Improving VBR Voice Performance In Integrated Services Broadband Wireless Networks

A THESIS
SUBMITTED IN PARTAIL FULLFILMENT
OF THE REQUIREMENTS FOR THE DEGREE
OF
MASTERS OF SCIENCE IN COMPUTER SCIENCE
IN THE
UNIVERSITY OF CANTERBURY
by
Mohammad Rezvan

University of Canterbury
1998

Abstract

The integration of multimedia applications into future wireless networks is expected to accelerate the demand for incorporating broadband infrastructure into wireless arena. Providing an efficient broadband wireless infrastructure capable to carry a mixture of different services brings a large set of new technical challenges. The current wireless networks that have been designed to carry low bit-rate voice and data are not able to carry bandwidth consuming and delay sensitive multimedia traffic. Consequently fundamental changes at different layers of current technology used in wireless networks are required.

In this thesis we focus on Medium Access Control protocols (MAC) suitable for broadband wireless networks. Wireless ATM (WATM) has been considered in this work. This is mainly because it is widely accepted that ATM is the foundation of future broadband networks and integrating ATM into wireless networks provides a seamless interface between wired and wireless environments. We investigate the major recent proposals for MAC protocols for broadband wireless networks and propose a new reservation mechanism for the reservation part of a FDD-based MAC protocol. This novel mechanism is called "***Dynamic Hybrid Partitioning with Adjustable Repeat***" which helps to improve the performance of the Variable Bit Rate (VBR) voice traffic in a broadband wireless network with integrated traffic.

Through a number of simulation experiments based on AKAROA2 [Ewi99], we analyze the different aspects of our proposed mechanism and show how it improves the performance of the VBR voice traffic sources in a network with different classes of traffic.

Acknowledgments

I would like to thank

- My supervisor *Associate Professor Dr. K. Pawlikowski* for his guidance, discussions and strong support throughout this work

- My co-supervisor *Associate Professor Dr. H. Sirisena* from the department of Electrical and Electronics Engineering.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Following significant advances in wireless communication and rapid market growth for wireless services in recent years, the need to support more sophisticated services such as interactive video and bandwidth consuming multimedia applications is anticipated [Cox95]. Supporting these new services requires fundamental changes in current wireless and cellular networks which have been designed to support circuit-switched voice and more recently, low bit rate data. These new services bring a broad spectrum of technical challenges, including the required provisioning of satisfactory bandwidth. This requires to move towards the concept of broadband wireless networks. This concept can be realized using cellular and ad hoc architectures.

The cellular architecture provides a public infrastructure at national or global level, similar to Global System for Mobile Communication (GSM). In this architecture, service areas are divided into cells. Each cell is controlled by a base station which handles bandwidth allocation, connectivity to wired backbone and mobility.

The ad hoc architecture provides an instantaneous connectivity pattern between several users [Eng95,Kav95-1]. This pattern varies with time as users move. This architecture does not need a fixed infrastructure which makes it suitable for places where access to fixed backbone is difficult due to physical or financial problems (e.g., rescue operation in remote areas, factories, and some of the third-world countries).

Two different terminology proposed in literature which should not be confused together: Wireless LAN (WLAN), and more recently, Wireless ATM (WATM) [Kav97]. While they are similar in many ways, in particular in those aspects which are related to physical layer, there are also fundamental differences between them. WLAN is a relatively mature technology with many commercial products. Its main goal is to provide more flexibility for office environments and currently available LAN applications by removing the wiring requirements. It targets ad hoc approach. On the other hand, WATM is a cellular-based architecture assumed to be offered by public service providers (e.g., NZ Telecom) as an extension of future Broadband ISDN (B-ISDN) infrastructure.

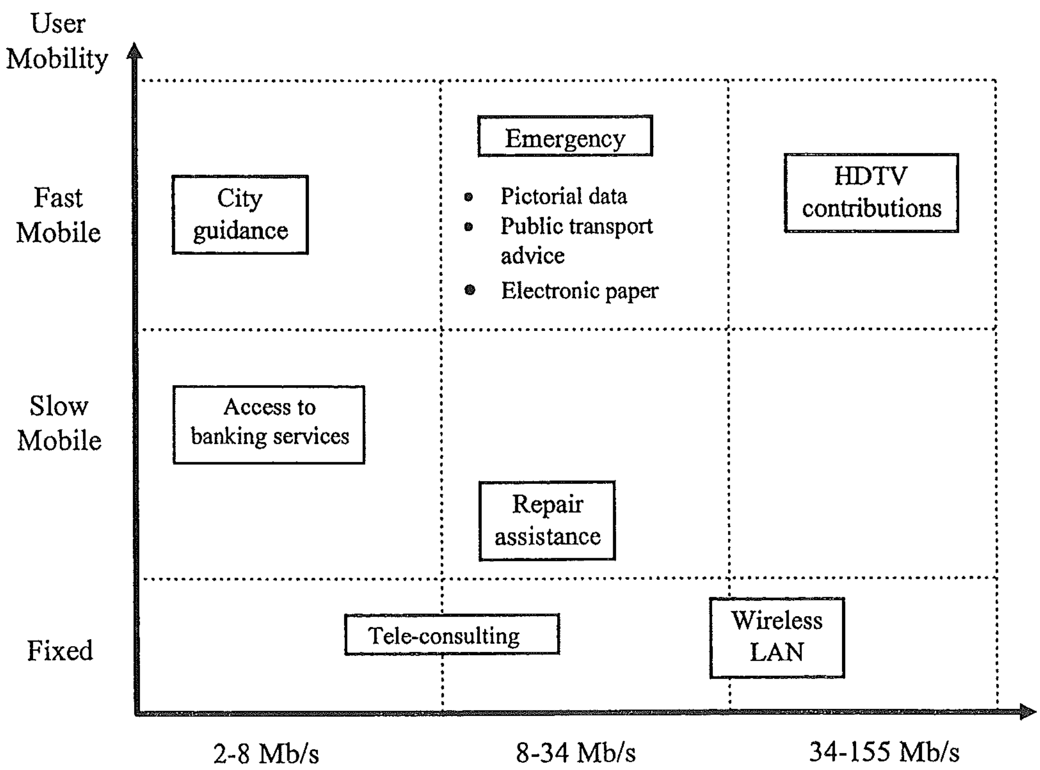

Figure 1.1 : Some examples of the possible wireless broadband applications

Although at the moment there is no clear view of future applications of broadband wireless networks, there is no doubt that they will deeply affect the way societies will live and operate. Users will be able to access a large variety of useful information services such as video-on-demand, city-guiding systems, video conferencing, remote education and shopping, and many other applications, some of which not easy-to-anticipate now. They could use all these services with absolute freedom of movement, not being attached to any fix access point. Figure 1.1 shows some of the possible broadband applications, according to mobility and data rate requirements.

1.1 Wireless Broadband : Principal Issues

1.1.1 Bandwidth Allocation/Frequency Administration Issues

At the moment, the total available radio spectrum has been divided into two parts: licensed and unlicensed. The licensed part of the spectrum is under control of major service providers and telecommunication authorities. Using this part, requires licensing agreements. The unlicensed part of radio spectrum has played a key role in developing emerging wireless technologies, and can be divided into following bands: [Aya96]

- **Industrial, Scientific, and Medical (ISM) Band**

This band is divided into several partitions as follows.

- 915 MHz (902-928 MHz , 26 MHz BW) is a very crowded partition in which many non-spread spectrum systems currently exist.

- 2.4 GHZ (2.4-2.4835 GHZ , 83.5 MHz BW) is lightly loaded. Interestingly, the main problem with this portion is interference created by microwave ovens which is an important problem for indoor environments.
- 5.2 GHZ (5.725-5.85 GHZ , 125 MHz BW) is lightly loaded. Some radar interference exists. Operating at these frequencies require GaAs(Gallium Arsenide) technology which is expensive.

The ISM band was the main driving force behind commercialization of WLAN technology and was released by FCC in the United States [Kav97]. This let the small companies to develop WLAN products without any concern about interfering with bands occupied by major service providers. Using this band is subject to adopting spread spectrum technology or using very low power. However, this band is not a promising candidate for broadband applications which need channels with high bit rate capacity. In addition, spread spectrum requirement of ISM band makes it unattractive for developing WATM.

- **Personal Communication Services (PCS) Band**

This band is divided into several partitions as follows.

- 1.9 GHz (1910-1920 GHz , 10 MHz BW) allocated for unlicensed packet-switched applications.
- 1.9 GHz (1920-1930 GHz , 10 MHz BW) allocated for unlicensed circuit-switched applications.

- **Unlicensed Band for Future Wireless Multi-media Networks**

The WINFORUM request for 250 MHz unlicensed bandwidth around 5 GHz, followed by Apple Computer Co. petition to FCC for 300 MHz bandwidth for deployment of National Information Infrastructure (NII), made FCC to release a 300 MHz partition around 5 GHz as follows.

- 5.15 - 5.25 GHz for indoor applications.
- 5.25 - 5.35 GHz (HIPERLAN compatible)
- 5.35 - 5.45 GHz to cover long range rural area with low interference

All the above discussed bands have defined power limits and can provide enough capacity to support bandwidth-consuming multimedia traffic.

1.1.2 Physical Infrastructure

Two basic technologies can be used to provide a wireless infrastructure: infra-red and radio. The infra-red technology is available in two different modes of propagation: diffused (using optical lens), and directed. The diffused infra-red has propagation characteristics similar to that of radio, but it has shorter range and more power fluctuation in comparison to the radio technology. The directed infra-red does not suffer from multi path fading problems and can provide very high bit rate capacity, but it requires Line of Sight (LoS) transmission which is difficult to achieve, in particular in crowded areas where objects and people move. The directed infra-red is useful for providing fast communication links between computers and peripherals such as remote printers in an office.

The radio technology has been used more widely than the infra-red. There are numerous technical challenges for developing efficient radio infrastructures. The major problems to be addressed are the spectrum limitations and poor and variable quality of wireless channels. Table 1.2 shows a summary of problems and potential solutions for radio technology. Interested readers can refer to [Kav95-1] for more details.

The spread spectrum technology helps to overcome traditional problems in the narrow band radio technology. The spread spectrum technology offers attractive features such as resistance against fading, and high degree of frequency reuse (useful in cellular networks). In addition, Code Division Multiple Access (CDMA)-based networks provide some degree of privacy because they use different spreading codes for different users. Furthermore, spread spectrum-based networks can be easily overlaid into currently operating cellular networks. The main disadvantages of spread spectrum technology (e.g., CDMA) are: power control requirement, and base station design complexity, and high chip rate requirement to achieve higher bit-rate capacity which is necessary to carry bandwidth-consuming multimedia traffic [Kav95-1].

At last, it is worth to mention that both radio and infra-red technologies have their own characteristics and can be adopted in different situations. One important disadvantage of infra-red technology is that it is limited to indoor environments because ambient light of outdoor environment can cause severe interference. Yet, infra-red technology is cheaper than radio technology mainly due to the fact that optical receivers detect amplitude (power) of optical signal, not their frequency or phase. This leads to much simpler transceiver architectures [Kav95-1]. In addition, using infra-red technology requires no licensing agreement. Finally, infra-red can be useful when electromagnetic interference is a concern, for example in radiography laboratories. When multiple access in a multi-user network is required, the radio technology is preferable.

Problem	***Reason(s)***	***Solution(s)***
Limited spectrum (Lower channel bit rate capacity)	• Telecommunication regulations • Licensing agreements	**1. Frequency reuse** • co-channel interference **2. Move to higher frequencies** • Propagation characteristics problem at higher frequencies • Technology problems
Poor and variable radio channel quality (High and variable Bit Error Rate)	• LOS power loss • Shadowing / fading • Pulse dispersion • Intersymbol interference • Frequency selective fading • Doppler spread	**1. DSP algorithms** • Power control • Synchronization • Pulse shaping • Equalization **2. Adaptive antennas** **3. Forward Error Correction (FEC) Automatic Repeat Requests (ARQ)** **4. Robust modulation techniques**

Table 1.2 : Major problems and potential solutions for radio technology

1.1.3 Broadband Applications

Multimedia applications are anticipated as dominate service in the future wireless broadband networks. In a general view, a multimedia application generates different types of traffic such as voice, video, and non delay-sensitive data (e.g., an ftp connection). Although at the moment characteristics of these applications are not fully understood, some estimates regarding their bandwidth requirements have been made [Hon97]. Table 1.3 shows a summary of some of the possible applications with their bandwidth requirements. As it shows, applications can generate two types of traffic pattern: symmetric or asymmetric. In symmetric case, both uplink and downlink are equally loaded, while in asymmetric case, down link carries the main part of the load. Video on Demand (VoD) is an example of an application with asymmetric traffic load, where downlink is heavily loaded (due to continuous transmission of video stream), but uplink carries only user requests and control information. Video conferencing is an application with symmetric load.

Broadband Service	*Downlink Bandwidth*	*Uplink Bandwidth*
Broadcast Video	1.5 to 6 Mbps	None
Interactive Video Video on Demand	64 Kbps to 6 Mbps	9.6 to 64 Kbps
Internet Access (www , FTP)	14.4 Kbps to over 10 Mbps	14.4 to 128 Kbps
Video Conferencing Desktop multimedia	9.6 Kbps to 2 Mbps	9.6 Kbps to 2 Mbps

Table 1.3 : Bandwidth requirements of some popular wireless applications

1.1.4 The Role of Coding

Coding plays a vital role in wireless broadband networks. This is mainly due to the bandwidth consuming nature of broadband multimedia applications such as real-time video. In wireless broadband networks, coding becomes even more vital as available bandwidth is limited and channel quality is low and time varying. Source and channel coding are two major types of coding. Source coding deals with issues such as data compression, for removing redundancy of data stream presented by sources, to lower bandwidth requirements of data streams. Channel coding is used for lowering channel Bit Error Rate (BER), by adding redundancy to data streams. From above discussion, it may look that these two types of coding are working against each other. Source coding aims to minimize redundancy, while channel coding adds extra redundancy. In practice, choice of efficient combination of source and channel coding schemes largely depends on the nature of targeted applications and is an important design issue, specially in a wireless network, where available bandwidth is limited. For example, more compressed data stream of say, video, needs less bandwidth for transmission, but it is more vulnerable to channel errors. This is because during compression, much of the temporal dependencies between data units (e.g., video frames) are deleted. Consequently, to combat channel errors, different types of error detection/correction schemes, such as FEC and/or ARQ, should be used, and they add redundancy to the data stream.

In a multimedia network built on top of the broadband wireless infrastructure, different classes of traffic with different QoS requirements co-exist. No single coding scheme is able to work efficiently for all types of traffic sources. In addition, channel condition is highly time varying. For these reasons, adaptive source/channel coding schemes capable to adapt themselves to traffic and channel variations are likely to be incorporated. Recently, adaptive source/channel and scalable/multi-resolution coding schemes have been studied [Nag97]. Another important issue in wireless networks is the complexity of coding schemes. For example, a complex coding/decoding algorithm may achieve better compression ratio, but it needs more memory and processing power, and this means more power consumption in mobile terminals with limited battery power. Algorithm complexity is most important for image and video compression, because video streams contain large amount of data that makes them very resourcehungry in terms of memory and processing power requirements. The Vector Quantization (VQ) technique [Kav95-1] has been found as a potential alternative for image/video compression in wireless networks, in particular for those applications with asymmetric traffic pattern, where down link in heavily loaded and mobile terminals have to execute decoding algorithm continuously. This is because VQ decoding algorithm is based on simple sequential table look-ups and does not need much memory and processing power.

At data link layer, emphasis is more on channel coding, to combat channel errors. Two major schemes that have been widely used are: the Forward Error Correction (FEC) and the Automatic Repeat Request (ARQ) techniques. FEC techniques add extra bits to data packets to help detecting and correcting errors, while ARQ is used to retransmit packets that contain errors without trying to correct errors at receiving end. FEC adds more redundancy to channel but does not causes much delay. ARQ causes more delay (due to retransmission) but helps reducing redundancy. Consequently, FEC and ARQ form two ends of a trade-off spectrum with more delay/less redundancy on one end, and less delay/more redundancy on the other end. Providing optimum/adaptive solutions based on hybrid FEC/ARQ schemes have been studied extensively [Aya96] .

1.1.5 Network Architectures

Wireless networks can be deployed using two different architectures: cellular and ad hoc. In cellular architecture, service area is divided into cells. Each cell is controlled by a base station that performs bandwidth allocation, call admission control functions, and looks after users mobility. The base station also acts as interface between wireless and wired parts of the network. Base stations can be connected together via wired backbone or microwave links. Cellular networks are usually deployed nation-wide, on a permanent basis. They are managed by big telecommunication companies as deployment, maintenance, and operation costs are high.

The ad hoc architecture requires no permanent infrastructure and can be deployed very fast without almost any deployment cost. Several users can be connected via radio links in a geographically limited area such as a classroom. As users move, architecture changes. If necessary, one or more terminals can provide connection to wired backbone.

There are several important characteristics that should be taken into account to decide which architecture should be adopted. Table 1.4 shows a comparative summary of these architectures.

Topic	*Cellular Architecture*	*Ad hoc Architecture*
Topology	Fixed	Time varying
Deployment Cost	High	Low (or nothing)
Spectrum Usage	Licensed frequencies Increased capacity via inter-cell frequency re-use patterns	Unlicensed frequencies No frequency re-use is possible
Management Network Control	Central Performed by base stations	Distributed Every station contributes
Reliability	High	Low
Range	Can be nation-wide or even global	Very limited area

Table 1. 4 : Cellular vs. ad hoc architecture

When considering architectural issues, the integration of wireless and wired networks becomes an important issue. The wireless and wired networks are inherently different [Acamp96]. The wired networks have usually plenty of relatively cheap bandwidth and low BER. In addition, channel condition is consistent. The wireless networks have very limited bandwidth and very high BER compared to wired networks. In addition, wireless channels have a variable nature due to user mobility and fading. Consequently, separate protocols should be designed for these two types of networks. Naturally, broadband wireless networks should be connected to wired backbone to let mobile terminals benefit from services provided by B-ISDN. The interface between these two types of networks should be designed carefully, to avoid performance bottleneck due to the protocol translation overhead. As Asynchronous Transfer Mode (ATM) is most likely to be used to deploy future B-ISDN, it sounds logical to incorporate ATM technology into broadband wireless networks. This lets seamless integration of wireless and wired networks. Although there is a discussion about the suitability of this approach, for the time being, most of the research activities have been focused on developing wireless networks based on ATM [Osa97].

1.2 Objective of this Research

In this project, we focus on MAC protocols for broadband wireless networks. WATM, which has been proposed in most recent potential candidates for broadband wireless networks, is the main target of our study. Although we consider cellular architecture, mobility and related issues have not been investigated. After studying recent proposals, a new reservation mechanism named "Dynamic Hybrid Partitioning with Adjustable Repeat" has been proposed for reservation part of Time Division Duplex (TDD)-based MAC protocols for WATM. The goal is to improve the performance of the real-time Variable Bit Rate (VBR) traffic (voice traffic has been considered) in a broadband wireless network with mixed voice/data traffic.

The performance and effects of the proposed scheme is evaluated through computer simulation. The simulation program is written in C++ and linked to AKAROA2 [Ewi99], a software package designed in the department of Computer Science, University of Canterbury, for automatic distributed precision control of simulation programs.

This thesis has been organized as follows. Chapter 2 details design principles and requirements of MAC protocols for broadband wireless networks, then reviews recent proposals for WATM. Chapter 3 details our proposed scheme. Chapter 4 presents comprehensive presentation and analysis of results obtained by simulation. Chapter 5 consists of summary and conclusions.

Chapter 2

Medium Access Control Protocols for Broadband Wireless Networks

2.1 Overview

By definition, MAC protocols are required when a single communication channel (wired or wireless) is shared by multiple users. In other words, it is essentially an arbitration mechanism that is responsible for the resolution of contentions if multiple users want to share the same transmission medium. This role becomes more critical in wireless networks as the available bandwidth is scarce, channel quality is low and highly variable. Furthermore, incorporation of multimedia applications brings more technical challenges, because more efficient statistical multiplexing schemes are required to support a variety of different QoS requirements simultaneously. These requirements distinguish the MAC protocols of broadband wireless networks from those used in commercial wireless networks now. In general, MAC protocols can be classified into few categories as follows:

- Fixed assignment protocols
- Random access protocols
- Demand-assignment protocols
- Combined protocols

Fixed assignment MAC protocols are widely used in traditional cellular networks to support circuit-switched voice transmission. The principal concept on top of which these MAC protocols have been designed is to assign a pre-defined, fixed portion of available transmission capacity to each user. Time Division Multiple Access (TDMA), Frequency Division Multiple Access (FDMA), and CDMA have been used to design fixed assignment MAC protocols. In TDMA, channel is usually divided into frames and within each frame, and each user is assigned a fixed number of time slots during which he/she can use the entire available bandwidth. In FDMA, channel is divided into different carrier frequencies and each user is assigned different frequency, which he/she can use to transmit all the time. Hybrid TDMA/FDMA MAC protocols have been designed as well. For example, during each time slot in a TDMA system, multiple carrier frequencies can be used, so that several users can transmit during the same time slot. CDMA provides channel sharing in a different way. Each user is assigned a different transmission code. Consequently, CDMA provides an environment similar to a multilingual meeting in which people talk in different languages all the time and one can detect only the language he understands!. Although all people talk all the time, there is no collision as long as they use different languages. In CDMA, this problem has been resolved by assigning special orthogonal codes to different users.

Intuitively, in TDMA, a user uses the ***entire*** channel bandwidth in a ***fraction of time*** while in FDMA a ***fraction*** of the channel bandwidth is used ***all the time***. In CDMA, user uses the ***entire*** channel bandwidth ***all the time***. TDMA has several advantages over FDMA, namely:

- TDMA offers more flexibility in multiplexing uplink and downlink traffics than FDMA because dealing with time slots is much easier than dealing with frequencies. In the presence of multimedia traffic (in broadband networks), this is a big advantage because available capacity can be partitioned ***dynamically*** in accordance to the load variations and QoS requirements of different services.

- In the presence of frequency selective fading, TDMA is likely to provide better performance. This is because in TDMA, user signal bandwidth is close to the coherent channel bandwidth (in a time slot, user uses the entire bandwidth). Consequently every user partly suffers from fading. But in FDMA, one or a few users can suffer significantly while others do not experience the fading at all.

- Time Division Duplex (TDD) provides some attractive features which makes it a good candidate for MAC protocols of broadband networks with multimedia traffic. In TDD, both uplink and downlink traffic are multiplexed on the same channel. We detail this issue in section 2.2.

CDMA does not require timing and ordering in the way it is required in TDMA and FDMA. Because CDMA is based on spread spectrum physical layer, it performs better in the presence of interference and fading. It also provides some level of data security because different users are assigned different spreading codes. In addition, CDMA does not put a ***hard limit*** on the total system capacity, in terms of number of users, contrary to TDMA and FDMA. TDMA puts a hard limit on the number of supportable users because it divides channel to time slots which are assigned to individual users. Once there are more users than the number of available slots, they have to wait for slots to be released. The same problem exists in FDMA systems regarding the number of carrier frequencies. However, in CDMA, every user is considered as a source of noise to the rest of the users. As long as this interference is at acceptable level, more users can be accepted into the network*.

Random access MAC protocols do not use any type of fixed assignment or channel splitting. Instead, as their name represents, they work in a random fashion. Every user can attempt to transmit whenever he/she has data. If a collision occurs, the user has to retransmit data. The nature of random access protocols makes them suitable for packet-switched communication networks, and because they are easy to implement, they have been widely used in practice. The most well-known protocols of this family are : ALOHA and its enhanced version known as such slotted-ALOHA, Carrier Sense Multiple Access (CSMA) and its enhanced versions such as CSMA with Collision Detection (CSMA/CD), and CSMA with Collision Avoidance (CSMA/CA). The main problem with random access protocols is their potential instability at higher loads which can lead to low channel utilization and large delays. The key point to alleviate instability is to avoid excessive collisions at higher loads. For example, in slotted-ALOHA, channel has been slotted to decrease vulnerable period to the duration of a single time slot [Kav95-1].

Demand-assignment MAC protocols are best to handle bursty traffic. They can achieve high channel utilization because they allocate resources dynamically depending on users' requests. As a result, they have an adaptive nature and can respond to channel load variations. Basically, these protocols consist of two phases: reservation and transmission. In the reservation phase, users send their channel access request packets indicating the amount of bandwidth needed. After receiving acknowledgment from the base station (we assume a centralized control), they transmit their data in the transmission phase. The key point is to provide a collision-free transmission mechanism.

Combined MAC protocols have been studied, specially for networks with multimedia traffic. The motivation behind these protocols is that no single protocol

---

* There are other factors involved in defining the capacity of a CDMA-based network such as the way pseudo codes have been designed. Here, our point is that CDMA does not put a hard limit on the capacity of the network.

performs well in all conditions. Fixed assignment protocols are good when traffic sources generate data at a steady and predictable pace. Random access protocols perform relatively well when channel load is low and traffic is bursty. Demand-assignment protocols can adapt to load variations but they are more complex. A combined protocol usually uses a mix and match set of mechanisms. For example, in a demand-assignment protocol, reservation and transmission phases can be designed using slotted-ALOHA and TDMA, respectively. Slotted-ALOHA provides a simple contention-based reservation phase that lets users transmit their requests. TDMA provides a collision-free transmission phase with high statistical multiplexing gain. Further, adaptive algorithms can be used to achieve better performance. We detail these issues in section 3.2.

2.2 Design principles and Requirements

MAC protocols play a major role in designing efficient communication networks and bring a variety of technical challenges. Supporting broadband multimedia services in a network brings even more technical challenges because MAC protocols contribute in providing guaranteed Quality of Service (QoS) for these services. Furthermore, when considering building a broadband multimedia network on top a wireless infrastructure, the role of the MAC protocol becomes much more important as bandwidth is scarce, and channel quality is highly variable. In practice, as Figure 2.1 shows, a MAC protocol designed for a broadband wireless network is under pressure from two sides, the physical layer and the multimedia application. The physical layer puts strict design constraints such as limited available bandwidth as well as high and variable BER. Multimedia applications require MAC protocol to provide guaranteed QoS for bandwidth-consuming services by means of efficient bandwidth allocation and traffic scheduling schemes in an error prone wireless network. No single MAC protocol can address all these technical challenges and restrictions simultaneously. A more feasible option is to use a combined MAC protocol which incorporates the best features of advantages of other protocols. Demand-assignment MAC protocols are strong alternatives, because they can offer better bandwidth utilization than other schemes. As mentioned, a typical demand-assignment MAC protocol is composed of two phases, reservation, and transmission. In reservation phase, users transmit their requests to the base station which allocates the requested bandwidth and acknowledges mobile terminals. In transmission phase, users transmit their data. There are important design alternatives for both phases which can greatly affect the performance of the MAC protocol. In the rest of this chapter, we discuss these alternatives, then review major recent proposals.

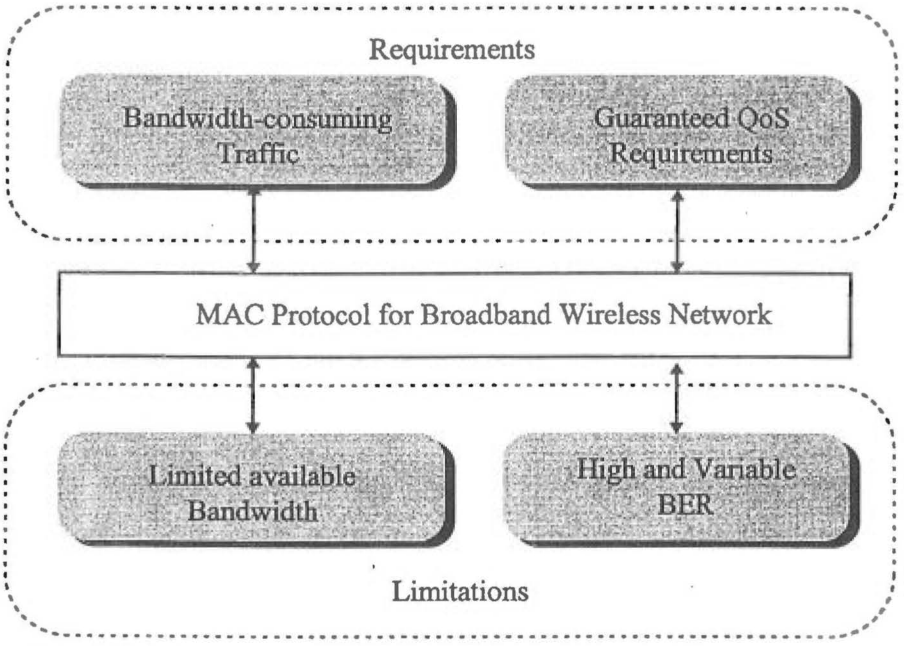

Figure 2.1 : The role of MAC protocol in a wireless broadband network

2.2.1 Frame Structure

Most wireless MAC protocols see the channel as a sequence of frames (except those protocols that do not use any frame at all), whose length can be fixed or variable. Each frame carries several data packets. Introduction of frames provides a flexible way of synchronization. In a broadband wireless network, frame should be short enough so that MAC protocol can respond to changes in multimedia traffic and channel conditions quickly. At the same time, frame should be long enough to avoid excessive overhead. Deep fades caused by movements of people and objects tend to occur at intervals of tens of hundreds of milliseconds [Fal96]. The onset of fade may last for a few milliseconds. Traffic variations, in particular for VBR video traffic, occur at tens of milliseconds. When considering these factors, a frame length of about a few milliseconds is considered as appropriate for a MAC protocol of a wireless broadband network. A typical demand-assignment MAC protocol has a frame that is usually divided into several major parts, such as:

- **Modem pre-amble**: is used for synchronizing transceivers in the mobile terminals. It is broadcasted to all mobile terminals and used as a training sequence by transceiver chips.
- **Header:** contains detailed information about the frame format such as the length of its different parts, the number and exact position of allocated slots.
- **Reservation / acknowledgment:** in this part of the frame, mobile terminals send their requests and receive acknowledgments about the results of their requests. To be efficient, this part should adapt itself to load variations. The functionality and important design issues concerning this part are detailed in Section 3.2.3.
- **Transmission:** in this part of the frame, those mobile terminals for which bandwidth has been successfully allocated transmit their data. This part has very important role to carry multimedia traffic efficiently. Its functionality and important design issues are briefly discussed in Section 3.2.4

2.2.2 Channel Structure

A wireless channel consists of two links, uplink (from a mobile terminal to the base station) and downlink (from the base station to a mobile terminal). These links can be maintained by single or separate carrier frequencies. In terms of channel structure, recent proposals can be classified into two categories [San97]: TDD and FDD. In TDD, both uplink and downlink use the same carrier frequency. In FDD, separate carrier frequencies are allocated for uplink and downlink. Both TDD and FDD have their own advantages and disadvantages. The main advantage of FDD over TDD is that it provides fast feedback about the result of bandwidth reservation to mobile terminals, because the base station can transmit acknowledgments via a different carrier frequency in downlink. On the other hand, TDD offers the following characteristics:

- Usage of a single carrier frequency simplifies the design of the transceivers. Less sophisticated transceiver architecture costs less, and, more importantly, consumes less power.

- TDD facilitates incorporation of adaptive/smart arrays of antenna. Using a single carrier frequency [Acamp96] makes antenna's placement easier. In a typical antenna array, individual antenna elements are placed at a distance about half-wavelength of carrier frequency from each other. In addition, TDD provides a symmetric bi-directional channel because uplink and downlink are being multiplexed on the same frequency but in different intervals. This makes antenna weight adjustments faster, following mathematical relationships between required weights for receive and transmit mode, respectively [Acamp96].

- In presence of asymmetric traffic load, when downlink is heavily loaded, TDD offers more flexibility and more efficient channel utilization. This is because the boundary between uplink and downlink can be dynamically adjusted.

- The main disadvantage of TDD is that it adds some extra delay due to turnovers between receiving/transmitting modes.

In this situation, there are advocates of both TDD and FDD. For example, cellular systems are almost exclusively FDD, while cordless systems are almost exclusively TDD.

2.2.3 Reservation

Achieving low delay is always considered as one of the most important design issues for wireless MAC protocols. The total delay experienced by a packet (or ATM cell) has two components: the channel access delay and the transmission delay. The channel access delay is the interval between the time a mobile terminal sends its reservation request and the time it can actually start transmitting data. The transmission delay is the time it takes for a packet to get to its final destination since its generation in a mobile station. The channel access delay should be minimized in reservation phase. The transmission delay should be minimized in transmission phase of the MAC protocol.

Random access protocols such as ALOHA and slotted-ALOHA are frequently implemented in the reservation phase of recently proposed MAC protocols for WATM [San97]. Although these protocols are simple to implement, they can be unstable at higher traffic loads. In addition, they do not adapt themselves to load variations. To alleviate these problems, the following techniques can be used:

A. Mini-slots. The concept of mini-slots has been incorporated, for example in a proposal by Raychaudhuri *et al.* [Rch96-1]. Because mini-slots have a fraction of the length of the data slot, in the case of a collision, less bandwidth is wasted. In addition, they lead to a shorter reservation phase and leave more bandwidth for transmission phase.

B. Fast collision resolution algorithms. These algorithms have been used to dynamically adjust the length of the reservation phase and adapt it to variations of the user requests. When more requests are generated, more collisions are experienced. The collision resolution algorithm can respond to this situation by rapidly allocating more mini-slots in reservation phase. On the other hand, when the number of requests decreases, unused mini-slots can be deleted.

C. Piggybacking. Under piggybacking, mobile terminals attach their subsequent channel access requests to the actual data packets. In this way, mobile terminals do not have to contend every time they need to access the channel. This, in particular, is very useful for VBR traffic. This technique has been used to provide a collision-free reservation, once a call is accepted for the first time.

Another important design issue in reservation phase, is to provide fast feedback by sending acknowledgment about the results of contention to mobile terminals. For this purpose, FDD-based MAC protocols outperform TDD-based ones, because they use separate carrier frequencies for uplink and downlink. In some of the MAC protocols (e.g., MASCARA, see section 2.3), reservation is performed on a slot-by-slot basis, as opposed to a frame-by-frame basis. These protocols usually have no frame structure and wireless channel simply has been divided into consecutive slots. Although they are demand-assignment protocols, there is no apparent distinction between reservation and transmission phases and each slot can be used either for reservation or transmission. This mechanism may present better performance in lower traffic load, because mobile terminals do not have to wait until the next frame to receive the acknowledgment from the base station. The handling of acknowledgments on a slot-by-slot basis is easier to implement in an FDD-based MAC protocol.

2.2.4 Transmission

Wireless broadband networks are expected to carry multimedia traffic with guaranteed QoS, based on user requirements. For most real-time applications, each packet can only be delayed up to its deadline, or it is of no use to the recipient and may be considered lost. As discussed in section 3.2.3, in a demand-assignment MAC protocol, the delay experienced by a packet has two components: the reservation delay and the transmission delay. Efficient scheduling algorithms and bandwidth allocation strategies are key solutions to minimize the transmission delay experienced by a packet. In this project, we have not focused on the transmission aspect of wireless MAC protocols. However, in following, some of the more widely-used algorithms are briefly discussed.

Several established scheduling algorithms have received wide attention. They can be used in both wireless and wired networks. Due to its simplicity, many current networks use the First In First Out (FIFO) algorithm which applies the same service discipline to all packets, independent of their performance objectives. According to FIFO, packets are served in the order in which they arrive. To allow simple differentiation of packets, some networks use the Static Priority (SP) algorithm. With SP, packets are given a level of priority before they enter the network, and the queued packet with the highest priority is always selected first for transmission. The SP algorithm allows the network to vary the service given to packets based on the importance and stringency of the performance objectives of their associated applications. However, SP does not consider the fact that the urgency of delivering a packet can vary with time, which is certainly the case for packets with deadlines. A simple variant of SP that many consider be appropriate for supporting traffic with deadlines is bandwidth reservation [Sri93]. With Bandwidth reservation, an application negotiates with the network, the right to transmit high priority packets at the desired rate. The Earliest Deadline First (EDF) algorithm was the first to consider the problem of scheduling jobs with deadlines [Peha90]. With this algorithm, deadlines are assigned to packets before they enter a queue, and the queued packet with the earliest deadline is selected for transmission, except any packets that have missed their deadlines.

Accurate separation and definition of different classes of traffic and their QoS requirements is of paramount importance. This issue has been addressed in the best way in ATM standards, where several major classes of traffic such as Constant Bit Rate (CBR), Variable Bit Rate (VBR), Available Bit Rate (ABR), and Unspecified Bit Rate (UBR) have been defined. Furthermore, VBR can be divided into real-time and non real-time subclasses. As a result, for designing a wireless MAC protocol based on ATM technology, channel capacity can be divided dynamically between these classes of traffic, starting from CBR and real-time VBR with the highest priority, to UBR with the lowest priority.

2.3 A Brief Survey on Recent Proposals

In this section, we briefly survey some of the recent proposals for TDMA-based MAC protocols of WATM. We divide our survey into two major groups, FDD-based and TDD-based proposals. Here, it is not intended to provide a comprehensive detailed review of these protocols. Instead, we want to give a basic but clear introduction to major principles used in these protocols. For more details, interested readers can refer to the stated references.

- **FDD-Based Proposals**

Distributed Queuing Request Update Multiple Access (DQRUMA) considers slotted channel without any frame structure [Kar95]. The uplink channel has been divided into mini-slots for request access, each followed by a slot for data transmission. If necessary, for example when number of mobile terminals is getting large, the base station can convert a transmission slot into M mini-slots, where M is a protocol parameter. Similarly, the downlink channel consists of mini-slots for acknowledgment of access requests, each followed by a slot for data transmission. As a result, there is no distinction between reservation and transmission phases. In DQRUMA, the mobile terminal is considered to be in one of three states: “empty”, “request” or “wait-to-transmit”. Piggybacking has been used to transmit the subsequent channel access requests once a mobile terminal gets initial access (during call set-up or hand-off). This protocol does not consider any bandwidth partitioning or priority scheme and treats both VBR and ABR traffic equally. This protocol has the advantage that the mobile terminal is able to receive the acknowledgment to its channel access request almost immediately because downlink uses a separate frequency and reservation is performed on a slot-by-slot basis, as opposed to frame-by-frame basis. In addition, this protocol has used piggybacking to provide contention-free channel access request once a mobile terminal gains access to wireless channel for the first time. The important disadvantage of this protocol is that it has not considered any kind of priority for different classes of traffic. Bringing in mind that this protocol has been designed to support ATM traffic, this could be a big disadvantage for real-time, delay-sensitive traffic.

Packet Reservation Multiple Access with Dynamic Allocation (PRMA/DA) considers a fixed-length frame format divided into fixed number of slots [Kim96]. This protocol has not used mini-slots and all slots are of the same size. A frame is divided into four types of partitions: data partition (used by non delay-sensitive data), VBR partition, CBR partition, and available partition (used for reservation). For CBR and real-time VBR calls, which have strict delay constraints, the unlimited repetition of contention procedures is worthless, so the protocol introduces a parameter called maximum setup time (W_{MAX}). If a contention procedure lasts for more than W_{MAX}, the call will be discarded. W_{MAX} has no limitation for other classes of traffic. This protocol considers a model with three states to represent the status of a mobile terminal: inactive, contending and reserving. A mobile terminal is initially in its inactive state. When a packet is generated, the mobile terminal switches to the contending state and tries to transmit a packet via contention, by reserving bandwidth. If the channel access procedure succeeds, the mobile terminal switches to the reserving mode. The base station controls the number of slots allocated to each partition, as well as determining the number of slots allocated to each mobile terminal. The downlink is considered as a contention-free TDM format. Available partition slots (fourth type) are used by mobile terminals for channel access request using slotted-ALOHA protocol. It uses a simple but efficient collision resolution algorithm to control the number of available slots. In this project, we have adopted this protocol (refer to section 4.2.2). Although PRMA/DA is an FDD-based MAC protocol, in comparison to DQRUMA, under PRMA/DA mobile terminals need usually more time to receive acknowledgments from the base station. This is because in PRMA/DA, uplink is divided into frames and the base station sends acknowledgments at the end of every frame, while in DQRUMA the base station sends the acknowledgments immediately after receiving the access request. This protocol allows mobile terminals to receive the result of their contention at the end of access contention period, but, in the case of a collision, the contending mobile terminals have to wait for the base station to announce how many available slots will be assigned for the next contention partition. This may lead to more delay in comparison to the scheme used in DQRUMA. The main advantage of this protocol is its fast collision resolution algorithm which helps to resolve the contention situation quickly. However, this protocol does not use mini-slots for contention. Instead, when a mobile terminal for the first time needs to gain access to the channel, it sends a channel access request packet accompanying by a data packet, so if a collision occurs, the effect on the throughput may be greater than if a small packet had been used.

Dynamic TDMA with Piggybacked Reservation (DTDMA/PR) uses a fixed-length frame, with mini-slots for reservation and ATM-packet-size slots for transmission of data [Rch94]. The uplink has been divided into three sub-frames (partitions): the first for reservation mini-slots, the second for long-term reservable slots, and the third for short-term reservable slots. The boundary between long-term and short-term reservable slots is movable and is defined dynamically. Three types of traffic has been considered in this protocol: CBR, VBR, and ABR. Reservation for CBR and VBR can happen only in the long-term sub-frame, while reservations for ABR can happen only in the short-term one. Mobile terminals contend in the reservation phase to send their channel access requests via mini-slots. At the end of the reservation phase, the base station sends a broadcast message that contains all necessary information about the result of contention: successful mobile terminals, and the number, type, and position of the allocated slots. The mobile terminals with CBR or VBR traffic that have been successful in reservation phase, can keep using the allocated slots as long as they have more packets to send, while mobile terminals with ABR traffic have to release the allocated slots as soon as they finish transmitting a packet. In fact, this protocol only lets CBR and VBR mobile terminals to use piggybacking. One advantage of this protocol is that it applies piggybacking, which is specially useful for delay sensitive VBR traffic. It also uses mini-slots in reservation phase. Another point about this protocol is that it considers two types of reservable sub-frames (partitions) with movable boundaries.

Dynamic Slot Assignment (DSA++) considers a variable-length frame structure called a signaling burst [San97]. The downlink signaling burst which is transmitted in broadcast mode, corresponds to an uplink signaling burst with the same length but with a time delay to compensate for the round-trip propagation delay. This protocol serves ATM traffic classes assuming that the priority of CBR is greater than VBR, is greater than ABR, is greater than UBR. The assignment of capacity is based on the priority calculation for each mobile terminal. The priority is determined according to a set of Dynamic Parameters (DP) such as the number of waiting ATM packets and their due dates, for each mobile terminals. The DPs are transmitted by each mobile terminal along with each ATM packet. The base station can ask a mobile station to update its DPs, via contention based on mini-slots. For this purpose, the base station uses an algorithm which calculates the number of mini-slots that must be available in the next signaling burst according to the following parameters:

- The probability of a new packet arrival at each mobile station in the contention mode since the last transmission of its DPs.
- The number of mobile terminals in the contention mode.
- The throughput achieved under the random access procedure (slotted ALOHA).

An advantage of this protocol is that it broadcasts the information that defines the next signaling period in a single downlink burst. This releases all other slots in the downlink signaling period, allowing the base station to implement a power control algorithm, if needed. This advantage is associated with the disadvantage of losing the broadcast packet which means that a whole signaling period would be lost. Otherwise the loss of a control packet addressed to a specific mobile station would not affect the throughput as much. Another advantage of this protocol is that it allows an uplink slot to be divided into up to four short slots to be used for access requests in contention mode. In fact, this the way this protocol has incorporated the concept of mini-slot.

- **TDD-Based Proposals**

Mobile Access Scheme Based on Contention and Reservation for ATM (MASCARA) is based on a variable-length frame format which is used both by uplink and downlink traffic [Bau96]. The downlink is divided into two partitions, the frame header and the downlink transmission. The header is used by the base station to broadcast to all mobile terminals: a descriptor of the current frame, the results of the contention procedure from the previous frame, and the slot allocation for each active mobile terminal. The uplink is also divided into two main partitions: contention-based reservation (based on slotted-ALOHA) and uplink transmission. All partitions have variable-length and can contain variable number of slots which is defined dynamically. For more efficient transmission, MASCARA defines the concept of "cell train" which is a sequence of ATM packets belonging to the same mobile terminal, with a common header. This protocol takes into account the service classes of the current active mobile terminals, the negotiated QoS requirements, the amount of generated traffic, and the number of reservation requests to determine the type and the volume of traffic that will be transmitted in the next frame. This information is kept in a slot map which specifies the size of the three different partitions: downlink, uplink, and contention, as well as the assignment of time slots in the current frame to each mobile terminal. The base station broadcasts the slot map within the frame header at the beginning of each frame. Using this slot map, each mobile terminal can determine if it will be allowed to either receive or transmit traffic in the current frame. This mechanism allows mobile terminals to perform power-saving procedure, such as entering a 'sleeping' mode when there is no traffic scheduled for them. MASCARA uses an algorithm called Priority Regulated Allocation Delay-Oriented Scheduling (PRADOS) to schedule transmissions. PRADOS combines the priorities of different ATM traffic classes with a Leaky Bucket Traffic Regulator (LBTR). The LBTR uses a token pool that is introduced for each mobile terminal (or connection). The generation of tokens happens at a fixed rate equal to the mean ATM packet rate of each traffic source. The size of the pool is equal to the maximum number of ATM packets that can be transmitted with a rate greater than the declared mean. Starting from the highest priority for CBR traffic to the lowest one for UBR traffic, the scheduler satisfies requests for uplink and downlink as long as tokens are available. For every slot allocated to a connection, one token is removed form the corresponding pool. This protocol is using "cell train" which can provide better channel utilization. A weak feature of this protocol is that it uses large slots for contention-based reservations. The size of a reservation slots equals to two ATM packets. Finally, the variable-length frame of MASCARA introduces extra difficulty in assigning capacity to CBR traffic. Assuming the case of a voice (64 kb/s) call, if the frame length is less than the time to fill an ATM packet (~ 6 milliseconds), there may be frames where no slots need to be assigned. Otherwise, if the frame length is longer than 6 milliseconds, it might be necessary to assign more than one slot in a frame for this call.

Packet Reservation Multiple Access with Adaptive Time-Division Duplex (PRMA/ATDD) has a fixed-length frame structure equal to 64 slots of equal size which are dynamically allocated for uplink and downlink [Pri96]. The first slot in a frame is used for synchronization, the second one for broadcast information, and the other 62 slots are used to carry data. The broadcast information includes:

- the number of slots in the downlink and uplink,
- the assignment of each slot for mobile terminals, and
- signaling related parameters.

This protocol maintains two different List Handlers (LH) at the base station for scheduling packets: Static List Handler (SLH) and Dynamic List Handler (DLH). The SLH determines the call static parameters to be stored in Static List (SL), which is updated only during call setup. Each record of the SL refers to a call in progress. The DLH maintains and controls a list of records, each one containing information about a specific ATM packet waiting in the base station, or in the mobile terminal buffer, to be transmitted on the air. Each record has a parameter indicating the last time at which the packet must be transmitted (packet deadline) in order to avoid its loss due to excessive waiting time. One advantage of this protocol is that it uses a fixed-length frame, which facilitates the Provision of CBR traffic by assigning fixed number of slots in each frame. Similar to PRMA/DA, a weak feature of this protocol is that it uses large reservation slots.

Dynamic TDMA with Time-Division Duplex (DTDMA/TDD) has a fixed-length frame [Rch96-1]. The downlink is handled in simple TDM format, transmitted in a single burst, and consists of two parts: control and acknowledgment signals, and data transmission. The uplink is dynamically divided into four partitions, a reservation partition based on mini-slots and slotted-ALOHA, a dynamic allocation partition which carries ABR and/or UBR traffic, a fixed and shared allocation partition which carries VBR traffic, and a fixed allocation partition which carries CBR traffic. Although a fixed-length frame structure is used, the boundary between uplink and downlink can change dynamically, according to instantaneous load variations. Functionally, this protocol is based on a master-slave architecture consisting of two main components, the Supervisory MAC (S-MAC) and the Core MAC (C-MAC). The S-MAC at the base station performs packet scheduling and dynamic bandwidth allocation for all classes of traffic. The C-MAC serves as the interface between the data link control and physical layer. According to QoS requirements of different classes of traffic, the C-MAC uses appropriate coding scheme at the data link level. One advantage of this protocol is the division of a frame into mini-slots (8 bytes), which allows it to use one mini-slot for random access transmission of control packets and to assign several mini-slots to form a the payload of different classes of ATM traffic. Another important advantage of this protocol is that it includes a data link control layer which complements the task of MAC layer. For example, to resolve the delay constraint imposed by CBR services, the data link layer uses a first-in-first-out (FIFO) buffer to ensure that ATM packet jitter will be kept under an acceptable limit. Another interesting idea is to let the data link layer handle the retransmission of corrupted CBR packets, by using the bandwidth allocated for ABR traffic, without disturbing the current flow of CBR traffic. Table 2.1 summarizes main features of the surveyed protocols.

2.4 Summary

The emergence of new services in wireless networks presents new technical challenges which distinguish MAC protocols of broadband wireless networks from those existing in current networks. In general, MAC protocols can be categorized into several major groups, such as: fixed assignment, random access, demand-assignment, and combined protocols. The presence of multimedia applications in broadband wireless networks demands new requirements, such as large amount of bandwidth and the need for guaranteed QoS for a variety of services. On the other hand, The nature of radio infrastructure has severe limitations, such as limited bandwidth and highly variable channel quality. The combination of these requirements and constraints makes traditional MAC protocols unsuitable for wireless broadband networks. A more feasible option is using combined MAC protocols, which incorporate the advantages of several protocols together. Demand-assignment MAC protocols provide the best channel utilization in the presence of bursty traffic. In addition, they can be designed to be adaptable to channel load variations. All recent proposals are based on demand-assignment scheme. The important aspects of these protocols are their frame structure, channel structure, reservation mechanism, packet scheduling and bandwidth partitioning policies which have been discussed in this chapter.

MAC Protocol	*Channel Structure*	*Frame Structure*	*Reservation Scheme*	*QoS*	*ACK Method*
DQRUMA	FDD	No Frame	S-ALOHA, mini-slot	VBR	slot-based
PRMA/DA	FDD	Fixed 6 ms	S-ALOHA, No mini-slot,	Voice, Video, Data	Frame-based
DSA++	FDD	Variable 8-15 slots	Splitting algorithm, mini-slot	CBR, VBR, ABR	Frame-based
DTDMA/PR	FDD	Fixed 16 ms	S-ALOHA, mini-slot	CBR, VBR, ABR	Frame-based
MASCARA	TDD	Variable	Not defined, mini-slot	CBR, VBR, ABR, UBR	Frame-based
PRMA/ATDD	TDD	Fixed 64 slots	S-ALOHA, mini-slot	Delay-based	Frame-based
DTMA/TDD	TDD	Fixed 2 ms	S-ALOHA, mini-slot	CBR, VBR, ABR	Frame-based

Table 2.1 : Summary of the main features of the surveyed MAC protocols

Chapter 3

A New Reservation Scheme : Dynamic Hybrid Partitioning with Adjustable Repeat

3.1 Overview

As mentioned in Section 2.4, although different proposed MAC protocols for WATM have successfully incorporated different techniques to achieve high channel utilization while providing guaranteed QoS requirements, they have not addressed, the so-called ***"idle-VBR"*** problem. Before discussing this problem, it is worth to have a brief overview on the main properties of VBR traffic. Analytically, a VBR traffic source can be modeled by a two state Markov Modulated Poisson Process (MMPP) [Fis92]. Such a source, if it is assumed to be inactive in one of its states, is referred to as an ON/OFF source. This model has been proven to be a good approximation for human speech. In this model, when the traffic source is in ON state, it generates data at a constant rate. In OFF state, the traffic source is "idle" and does not generate any output.

The time spent in each state, has exponential distribution with a given mean value. Figure 3.1 shows a state diagram for such a model.[1]

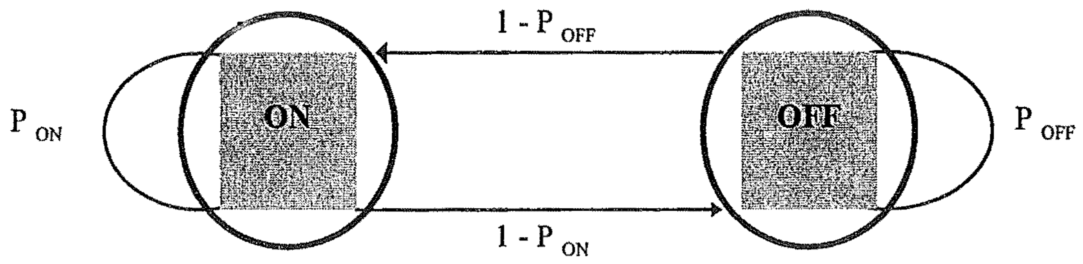

Figure 3.1 : Two state ON/OFF model for a VBR source

Considering above explanation, the "idle-VBR" problem can be clearly explained. In most proposals, the reservation phase of the proposed MAC protocol, is purely based on contention. Consequently, all ***currently accepted*** delay-sensitive VBR traffic (in this project, VBR voice has been considered) sources that are in ***idle-mode***, new arrivals and non delay-sensitive traffic sources such as UBR, have to compete equally to send their reservation requests. Every time an ***idle-VBR*** traffic source wakes up from idle-mode, it has to gain access to channel through contention. This can lead to large channel access delays and channel access delay variation, specially at higher loads. In the rest of this chapter, we propose and detail an easy-to-implement scheme to address this problem. The flowchart in Figure 3.2 depicts a general mechanism currently used in most proposals. (Note that some proposals may have not used piggybacking)

---

[1] Although here we concentrate on VBR voice traffic, it is worth to mention that the behavior of a VBR video stream can be modeled by superimposing of multiple ON/OFF sources. Work done in [Lng96] presents a practical example of such a model used for simulation of VBR video traffic.

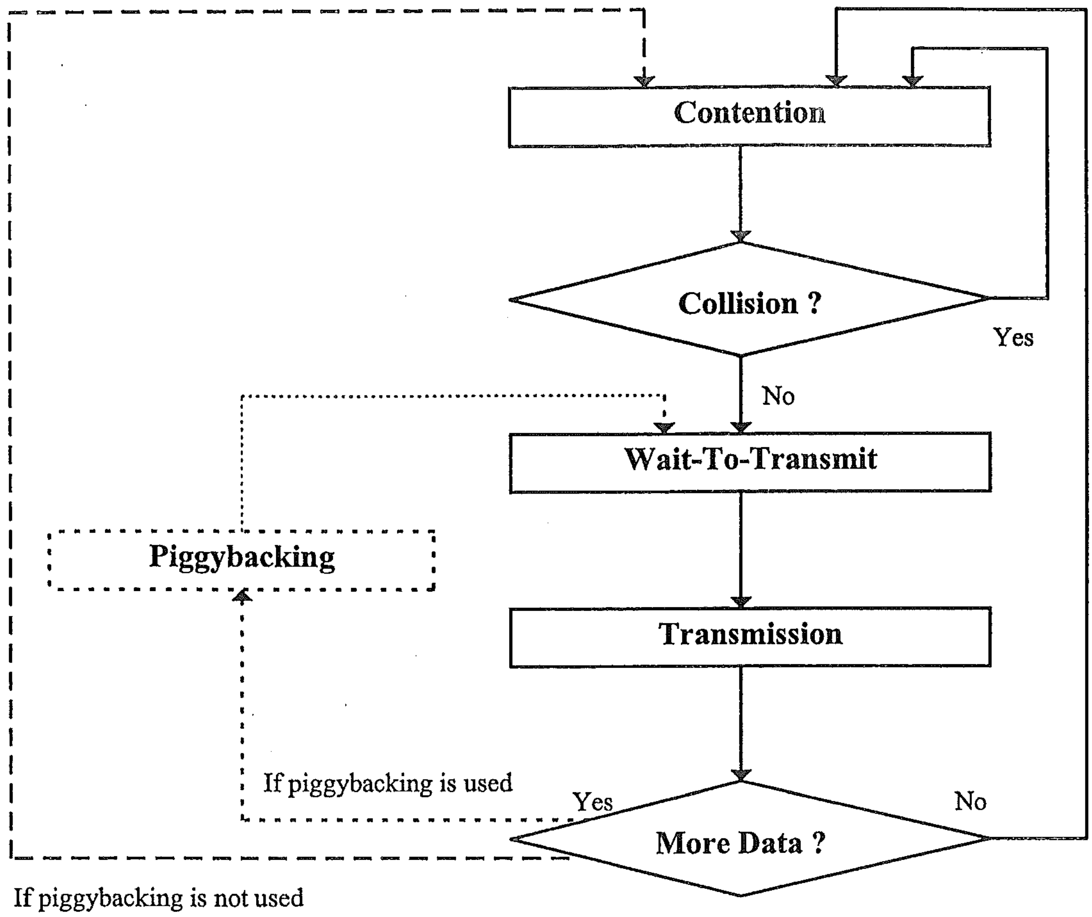

Figure 3.2 : A general mechanism used in most recent proposals for MAC protocols of WATM

3.2 Detailed Description of Scheme

In this section, we introduce our proposed solution to the "idle-VBR" problem. We propose a general mechanism which can be incorporated by all MAC protocols based on demand-assignment, whose reservation phase in based on pure-contention.

3.2.1 Reservation Partitions

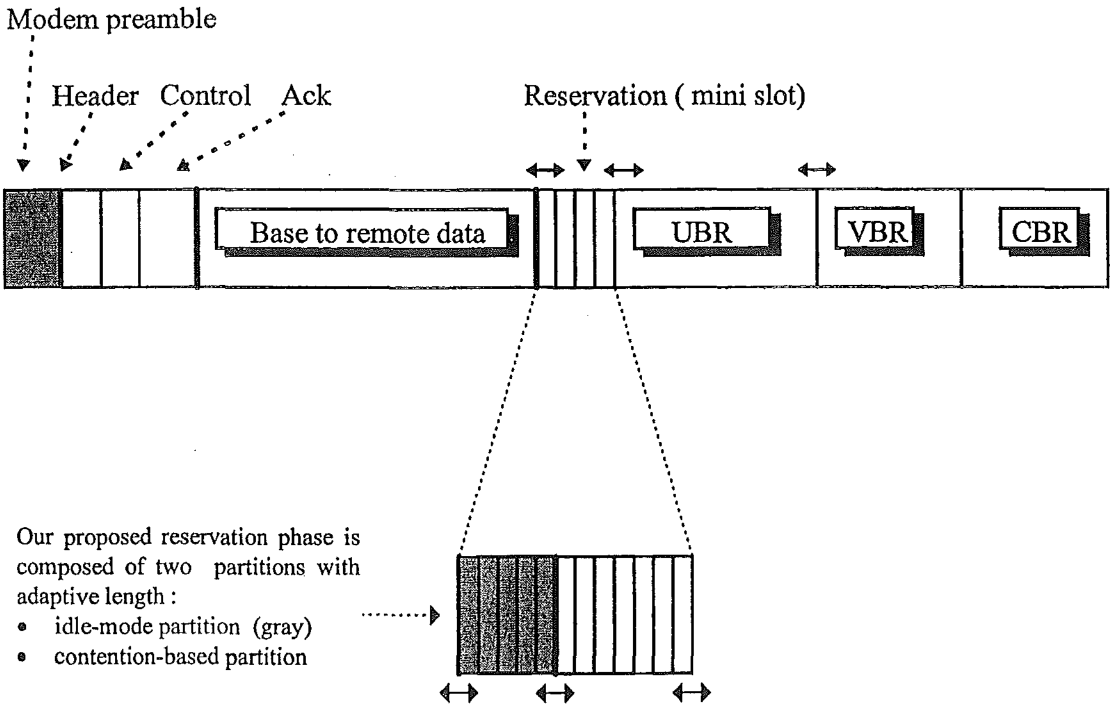

Figure 3.3 : Frame structure of a TDD-based MAC protocol for WATM (up) and our proposed dynamic hybrid partitioning for reservation part of the protocol (down)

Under our proposed scheme, the reservation phase of a TDD-based MAC protocol is divided into two partitions, the idle-mode and the contention-based partition. Both partitions are based on mini-slots. The length of both partitions is controlled by adaptive algorithms. The length of idle-mode partition can be zero but contention-based partition always maintains a non-zero length. This is because contention-based partition is used by new call arrivals. The idle-mode partition is used by currently accepted, but idle VBR traffic sources. The contention-based partition is used by new call arrivals of all traffic classes, and by currently accepted non delay-sensitive data such as UBR traffic sources. This partitioning approach provides a contention-free reservation mechanism for currently accepted delay-sensitive VBR traffic sources. Consequently, when a VBR traffic source wakes up from idle-mode to transfer data (e.g., at the beginning of a new talk spurt), it can transmit its reservation request using a pre-allocated mini-slot in the idle-mode partition without contention. Although this scheme decreases channel access delay, allocating one mini-slot for every accepted VBR traffic source all the time, wastes the scarce wireless bandwidth. To alleviate this problem, the length of the idle-mode partition is adjusted dynamically by an adaptive algorithm, in conjunction with a "repeat-parameter" which have been detailed in section 3.2.2. Figures 3.3 shows a simple view of our proposed, and purely contention-based reservation phase of a TDD-based MAC protocol for WATM.

3.2.2 Description of Algorithms

Figure 3.4 depicts a general view of our proposed solution for the idle-VBR problem. The dashed part of the flowchart in this Figure represents the proposed scheme. As it shows, under the proposed scheme, VBR terminals do not have to contend to access wireless channel once they are accepted into the network.

We are assuming that the idle-mode and contention-based partitions are controlled by separate adaptive algorithms. The contention-based partition is based on slotted-ALOHA. The length of this partition is controlled by a fast collision resolution algorithm adopted from [Kim96]. It works as follows. Assume N_c is the number of allocated mini-slots in contention-based partition and at the beginning $N_c = 1$. If the base station detects a collision, it assumes that at least two mobile terminals are contending and sets $N_c = 2$ for the next frame. In the next frame, if the base station detects collisions in both mini-slots, it assumes that at least four mobile terminals have been contending and sets $N_c = 4$, and so on. If the base station does not hear anything in a mini-slot, and do not detect any collision in contention phase, it assumes that the mini-slot is redundant and removes it in the next frame. The flowchart in Figure 3.5 presents this algorithm in more detail. As it shows, at the beginning, the length of contention-based partition equals to 2 mini-slots. At the beginning of a new frame, the base station recalculates (and broadcasts in the frame header) the length of the contention-based partition (L) as follows. At the beginning, the length of the contention-based partition equals to two mini-slots. The array l that has two elements : l(1) and l(2). At the beginning l(1) = 1 and l(2) = 1. If the base station detects a collision in the first or second mini-slot of the contention-based partition, it doubles the value stored in l(1) or l(2) respectively. On the other hand, if silence is detected in the first or second mini-slot, the corresponding values stored in l(1) and l(2) will be divided by 2.

The idle-mode partition is controlled by an adaptive algorithm and a user-defined parameter R, named "repeat-parameter". They work as follows. The maximum length of the idle-mode partition is given. When an active VBR terminal goes to idle-mode in a given frame, then, if the current length of the idle-mode partition is less then maximum allowed length, one mini-slot is allocated for it in the idle-mode partition, starting from the next frame. If the current length of idle-mode partition equals to the maximum allowed value, two different strategies can be adopted:

Strategy 1. VBR terminal waits until a mini-slot can be allocated in idle-mode partition. We call this strategy ***blocking*** because VBR terminal is blocked by the base station until a mini-slot can be allocated for it in the idle-mode partition. This strategy called ***blocking*** because if there is no room in the idle-mode partition for more mini-slots to be allocated, the VBR terminal has to ***wait*** until a mini-slot can be allocated. The flowchart in Figure 3.6 (page 42) shows this strategy. It shows that if there is no room in the idle-mode partition, the mobile terminal has to wait.

Strategy 2. VBR terminal switches to the contention-based partition and transmits its channel access request through contention. We call this strategy ***non-blocking*** because VBR terminal does not wait and uses contention-based partition. The flowchart in Figure 3.7 (page 43) presents this strategy. We see that when the current length of idle-mode partition does not allow for more mini-slots to be allocated, the mobile terminal

If ***blocking*** strategy is used, after allocating a mini-slot in idle-mode partition, the base station sends the information about the position of allocated mini-slot within idle-mode partition in the header of the next frame. When that VBR terminal wakes up from the idle-mode (e.g., at the beginning of a talk spurt), it uses the allocated mini-slot to send its channel access request without contention. Having received the request, the base station releases the allocated mini-slot (decreases the size of idle-mode partition by one mini-slot). In addition, the algorithm that controls the idle-mode partition uses a repeat parameter R (which can be defined by each VBR terminal independently) to determine how often a mini-slot should be assigned to an idle-VBR terminal before it wakes up. It works as follows. Let R_i represents the repeat parameter for i^{th} VBR terminal. If $R_i = 1$, then whenever the terminal goes to idle-mode, one mini-slot is allocated for it in idle-mode partition in every frame. If $R_i = 2$, one mini-slot is allocated in every 2 frames, and so on. It is obvious that the repeat parameter provides a trade-off between bandwidth efficiency and channel access delay. $R_i = 1$ provides the least access delay but may waste more bandwidth if the call does not wake up during multiple frames. In contrast, assigning a higher value to R leads to larger access delay because when a VBR terminal wakes up, it may have to wait several frames to send its access request in the allocated mini- slot. At the same time, a higher value for R saves some bandwidth (specially in high traffic load and dense areas, where there are many users in a cell) because mini-slot is not allocated in every frame.

If ***non-blocking*** strategy is used, the base station lets the VBR terminal to send its channel access request packet via contention in contention period. In Section 4.2.5 we have analyzed the effects of ***blocking*** and ***non-blocking*** strategies on the performance of the protocol.

The flowchart in Figure 3.4 presents how these two strategies fit into our proposed scheme.

3.3 Summary

In this chapter, after a brief introduction to the "idle-mode VBR" problem, an enhanced reservation scheme for reservation phase of a TDD-based MAC protocol for WATM has been proposed. We called it as the ***Dynamic Hybrid Partitioning with Adjustable Repeat.*** Under this scheme, the reservation phase of the MAC protocol is divided into two separate adjacent partitions called the idle-mode and the contention-based partitions, respectively. The idle-mode partition is used by currently accepted delay-sensitive idle VBR terminals, when they wake up from idle mode. The contention-based partition is used by non-delay sensitive data terminals and new arrivals. Separate adaptive algorithms control these partitions. In addition, we discussed how a repeat parameter R has been associated with our scheme to provide more flexibility.

We expect this scheme to improve the channel access delay for idle VBR traffic while putting negligible redundancy on wireless channel. In addition, a load independent (flat) delay characteristic for VBR traffic is expected. This is, in particular, of paramount importance in higher traffic loads and crowded areas, where simple contention-based reservation schemes may suffer from fast performance degradation. In the next chapter, we evaluate various aspects of this scheme through computer simulation.

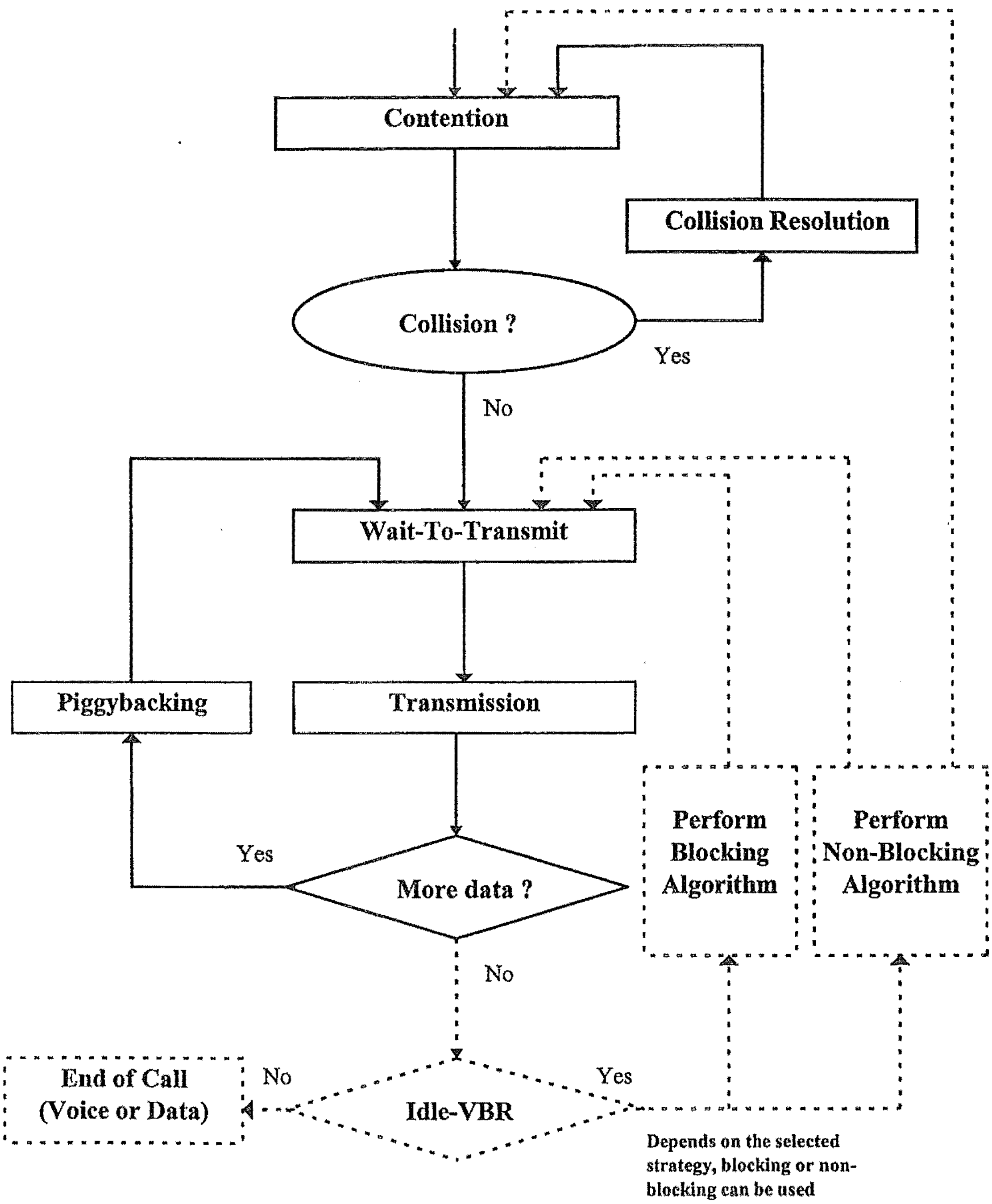

Figure 3.4 : Flowchart of our proposed scheme called "Dynamic Hybrid Partitioning with Adjustable Repeat".

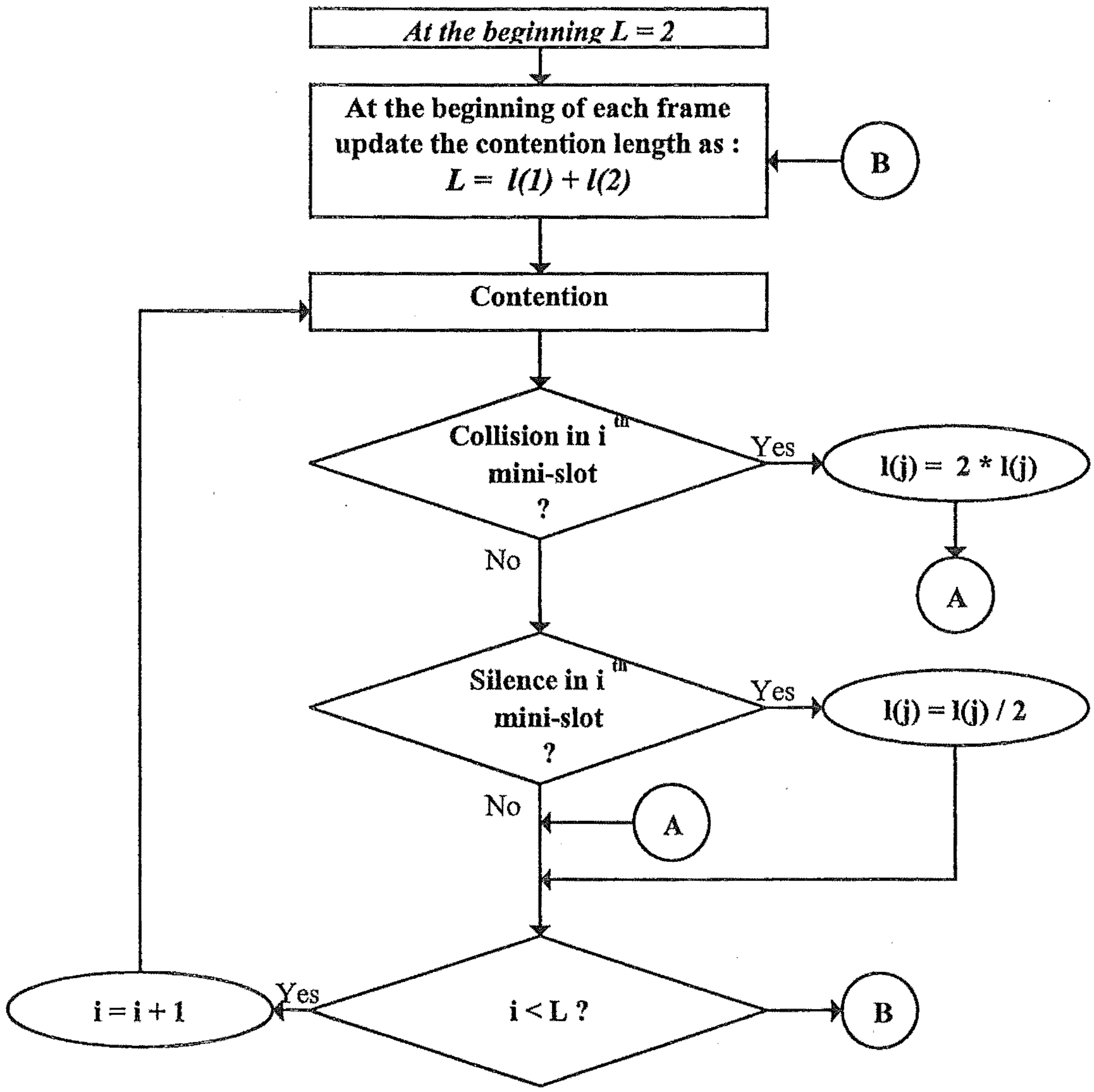

Figure 3.5 : Flowchart of the algorithm that controls length of the contention-based partition

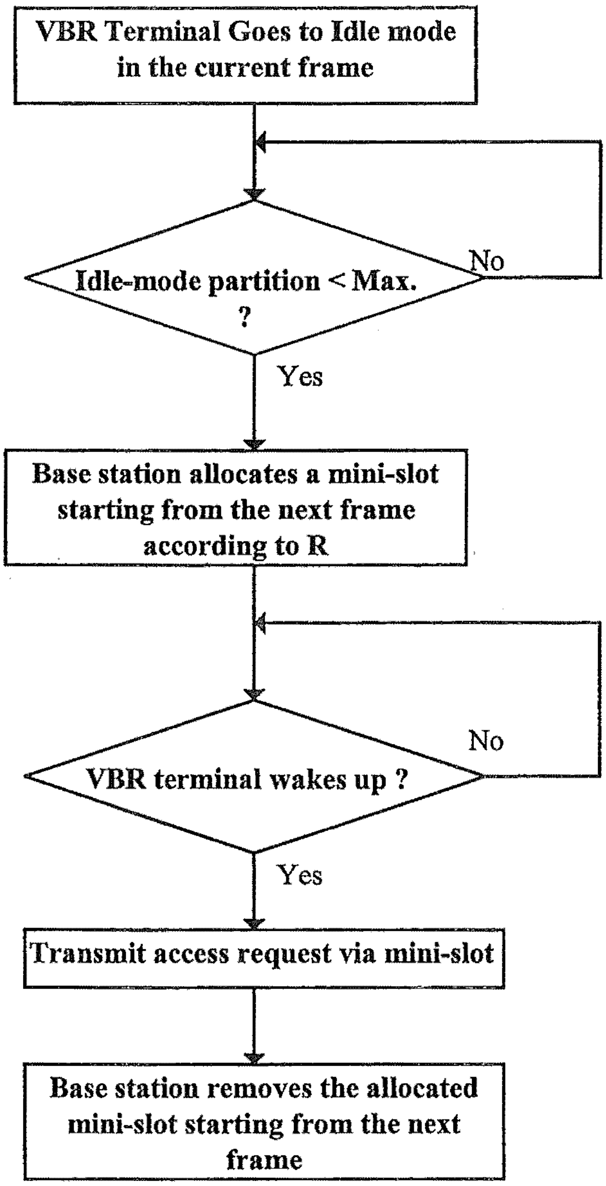

Figure 3.6 : Flowchart of blocking strategy for the algorithm that controls the idle-mode partition

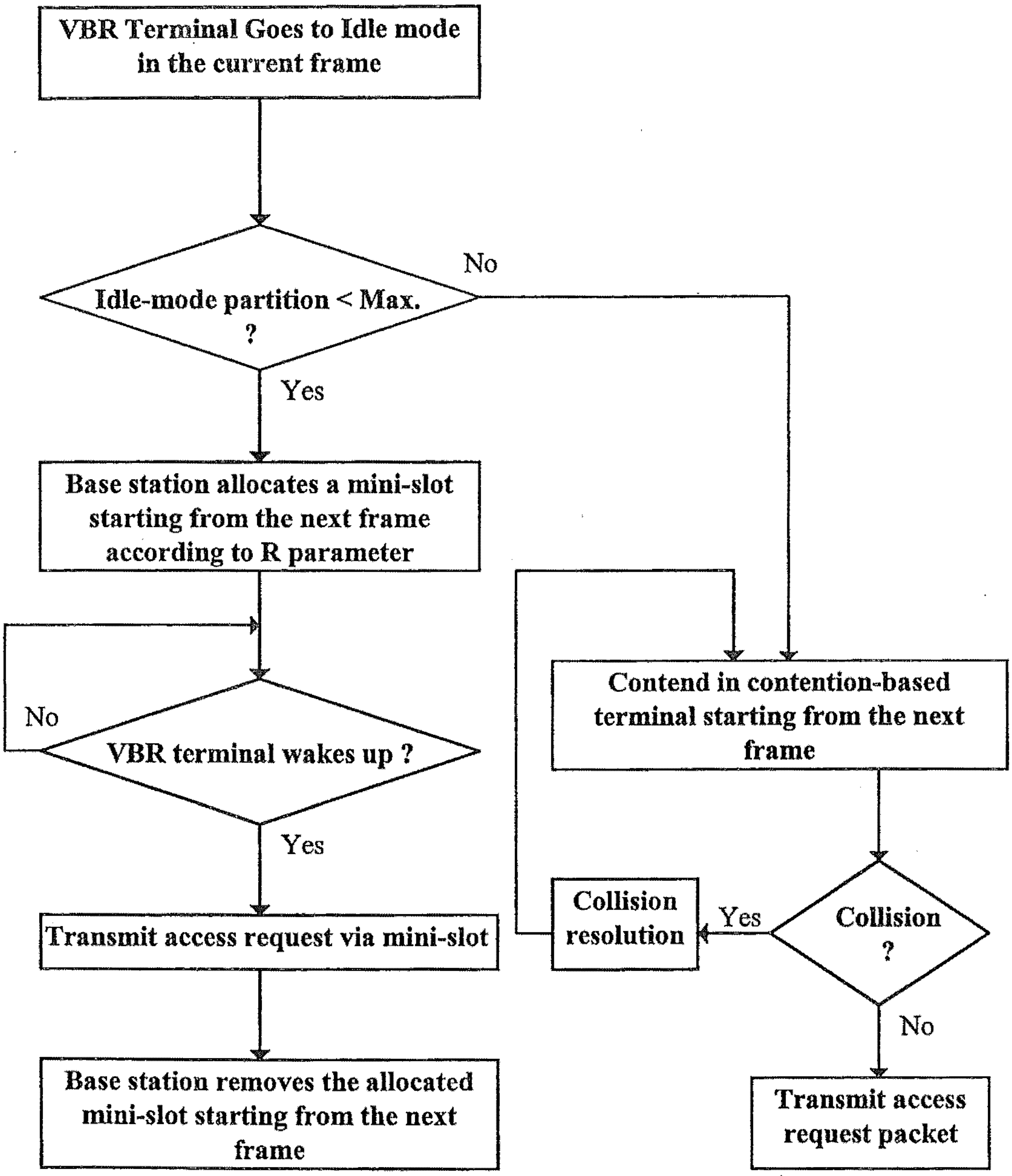

Figure 3.7 : Flowchart of non-blocking strategy for the algorithm that controls the idle-mode partition

Chapter 4

Performance Evaluation and Comparison

4.1 Overview and Assumptions

In this chapter, we evaluate the performance of the Dynamic Hybrid Partitioning scheme through computer simulation. The simulation program has been written in C++, linked to AKAROA2 [Ewi99], and run on a network of UNIX workstations. AKAROA2 transparently transforms a sequential simulation program into one for parallel execution on a network of UNIX workstations and automatically stops the simulation when the steady-state estimates of performance measures obtain the required relative precision.

To provide a clear view of various aspects of our proposed scheme, the performance evaluation has been done assuming several scenarios, each one focused on a different performance parameter. The following performance parameters were distinguished: the number of VBR voice terminals, the number of data terminals, the repeat parameter, the maximum length of idle-mode partition, and alternative strategies in the algorithm that controls the idle-mode partition. In addition, we briefly present the effect of simulation precision on our final results. In all simulation scenarios, a mixed voice/data environment has been assumed. We have not considered the video traffic because our proposed scheme does not affect the performance of VBR video. This is because a VBR video stream does not go to the idle-mode. Yet we have considered a reasonably high bit-rate capacity (20 Mbps) for wireless channel so that VBR video can coexist in our assumed environment.

Additional assumptions, specially considered with each simulation scenario, have been specified in the corresponding graphs. The following general assumptions have been made for all scenarios :

A1. TDD-based MAC protocol with a frame length of 2 milliseconds has been assumed.

A2. ON/OFF model [Fis92] has been used to approximate the behavior of VBR voice and UBR data. The specified numeric parameters associated with ON/OFF models described in table 4.1 have been used in many other experiments such as [Fis92]. In the rest of this chapter, for simplicity the term "data terminal" has been used for "UBR data terminal".

A3. In all simulation scenarios, a mixed voice/data traffic has been considered.

A4. For simplicity, a symmetric traffic pattern in which uplink and downlink are equally loaded, has been considered.

A5. Due to its simplicity, First-In-First-Out (FIFO) packet scheduling policy has been considered in the base station.

A6. All simulation results characterize the steady state behavior of the protocol. They are repeated with a relative simulation precision of 5% (except in Section 4.2.6, where we study the effect of relative precision on the quality of simulation results).

A7. In the rest of this chapter we use R and L_{MAX} to refer to the repeat parameter, and the maximum allowable length of the idle-mode partition respectively.

Frame length	2 ms
Size of data slot	56 bytes as follows : 48 bytes payload 8 bytes ATM and wireless header + SYNC + FEC
Size of mini slot	8 bytes
Max. length of contention partition	24 mini slots
Max. length of idle-mode partition	24 mini slots (holds for all scenarios except when we analyze the effects of this parameter on the performance of MAC protocol)
Packet scheduling policy	First-In-First-Out (FIFO)
Idle-mode partition algorithm	Blocking (unless otherwise has been mentioned)
Architecture	Cellular (centralized control at base station)
Traffic characteristics generated by a VBR voice terminal	Two state ON/OFF model ON = 1 sec , Rate = 8 kbps OFF = 1.3 sec
Traffic characteristics generated by a UBR data terminal	Two state ON/OFF model ON = 0.1 sec , Rate = 64 kbps OFF = 1.9 sec
Relative simulation precision	5 % (unless otherwise has been mentioned)

Table 4.1 : Summary of the main simulation assumptions

4.2 Simulation Scenarios and Results

4.2.1 Dependence of Performance on the Number of VBR Terminals

In this section, the effects of the number of VBR terminals on the mean channel access delay, and on the mean channel access delay variation experienced by idle VBR terminals, have been studied. Here, by the channel access delay we mean the time elapsed since an idle-mode VBR terminal wakes up until it can successfully transmits its channel access request packet to the base station. The mean channel access delay variation can be defined as follows:

$$\text{mean\_variation} = \left(\sum_{i=1}^{i=N} \left| D_i - ED \right| \right) / N$$

Where $\mathbf{D_i}$ is the channel access delay experienced by the i^{th} mobile terminal, **ED** is the mean channel access delay experienced by mobile terminals, and **N** is number of mobile terminals in the cell. The mean channel access delay variation is important because it contributes to the final Cell Delay Variation (CDV) of ATM cells. In addition, the effect of the length of reservation partitions vs the number of VBR terminals has been studied to determine how much redundancy is added on wireless channel by our scheme.

Figures 4.1 shows the effect of the number of VBR terminals on the mean channel access delay experienced by idle-VBR terminals in a mixed voice/data scenario, for R=1. As it shows, under our proposed scheme, the channel access delay experienced by idle-VBR terminals is constant (slightly more then 1 millisecond), independent from the number of VBR or data terminals in a cell. This is because under our proposed scheme, idle-VBR terminals use idle-mode partition to transmit their channel access request to the base station without contention, each time they wake up. In contrast, when using a pure contention-based reservation scheme, as the graph shows, the experienced mean channel access delay by idle-VBR terminals is increasing with the number of VBR or data terminals in the cell. The difference in the experienced mean channel access delay is more significant at higher loads because of more collisions.

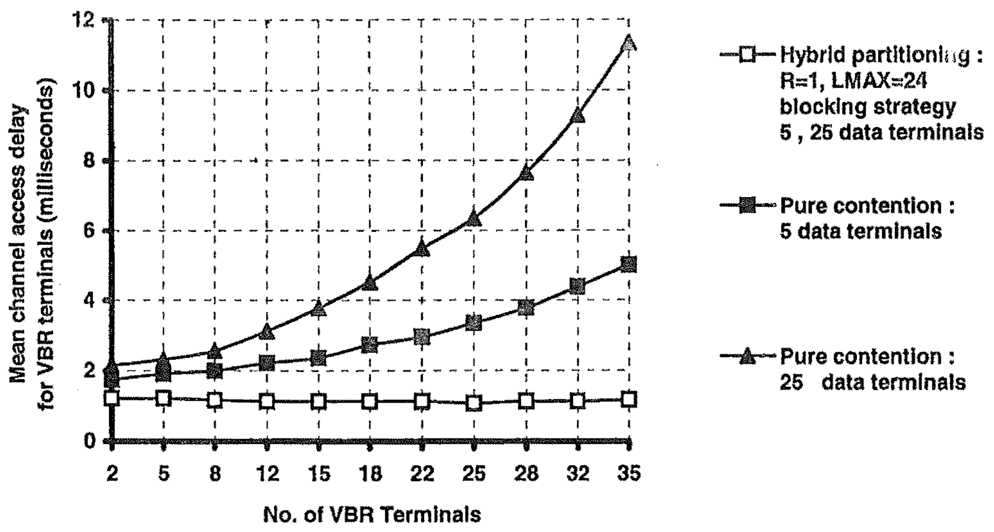

Figure 4.1 : Mean channel access delay for VBR terminals vs the number of VBR terminals.

Figure 4.2 shows the mean channel access delay variation vs the number of VBR terminals, for two different numbers of data terminals and R =1. As mentioned, this variation is important because it contributes to the final CDV of ATM cells which is an important QoS parameter. It is observed that under our proposed scheme, the mean channel access delay variation for VBR terminals is constant (about 0.8 millisecond) and independent from the number of VBR or data terminals in the cell. As it shows, using a pure contention-based reservation mechanism leads to higher variations specially at higher traffic load. The reason for the channel access delay variation experienced by VBR terminals is as follows. When an active VBR terminal goes to the idle mode in a given frame, the base station tries to allocate a mini-slot in the idle-mode partition starting form the next frame. Depending on the position in which an idle VBR terminal wakes up in a frame, it experiences different time intervals before transmitting its channel access request packet to the base station. For example, if an idle-VBR terminal wakes up just before the uplink part of the MAC frame, it can transmit its channel access request almost without any delay. In contrast, if an idle-VBR terminal wakes up just after the idle-mode partition, and R =1, it has to wait a period equal to the length of the MAC frame, which is 2 milliseconds in our study . When considering a pure contention-based reservation mechanism, the mean channel access delay variation is larger, because VBR terminals have to transmit its channel access request through contention, each time they wake up. In the case of collision, VBR terminal has to wait until the next frame, and then try again.

Above explanation has been reflected in Figure 4.2, where one can observe the existence of an upper bound for the mean channel access delay variation under our scheme, while the mean channel access delay variation monotonically increases under a pure contention-based reservation mechanism.

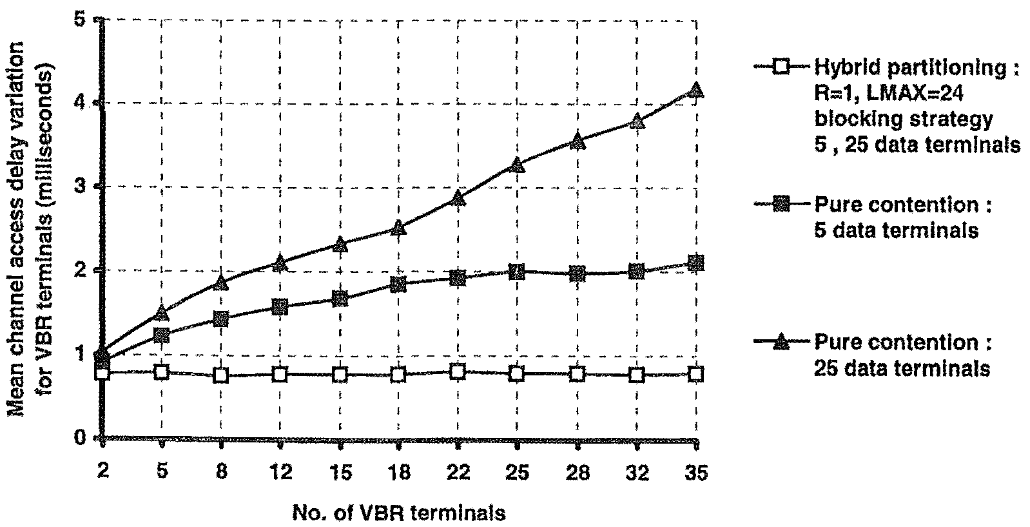

Figure 4.2 : Mean channel access delay variation for VBR terminals vs the number of VBR terminals.

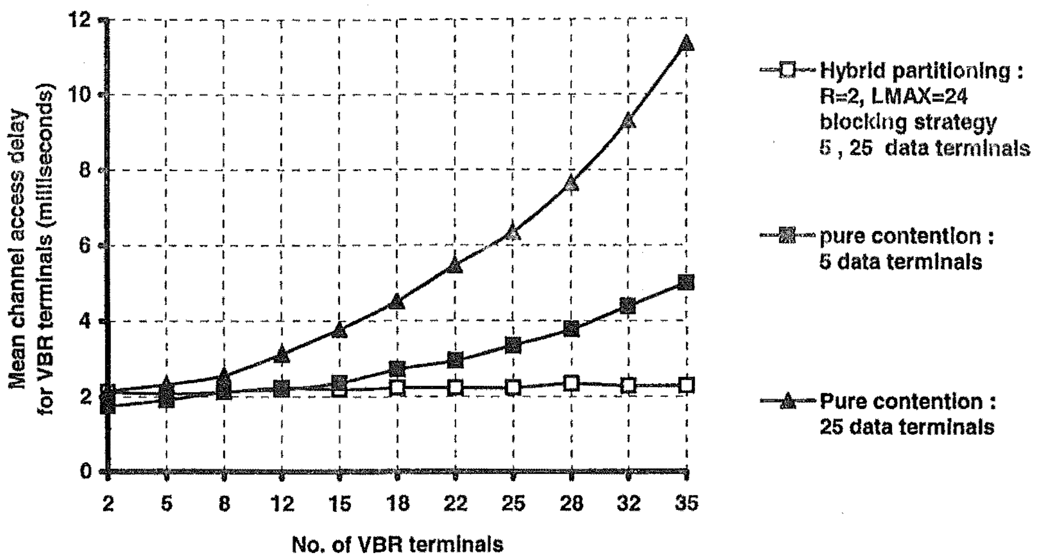

Figure 4.3 : Mean channel access delay for VBR terminals vs the number of VBR terminals.

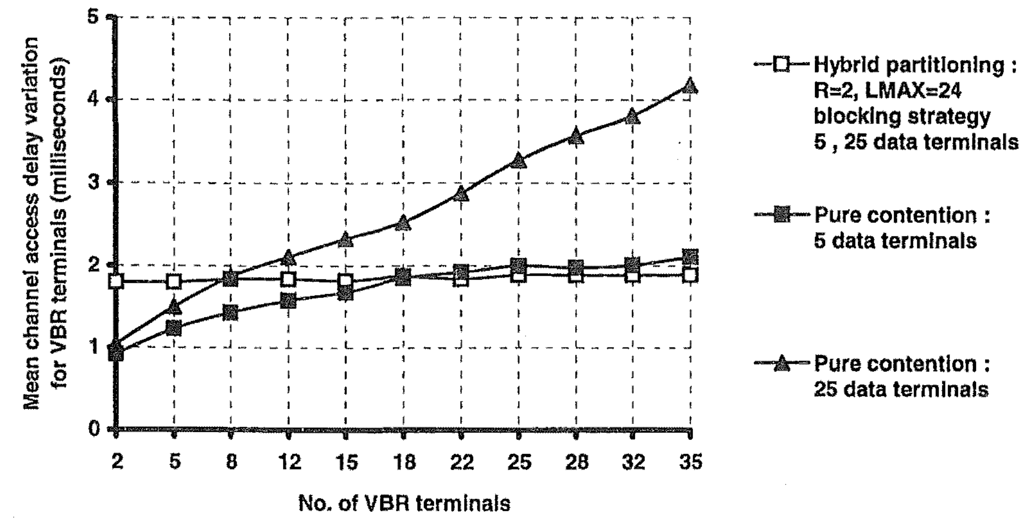

Figure 4.4 : Mean channel access delay variation for VBR terminals vs the number of VBR terminals.

Figure 4.3 and 4.4 were attained for simulation scenarios similar to those assumed in Figure 4.1 and 4.2, respectively, but with R = 2. The effects of repeat the parameter (R) on the performance of our proposed scheme are discussed in full detail in Section 4.2.3. Here, we only show that while under our proposed scheme, the mean channel access delay is constant and independent from the number of VBR terminals in the cell, its value changes with R, in comparison to the results in Figure 4.3, for R=2. The reason for the shift is that with R = 2, after a VBR terminal goes to the idle mode, the base station allocates one mini-slot in the idle-mode partition ***every each*** frame. Consequently, after waking up, the VBR terminal is likely to wait a longer period to be able to transmit its channel access request packet (in the worst case, the VBR terminal has to wait 4 milliseconds = 2 frame length). Figure 4.4 shows the mean channel access delay variation experienced by VBR terminals when R = 2. Again it is obvious that under our scheme, mean access delay variation is constant and has been shifted up.

Figures 4.5 and 4.6 show the mean length of different partitions vs number of VBR terminals in a cell. Both graphs are obtained under the same scenario but with different numbers of data terminals. In Figure 4.5, we observe that, under our scheme, the length of contention-based partition is independent from the number of VBR terminals in the cell. This is because VBR terminals do not use contention based partition to transfer their channel access request packet. Figure 4.5 also shows the redundancy put by our scheme on the wireless channel. As it shows, even at higher loads, this redundancy is trivial and is smaller than one data slot. Figure 4.6 shows results for the same simulation scenario with number of data terminals equal to 25. We observe that by increasing the number of data terminals, the mean length of contention-based partition gets shifted upward but still is independent from the number of VBR terminals. The reason for this vertical shift is that by increasing the number of data terminals, the base station experiences more collisions and increases the length of the contention-based partition.

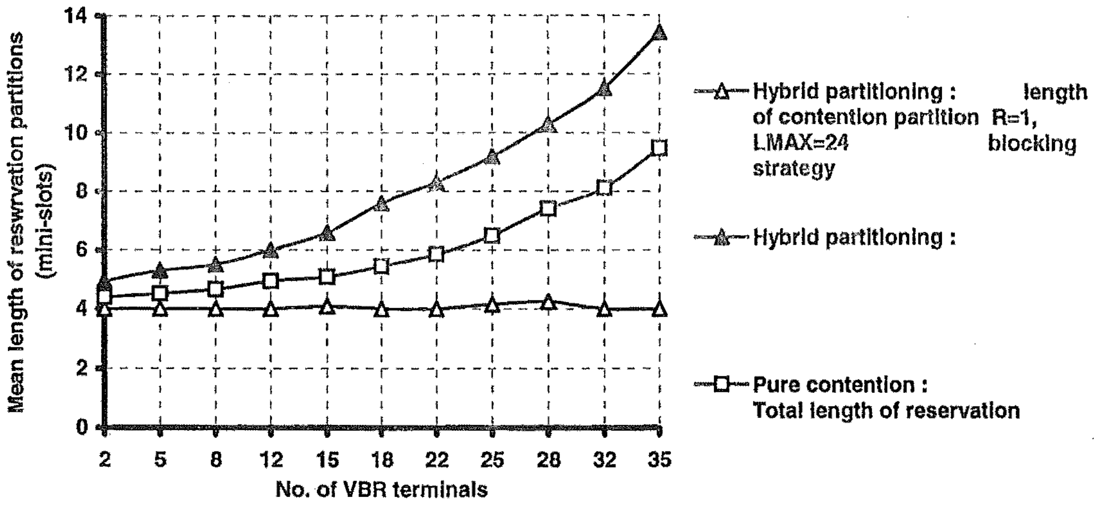

Figure 4.5 : Mean length of reservation partitions vs the number of VBR terminals. (10 data terminals)

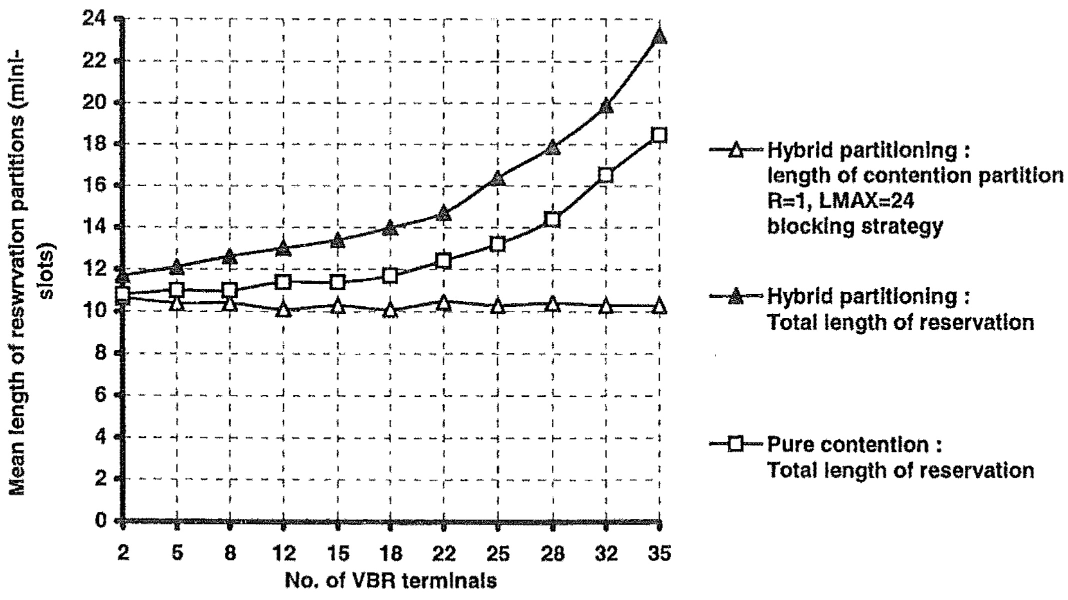

Figure 4.6 : Mean length of reservation partitions vs the number of VBR terminals. (25 data terminals)

4.2.2 Dependence of Performance on the Number of Data Terminals

In this section, the effects of number of data terminals on the mean channel access delay and the mean channel access delay variation experienced by VBR and data terminals is discussed. Figure 4.7 shows the mean channel access delay experienced by VBR terminals vs the number of data terminals in a cell, assuming R=1. It is obvious that under our scheme, the delay experienced by VBR terminals is constant and independent from the number of data terminals. This is because under our scheme, VBR terminals use pre-allocated mini-slots in the idle-mode partition to transmit their channel access request packets each time they wake up from the idle-mode. In contrast, when reservation phase is based on pure contention, the mean channel access delay experienced by VBR terminals increases with increase of the number of data terminals, as a consequence of more collisions between contending terminals.

Figure 4.8 shows the mean channel access delay variation experienced by VBR terminals vs the number of data terminals in a cell, for R =1. As it shows, under our scheme, the experienced mean channel access delay variation by VBR terminals is independent from the number of the data terminals. This is because the contention-free transmission of channel access request packets in idle-mode partition, provides a guaranteed upper bound on the mean channel access delay variation. On the other hand, using a pure contention-based reservation scheme, the variation experienced by VBR terminals rapidly grows increase of the number of data terminals.

Figures 4.9 and 4.10 present the mean channel access delay and mean channel access delay variation experienced by data terminals vs the number of data terminals. One can see that the mean channel access delay and mean channel access delay variation experienced by data terminals have been reduced under our proposed scheme. Yet we do not have a flat delay characteristic because data terminals have to contend after waking up from idle mode.

One can argue that measuring delay for non-delay sensitive traffic and presenting it as an advantage of a MAC protocol has no benefit. Although we agree with this argument in general, the reason that we look at this delay here is that the improved delay performance for data terminals can be considered as an extra bonus offered by the proposed scheme. For example, transferring data files at higher speed with less delay is definitely desirable, although this delay is not a critical factor.

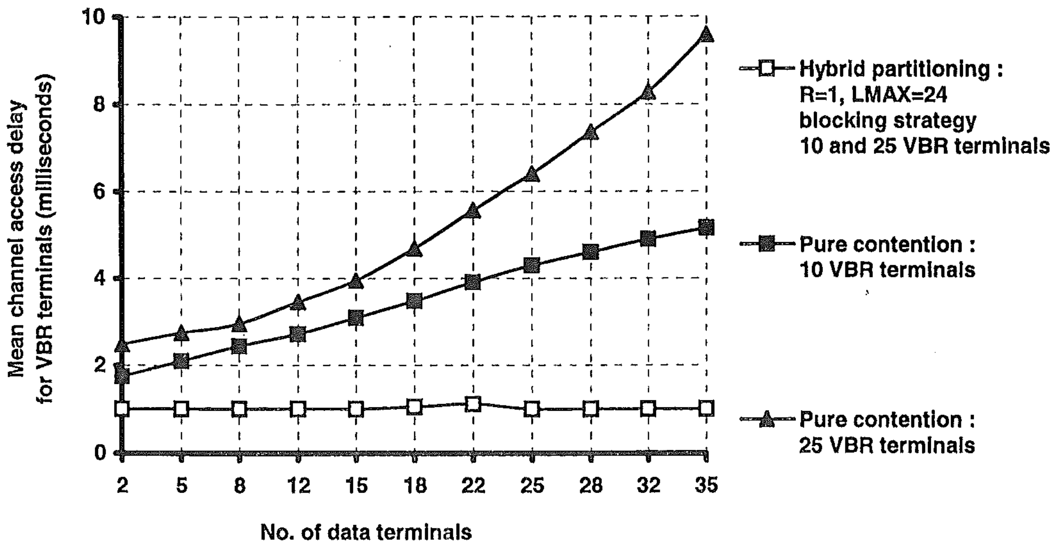

Figure 4.7 : Mean channel access delay for VBR terminals vs number of data terminals.

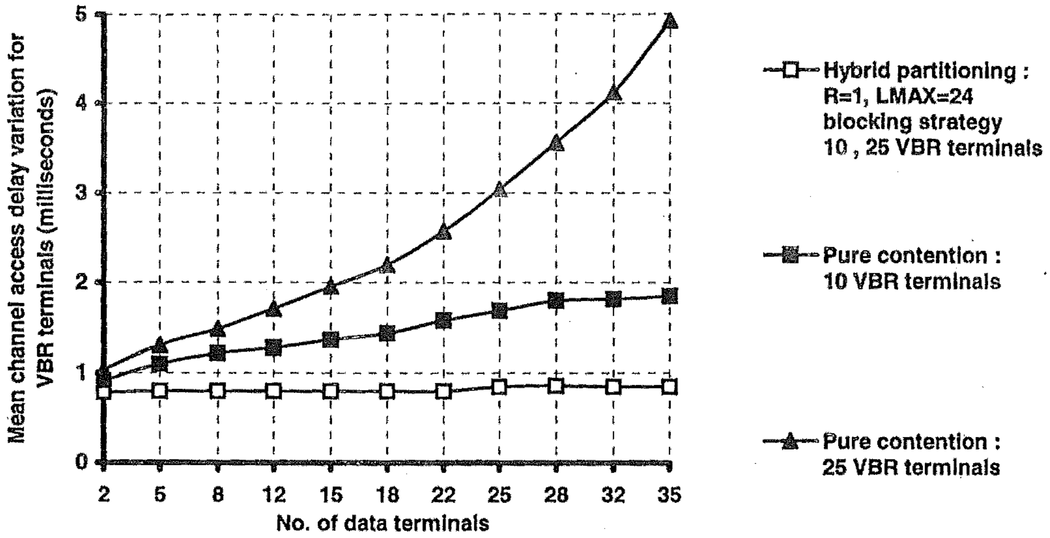

Figure 4.8 : Mean channel access delay variation for VBR terminals vs number of data terminals.

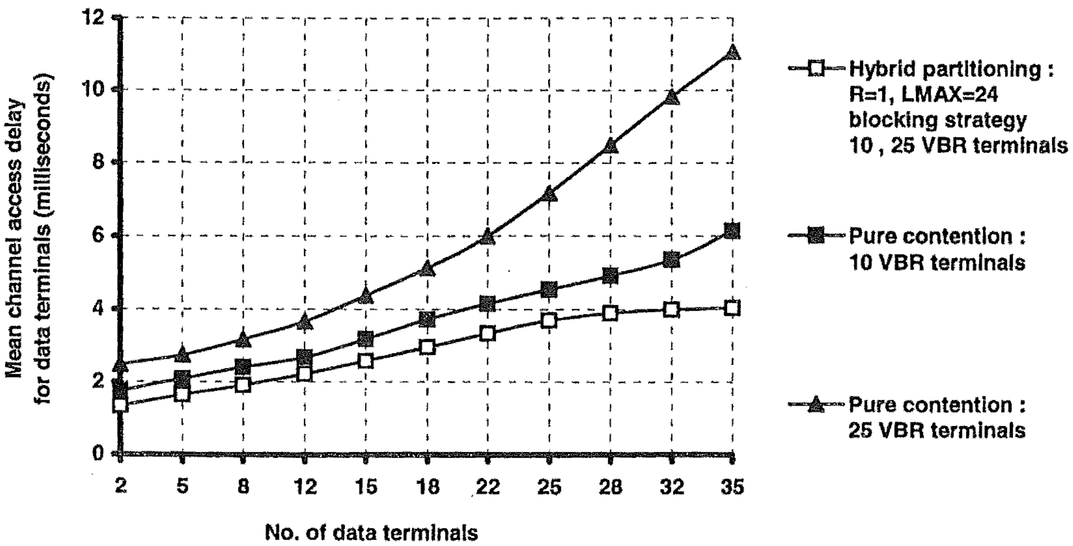

Figure 4.9 : Mean channel access delay for data terminals vs number of data terminals.

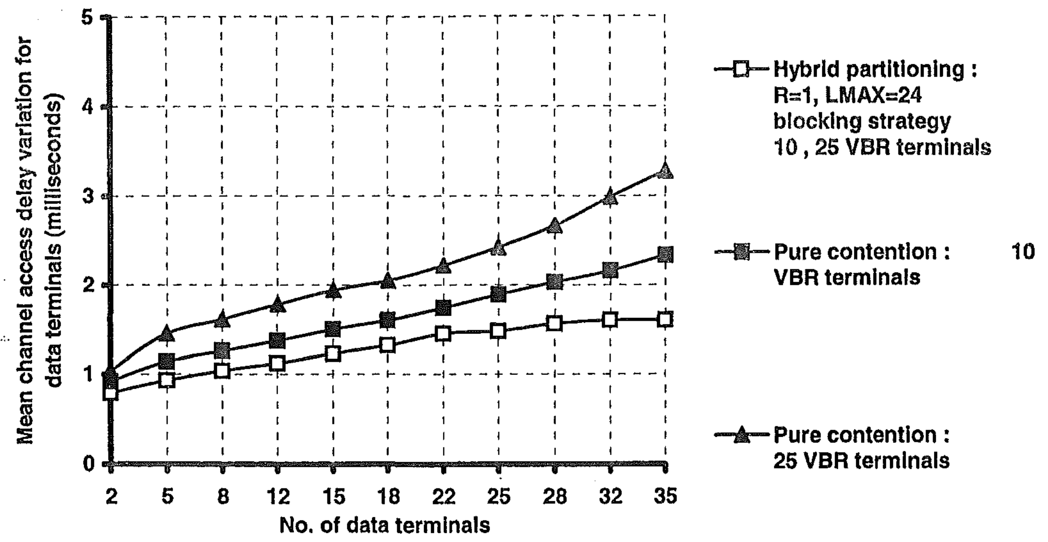

Figure 4.10 : Mean channel access delay variation for data terminals vs number of data terminals.

4.2.3 Dependence of Performance on the Repeat Parameter

In this section, effects of the repeat parameter (R) on the mean channel access delay, and mean channel access delay variation, experienced by VBR terminals have been studied.

Figure 4.11 presents the mean channel access delay experienced by VBR terminals for different values of R. One can observe that experienced delay is independent from the number of VBR terminals for different values assigned to R. Yet, assigning a higher value to R increases the mean channel access delay for VBR terminals. This effect is easily seen in Figure 4.11. With a larger value for R, when an idle-VBR terminal wakes up, it may have to wait a longer interval (several frames) to have the chance to transmit its channel access request packet. Figure 4.11 also shows the mean channel access delay for VBR terminals when reservation phase is only based on contention. As it shows, the mean channel access delay rapidly increases when we increase the number of VBR terminals, and this is a consequence of more packet collisions. Figure 4.12 was attained by studying a scenario similar to that discussed in Figure 4.11, but with smaller number of data terminals. It clearly shows that under our scheme, assuming different values of R, the mean channel access delay experienced by VBR terminal is always independent from the number of data terminals. On the other hand, we observe that decreasing the number of data terminals in a cell, causes VBR terminals to suffer shorter channel access delays than under a pure contention-based reservation scheme.

Figures 4.13 presents the mean channel access delay variation for VBR terminals vs number of VBR terminals, for different values of R. It shows that R has a significant effect on the mean channel access delay variation. A larger value for R results in a larger mean channel access delay variation. This is because with a larger R an idle-VBR terminal may have to wait a longer period before being able to transmit its channel access request each time it wakes up. Figure 4.14 presents similar results for smaller number of data terminals. We observe that by decreasing the number of data terminals, under pure contention scheme, VBR terminals experience shorter channel access delay because fewer collisions happen. At the same time, it shows that, under our scheme, the mean channel access delay for VBR terminals is independent from the number of data terminals.

Figures 4.15 and 4.16 depict mean length of the reservation partitions vs the number of VBR terminals for several values of R. They clearly show that by increasing R from 1 to 2, the total length of reservation phase becomes slightly shorter (less redundancy on wireless channel) at the price of longer access delays.

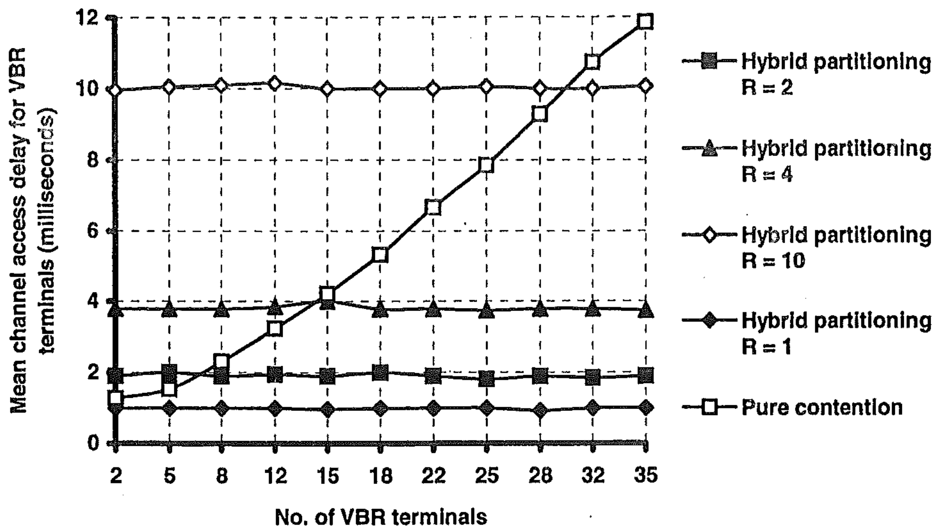

Figure 4.11 : Mean channel access delay for VBR terminals vs number of VBR terminals using pure contention and hybrid partitioning (L_{MAX}=24, blocking strategy , 35 data terminals)

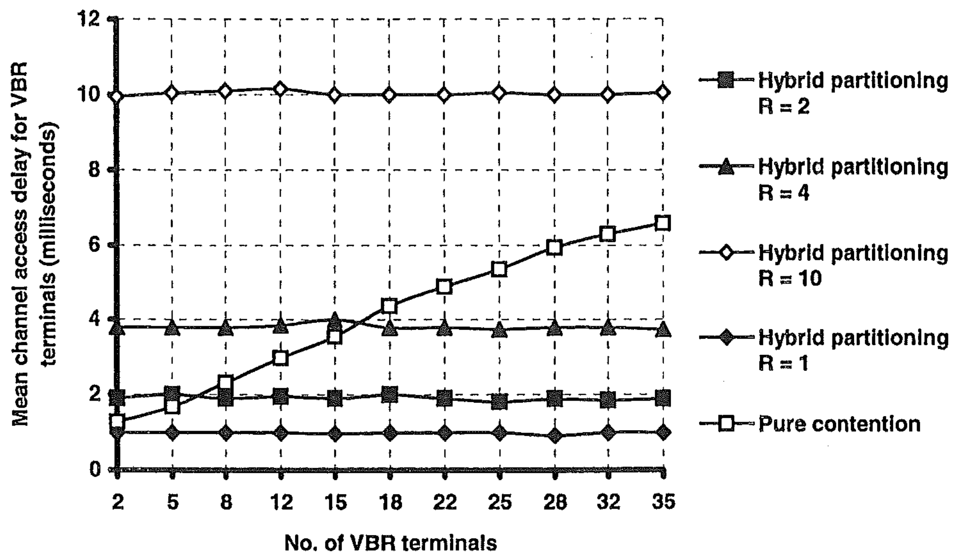

Figure 4.12 : Mean channel access delay for VBR terminals vs number of VBR terminals using pure contention and hybrid partitioning (L_{MAX}=24, blocking strategy, 10 data terminals)

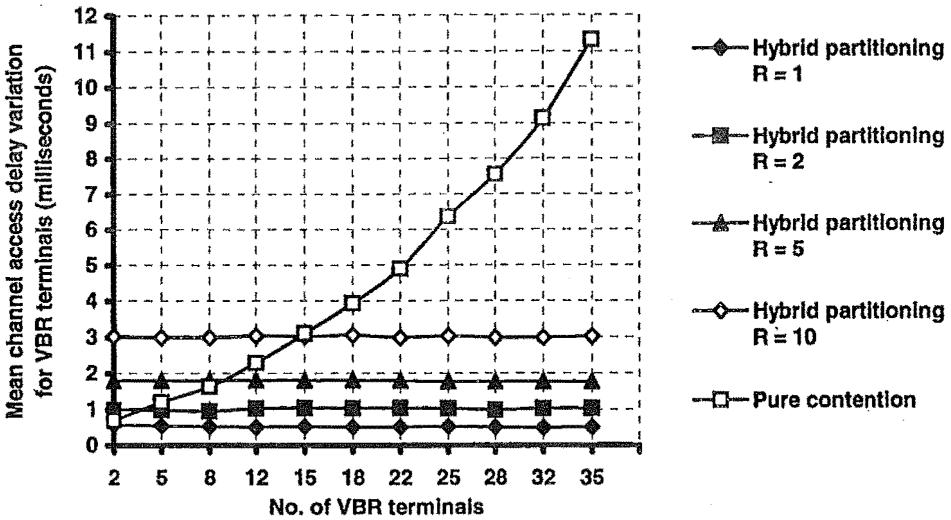

Figure 4.13 : Mean channel access delay variation for VBR terminals vs number of VBR terminals using pure contention and hybrid partitioning (L_{MAX}=24, blocking strategy, 35 data terminals)

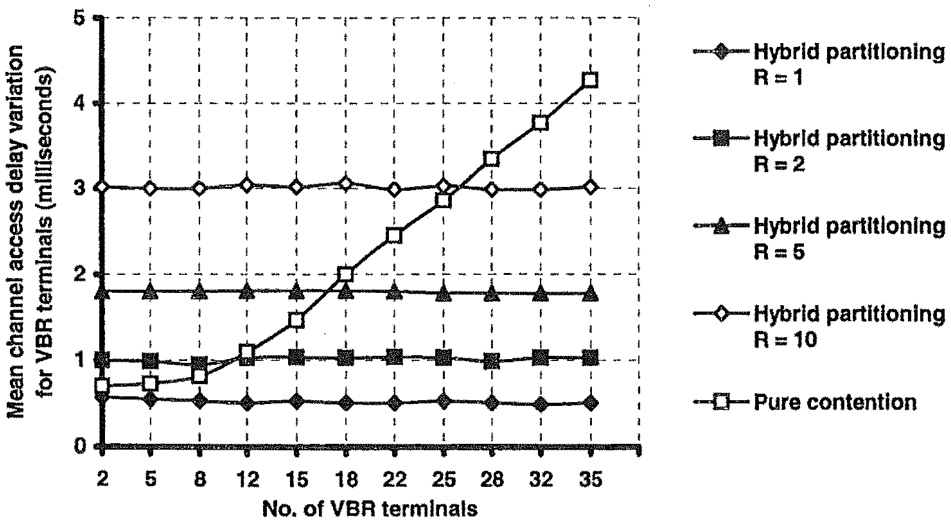

Figure 4.14 : Mean channel access delay variation for VBR terminals vs number of VBR terminals using pure contention and hybrid partitioning (L_{MAX}=24 , blocking strategy)

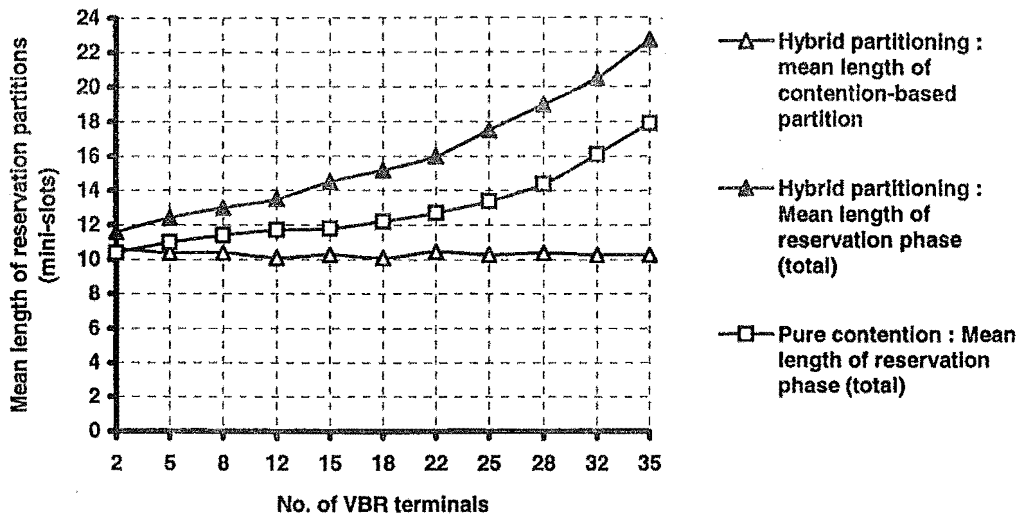

Figure 4.15 : Mean length of reservation partitions vs number of VBR terminals. (R=1, L_{MAX}=24, blocking strategy, 25 data terminals)

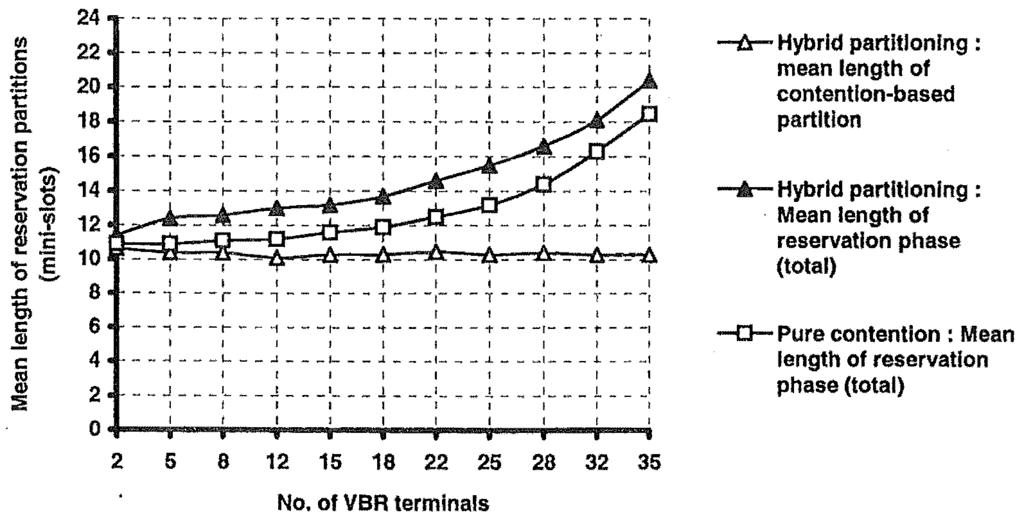

Figure 4.16 : Mean length of reservation partitions vs number of VBR terminals. (R=2, L_{MAX}=24, blocking strategy, 25 data terminals)

4.2.4 Dependence of Performance on the Maximum Length of Idle-Mode Partition

In this section, effects of maximum allowed length of idle-mode partition (L_{MAX}) on the mean channel access delay and the mean channel access delay variation experienced by VBR terminals, assuming ***blocking strategy*** for idle-mode partition, have been studied. In all previous experiments, we considered a large value of L_{MAX} because our purpose was to focus on other aspects of our scheme, without being worried about the effect of this parameter.

Figure 4.17 shows the mean channel access delay experienced by VBR terminals vs L_{MAX} . Different curves in Figure 4.17 have been obtained by changing the number of VBR terminals. As Figure 4.17 shows, after increasing L_{MAX} to 12 or more, the number of VBR terminals in a cell has almost no effect on the mean channel access delay. This is because even when the number of VBR terminals is reasonably high at any given time, only a fraction of them is in idle-mode or has just waked up. Consequently, there is no need to consider a very large upper limit for L_{MAX} . In fact, as Figure 4.17 shows, considering an upper limit between 12 to 16 mini-slots for L_{MAX} , even at reasonably high load, provides a good balance between delay performance of VBR terminals and added redundancy on wireless channel.

Figure 4.18 shows the mean channel access delay variation experienced by VBR terminals vs L_{MAX} . It shows similar results to those presented in Figure 4.17. As we see, for a L_{MAX} of about 12 or more mini-slots, the mean channel access delay variation is almost constant.

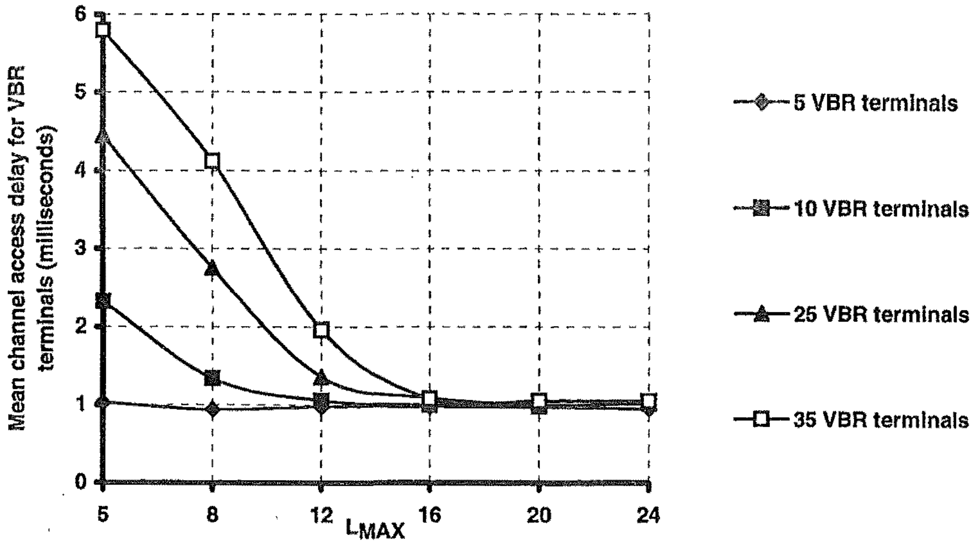

Figure 4.17 : Mean channel access delay for VBR terminals vs L_{MAX} assuming hybrid partitioning. (R=1, blocking strategy)

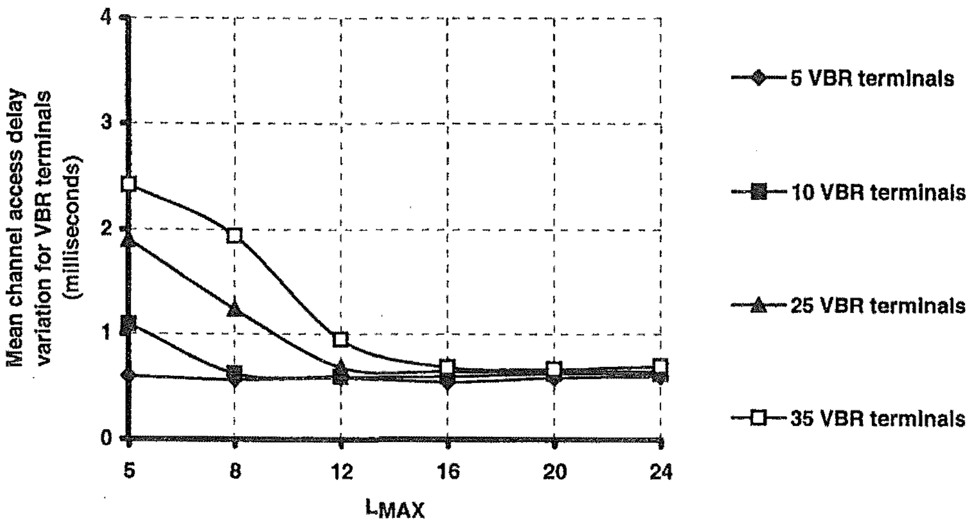

Figure 4.18 : Mean channel access delay variation for VBR terminals vs L_{MAX} assuming hybrid partitioning. (R=1, blocking strategy)

4.2.5 Blocking vs Non-blocking Strategies for Idle-Mode Partition

As discussed in Section 3.2.2, when a VBR terminal goes to the idle-mode and there is no room in the idle-mode partition to allocate mini-slot, ***blocking*** or ***non-blocking*** strategy can be adopted. As a brief reminder, under blocking strategy, when there is no room in the idle-mode partition to allocate a mini-slot for a VBR terminal that has just gone to the idle-mode, the VBR terminal has to wait (even if it wakes up) until the base station can allocate a mini-slot. In contrast, under non-blocking strategy, when a VBR terminal goes to the idle-mode and it is not possible for the base station to allocate a mini-slot for it, the VBR terminal switches to contention partition after waking up, and transmits its channel access request packet via contention. In this section, we compare and contrast these two strategies based on the results obtained by simulation.

Figures 4.19 to 4.22 present the effects of these two strategies on the mean channel access delay for VBR terminals vs the number of VBR terminals, assuming that R =1. In Figure 4.19, we have assumed that L_{MAX} equals 5 mini-slots. As it shows, when non-blocking strategy has been used, increasing the number of data terminals causes VBR terminals to suffer from larger channel access delay. In contrast, when blocking strategy is used, the number of data terminals has no effect on the mean channel access delay of VBR terminals. This is an expected result because under blocking strategy, VBR terminals do not use contention-based partition at all. Figures 4.20 to 4.22 present similar investigations, done for different values for L_{MAX}. They clearly show that by increasing L_{MAX} the results obtained assuming blocking or non-blocking strategy get closer, so changing L_{MAX} from 12 to 16, causes negligible difference.

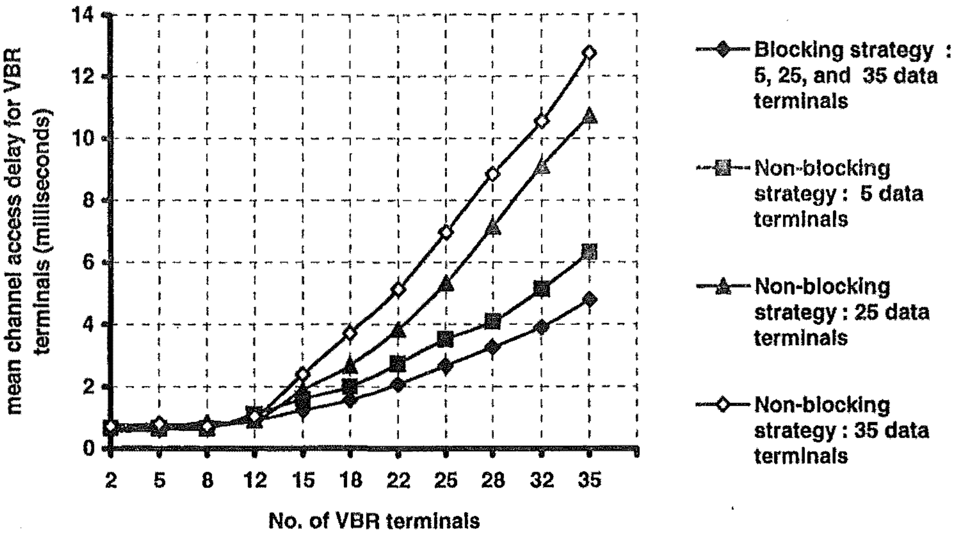

Figure 4.19 : Effect of L_{MAX} on the mean channel access delay for VBR terminals. (R=1 , L_{MAX} = 5)

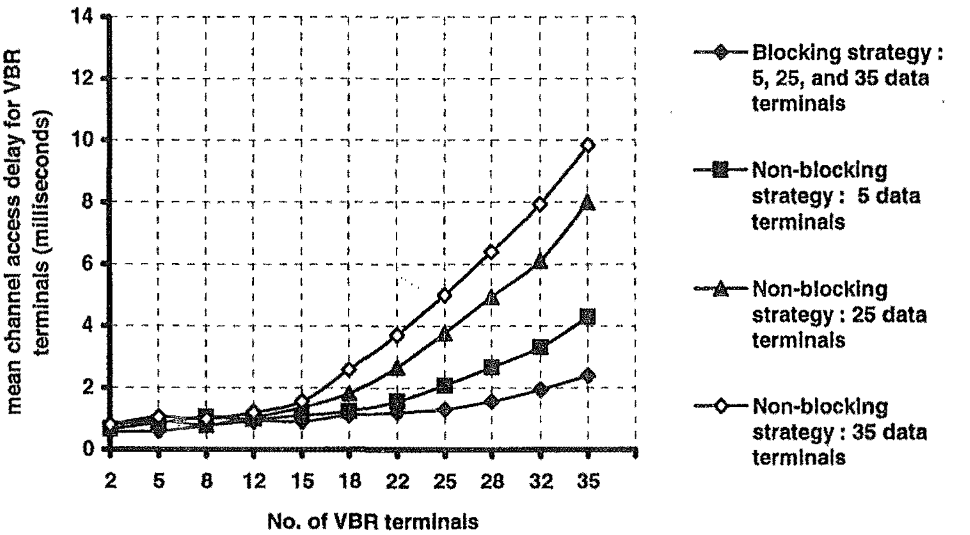

Figure 4.20 : Effect of L_{MAX} on the mean channel access delay for VBR terminals. (R=1 , L_{MAX} = 8)

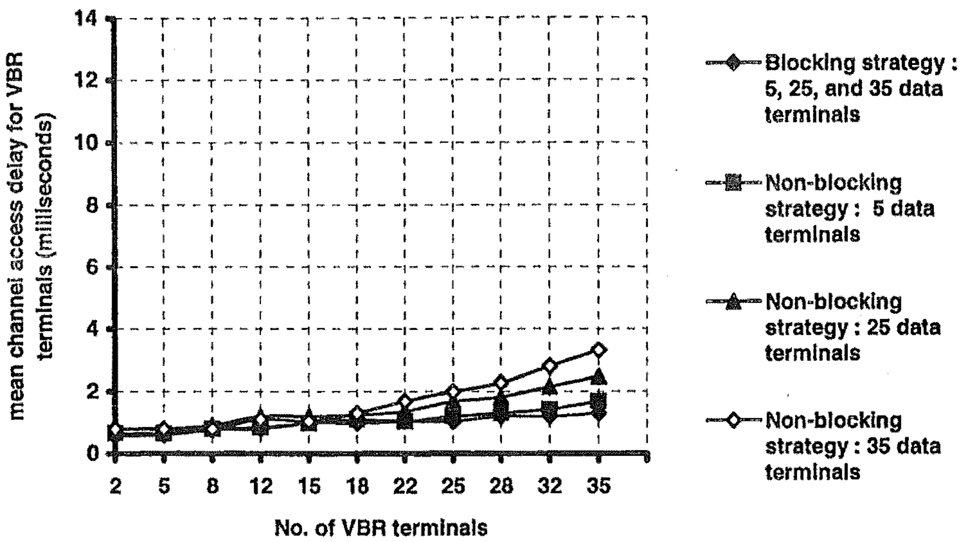

Figure 4.21 : Effect of L_{MAX} on the mean channel access delay for VBR terminals. (R=1 , L_{MAX} = 12)

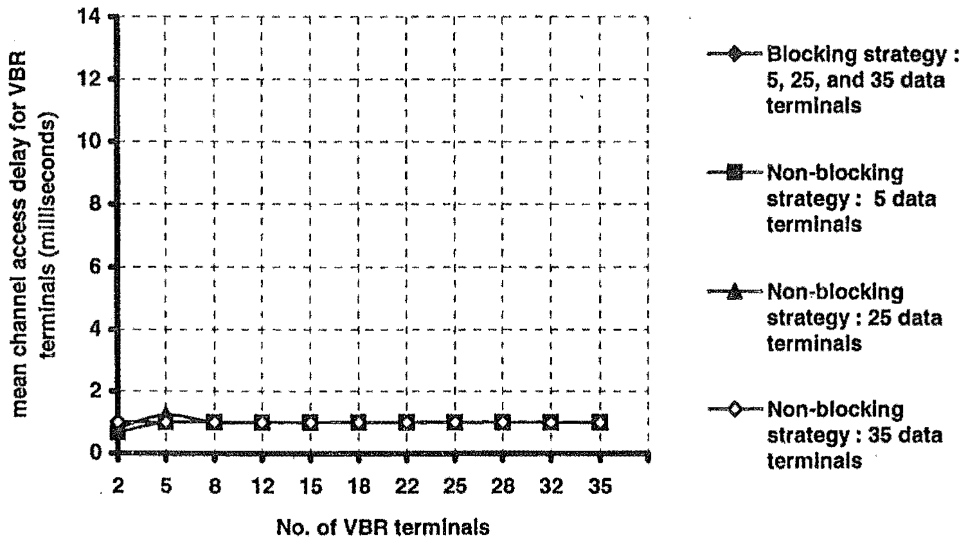

Figure 4.22 : Effect of L_{MAX} on the mean channel access delay for VBR terminals. (R=1 , L_{MAX} = 16)

4.2.6 The Issue of the Precision of the Final Simulation Results

The influence of the relative precision of the simulation results is demonstrated in Figures 4.23 and 4.24.

Figure 4.23 (a) shows the mean channel access delay for VBR terminals vs the number of VBR terminals in a cell, assuming a relative precision of 5%. The flat line on the bottom corresponds to our proposed scheme for reservation phase. Other curves have been obtained assuming different number of data terminals. Figure 4.23 (b) presents the same simulation results but with a relative precision of only 10%. As we see, it shows big variations in all curves.

Figure 4.24 (a) shows the mean length of reservation partition vs number of VBR terminals in a cell, assuming a relative precision of 5% and 35 data terminals in the cell. The flat line in the middle corresponds to our proposed scheme and shows that the mean length of the contention-based partition is not sensitive to the number of VBR terminals because VBR terminals use only idle-mode partition. Two other curves show the total length of reservation phase under our scheme and when reservation is only based on contention. It shows that our proposed scheme puts some redundancy on wireless channel. Figure 4.24 (b) corresponds the same scenario but with a relative precision of 10%. In addition to local variations in the derived curves, we observe that for number of VBR terminals less than 12, the total length of reservation phase under our proposed scheme is shorter than that of for pure contention. This shows that one has to be careful with collecting simulation results with appropriate (high) precision, since there is a danger of drawing wrong conclusions about the performance of protocols on the basis of unsatisfactory accurate estimates.

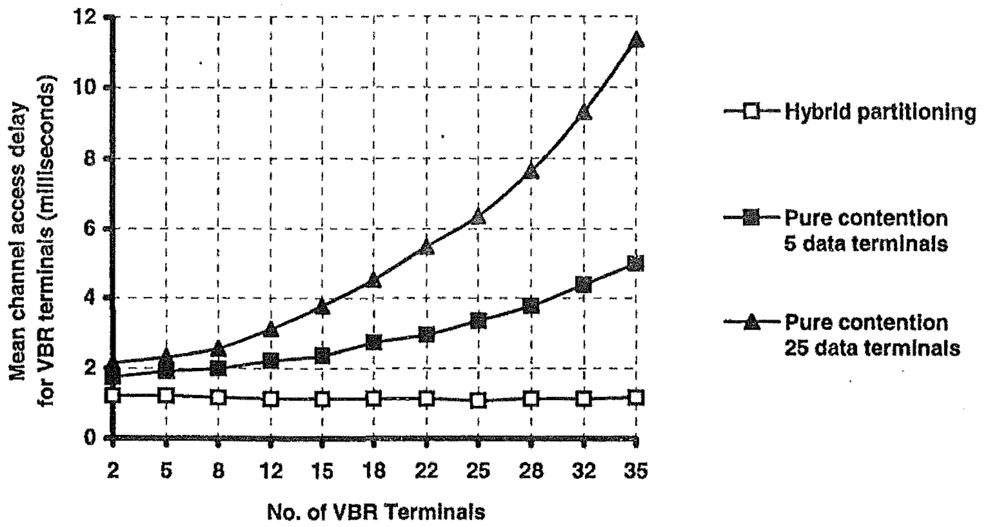

Figure 4.23 - (a) : Mean channel access delay for VBR terminals vs number of VBR terminals. (R=1, L_{MAX}=24, blocking strategy) Relative simulation precision = 5 %

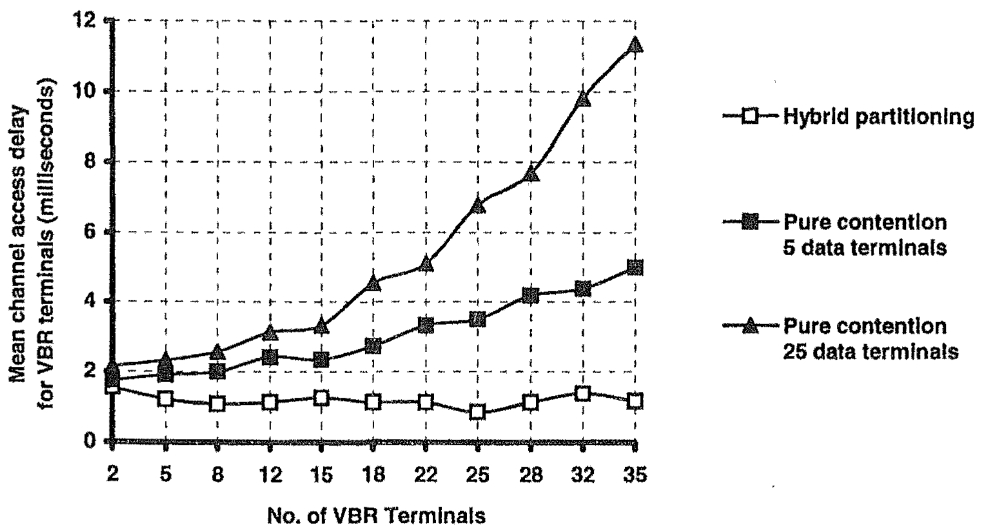

Figure 4.23 - (b) : Mean channel access delay for VBR terminals vs number of VBR terminals. (R=1, L_{MAX}=24, blocking strategy) Relative simulation precision = 10%

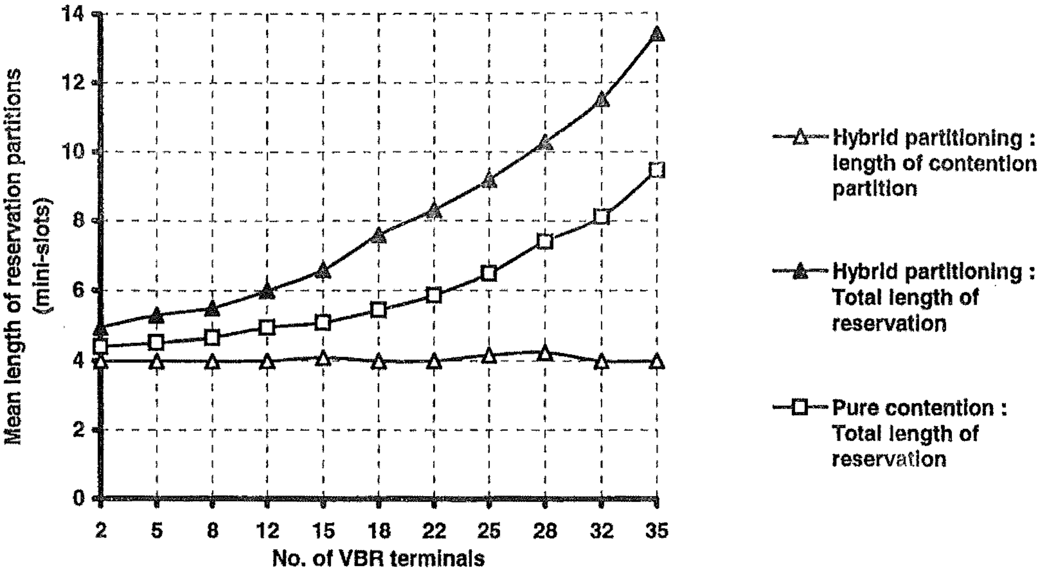

Figure 4.24 - (a) : Mean length of reservation partitions vs number of VBR terminals. (R=1, L_{MAX}=24, data terminals=10, blocking strategy) Relative simulation precision = 5%

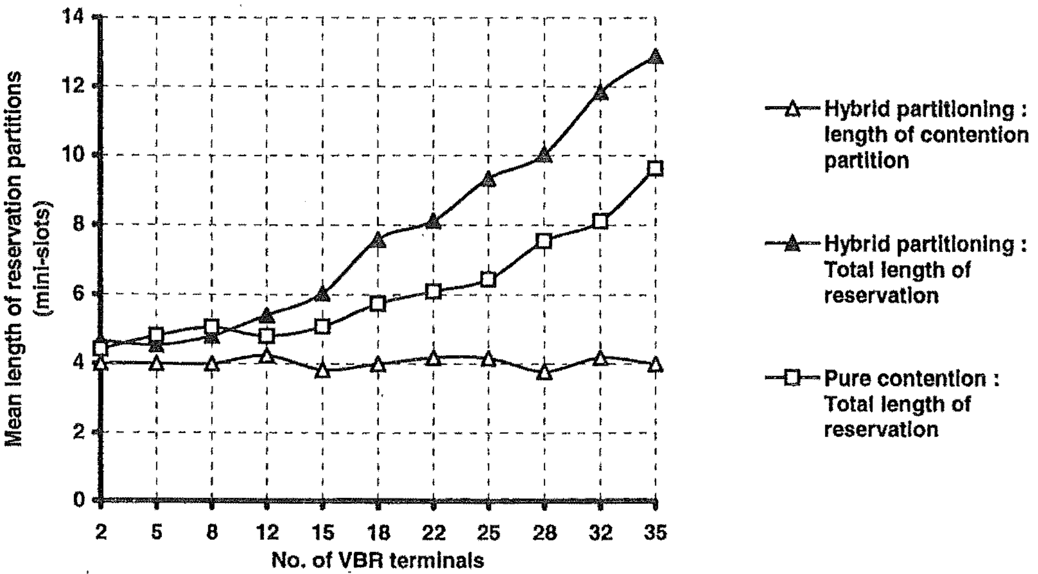

Figure 4.24 - (b) : Mean length of reservation partitions vs number of VBR terminals. (R=1, L_{MAX}=24, data terminals=10, blocking strategy) Relative simulation precision = 10%

4.3 Summary

In this chapter, the performance of our scheme for the reservation part of a TDD-based MAC protocol for WATM has been evaluated through quantitative stochastic computer simulation. To have a clear picture, the performance evaluation was divided into several cases, with each one focused on a different important performance factor.

In section 4.2.1, the effects of number of VBR terminals on the mean access delay and on the mean access delay variation experienced by VBR terminals, and the mean length of the reservation partitions under our scheme were evaluated and compared to those of pure contention-based reservation phase. The results show that, under our scheme, the mean access delay and mean access delay variation for VBR terminals are independent of the number of VBR terminals. This is an important feature of our scheme, in particular in higher loads, when there are too many mobile terminals in a cell. The mean access delay variation is important because it contributes to the final CDV of ATM cells. In addition, we showed that our scheme puts negligible overhead on wireless channel.

In Section 4.2.2, the effects of number of data terminals on the mean channel access delay and the mean channel access delay variation for both VBR and data terminal were studied. The results show that under our scheme, number of data terminals has no effect of the mean access delay and the mean channel access delay variation for VBR terminals. This is of paramount importance in mixed voice/data wireless networks. In addition, we observed that the mean channel access delay for data terminals has been slightly improved. Although this is a less important result, but one can consider it as an extra bonus offered by the proposed scheme.

In Section 4.2.3, the effects of repeat parameter (R) were studied. It was shown that a larger value for R leads to larger access delay of VBR terminals. Yet, we showed that even R = 1 does not put too much redundancy on wireless channel while minimizes the mean channel access delay and the mean channel access delay variation experienced by VBR terminals. Therefore, R=1 can be considered as optimal choice for the repeat parameter. This simplifies the implementation of our proposed mechanism because with R=1 the "***adjustable repeat***" part of our proposed scheme is not needed.

In Section 4.2.4, the effects of maximum allowable length of the idle-mode partition (L_{MAX}) were studied. It was shown that a value between 12 to 16 mini-slots for L_{MAX} is an optimal value.

In Section 4.2.5, the effects of blocking and non-blocking strategies for idle-mode partition were studied. We observed that for $L_{MAX} < 12$ mini-slots, the non-blocking strategy shows sensitivity to the number of data terminals. For $L_{MAX} > 12$ mini-slots, it was shown that both strategies present almost the same performance. In this case, blocking algorithm is preferred because of its simplicity.

Finally, in Section 4.2.6, we showed the importance of satisfactory high precision of estimates obtained by stochastic simulation.

Chapter 5

Summary and Conclusions

Following significant advances in wireless communication and rapid market growth for wireless services in recent years, the need to support more advanced multimedia services such as interactive video, and wireless access to the Internet is anticipated. Broadband wireless networks can satisfy the bandwidth requirements of these new services. A broadband wireless network can be realized using cellular or ad hoc architecture, each one its own advantages.

The primary step to deploy a broadband wireless network, is to have a good understanding of the physical layer and frequency administration issues. In chapter 1, we have briefly discussed the importance of frequency administration and regulations. Then we have briefly reviewed such principal issues as: the licensed and unlicensed bands and their characteristics, the infra red and the radio technologies, multimedia applications and their characteristics, the role of source and channel coding and their importance in wireless networks, and alternative wireless network architectures. The WATM has been considered in our research because it has been targeted by many other research projects around the world. WATM facilitates the integration of wireless services into future B-ISDN and provides an end-to-end QoS standard for future wireless applications.

This research project has been focused on the studying and improving the MAC protocols for broadband wireless networks. In accordance with this goal, in chapter 2, we have focused on MAC protocols for broadband wireless networks. First, a review of traditional MAC protocols has been presented. In a broadband wireless network designed to carry multimedia traffic, MAC layer has to deal with bandwidth consuming nature of multimedia applications and different QoS requirements, while dealing with limited bandwidth and high BER of wireless channel. Consequently, traditional MAC protocols can not be used in wireless broadband networks. A combined mechanism which incorporates advantages of several MAC protocols is likely to be more suitable. Different aspects of a typical MAC protocol designed for wireless broadband networks have been discussed namely: frame structure, channel structure, packet scheduling, and bandwidth allocation strategies. A survey on recent proposals has been presented in which we have compared and contrasted these proposals. Finally, we have discussed a general problem that has not been addressed in most recent proposals adequately. We refer to this problem as "idle-mode VBR" problem.

In chapter 3, after a brief introduction to the "idle-mode VBR" problem, an enhanced reservation scheme for reservation phase of a TDD-based MAC protocol for WATM has been proposed. This scheme is called ***"dynamic hybrid partitioning with adjustable repeat"***. Under our scheme, the reservation phase of the MAC protocol is divided into two separate adjacent partitions called the idle-mode and the contention-based partitions, respectively. The idle-mode partition is used by currently accepted delay-sensitive idle VBR terminals, when they wake up from idle mode. The contention-based partition is used by non-delay sensitive data terminals and new arrivals. Separate adaptive algorithms control these two partitions. In addition, we discussed how a repeat parameter has been associated with our scheme to provide more flexibility. The algorithms have been presented with their detailed flow charts.

In chapter 4, we have evaluated the performance of our proposed solution to the idle-VBR problem. The ON/OFF source model has been adopted to simulate the delay-sensitive VBR and the non delay-sensitive data traffic sources. Simulation program has been written in C++ and linked to AKAROA. To present the results in a coherent fashion, the evaluation has been divided into several simulation scenarios, each one focused on one particular performance measure. The mean channel access delay, and the mean channel access delay variation have been considered as the primary performance measures. In addition, the mean length of the reservation partitions has been studied to measure the extra overhead put by our scheme on the wireless channel. Also the importance of simulation precision to obtain reliable results has been presented.

The obtained results suggest that the proposed scheme can improve the performance of VBR voice traffic in a broadband wireless network, in the presence of multimedia traffic. The mean channel access delay for VBR terminals has been significantly improved in all investigated scenarios. In addition, under our scheme, the mean channel access delay experienced by VBR terminals, is independent from the number of VBR terminals. In contrast, when reservation phase is purely based on contention, the experienced mean channel access delay increases with increasing the number of VBR terminals. Similar results have been obtained for mean channel access delay variation.

The results also indicate that even assuming a repeat parameter R=1, and a maximum length of idle-mode partition L_{MAX}=24 mini-slots, our scheme does not add much redundancy to wireless channel. Therefore, the "***adjustable repeat***" part of our proposed scheme is not needed. This simplifies our proposed scheme as ***"dynamic hybrid partitioning"***. In addition, assigning a value more than 12 to 16 mini-slots to L_{MAX}, does not provide a better performance. We have also shown that when we increase the L_{MAX} to more than 12 mini-slots, blocking and non-blocking strategies for idle-mode partitions tend to offer similar behavior. Considering that blocking strategy is in general, easier to implement, we can conclude that a value of R =1, L_{MAX} = 12, combined with blocking algorithm for idle-mode partition, can provide a good performance with minimum redundancy.

Although in this project, we integrated our proposed solution for idle-VBR problem into TDD-based MAC protocols, incorporation of our solution into FDD-based MAC protocols can be considered as another research work. Also it is worth to mention that our proposed scheme can be considered as a general one, and can be incorporated in any demand-assignment MAC protocol whose reservation phase has been based on pure contention.

Bibliography

[Acamp96] A. Acampora, "Wireless ATM: A Perspective on Issues and Prospects", *IEEE Personal Communication Magazine*, August 1996, pp. 8-18.

[Agr96] P. Agrawal et al., "SWAN: A Mobile Multimedia Wireless Network", *IEEE Personal Communication Magazine*, April 1996, pp. 18-34.

[Aik96] S. Aikawa, et al., "Forward Error Correction for Wireless ATM Systems", *Proceedings of ICC'96*, Dallas, Texas, June 1996, pp. 454-458.

[Aya96] E. Ayanoglu, K. Y. eng, and M. J. Karol, "Wireless ATM: Limits, Challenges, and Proposals", *IEEE Personal Communication Magazine*, August 1996, pp. 18-36.

[Bah97] P. Bahl et al., "Optimising Resource Utilisation in Wireless Multimedia Networks", *Proceedings of ICC'97*, Montreal, Canada, June 1997, pp.1432-37.

[Bam95] N. Bambos, S. C. Chen, and D. Mitra, "Channel Probing for Disrtibuted Access in Wireless Communication Networks", *Proceedings of GLOBCOM'95*, Singapore, November 1995, pp. 322-27.

[Bau96] F. Bauchot et al., "MASCARA, a MAC protocol for Wireless ATM", *Proceedings of ACTS Mobile Summit'96*, Granada, Spain, Nov. 1996, pp. 17-22.

[Ben97] G. Benelli et al., "A Data Link Layer protocol for Wireless ATM", *Proceedings of ICC'97*, Montreal, Canada, June 1997, pp. 1438-42.

[Blo92] C. Blondia, and O. Casals, "Performance Analysis of Statistical Multiplexing of VBR Sources", *Proceedings of INFOCOM'92*, 1992, pp. 828-38.

[Chu97] W. Chung Chan, and E. Geraniotis, "A Medium Access Protocol for Interconnecting ATM and Wireless Networks", *Proceedings of ICC'97*, Montreal, Canada, June 1997, pp. 1448-53.

[Cox95] D. C. Cox, "Wireless Personal Communications: What is it?", *IEEE Personal Communication Magazine*, April 1995, pp. 20-36.

[Eng95] K. Y. Eng et at., "BAHAMA: A Broadband Ad Hoc Wireless ATM Local Area Network", *Proceeding s of ICC'95*, 1995, pp. 1216-23.

[Ewi99] G. C. Ewing, K. pawlikowski, D. Mcnickle, "Akaroa2: Exploiting Network Computing by Distributing Stochastic Simulation", *Submitted for the Australian Computer Science Conference to be held in Auckland, January 1999.*

[Fal96] D. Falconer, "A System Architecture for Broadband Millimeter-Wave Access to an ATM LAN", *IEEE Personal Communication Magazine*, August 1996, pp. 36-42.

[Fis92] W. Fischer, K. Meier-Hellstern, "The Markov-modulated Poisson process (MMPP) cookbook", Performance Evaluation, pp.149-171.

[Gao96] C. Gao, and J. S. Meditch, "Two-Layer Video coding and Priority Statistical Multiplexing over ATM Networks", *Proceedings of ICC'96*, Dallas, Texas, June 1996, pp. 127-33.

[Ho96] S. Ho Lin et al., "Cell Loss Reduction by Cell Unit Interleaving in Wireless ATM Networks", *Proceedings of ICC'96*, Dallas, Texas, June 1996, pp. 449-53.

[Hon97] W. Honcharenko, et al., "Broadband Wireless Access", *IEEE Communication Magazine*, January 1997, pp. 20-26.

[Jab96] B. Jabbari, "Teletraffic Aspects of Evolving and Next-Generation Wireless Communication Networks", *IEEE Personal Communication Magazine*, December 1996, pp. 4-10.

[Kar95] M. J. Karol, et al., "Distributed-Queuing Request Update Multiple Access (DQRUMA) for Wireless Packet (ATM) Networks", *Proceedings of INFOCOM'95*, 1995, pp.1224-31.

[Kat96] I. Katzela, and M. Nagshineh, "Channel Assignment Schemes for Cellular Mobile Telecommunication Systems: A Comprehensive Survey", *IEEE Personal Communication Magazine*, June 1996, pp. 10-32.

[Kav95-1] Kaveh Pahlevan, Allen H. Levesque, "Wireless Information Networks", *Willey series In Telecommunications and Signal Processing, John Willey & Sons, INC.*, 1995.

[Kav95-2] Kaveh Pahlevan, T. H. Probert, and M. E. Chase, "Trends in Local Wireless Networks", *IEEE Communication Magazine*, March 1995, pp. 88-96.

[Kav97] Kaveh Pahlevan, Ali Zahedi, and Prashant Krishnamurthy, "Wideband Local Access: Wireless LAN AND Wireless ATM", *IEEE Communication Magazine*, November 1997, pp. 34-42.

[Kim96] J. G. Kim and I. Widjaja, "PRMA/DA: A New Media Access Control Protocol for Wireless ATM", *Proceedings of ICC'96*, Dallas, Texas, June 1996, pp. 240-44.

[Lag97] Xavier lagrange, "Multitier Cell Design", *IEEE Communication Magazine*, August 1997, pp. 60-66.

[Lng96] T. Ling, and N. Shroff, "Scheduling Real-Time Traffic In ATM Networks", *Proceedings of INFOCOM'96*, 1996, pp. 198-206.

[Lo96] K. Lo, and S.-M. Cheng, "Dynamic Codebook Adaptive Vector Quantization for Image Compression", *Proceedings of ICC'96*, Dallas, Texas, June 1996, pp. 137-41.

[Mer93] P. Mermelstein, A. Jalali, and H. Leib, "Integrated Services on Wireless Multiple Access Networks", *Proceedings of ICC'93*, Geneva, Switzerland, May 1993, pp. 863-67.

[Mov95] N. Movahhedinia, et al., "Slot Assignment Protocol for Indoor Wireless ATM Networks Using the Channel Characteristics and the Traffic Parameters", *Proceedings of GLOBCOM' 95*, Singapore, November 1995, pp. 327-31.

[Nag96] Mahmoud Naghshineh, and A. S. Acampora, "Design and Control of Micro-Cellular Networks with QoS provisioning for Data Traffic", *Proceedings of ICC'96*, Dallas, Texas, June 1996, pp. 464-467.

[Nag97] Mahmoud Naghshineh, and Marc Willebeek-LeMair, "End-To-End QoS provisioning In Multimedia Wireless/Mobile networks Using An Adaptive Frame Work", *IEEE Communication Magazine*, November 1997, pp. 72-82.

[Nik97] Nikos Passas, Sarantis Paskalis, Dimitra Vali, and Lazaros Merakos, "Quality-of-Service-Oriented Medium Access Control For Wireless ATM Networks", *IEEE Communication Magazine*, November 1997, pp. 42-52.

[Oli96] C. Oliveire et al., "Quality-of-Sewrvice Guarantee in High-Speed Multimedia Wireless Networks", *Proceedings of ICC'96*, Dallas, Texas, June 1996, pp. 728-34.

[Osa97] Osama Kubbar, and Mouftah, "Multiple Access Control Protocols For Wireless ATM: Problems, Definition, And Design Objectives", *IEEE Communication Magazine*, Nov. 1997, pp. 93-100.

[Peha90] J. M. Peha, F. A. Tobagi, "Evaluating Scheduling Algorithms for Traffic with Heterogeneous Performance Objectives", *Proceedings of GLOBCOM'90*, San Diego, California, pp. 21-27.

[Pri96] F. D. Priscoli, "Medium Access Control for the MEDIAN System", *Proceedings of ACTS Mobile Summit'96*, Granada, Spain, 1996, pp. 1-8.

[Ray94] Raymond Steele, "Mobile Radio Communications", *Pentech Press Publishers*, London, 1994.

[Rch93] D. Raychaudhuri, and N. Wilson, "Multimedia Transport in Next-Generation Personal Communication Networks", *Proceedings of ICC'93*, Geneva, Switzerland, May 1993, pp. 858-62.

[Rch94] D. Raychaudhuri, and N. D. Wilson, "ATM-Based Transport Architecture for Multiservice Wireless Personal Communications Networks", *IEEE Journal of Selected Areas in Communication*, September 1994.

[Rch96-1] D. Raychaudhuri et al., "WATMnet: A Prototype Wireless ATM System for Multimedia Personal Communication", *Proceedings of ICC'96*, Dallas, Texas, June 1996, pp.469-77.

[Rch96-2] D. Raychaudhuri, "Wireless ATM Networks: Architecture, System Design and Prototyping", *IEEE Personal Communication Magazine*, August 1996, pp. 42-50.

[San97] J. Sanchez et al., "A Survey of MAC protocols Proposed for Wireless ATM", *IEEE Network Magazine*, November 1997, pp. 52-62.

[Sch95] M. Schwartz, "Network Management and Control Issues in Multimedia Wireless Networks", *IEEE Personal Communication Magazine*, June 1995, pp. 8-18.

[Ser96] M. Serizawa et al., "Broadband Multi-media Communication System with Asymmetric Wireless Access Link, SDL-Net", *Proceedings of ICC'96*, Dallas, Texas, June 1996, pp. 735-39.

[Sri93] K. Sriram, "Methodologies for Bandwidth Allocation, Transmission, and Congestion Avoidance in Braodband ATM Networks,", *Computer Networks and ISDN Systems*, vol. 26, September 1993.

[Toh96] C. K. Toh, "Performance Evaluation of Crossover Switch Discovery Algorithms For Wireless ATM LANS", *Proceedings of INFOCOM '96*, March 1996, pp. 1380-87.

[Wal96] B. Walke, D. Petras, and D. Plassmann, "Wireless ATM: Air Interface and Network Protocols of the Mobile Broadband System", *IEEE Personal Communication Magazine*, August 1996, pp. 50-56.

[Wu97] X. Wu et al., "Dynamic Slot Allocation Multiple Access protocol for Wireless ATM Networks", *Proceedings of ICC'97*, Montreal, Canada, 1997, pp. 1560-66.

[Zan97] Jens Zander, "Radio Resource Management in Future Wireless Networks: Requirements and Limitations", *IEEE Communication Magazine*, August 1997, pp.30-38.