE Network* 1995; **9**(6): 12–16.

18. Heyman D, Tabatabai A, Lakshman TV. Statistical analysis and simulation study of video teleconference traffic in ATM networks. *IEEE Transactions on Circuits and Systems for Video Technology* 1992; **2**(1): 45–59
19. Heyman DP, Lakshman TV. Source models for VBR broadcast-video traffic. In *Proc. IEEE Infocom '94*, June, 1994; pp. 664–671.
20. Ismail MR, Lambadaris IE, Devetsikiotis M, Kaye AR. Modelling prioritized MPEG video using tes and a frame spreading strategy for transmission in ATM networks. In *Proc. IEEE Infocom '95*, 1995; 762–770.
21. Garrett MW, Willinger W. Analysis, modeling and generation of self-similar VBR video traffic. In *Proc. ACM Sigcomm '94*, August, 1994; pp. 269–280.
22. Zhang Z, Habib I, Saadawi T. A bandwidth reservation multiple access protocol for wireless ATM local networks. *International Journal of Wireless Information Networks* 1997; **4**(3): 147–160.
23. Kurose JF. On computing per-session performance bounds in high-speed multi-hop computer networks. In *Proc. 1992 ACM Sigmetrics and Performance '92*, June, 1992; pp. 128–139.
24. Goodman DJ. Packet reservation multiple access for local wireless communications. *IEEE Transactions on Communications* 1989; **37**(8): 885–890.
25. Kim JG, Widjaja I. PRMA/DA: a new media access control protocol for wireless ATM. In *ICC '96*, June, 1996; 1–19.
26. Petras D, Kramling A. Mac protocol with polling and fast collision resolution for an ATM air interface. In *IEEE ATM Workshop*, August 1996.
27. CCITT, CCITT draft recommendation i.371. In *Traffic Control and Resource and Resource Management in B-ISDN*, December 1991.
28. Cruz RL. A calculus for network delay, part II: network analysis. *IEEE Transactions on Information Theory* 1991; **37**(1): 132–141.
29. Golestani SJ. A framing strategy for congestion management. *IEEE Journal on Selected Areas In Communications* 1991; **9**(7): 1064–1077.
30. Ferrari D. Client requirements for real-time communication services. *IEEE Communications Magazine* 1990; **28**(11): 65–72
31. Reibman AR, Berger AW. Traffic descriptors for VBR video teleconferencing over ATM network. *IEEE/ACM Transactions on Networking* 1995; **3**(3): 329–339.
32. Rathgeb EP. Modeling and performance comparison of policing mechanisms for ATM networks. *IEEE Journal on Selected Areas in Communications* 1991; **9**(4): 325–334.
33. Dittman L, Jacobsen SB, Moth K. Flow enforcement algorithms for ATM networks. *IEEE Journal on Selected Areas in Communications* 1991; **9**(3): 343–350.
34. Ferrari D, Verma DC. A scheme for real-time channel establishment in wide-area networks. *IEEE Journal on Selected Areas in Communications* 1990; **8**(3): 368–379.
35. Zhang H, Ferrari D. Rate-controlled static-priority queueing. In *Proc. IEEE Infocom '93*, April, 1993; pp. 227–236.
36. Passas N. Mac protocol and traffic scheduling for wireless ATM networks. *ACM Mobile Networks and Applications Journal* 1996; **3**: 43–49

Authors' Biographies



Mounir Hamdi received the B.S. degree in Computer Engineering (with distinction) from the University of Louisiana in 1985, and the MS and PhD degrees in Electrical Engineering from the University of Pittsburgh in 1987 and 1991, respectively. He has been a faculty member in the Department of Computer Science at the Hong Kong University of Science and Technology since 1991 where he is now Associate Professor of Computer Science and the Director of the Computer Engineering Programme. In 1999 and 2000 he held visiting professor positions at Stanford University and the Swiss Federal Institute of Technology. His general areas of research are in Networking and Parallel computing, in which he has published more than 150 research publications, and for which he has been awarded more than 10 research grants. He has graduated more than 10 MS/PhD students in the area of study. Currently, he is working on high-speed networks including the design, analysis, scheduling, and management of high-speed switches/routers, wavelength division multiplexing (WDM) networks/switches, and wireless networks. Dr. Hamdi is/was on the Editorial Board of *IEEE Transactions on Communications*, *IEEE Communication Magazine*, *Computer Networks*, *Wireless Communications and Mobile Computing*, and *Parallel Computing*, and has been on the program committees of more than 50 international conferences and workshops. He was a guest editor of *IEEE Communications Magazine* and *Informatica*. He received the best paper award at the International Conference on Information and Networking in 1998 out of 152 papers. He received the best 10 lecturers award and the distinguished teaching award from the Hong Kong University of Science and Technology. He is a member of IEEE and ACM.

Felix K. L. Lee received the B.S. Degree and the MPhil degree in Computer Science from the Hong Kong University of Science and Technology. His areas of interest are in networking with special emphasis on the performance evaluation, medium access control protocols, and QoS provision in wireless environment. He is currently a networking engineer at the Hong Kong and Shanghai Banking Corporation.