QUALITY OF SERVICE SUPPORT FOR WIRELESS DATA SYSTEMS OVER HFC CABLE TV NETWORKS

by

ANWAR A. ELFEITORI

B. Sc., University of Garyounis, Libya

M. A. Sc., University of Waterloo, Canada

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

We accept this thesis as conforming

THE UNIVERSITY OF BRITISH COLUMBIA

July 2003

© Anwar Elfeitori, 2003

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of _Electrical and Computer Engineering_

The University of British Columbia
Vancouver, Canada

Date _July 31, 03_

TO MY PARENTS, WIFE, AND CHILDREN:

WITH ALL MY LOVE AND GRATITUDE

Abstract

We propose a cost effective deployment of Third Generation (3G) Wireless systems, and Wireless LANs (WLANs) over Hybrid Fiber Coaxial (HFC) CATV networks. WLANs and 3G networks are viewed as two wireless data technologies that complement each other in different environments. The main goal is to facilitates 3G/WLAN integration over the existing CATV plant, and reduce the huge cost for building a dedicated last mile infrastructure for 3G and WLAN access networks. For example, the last mile infrastructure cost for 3G systems accounts for more than 20% of a 3G system total cost. This is a major cost considering the tens of billions of dollars to be invested in 3G networks world-wide. Our proposal reduces the last mile cost by sharing the existing CATV network, and using the standard equipment and protocols of Data-Over-Cable Systems Interface Specifications (DOCSIS). This allows rapid deployment of wireless data systems, facilitates convergence of wireless and wireline networks, and paves the way towards all IP wireless networks. Special features and enhancements to DOCSIS Medium Access Control (MAC) protocol must be implemented in order to support Quality of Service (QoS) guarantees for Wireless users' data and signalling traffic.

First, we conduct a traffic study for estimating the wireless data traffic loads in different geographical areas. We use the results of this study to ensure that CATV networks have adequate capacity for supporting wireless data users in addition to the existing services like CATV broadcasting and high-speed Internet access. The traffic study results are also used for mapping wireless network elements to an HFC CATV network.

Second, we define a network architecture for deploying 3G systems over HFC networks. We identify the required components and functionalities for 3G over HFC deployment. Rather than

defining a generic 3G/HFC architecture, we choose to define a specific 3G/HFC deployment. We select deploying 3G Universal Mobile Telecommunications System (UMTS) over DOCSIS CATV networks. Also, we define the required protocol stack for our UMTS over HFC deployment. Despite our specific UMTS over HFC deployment, our proposal can be extended to integrate other wireless data networks.

Third, we propose a network architecture for interworking WLANs and UMTS networks over DOCSIS CATV networks. The proposed architecture is independent of the WLAN/3G coupling chosen (e.g., loose coupling, or tight coupling). Different WLAN/3G interworking scenarios can be accommodated. We identify the network elements needed, and define the protocol stack for our deployment.

Fourth, we present a QoS framework for DOCSIS CATV networks. The proposed QoS framework identifies, and discusses the required QoS functions in DOCSIS Cable Modem Termination System (CMTS), and Cable Modems (CMs). These functions are proposed for deployment in the upstream, and the downstream of DOCSIS transmission plane.

Fifth, we present an implementation of DOCSIS MAC protocol that facilitates the proposed 3G/HFC deployment. Also, we enhance the DOCSIS MAC protocol with features that enable it to support QoS guarantees for multiple traffic classes. We focus on real-time traffic that requires very strict delay guarantees. The proposed MAC enhancements can be used for supporting real-time wireless traffic as well as other real-time wireline Internet traffic. Performance analysis for the proposed MAC protocol shows that QoS guarantees for real-time

traffic is assured.

Sixth, we propose a Weighted Fair Queueing (WFQ) based MAC Scheduler for DOCSIS networks. We identify scheduling at the MAC layer as crucial for QoS guaranties for both wireline and wireless traffic. Our WFQ scheduler guarantees QoS for both wireline and wireless traffic. Real-time traffic is provided with QoS guarantees without sacrificing QoS for other traffic classes with less stringent QoS requirements.

Table of Contents

List of Figures

List of Tables

Acknowledgments

Foremost, I am obediently thankful to Almighty Allah, praise and glory to be Him, for His uncountable bounties. Without Allah's guidance, this work would not have been possible.

I would like to express my sincere thanks to my thesis supervisor, Dr. Hussein Alnuweiri, for his generous support, encouragement, advice, patience and friendship. His friendship, and kindness are beyond what a student expects from his supervisor.

Thanks and appreciation to my thesis committee for their valuable time, and constructive feedback.

The financial support provided by the Ministry of Education and Scientific Research of Libya, Motorola Inc., and NSERC, is greatly appreciated.

I would like to thank the many people who have made my stay at UBC such a pleasant experience, especially my fellow students in Dr. Alnuweiri's research team, and my close friends at UBC and Richmond.

The technical help of my friends Ayman, Haitham, and Tamer is appreciated. Words can not express my gratitude towards them.

A special thank-you to my wife, Zainab, for her love, support, and patience. She stood behind me during difficult times. Special thanks for my lovely children: Sarah, Fatimah, Mohammad, and Younus. They have been denied lots of my time. Sorry about that.

No words could possibly express how much I owe my parents for their unconditional love,

constant support, and persistent encouragement throughout the years. They have instilled within me the thirst for knowledge and quest for excellence.

The love, support, and encouragement of my brothers and sisters is appreciated. Thank you so much.

Chapter 1

Introduction

1.1 Background

The recent years have witnessed a rapid development of communication systems that provide

mobile services. As the public has become more aware of the benefits of mobile services,

more advanced forms of services are being demanded. The next step after the first generation

analog cellular (e.g., AMPS) and the second generation digital cellular (e.g., GSM), is the

Third Generation (3G) systems. Examples of 3G systems include UMTS, IMT 2000, and

CDMA 2000. 3G systems are aimed at providing truly mobile, cost effective multimedia

communications with guaranteed Quality of Service (QoS).

The transition from second generation digital cellular (2G) systems to 3G systems requires

large investment of money in installing new equipment for 3G, building backhaul networks,

1

and purchasing 3G licenses. Due to the fact that many mobile operators have recently invested considerably in GSM makes it difficult to throw away all their second generation cellular (2G) equipment and start all over. Also, the high cost of air licensing fees for 3G systems make it hard for operators to move to 3G. A smooth transition is believed to occur through 2.5G digital cellular systems. Examples of 2.5G are General Packet Radio Service (GPRS) and its derivative Enhanced Data Rates for Global Evolution (EDGE). GPRS and EDGE have been adopted for deployment over GSM and IS-136. GPRS/EDGE is supposed to reuse the GSM network infrastructure and protocols with the addition of new network components, namely Packet Control Units (PCUs), and GPRS Support nodes (GSNs).

Recently, the success of Wireless LANs (WLANs), namely IEEE 802.11, made wireless Internet access available in places such as Hotels, Airports, Hospitals and residential premises. Even though WLANs Internet access does not provide true mobility, users are attracted to this option because of the low (or even zero) cost incurred. Therefore, different access technologies are expected to coexist for proving mobile wireless Internet access. Among these: GPRS, UMTS, and WLANs are dominant technologies. GPRS is currently being deployed world-wide. UMTS is replacing GPRS where higher speed and better quality is needed. WLANs provide much higher speeds in residential premises, and hot spots like Hotels and Airports at low or no cost.

The fiber rich HFC network architecture is being adopted by many CATV companies for its superiority over traditional tree-and-branch coaxial architecture. In this topology, optical

fibers (OFs) run from the Headend (HE) to Fiber Nodes (FNs) that service clusters of 500 homes [79]. Recently, considerable interest was shown in using CATV networks for delivering integrated multiple services (audio/data/image/video). This is motivated by emerging trends in telecommunication deregulation in many countries that will allow CATV companies to get into the telecommunication business. Recognizing the above, CableLabs has formed the Multimedia Cable Network System (MCNS) working group for the purpose of standardizing DOCSIS MAC protocol, and Physical Layer (PHY). DOCSIS networks are aimed to provide high-speed Internet access with QoS guarantees over HFC networks.

HFC CATV networks have several assets that make them an attractive platform for supporting wireless data systems [73]. These include: excess bandwidth, real estate for electronics equipment, pole attachment rights, rights of way, and shared maintenance and customer services staff. Many CATV companies are rushing to enter the Wireless Personal Communication Services (PCS) business. In North America, many PCS licensees also own CATV systems.

This thesis proposes deploying 3G wireless systems over HFC CATV networks. We propose a new wireless data over CATV network architecture, DOCSIS MAC enhancements, and QoS mechanisms for facilitating the proposed deployment. Rather than defining a generic 3G/HFC deployment, we choose to define a specific 3G/HFC deployment. We select deploying 3G UMTS networks over DOCSIS CATV networks as it is the most popular 3G system (see Figure 1.1). Our proposal can be extended to deploy other wireless data systems over HFC

networks. In addition, we propose a network architecture for interworking 3G with WLAN

systems over the CATV plant.
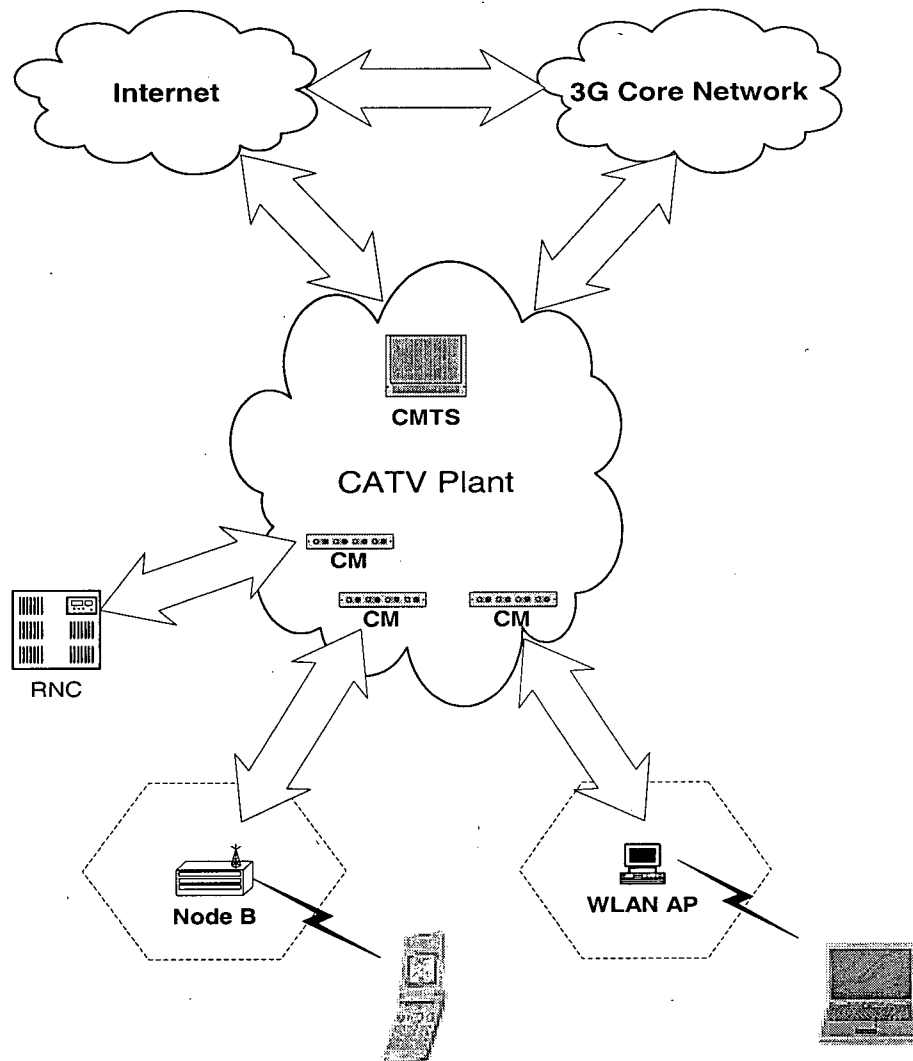
Figure 1.1 Proposed wireless data systems over CATV deployment

In summary, this research work is motivated by the following emerging issues:

4

(1) The increased attention given to wireless data systems in general, and 3G and WLAN systems in particular to support ubiquitous communication coverage at a low cost.

(2) Various assets of the HFC CATV networks that make them an attractive platform for supporting wireless data systems. These include: excess bandwidth, real estate for electronics equipment, pole attachment rights, right of way, and shared maintenance and customer services staff. Deploying wireless data systems over HFC Cable TV networks represents a cost effective implementation.

(3) The lack of a comprehensive practical deployment of wireless data over HFC CATV networks, in the literature, that addresses the related issues of network architecture, MAC, and QoS, in a cost effective manner.

(4) The emergence of the DOCSIS as a MAC, and PHY standard for HFC networks that enables them to support high-speed Internet Access. The adoption of the standard represents a way for ensuring interoperability, and to put together the efforts of many vendors to bring forth the most effective solutions.

1.2 Thesis Objectives

The main and broad objective of this research is to propose a new architecture, MAC protocols, and QoS mechanisms for the deployment of wireless data systems over HFC Cable TV networks. The proposed deployment is aimed to provide users with ubiquitous communication coverage at a reduced cost, by reusing existing CATV infrastructure. The specific

objectives of this research are:

(1) Propose a novel and effective 3G systems deployment over DOCSIS based HFC CATV networks. The proposed network architecture is aimed to support wireless data traffic with diverse Quality of Service (QOS) requirements in a scalable manner.

(2) Enhance the DOCSIS MAC Protocol by developing new mechanisms that enable supporting QoS guarantees for multimedia applications. The proposed enhancements should benefit both wireless and wireline users, and provide guaranteed QoS for large numbers of real-time flows.

(3) Propose a network architecture for interworking 3G and WLAN systems over HFC CATV networks. This network architecture should work regardless of the WLAN/3G interworking scenario chosen.

1.3 Thesis Contributions

This thesis proposes a new architecture, DOCSIS MAC enhancements, and QoS mechanisms for the deployment of wireless data systems over existing DOCSIS CATV networks. For example, the cost for building a backhaul network for interconnecting 3G cell sites accounts for more than 20% of a 3G system total cost. Considering the tens of billions of dollars to be invested in 3G networks world-wide, the backhaul network accounts for a major part of the total cost. Using the existing HFC CATV network as a backhaul network for 3G and WLAN systems provides a true cost-saving approach for building the 3G infrastructure.

The main contributions of this thesis can be summarized as follows:

(1) Proposal of a novel architecture for 3G systems deployment over HFC CATV networks. Related contributions include defining the necessary network elements and protocol stack.

(2) Proposal of a QoS framework for DOCSIS networks that facilitates QoS guarantees for wireless and wireline traffic. The framework defines the QoS mechanisms needed in DOCSIS network elements in the upstream, and the downstream traffic paths.

(3) Proposal of a new DOCSIS MAC implementation for facilitating transporting wireless data traffic over HFC CATV networks. Enhancements for DOCSIS MAC protocol are proposed in order to improve QoS guarantees for different traffic classes. Our focus is real-time traffic such as 3G wireless traffic, and wireline conversational traffic.

(4) Proposal of a novel control-plane reservation method based on a superior Weighted Fair Queueing MAC scheduler for DOCSIS networks. The proposed scheduler provides different priority traffic streams with the required fair share of the bandwidth and delay bounds. High priority traffic like 3G wireless traffic is guaranteed low delay while providing fair services for lower priority best-effort traffic. The proposed scheduler is essential for the 3G/WLAN deployment over CATV networks, as well as for the existing CATV high-speed Internet subscribers.

Other contributions of this thesis are:

(1) A traffic study for estimating the network capacity, and coverage of the wireless data systems over HFC networks deployment. This traffic study facilitates the mapping wireless network elements to HFC CATV network architecture.

(2) Proposal of a new network architecture for interworking 3G and WLAN systems over HFC CATV networks. The proposed architecture is independent of the WLAN/3G coupling chosen (e.g., loose coupling, or tight coupling), and can accommodate different WLAN/3G interworking scenarios.

(3) Development of a simulation tool for evaluating the performance of the proposed architecture. This simulation tool is developed based on CableLabs OPNET Common Simulation Framework (CSF).

1.4 Thesis Overview

This thesis is organized as follows. Chapter 2 briefly reviews the background for the material researched in this thesis, and introduces UMTS, WLAN, and HFC network architectures. The DOCSIS MAC protocol features related to this thesis are summarized. Previous work for wireless data systems deployment over CATV networks. is surveyed.

In Chapter 3, we present and discuss our proposed 3G over DOCSIS network architecture. Also, we present an architecture for WLAN and 3G systems interworking over DOCSIS networks. At the end of this chapter, we outline a QoS framework for DOCSIS networks.

Chapter 4 explains a DOCSIS MAC implementation for supporting 3G wireless traffic along

with existing wireline traffic. DOCSIS MAC enhancements that improve QoS guarantees for delay sensitive traffic are detailed. Performance results are presented and discussed.

The WFQ based MAC scheduler for DOCSIS networks is detailed in Chapter 5. Supporting QoS mechanisms are presented. The performance of the proposed scheduler is evaluated and performance results are compared with that for a priority based scheduler.

Finally, Chapter 6 gives the conclusions and summarizes the contributions of this thesis. Possible directions for future research are also outlined in this chapter.

Chapter 2

Background and Previous Related Work

2.1 Introduction

This chapter gives background information for material related to this thesis. It presents an overview of related subjects, and refers the reader to further readings for detailed information. Background on the two wireless data systems (UMTS, and WLANs), which are the focus of this thesis, is given in Section 2.2, and Section 2.3, respectively. DOCSIS CATV networks over which we propose to deploy wireless data systems are discussed in Section 2.4. Finally, previous work on wireless data systems over CATV networks is summarized in Section 2.5. Readers familiar with UMTS, DOCSIS, WLANs, and previous proposed wireless data systems deployments over CATV may safely skip this chapter and proceed to Chapter 3.

2.2 UMTS Concepts

Wireless Mobile communications systems went through several stages in their evolution towards Personal Communication Services (PCS). These are the pre-prevailing stage, the first generation analog cellular system, the second generation digital cellular system, and the 3G system [67]. During the pre-prevailing stage, the application of radio services was very limited because of the limited capacity and the bulky and expensive radio equipment. The demand for more communication services urged the need for more channel capacity. The concept of cellular mobile networks was introduced to solve this problem through allowing frequency reuse in different coverage areas.

In a cellular mobile network, the service area is divided into smaller areas (cells) [89]. Each cell is serviced by a Base Station (BS) that transmits at a low power using an antenna located at a relatively low location. The cellular concept was adopted in the first generation analog cellular system. As the demand for more channel capacity was increased, the second generation digital cellular system was introduced. The main features of this generation is the adoption of the digital technology for achieving higher capacity, better quality, more service features, and lower service cost than that for the first generation analog cellular system.

The 3G system is some times referred to as PCS. Throughout this document, we use the terms 3G, and PCS interchangeably. PCS is aimed to provide users with the ability of timely exchanging various kinds of information (voice/data/image/video) anywhere, with anyone, at any time, from any device, at a low cost, using a single Personal Telecommunication Number (PTN) [70].

UMTS is a dominant standard for 3G networks that promises high data rates (144 Kbps - 2 Mbps). European Telecommunications Standards Institute (ETSI) was responsible for UMTS standardization process in Europe. The Third Generation Partnership Project (3GPP) was formed in 1998 in order to continue the technical standardization process for UMTS. In this section, we briefly give background information related to UMTS. For detailed information, please refer to [1], and [85].

2.2.1 UMTS Network Architecture

Figure 2.1 shows a simplified UMTS/GPRS network architecture. The UMTS transmission and control plane protocol stack is shown in Figure 2.2, and Figure 2.3, respectively. The UMTS reuses GPRS Core Network elements but has a different access network and air interface. However, GPRS Core Network elements are modified to implement new UMTS capabilities specified by 3GPP.

A UMTS network is comprised of three main interacting domains [94]. These are the User Equipment (UE), the UMTS Terrestrial Radio Access Network (UTRAN), and the Core Network (CN). The architecture and functionality of these domains is summarized in the following.

2.2.1.1 The User Equipment

A UMTS UE is a user's terminal that communicates with a Node B at a cell site using UMTS air interface. It has many types of identities that it uses during its operations. These include the International Mobile Subscriber Identity (IMSI), the Temporary Mobile Subscriber Identity
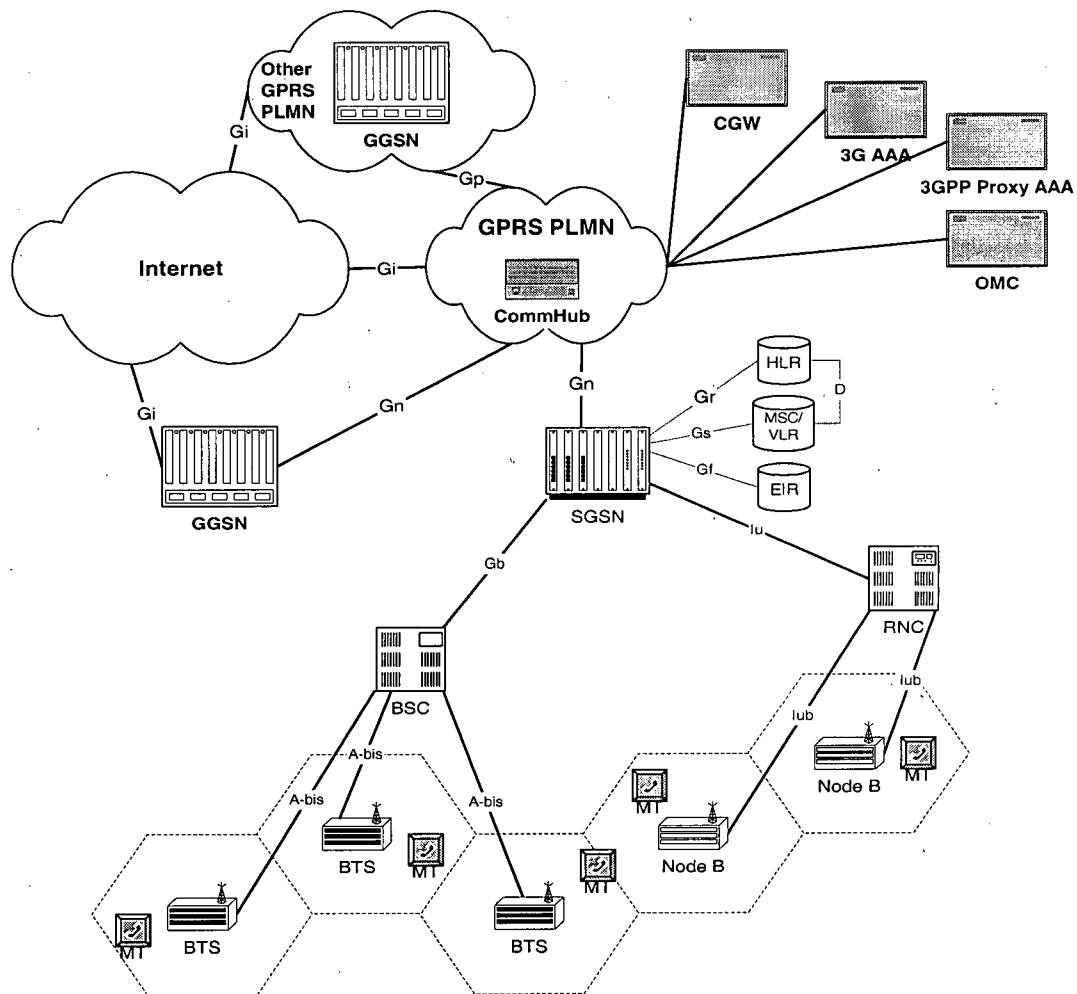
Figure 2.1 GPRS/UMTS 3GPP network architecture

(TMSI), and the Temporary Logical Link Identity (TLLI).

2.2.1.2 The UTRAN

The UTRAN provides UE access to a UMTS network using UMTS air interface. Figure 2.4 shows a simplifies UTRAN network architecture [5]. A UTRAN consists of one or more Radio Network Subsystems (RNSs). An RNS contains a Radio Network Controller (RNC) interconnected to a number of Node Bs using Iu-b interfaces. A Node B is equivalent to a

MS

Application
E.g., IP, PPP
PDCP
RLC
MAC
L1

RNS

Relay	
PDCP	GTP-U
RLC	UDP/IP
MAC	L2
L1	L1

3G SGSN

Relay	
GTP-U	GTP-U
UDP/IP	UDP/IP
L2	L2
L1	L1

3G GGSN

E.g., IP, PPP
GTP-U
UDP/IP
L2
L1

└────Uu────┘   └────lu-Ps────┘   └────Gn────┘
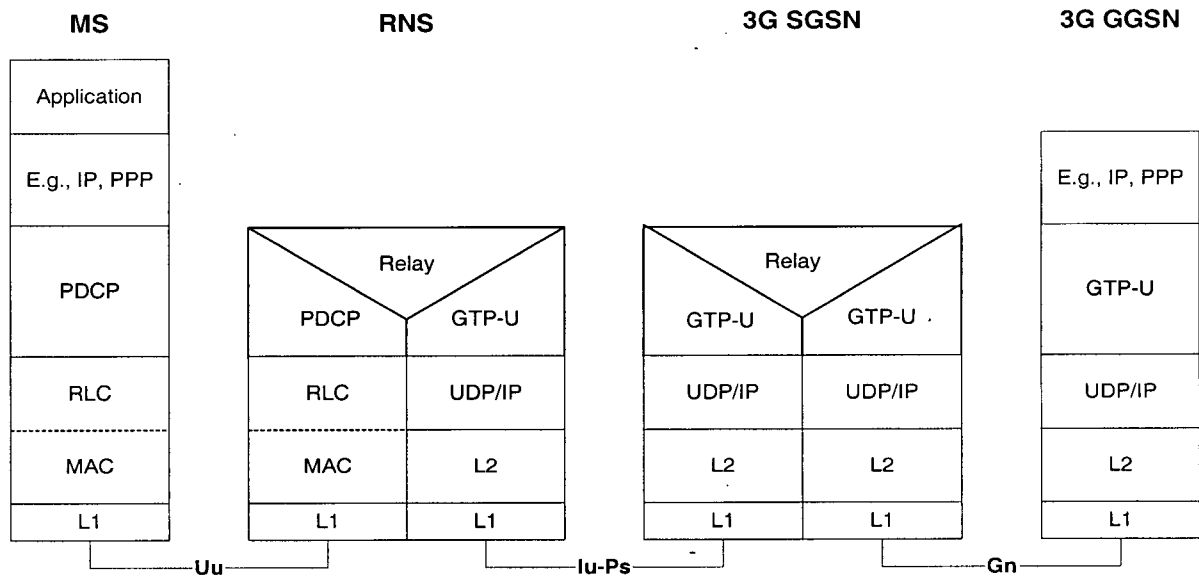
Figure 2.2 UMTS Transmission Plane

MS

GMM/SM/ SMS
RRC
RLC
MAC
L1

RNS

Relay	
RRC	RANAP
RLC	SCCP
MAC	Signaling Bearer
L1	L2
	L1

3G-SGSN

GMM/SM/ SMS
RANAP
SCCP
Signaling Bearer
L2
L1

└────Uu────┘   └────lu-Ps────┘
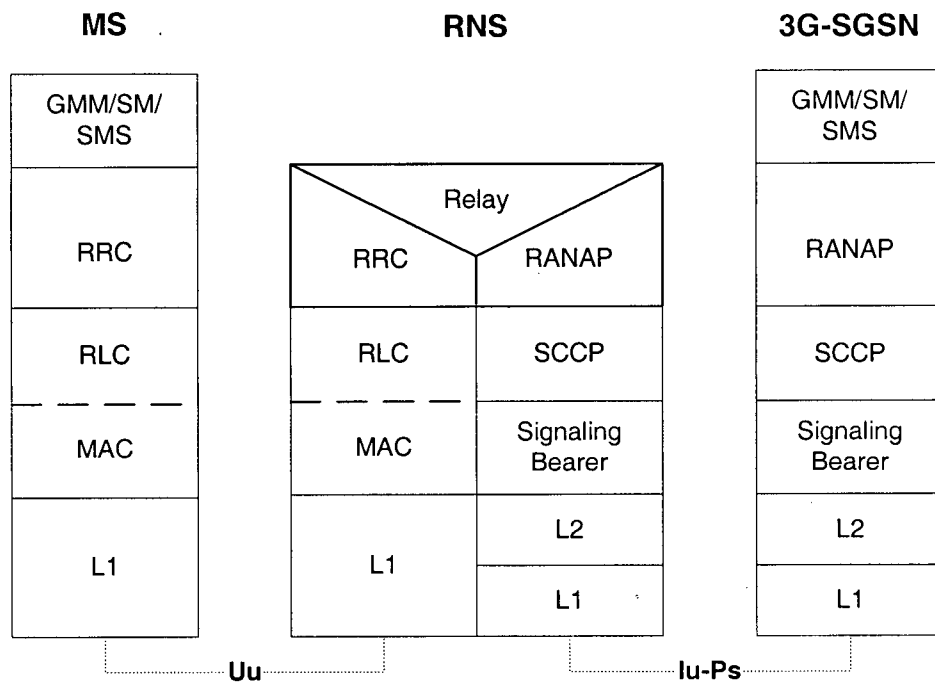
Figure 2.3 UMTS Control Plane

14

Base Station in other cellular networks. It communicates with UEs within one or more cell sites using UMTS air interface.

Within a UTRAN, RNCs of different Radio Network Subsystems can be interconnected together through the Iu-r interface. Iu-b, and Iu-r interfaces are logical interfaces that can be transported over either direct physical connection or virtual networks using any suitable transport network. For more information about Iu-b and Iu-r interfaces, refer to [6], and [7]. A simplified protocol stack for these interfaces is presented in Appendix D. In this thesis, we propose transporting Iu-b, and Iu-r interfaces traffic over HFC CATV networks, as detailed in Chapter 3.
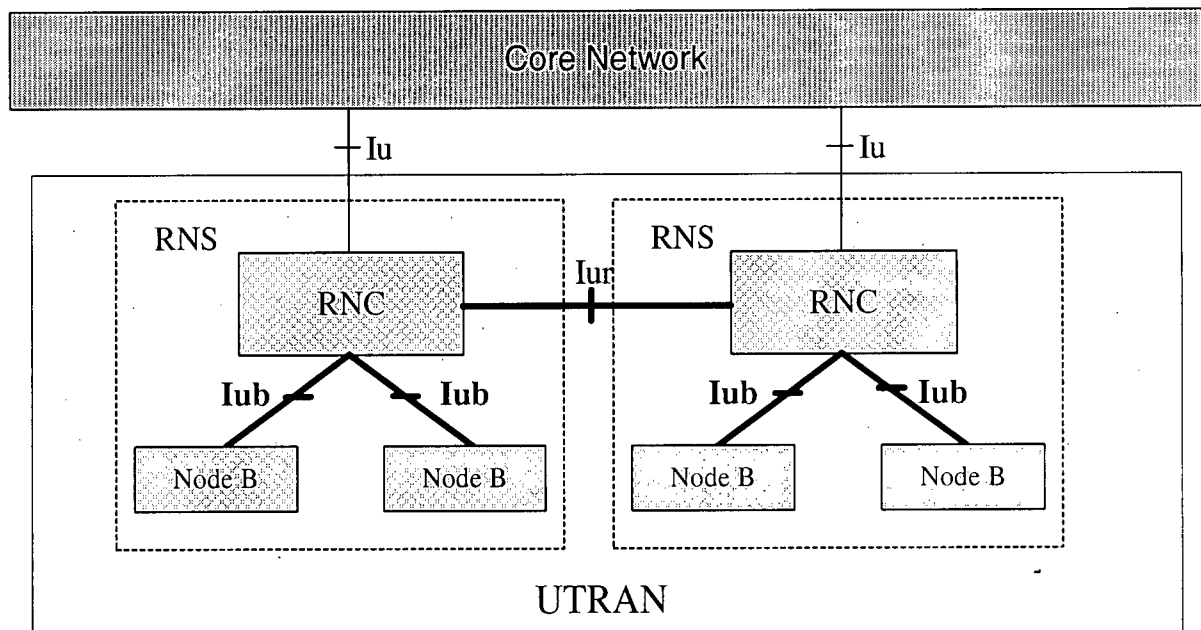
Figure 2.4 UTRAN network architecture

2.2.1.3 The Core Network

The Core Network deploys two main network elements for serving both UMTS and GPRS users. These are the Serving GPRS Support Node (SGSN), and the Gateway GPRS Support Node (GGSN). The main aim of the Core Network is to provide routing, mobility support, QoS support, charging, and security for UMTS users. It provides access to the Internet, as well as GSM Circuit Switched (CS) domain. It either deploys or provides access to databases like the Home Location Register (HLR), and Visitor Location Register (VLR) for handling user mobility. The main functional elements of the Core Network are summarized in the following sections.

2.2.1.4 SGSN

The SGSN is a key network element in the Core Network. It communicate with various network elements using 3GPP defined interfaces. Specifically, the SGSN uses the Gn to interface to local SGSNs, and GGSNs, the Gp interface to interface to GGSNs in other networks, the Iu-PS to interface to the RNC, the Gb to interface to GPRS Packet Data Units (PCUs), the Gr to interface to the HLR, the Gs to interface to the MSC/VLR, and the Ga to interface to the Charging Gateway Functionality (CGF). Also, the SGSN handles Short Messaging Services (SMS) through the Gd interface.

The main responsibilities of the SGSN are routing, mobility management, Packet Data Protocol (PDP) activations/deactivations/modification, and SMS delivery to/from the CS domain. Users in an area attach to the SGSN by exchanging various control messages. Connections can be established by having the user of the SGSN activate a PDP context which

is used for routing and mobility purposes. PDP context activation also makes the user known in the corresponding GGSN, and allows interworking with the data network. The QoS offered to the PDP context is agreed upon with the user through negotiations. Users with activated PDP context can exchange information with other users through the SGSN. The SGSN keeps a record of PDP contexts which it can update or exchange with other SGSNs in case of user mobility.

2.2.1.5 GGSN

The GGSN is the downstream ingress boundary for the UMTS network. It interfaces to various network elements using 3GPP defined interfaces. The GGSN routes information between the SGSN, and the packet data networks using the PDP context information. It stores routing information for PS-attached users, which it can use to page and deliver downstream packets to user through an SGSN. Various QoS functions are performed by the GGSN. These include admission control, shaping, policing, and scheduling.

The GGSN, as well as the SGSN generate periodic Charging Data Records (CDRs) for established PDP context and transfer the CDRs to the CGF. The CGF processes these CDRs for billing purposes.

2.2.1.6 Operation and Maintenance Center

An Operation and Management Center (OMC) is deployed at the Core Network in order to perform Network Management function for the Core Network, and the UTRAN. In this section, we briefly address network management architecture for the UTRAN, as it is relevant

to our proposed UMTS/HFC network architecture in Chapter 3.

Figure 2.5 shows a typical UTRAN Network Management architecture [9]. A remotely located OMC handles configuration management, fault management, and performance management in the UTRAN. The OMC incorporates RNC, and Node B Element Managers. Different management messages need to be transported between these OMC Element Managers on one side, and RNCs and Node Bs on the other side. In order to carry out these tasks, the Itf-R interface is deployed between RNCs and their OMC Element Manager, while the Itf-B interface is deployed between Node Bs and their Element Manager [8]. Both of these interfaces are logical interfaces. The Itf-R interface is a direct connection, while the Itf-B interface could be either a direct connection or via the RNC.

2.2.2 UMTS QoS Support

UMTS networks deploy various mechanisms for supporting QoS for different types of traffic. These QoS mechanisms are deployed in the transmission, control, and management planes of the access, and the Core networks. For a summary of UMTS QoS support, refer to Appendix A.

2.3 Wireless LAN - Concepts

Wireless LANs (WLANs) are providing access to data networks in places such as hotels, hospital, airports, and residential premises. The use of WLANs offers a number of benefits . that include mobility, flexibility, and fast access, which in turn improve productivity and enhance competitiveness. Recognizing these benefits an increased number of individuals, and
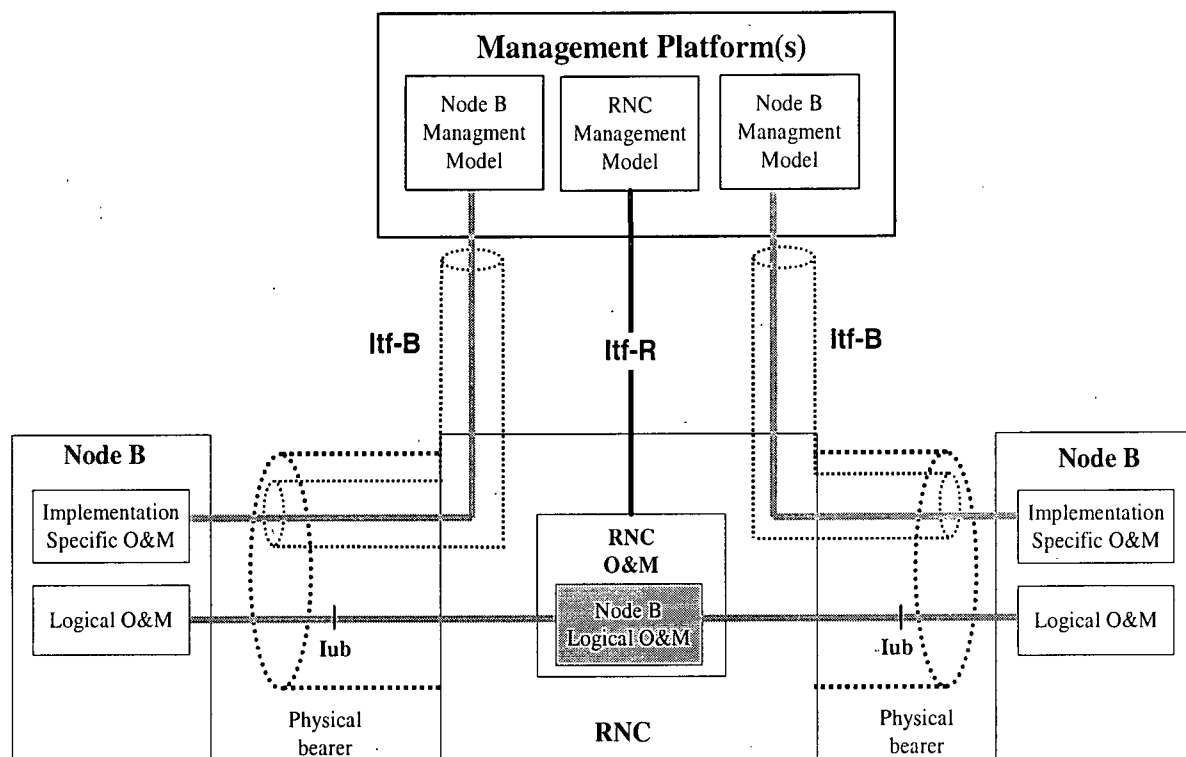
Figure 2.5 UTRAN OMC architecture

organizations started deploying WLANs.

There are two main WLAN standards: the IEEE 802.11, and the HIPERLAN/2. HIPERLAN/2, which stands for High Performance Radio Local Area Network, is developed by the Broadband Radio Access Networks (BRAN) division of the European Telecommunications Standards Institute (ETSI). In our discussion of WLANs throughout this thesis, we focus on the IEEE 802.11 based WLANs, as they are the dominant WLANs in the market as of today and likely to be in the future.

The IEEE 802.11 standard identifies several main components for a WLAN architecture.

These are Station (STA), Access Point (AP), Basic Service Set (BSS), Independent Basic Service Set (IBSS), Distribution System (DS), Extended Service Set (ESS), and Portal [13]. The standard defines a general framework instead of imposing a specific physical network architecture.

The wireless STA is the user terminal that contains an embedded device, or an adapter card to provide wireless connectivity. An IBSS can be formed in an ad-hoc basis by establishing communication between two or more STAs in a BSS, without an AP and a DS. Alternatively, a WLAN can contain several BSSs.

Figure 2.6 shows a typical WLAN architecture. A BSS, which is similar to a cell in mobile nomenclature, is the basic building block for a WLAN network. It is controlled by an AP which interconnects one or more BSSs to the DS. The DS could be implemented by a wired LAN like an Ethernet or other means. An ESS can be formed by interconnecting several BSSs to a DS using APs. The whole interconnected WLAN in an ESS is seen as a single 802 network to the upper layers of the OSI model. A WLAN interconnects to other wired LANs through a device called a portal. The standard does not preclude that an AP be combined with a portal in the same physical entity. The protocol stack for IEEE 802.11 is shown in Figure 2.7. For detailed information about this protocol stack or protocols used at the MAC and PHY layers, please refer to [13].

2.4 DOCSIS CATV Networks - Background

There has been a lot of interest in using HFC CATV networks for providing residential
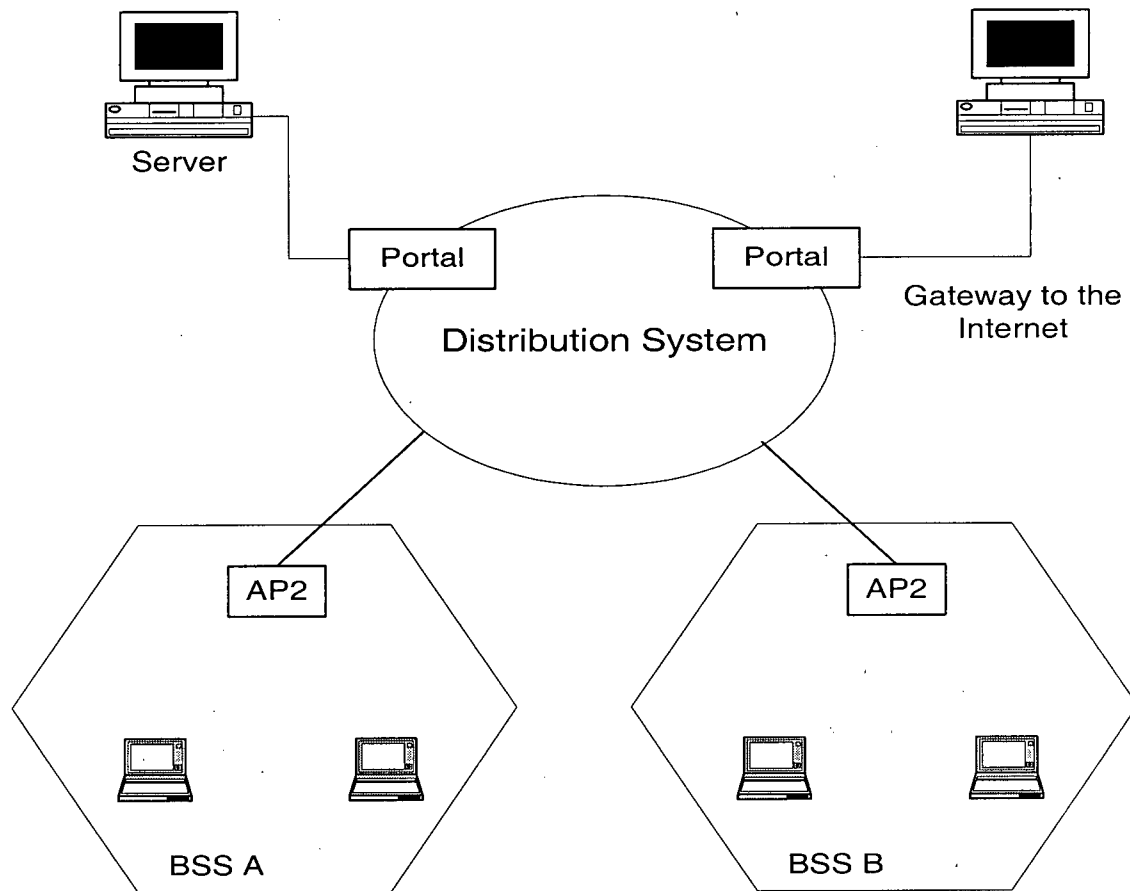
Figure 2.6 Wireless LAN architecture

Internet access. Two competing standardization efforts were initiated by the IEEE, and

CableLabs in 1994, and 1995, respectively. These efforts resulted in two Data over CATV

standards: the IEEE 802.14 [56], and CableLab's DOCSIS [25]. By 1998, it was clear that

DOCSIS was the winner and standardization efforts in IEEE 802.14 winded down. Today,

DOCSIS is the standard for supporting multimedia applications over HFC networks. Our

proposed wireless over CATV is based on DOCSIS. In this section, we review the DOCSIS

HFC CATV network architecture, and DOCSIS MAC protocol.

WLAN STA   **WLAN AP**

Application
TCP
IP
802.2 LLC
802.11 Link Security
802.11 MAC
802.11 PHY

IP
802.2 LLC
802.11 Link Security
802.11 MAC
802.11 PHY
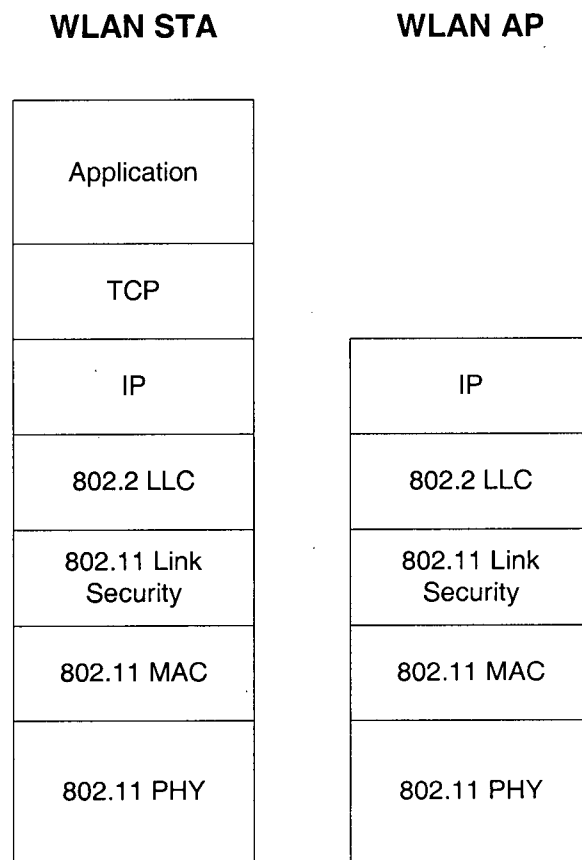
Figure 2.7 Wireless LAN Transmission Plane

2.4.1 HFC CATV Network Architecture

CATV was originally designed for the purpose of delivering broadcast signals in areas where

off-air broadcast signals were not available or received with poor quality [33]. Since CATV

was a specialized system for broadcasting numerous television channels in a sealed spectrum

rather than a general purpose communication system, the topology of the network was

customized for maximum efficiency. The CATV topology that has evolved over the years is a

tree-and-branch architecture. This architecture consists of five major parts: the Headend, the

trunk cable, the distribution cable, the drop cable, and the in-home wiring and terminal equipment.

The Headend is the regional origination point for broadcast signals in the CATV system. It consists of five major components: a satellite antenna for receiving satellite-delivered program signals, high-gain directional antennas for receiving distant TV broadcast signals, directional antennas for receiving local signals, machines for playback of taped programs, and studios for local origination and community access programming [33].

Prior to 1987, coaxial cables were used in the trunk, distribution and drop portions of the CATV distribution plant. The HFC network architecture was introduced in order to increase the reliability, improve the quality of service, and increase the channel capacity of the tree-and-branch coaxial topology. In this topology, large segments of the coaxial cables along with the attached broadband amplifiers were replaced with OFs. Each OF runs from the Headend to a FN that services a cluster of 500 homes [79]. This has the effect of bringing more capacity to the neighborhood of small groups of customers. As the need for more capacity arises, the area served by a FN can be partitioned into smaller areas that are served by dedicated OFs.

2.4.2 DOCSIS Network Architecture

DOCSIS utilizes the bandwidth-rich HFC CATV plant. A typical DOCSIS HFC CATV network is shown in Figure 2.8. The DOCSIS Cable Modem Termination System (CMTS) deployed at CATV Headend is in control of downstream and upstream transmissions to/from Cable Modems (CMs) deployed at customer premises. DOCSIS MAC and PHY protocols are

used for CMTS - CM intercommunication. The CMTS interfaces to other CATV networks, the Internet, and the Public Switched Telephone Network (PSTN) using CMTS Network Side Interface (CMTS-NSI).
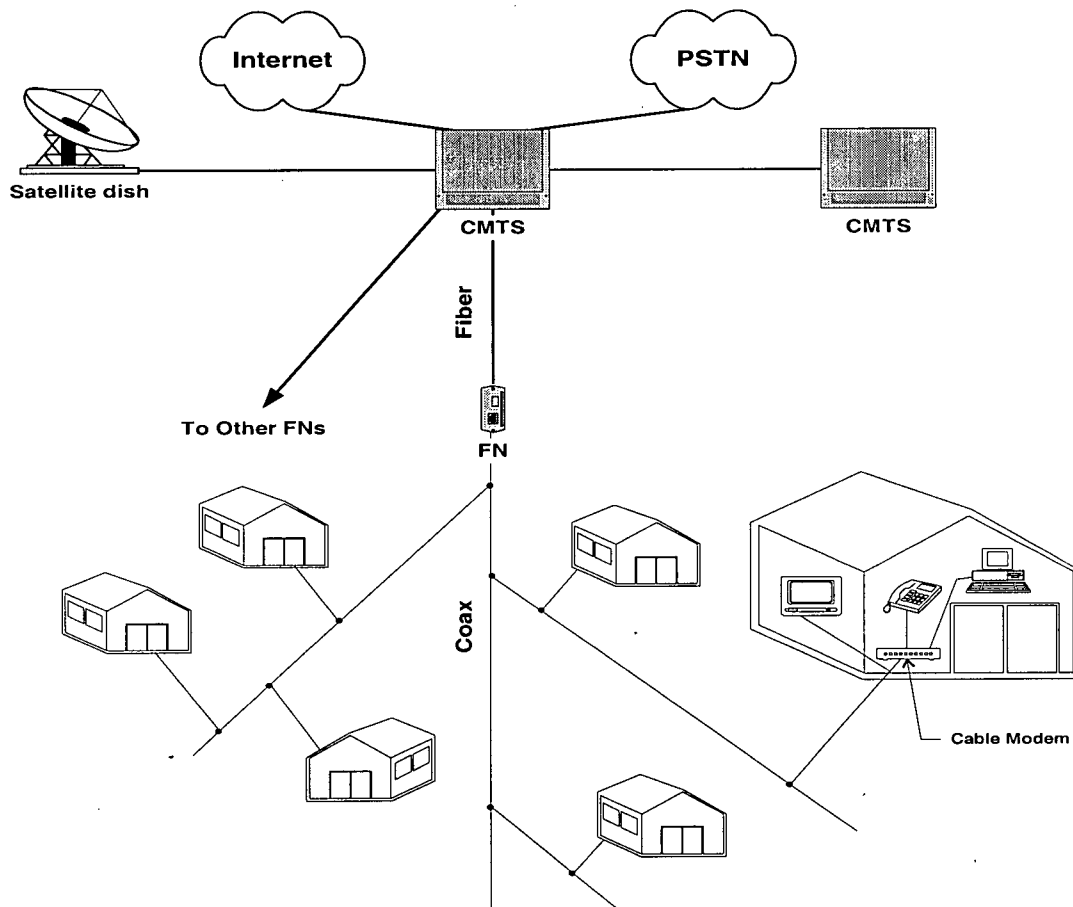
Figure 2.8 DOCSIS network architecture

Figure 2.9 shows DOCSIS protocol stack. DOCSIS MAC protocol is explained in Section 2.4.3. For information on other protocols in the DOCSIS protocol stack, refer to [27].

Cable Modem

IP

802.2/DIX LLC

Link Security

Cable MAC

DS TC layer	Upstream Cable PMD
Cable PMS	

Cable modem termination system

IP

802.2/DIX LLC

Link Security

Cable MAC

DS TC layer	Upstream Cable PMD
Cable PMS	

└─Cable Network Transmission─┘
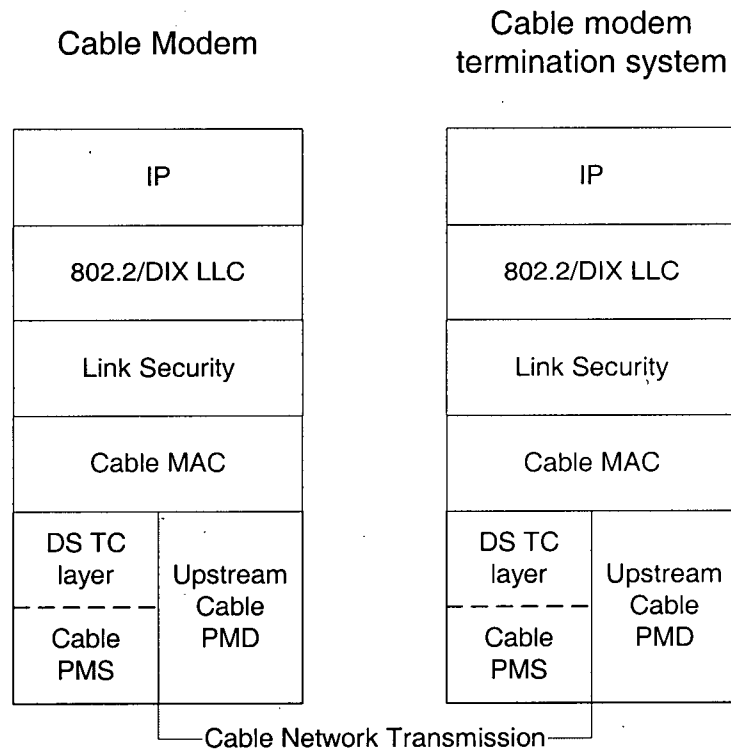
Figure 2.9 DOCSIS protocol stack

2.4.3 DOCSIS MAC Protocol - Overview

The logical network topology upon which the standard is based is that of a tree-branching bus. The DOCSIS standard specifies the MAC, and PHY protocol that provide two way data communications services over the HFC networks.

The DOCSIS MAC protocol is aimed at describing the behavior of subscriber's stations (CMs), and protocol flows across the external interface between the subscriber station and the CMTS located at the Headend. DOCSIS standard defines a framework for MAC protocols, however, specific implementation details are left to be defined by vendors. This ensures interoperability while offering flexibility in implementation.

25

DOCSIS MAC is a centralized reservation based mechanism, where the protocol resides in both DOCSIS CMs and CMTS. The available spectrum on the CATV plant is partitioned into upstream and downstream using Frequency Division Multiplexing (FDM). The Upstream path is from CMs to a CMTS, while downstream is in the opposite direction. The upstream and downstream are divided into channels, where a CM or a group of CMs share a single channel.

In the upstream, the MAC provides for partitioning a given upstream channel into multiple contention and reservation mini-slots, each with an implicit number that goes from 0 to a CMTS defined maximum. A mini-slot is a power-of-two multiple of 6.25μs increments: 1, 2, 4, 8, 16, 32, 64, or 128 times 6.25μs. Contention mini-slots are used to transmit request mini-PDUs (mini protocol data units), or short data packets. Reservation mini-slots are used to transmit users' data and various MAC messages (see Figure 2.10). Integral number of mini-slots are used to transmit either variable length packets or ATM cells (note that the support for ATM is a future issue).

An active CM joins the systems by undergoing a synchronization and a ranging process, where the CM is assigned a Local Identifier, and the local time for the station is adjusted according to its distance from the CMTS. The later ensures that when CMs, located at different distances from the CMTS, transmit at the same local time, their transmissions arrive at the CMTS at the same CMTS time.

The CMTS periodically broadcasts downstream bandwidth allocation MAPs that inform CMs
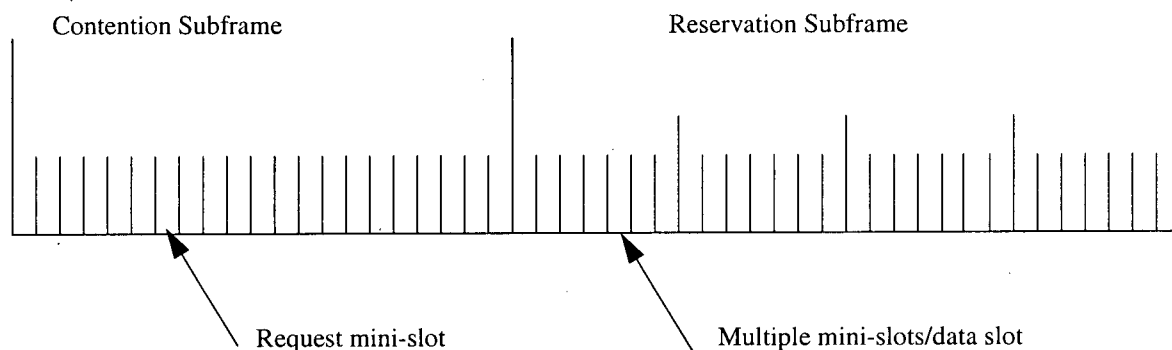
Figure 2.10 DOCSIS upstream frame structure

of the allocation of upstream mini-slots (contention, data, etc.), in a given time frame. A scheduler located at the CMTS takes care of mini-slots assignment. The DOCSIS standard does not mandate any specific scheduler, rather it is implementation dependent.

CMs that have some traffic for transmission, send their request mini-PDUs in the contention mini-slots. The CMTS uses MAP messages to indicate to various CMs the status of transmissions in the previous contention mini-slots (e.g. successful, collision). If this request is in collision, then the station would apply the Binary Exponential Backoff Contention Resolution algorithm until the collisions are resolved. Based on the successfully received requests in the upstream, the CMTS allocates reservation mini-slots for various stations, and informs them via broadcasting MAP messages in the downstream. CMs can send their data packets in the allocated upstream mini-slots. Depending on transmission policy, CMs could piggyback requests for extra mini-slots on their transmitted data packets using the Extended Header field

27

(EHDR).

DOCSIS 1.1 standard [26] specifies Quadrature Phase Shift Keying (QPSK), and 16 Quadrature Amplitude Modulation (16QAM) modulation schemes for the upstream. In the downstream, 64QAM, and 256QAM modulation scheme are specified. Refer to [26] for details about DOCSIS Physical layer.

2.4.4 QoS Support over DOCSIS

DOCSIS 1.1 [26] specifications introduces a number of new Quality of Service (QoS) related features not present in [25]. These include:

- Packet Classification & Flow Identification

- Service Flow QoS Scheduling

- Dynamic Service Establishment

- Fragmentation

- Two-Phase Activation Model

More information on any of the above QoS features can be found in [27]. Our work focuses on Service Flow QoS Scheduling and MAC enhancements for real-time traffic support, in the upstream. Therefore, we present some background on these issues in Appendix B.

2.5 Previous Wireless Data Systems over HFC Networks Proposals

A number of researchers have proposed network architectures for supporting wireless data networks over HFC CATV networks. Most of these proposals address PCS deployment over HFC networks rather than focusing on a specific PCS/3G network. Also, most proposals do not address the fact that DOCSIS MAC is in sole control of data communications over CATV networks. Statistical multiplexing of wireless traffic, and DOCSIS high-speed Internet access traffic was not a concern in previous publications. We will summarize these proposals in the following.

Lipoff [73] proposed a centralized PCS/CATV network architecture, where the Base Stations (BSs) are collocated with the PCS control and switch at the Headend. A Remote antenna interfaced to the cable by a Remote Antenna Driver (RAD) is located at the center of each microcell. RADs relay bidirectional RF signals between the Mobile Terminals and the centrally located BSs via the cable plant. Additional BSs can be added at the Headend in order to provide for the system growth. The above proposal addressed the economical advantages of deploying PCS approach over CATV networks rather than dealing with the details of functional network architecture along with the performance analysis.

Beasley [16] proposed providing wireless access to the CATV plant using discrete-element distributed antennas. Each of these antennas is interfaced to the cable by a RAD. RADs are located inside residence, multistory parking lots, office buildings, or mounted at utility poles. Also, the proposal describes microcell extenders that extend the coverage area of a RAD or a

BS in areas where there is no cable plant. These extenders connect to the BSs or RADs using dedicated coax or fibers. Users can roam within the coverage area of any RAD without the need for handoff. RADs are controlled by a Remote Antenna Signal Processor (RASP) which is located at a central site. Also, a RASP interfaces the BSs to the cable plant. BSs interface directly to twisted pair telecom lines and can be either centrally located or distributed at cell sites. The proposal states that by centrally locating the BSs, the cell site equipment is simplified to a RAD which reduces the cost for implementing PCS. However, the proposal does not specify a detailed functional network architecture (in terms of MSC, BSCs, switching, etc.)

Donaldson et al [37],[38] analyzed the simulcast effect and other effects associated with providing wireless access to the CATV plant using discrete-element distributed antennas. In this proposal, RADs along a specific cable segment are separated by 250 m. These RADs transceive on the same radio channel, using simulcast mode, and they are controlled by a RASP. A RASP interconnect to the PSTN or any other broadband network. In the case where cell partitioning is needed, each cell can be served by a RAD which can be controlled by a dedicated RASP. This simplifies the BS equipment to a RAD and allow the RASP to be centrally located at the interface point to PSTN. The researchers indicated that, communication is feasible even in the simulcast interference zone. However, this proposal was aimed in analyzing simulcast effect rather than proposing a detailed functional network architecture.

Mueller [78] proposed a centralized approach where all electronics are centrally located. BSs which are located at the FNs are simplified to just frequency converters that translate the off-

air frequencies to cable distribution frequencies. However, the proposal was generalized in the sense that neither a specific network architecture nor a performance analysis was given.

A Personal Communication Network (PCN) over CATV architecture, based on IEEE 802.6 MAN, was proposed by a number of researchers [48][68][69][75][76]. Despite the differences between these proposal, IEEE 802.6 MAN was proposed for interworking PCS network elements. For example, in [68], interconnection of various elements of the PCS to the dual bus of the 802.6 MAN is accomplished using several nodes. These include Access Nodes, Functional Nodes, and Interworking Nodes. An Access Node interfaces a BSC to the MAN. A BSC controls several Radio Transceivers (RTS), or BSs in other terminologies. Functional Nodes include the Bandwidth Manager (BWM), the Data Base Manager (DB), the Signaling Termination (ST). Interworking Nodes are used for interconnecting the MAN to PSTN or other MANs. The above PCN/MAN architecture is mapped to CATV network in such a manner that the distribution and drop cables are used for the dual bus of 802.6 MAN and BSC-RTS links, respectively. The Master Station (MS, equivalent to MSC in other terminology), and the BSCs are mapped to the Distribution Center (DC), and the Drop Points (DPs), respectively. This proposal does not define mechanisms for communications between RTS and BSCs, where other wireline applications share the coaxial cables in a multiple access environment. All IEEE 802.6 based PCN proposals do not consider that DOCSIS is the choice for CATV because they were prior to DOCSIS era.

An ATM based PCN/HFC overlay was proposed in [44], and [45]. In this proposal, a hierar-

chical switching architecture is proposed where small, medium and large size ATM switches are located at FNs, DPs, and DCs, respectively. BSs are located on coaxial drop cables. This proposal lacks definition of protocols for communications between BSs and ATM switches, in a multiple access situation.

Other architectures for the deployment of wireless systems over HFC CATV networks were proposed in [31], [54], [72], [74], and [86]. As in the other previous proposals discussed earlier in this section, the use of the CATV de facto standard DOCSIS was not considered.

A Multipurpose Fiber Access Network for interworking CATV, PCS, Wireless Local Loop (WLL), and Local Multipoint Communication Services (LMDS) was proposed in [51] and [52]. This proposal assumes the availability of Fiber Optic networks at cellular Base Station sites, which is not always the case. Extending fiber links to the large number of cell sites in 3G networks is an expensive choice.

2.6 Summary

In this chapter, we introduced background information related to this thesis. We introduced UMTS, WLAN, and DOCSIS networks. More information on these topics is provided appendices and cited references. Previous proposal for wireless data over CATV deployment were reviewed. Our review concludes that non of the previous proposals can provide a feasible complete solution that address all related issues of network architecture, MAC, and QoS.

Chapter 3

Proposed Network Architecture

3.1 Introduction

In UMTS networks, a large number of Node Bs (equivalent to Base Stations in other cellular systems) will be deployed world wide. A significant investment is needed to build an infrastructure that interconnects the deployed Node Bs to 3G RNCs (equivalent to BSCs in GSM). In this chapter, we propose a network architecture for deploying 3G networks over existing CATV plant. Our 3G/ HFC deployment does not require building a new last mile infrastructure for interconnecting 3G Access network elements. Instead, we utilize the existing CATV plant and DOCSIS equipment and protocols. This translates into a huge saving in 3G cost, rapid deployment of 3G networks, convergence of wireless and wireline networks, migration towards all IP wireless networks, and additional revenue to Cable operators. In addition, we propose interworking 3G and WLANs over CATV networks.

In Section 3.2, we present a traffic study. The proposed UMTS network architecture over the HFC CATV network is presented in Section 3.3. Section 3.4, presents the proposed WLAN/ 3G interworking architecture over CATV networks. Our QoS framework for DOCSIS network is explained in Section 3.5.

3.2 Traffic Study

In order to map a UMTS network to an HFC infrastructure, a traffic study is conducted. This study involves estimating the generated traffic in different environments, capacity calculations of a wireless system that can handle the generated traffic, and mapping the wireless system to the HFC network, as detailed below. These estimates will be used in our simulations in Chapter 4, and Chapter 5 to demonstrate the performance of our techniques under very realistic conditions.

3.2.1 Generated Traffic in Different Environments

The number of potential users per Km^2, in different environments was estimated in a previous study [35]. Table 3.1 lists these results in downtown, urban, and suburban environments.

Table 3.1: Number of potential users per Km^2 in different environments

Environment	Pedestrians	Private Cars	Public transport passengers
Downtown	10,000	800	2,000
Urban	500	300	200
Suburban	10	50	10

34

The traffic demand in Erlangs generated in different environments is shown in Table 3.2. These results are obtained based on the following assumptions:

- Each voice user makes 2 calls/hour of 2 minutes duration.

- Each data user transmits 0.1 messages per second with 10 Kbits per message.

- Both voice and data use circuit switching. Note that in CDMA neither a frequency band nor a time slot is reserved for the whole duration of a circuit switched call.

- The penetration rate is 50%.

Table 3.2: Traffic demand in Erlangs per Km^2 in different environments

Environment	Pedestrians		Private cars		Public transport		Total
	Voice	Data	Voice	Data	Voice	Data	
Downtown	300	500	24	40	60	100	1024
Urban	15	25	9	15	6	10	80
Suburban	0.3	0.5	1.5	2.5	0.3	0.5	5.6

3.2.2 Radio Cell Coverage

In this section, we find the cell coverage in different environments. We calculate the cellular capacity of a given wireless system from which the cell radii in different environments is obtained. We consider the use of Code Division Multiple Access (CDMA) in the wireless links because its is the choice for UMTS air interface. Our proposal is generally applicable regardless of the air interface standard employed.

The capacity of a CDMA cellular system can be calculated using the following equation [82].

$$C = \frac{BW}{R} \cdot \frac{1}{E_b/N_o} \cdot \frac{1}{d} \cdot F \cdot S_G \tag{3.1}$$

Where:

- C - Number of channels per radio cell

- BW - Spread spectrum bandwidth (assumed 2.5 MHz)

- R - Channel rate (assumed 8 and 9.6 Kbps for voice and data respectively

- E_b/N_o - Bit energy/noise power spectral density (assumed 7.0 dB)

- d - Voice duty cycle (assumed 40%)

- F - Frequency reuse efficiency (assumed 60%)

- S_G - Sectorization gain (assumed 2.55)

Using (3.1), and the above assumptions, a Node B can support up to 75 Voice channels and 50 data channels. Now, we use Erlang B formula to find the traffic intensity in Erlangs that can be supported by a single Node B.

$$GOS = P[blocking] = \frac{\dfrac{A^C}{C!}}{\displaystyle\sum_{k=0}^{C} \dfrac{A^k}{k!}} \tag{3.3}$$

Assuming 0.01 blocking probability, a single Node B can support up to 98 Erlangs of voice and data traffic. The calculations were made using the assumption that the soft capacity feature of CDMA is not used.

Given the offered traffic intensity per Km^2 (see Table 3.2) and the capacity of each Node B, the radio cell radii are 170, 624, and 2,360 m for downtown, urban, and suburban areas, respectively.

3.2.3 Mapping Cell Sites to HFC Fiber Nodes

In this section, we associate the radio cell sites in different environments with the HFC FNs. The average area inhabited by 500 homes (a FN service area) in downtown, urban, and suburban areas are found to be 0.334, 0.498, and 1.838 Km^2, respectively [91]. Taking into consideration the radio cell radii in different environments (see Section 3.2.2), we can find the areas of radio cells in different environments. Our calculations are based on a 50% radio cell overlap [96]. The hexagon radio cell scheme is used to calculate the cell coverage area, and the number of cells in a FN service area.

Table 3.4 shows the radio cells coverage areas in different environments and the number of radio cells within a FN service area. This table shows that in downtown areas, more Node Bs per FN service area are needed (4 Node Bs), whereas in urban and suburban areas, only 1 Node B per a number of FNs is needed. Assuming 4 coaxial legs per FN, only one Node B per coaxial distribution cable is needed, even for downtown areas.

Table 3.4: Radio cell coverage area in different environments and the number cell in a FN service area

Environment	Radio cell coverage area Km²	Number of cells per FN service area
Downtown area	0.0727	4
Urban area	0.983	1 per 2 FNs
Suburban area	17,067	1 per 10 FNs

Now, we determine whether the capacity of CATV coaxial distribution cables is sufficient for supporting 3G Node Bs. The answer to this question depends on the CATV spectrum allocation, shown in Figure 3.1 for a typical CATV system [29]. In this spectrum allocation, TV analog broadcast channels occupy the spectrum from 54 to 550 MHz. The spectrum from 5 to 40 MHz is reserved for the upstream traffic, that includes telephony, data, and signaling. The spectrum from 600 to 750 MHz is reserved for downstream traffic that includes digital video, telephony, data, and signalling.

```
5        40   54                              550     600          750

Freq.      (MHz)

5 to 40 MHz              54 to 550 MHz          600 to 750 MHz
Digital Upstream Spectrum   Broadcast Spectrum    Digital Downstream Spectrum
```
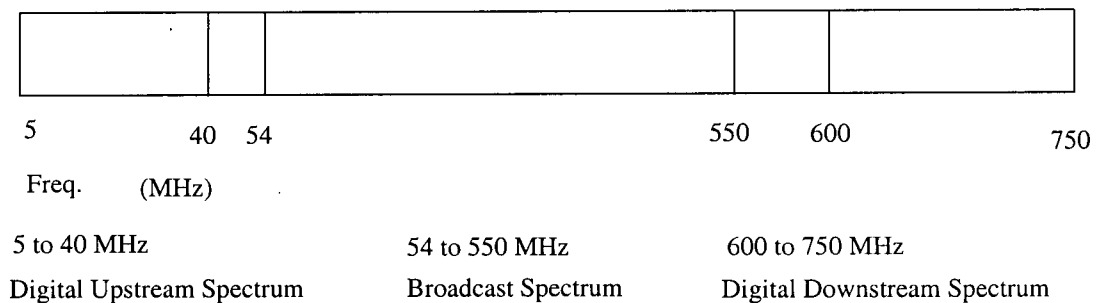
Figure 3.1 Spectrum allocation for CATV

Since FNs are interconnected to the Headend using optical fibers, the only apparent bottleneck is in the coaxial distribution cables, especially in the upstream (35 MHz). However, this amount of spectrum can support up to 56 Mbps, for example using QPSK modulation scheme. Details of the physical layer are beyond the scope of this work.

From Section 3.2.2, it is clear that a single Node B can support up to 75 voice channels, and 50 data channels. This corresponds to a total bit stream of 1.08 Mbps. It is clear that the upstream capacity of the coaxial distribution cables (56 Mbps) can easily accommodate the traffic generated by UMTS Node Bs.

3.3 Proposed 3G/DOCSIS Network Architecture

We propose a UMTS deployment over DOCSIS based HFC networks. CATV operators worldwide deploy DOCSIS equipment and protocols over their HFC networks. This will make it difficult for any 3G/HFC proposal that does not comply with DOCSIS standards to be accepted worldwide. We expect our proposal to be dominant over other proposals, since it is to be deployed over the existing HFC networks, and DOCSIS equipment and protocols. The proposed UMTS/HFC network architecture is shown in Figure 3.2.

In our proposed 3G/CATV deployment, we use the existing CATV plant, and DOCSIS equipment and MAC protocol to interconnect the large number of UMTS Node Bs to be deployed with RNCs. Also, RNCs communicate with each other using the CATV plant. These Node Bs and RNCs interface to the CATV plant using DOCSIS Cable Modems. Bidirectional data transfer between 3G Node Bs and RNCs take place through a DOCSIS CMTS using a
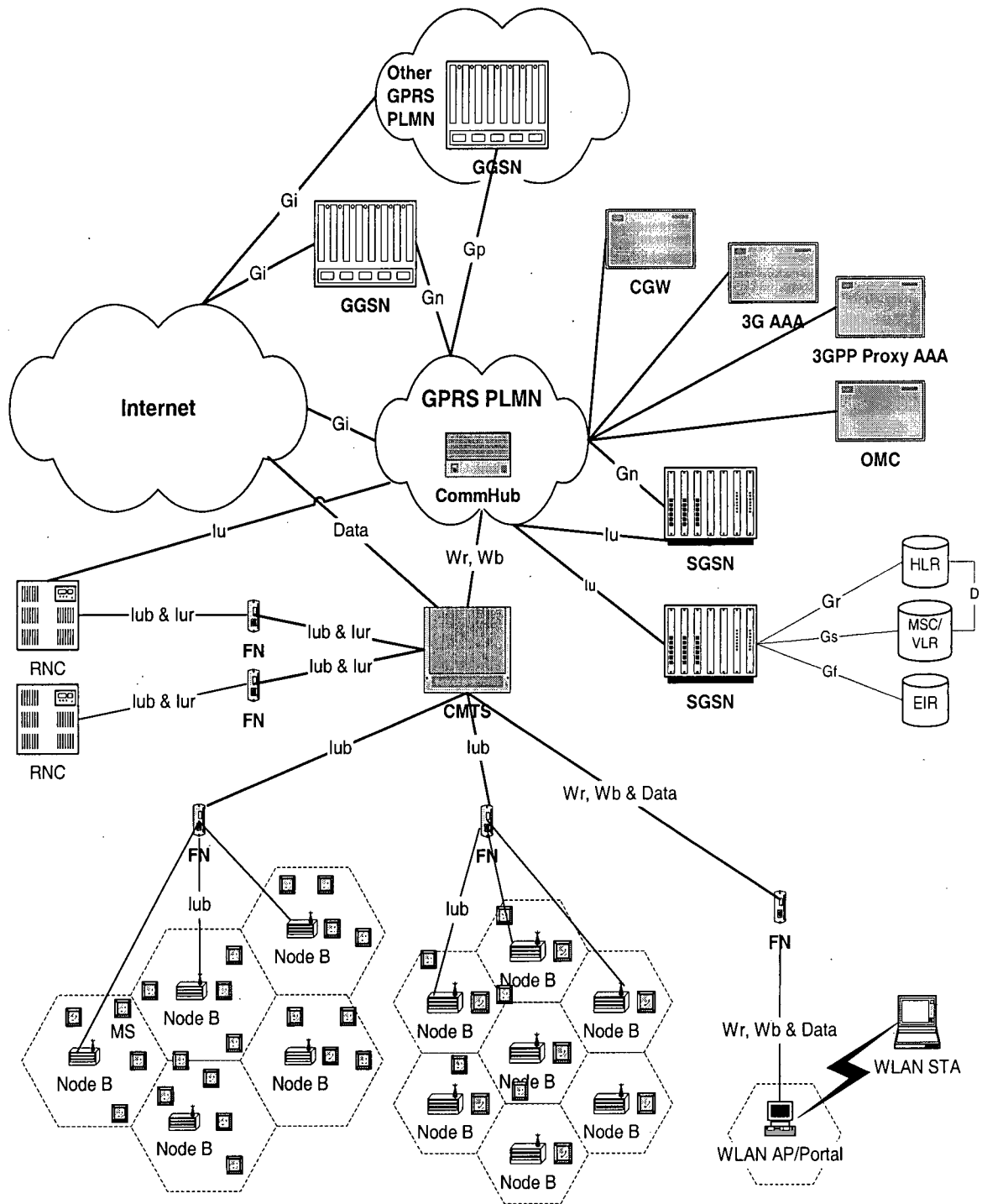
Figure 3.2 Proposed Wireless Data over CATV deployment

DOCSIS MAC and PHY protocols.

In this section, we discuss various functional elements and protocol stack of our proposed 3G/ CATV network architecture. DOCSIS MAC protocol along with enhancements to support 3G traffic will be discussed in Chapter 4, and Chapter 5.

3.3.1 Iu-b Interface over DOCSIS

The Iu-b interface interconnects Node Bs to RNCs. UMTS Node Bs cover pico-, micro- and macro-cells in urban, sub-urban, and rural areas. An RNC controls a number of Node Bs within an RNS coverage area (refer to Chapter 2 for more details). Node Bs and RNCs interface to the cable plant using DOCSIS Cable Modems (see Figure 3.2).

Figure 3.3 and Figure 3.4 show the proposed Iu-b over DOCSIS network architecture and protocol stack, respectively. In the upstream, multipath components of received signals from each mobile user are combined, detected, framed, and converted to IP packets at Node Bs. These IP packets are transmitted by Node Bs CMs to one or more of the RNCs CMs, using the DOCSIS MAC and PHY protocols. In the downstream, packets received at RNCs from SGSNs are forwarded by the RNCs CMs to one or more of Node Bs CMs, using DOCSIS PHY and MAC protocols. Note that all transmissions in DOCSIS pass through and are controlled by DOCSIS CMTS. The CMTS uses IP addresses, MAC addresses, and SIDs for relaying packets between CMs. The IP addresses are used to relay packets to external network elements like the 3G Core Network HUB (see Section 3.3.5).
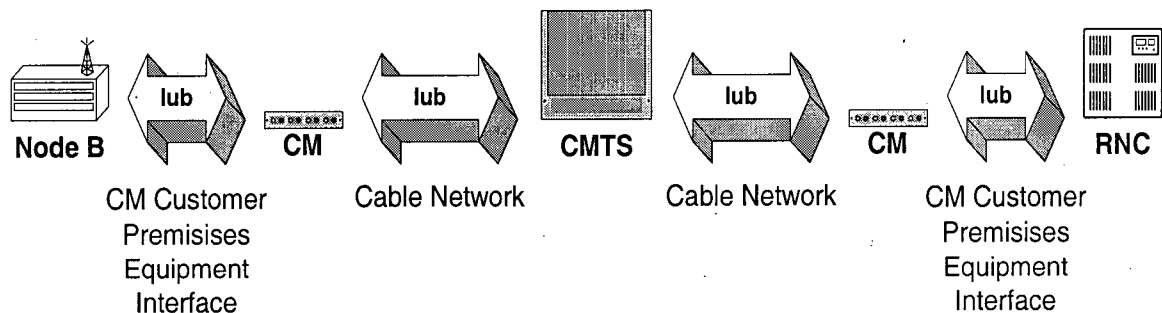
Figure 3.3 Proposed Iu-b over DOCSIS deployment

Node B	CMTS	RNC

Frame Protocol			Frame Protocol
UDP			UDP
IP	IP	IP	IP
802.2/DIX LLC	802.2/DIX LLC	802.2/DIX LLC	802.2/DIX LLC
Link Security	Link Security	Link Security	Link Security
Cable MAC	Cable MAC	Cable MAC	Cable MAC
DS TC / Up- strm Cable PMD / Cable PMD	DS TC / Up- strm Cable PMD / Cable PMD	DS TC / Up- strm Cable PMD / Cable PMD	DS TC / Up- strm Cable PMD / Cable PMD

Cable Network Transmission   Cable Network Transmission

Figure 3.4 Proposed Iu-b over DOCSIS protocol stack

42

3.3.2 Iu-r Interface over DOCSIS

The Iu-r interface interconnects RNCs within an RNS (refer to Section 2.2.1.2 for more details). RNCs interface to the cable plant using DOCSIS Cable Modems (see Figure 3.2).
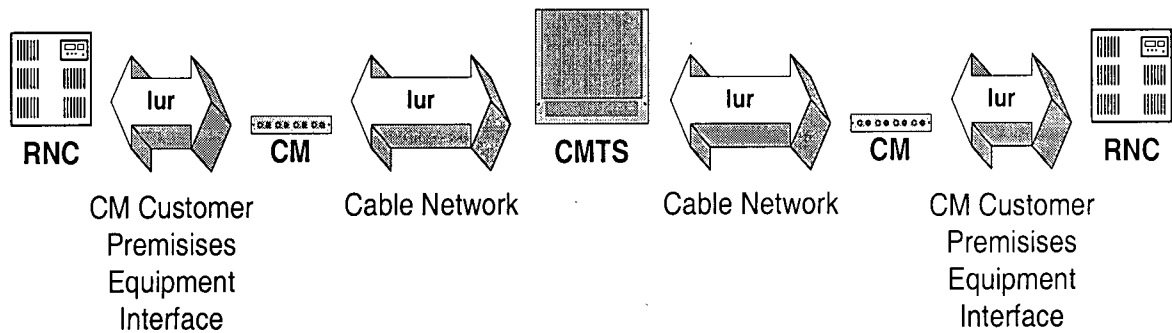
Figure 3.5 Proposed Iu-r over DOCSIS deployment

Figure 3.5 and Figure 3.6 show the proposed Iu-r over DOCSIS network architecture and protocol stack, respectively. Bidirectional 3G traffic between RNCs is framed at the corresponding RNCs and relayed by their CMs. The IP addresses of the communicating RNCs are used to relay IP packets between them. Note that all packets in transit between RNCs pass through DOCSIS CMTS, which translates IP address to CM MAC addresses and SIDs for relaying these packets. DOCSIS MAC and PHY protocols are used for the data transfer between the communicating RNCs.

3.3.3 Iu Interface over DOCSIS

The Iu interface interconnects RNCs to SGSNs. Fewer communication links are needed for the Iu interface than that needed for the Iu-b interface. This is due to the fact that much smaller number of SGSNs exist in the network compared to Node Bs. However, higher capacity is

43

RNC	CMTS		RNC
Frame Protocol			Frame Protocol
UDP			UDP
IP	IP	IP	IP
802.2/DIX LLC	802.2/DIX LLC	802.2/DIX LLC	802.2/DIX LLC
Link Security	Link Security	Link Security	Link Security
Cable MAC	Cable MAC	Cable MAC	Cable MAC
DS TC Cable PMD	Up-strm Cable PMD	DS TC Cable PMD	Up-strm Cable PMD	DS TC Cable PMD	Up-strm Cable PMD	DS TC Cable PMD	Up-strm Cable PMD

└Cable Network Transmission┘   └Cable Network Transmission┘

Figure 3.6 Proposed Iu-r over DOCSIS protocol stack

needed in each of the Iu interface links. If an Iu interface requires a higher capacity than that available in a DOCSIS network, we recommend to have dedicated communication links other than the CATV cables for relaying the Iu interface traffic, as shown in Figure 3.2. This takes Iu interface out of our scope.

The second alternative, if the capacity of a DOCSIS network suffice for relaying the Iu interface traffic, the CATV plant can be used for relaying this traffic. We expect that this is likely to be the case. At the time of this writing, a typical GPRS SGSN handles a maximum of

1.6 Mbps in each direction. The capacity of a UMTS SGSN at its initial deployment is not expected to be much more than that. Considering that a single SGSN controls a number of RNCs, the bandwidth required per an Iu interface can be handled by a DOCSIS network.

The second alternative is illustrated in Figure 3.7 and Figure 3.8, which show Iu interface over DOCSIS network architecture, and protocol stack, respectively. On one side of Iu interfaces, the RNCs interface to the CATV plant using DOCSIS CMs in a similar fashion as in the Iu-b, and Iu-r interfaces. On the other side of Iu interfaces, SGSNs (via a Core Network Hub as explained in Section 3.3.5) interface to the CATV plant through the CMTS Network side Interface (CMTS-NSI). We choose to interface SGSNs to CMTSs through the optional IP CMTS-NSI (see Figure 3.7).

RNC   CM   CMTS   CN HUB   SGSN

CM Customer   Cable Network   CMTS   GPRS PLMN
Premisises   Network Side
Equipment   Interface
Interface
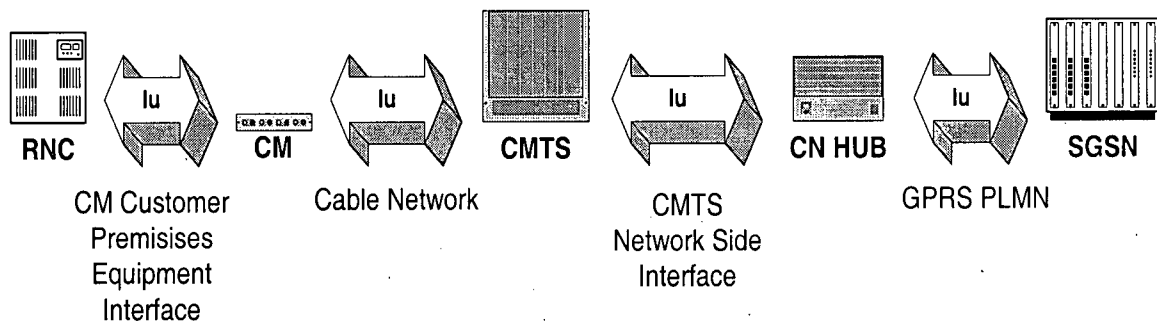
Figure 3.7 Proposed Iu over DOCSIS deployment

In the upstream, packets are aggregated at an RNC CM and transmitted to an affiliated SGSN using its IP address. DOCSIS MAC and PHY protocols are used between the DOCSIS CM and the CMTS. These packets are forwarded to the SGSN on the IP CMTS-NSI side through the Core Network Hub. In the downstream, packets received from a GGSN are forwarded by

45

RNS **CMTS** **CN Hub** **3G-SGSN**

RNS		CMTS		CN Hub		3G-SGSN	
Relay						Relay	
PDCP	GTP-U					GTP-U	GTP-U
RLC	UDP	IP	IP	IP	IP	UDP	UDP
	IP					IP	IP
	802.2/DIX LLC	802.2/DIX LLC					
MAC	Link Security	Link Security	L2	L2	L2	L2	L2
	Cable MAC	Cable MAC					
L1	DS TC Cable PMD / Up-strm Cable PMD	DS TC Cable PMD / Up-strm Cable PMD	L1	L1	L1	L1	L1

Cable network
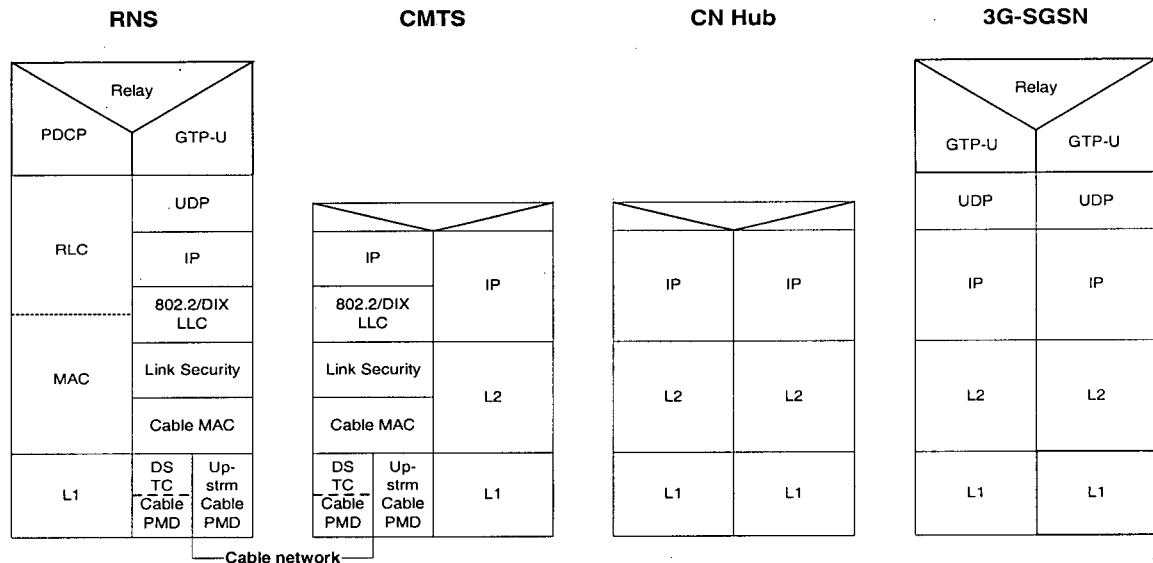
Figure 3.8 Proposed Iu over DOCSIS protocol stack

the SGSN to one or more RNCs using their IP addresses. These packets go through the CN Hub to the CMTS where they are forwarded to their destination RNC CM.

3.3.4 OMC Interface over DOCSIS

As mentioned in Chapter 2, a remotely located Operation and Maintenance Center (OMC) handles configuration management, fault management, and performance management in the UTRAN. Element Managers at the OMC communicate with Node Bs, and RNCs using the Itf-B, and Itf-R interfaces, respectively. While the Itf-R interface is a direct connection, the Itf-B interface could be either a direct connect or via the RNC [8].

We propose a deployment of the Itf-B interface over DOCSIS CATV networks, as shown in Figure 3.9. We choose the option of having direct connections between Node Bs and the OMC for the Itf-B interface, as defined in 3GPP TR 32.800 [10]. The Itf-B interface traffic between

Node Bs and the OMC will be transported over the CATV plant. In the upstream, the Itf-B interface traffic is framed at Node Bs, packetized into IP packets, and transmitted by DOCSIS CMs using DOCSIS MAC and PHY protocols to the DOCSIS CMTS. The CMTS uses the destination address of these packets to forward them to the OMC via the 3G CN HUB. In the downstream, IP packets that carry the Itf-B interface traffic received at the CMTS from the OMC via the CN HUB are forwarded to their destination Node B CMs. The IP addresses of these packets are translated by the CMTS into MAC addresses and SIDs and are used to forward these packets to their destination Node B CMs. The CMTS carries out this task without any concern for the contents of these packets.
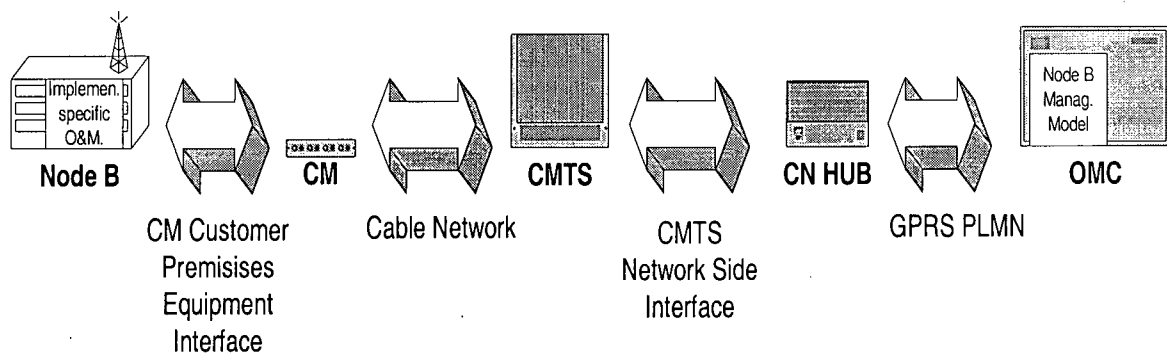
Figure 3.9 Proposed Node B OMC over DOCSIS deployment

The Itf-R interfaces traffic could be transported in either of two options, as follows. It is either carried by dedicated communication links or over the CATV plant. This is decided based on which of the two options for transporting the Iu traffic, presented in Section 3.3.3, is chosen. In other words, If dedicated links are deployed for the Iu interface traffic, the Itf-R interfaces traffic is transported using these dedicated links. Otherwise, if the Iu interface traffic is carried

over the CATV plant, then similarly the ItB-R interfaces traffic is carried over the CATV plant. The CMTS forwards bidirectional Itf-R interface traffic between the OMC and RNCs in a similar manner as in the case of Itf-B interface traffic, as explained above (see Figure 3.10).
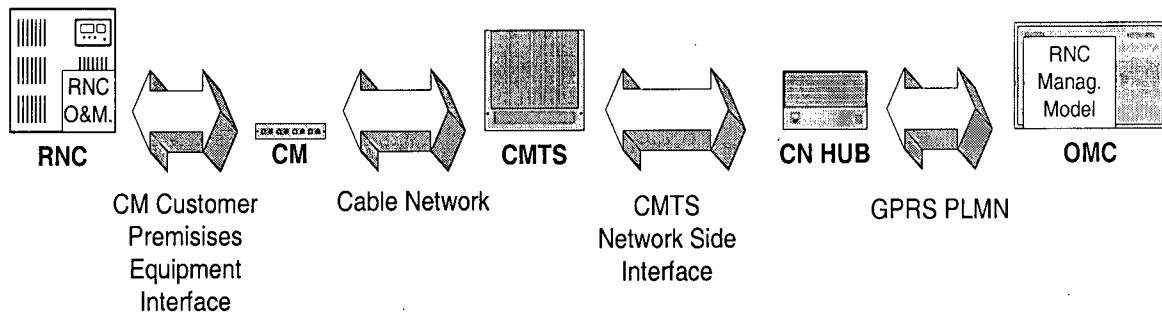
Figure 3.10 Proposed RNC OMC over DOCSIS deployment

3.3.5 UMTS Core Network HUB

A 3G Core Network (CN) HUB interconnects various CN functional elements. These include: SGSNs, GGSNs, an Authentication, Authorization and Accounting (AAA) server, an OMC, and the Charging Gateway (CGW). It is also the single point of communication between a UMTS CN, and the outside world. It interconnects a UMTS CN to RNSs, and the Internet.

The DOCSIS CMTS interfaces to the CN HUB using the optional IP based CMTS-NSI. In the downstream, the CN HUB routes users' data, signalling, and network management traffic addressed to RNCs and Node Bs located within the CATV plant to the CMTS. In the upstream stream, the CMTS forwards 3G traffic from RNCs and Node Bs located within the CATV plant, which is destined to the OMC and the SGSNs, to the CN HUB.

3.4 Proposed WLAN Over DOCSIS Network Architecture

WLANs made high-speed wireless Internet access available in hot spots such as Hotels, Airports, Hospitals and residential premises. WLANs Internet access does not provide true mobility such as that in wide area wireless network. On the other hand, wide area wireless systems such as UMTS systems offer true user mobility but at lower data rates than WLANs. It is believed that both WLANs and 3G systems are going to co-exist and complement each other.

There has been significant research in the area of WLAN/3G interworking. In 3GPP standards group, a number of scenarios for 3GPP and WLAN interworking have been considered [4]. These range from common billing (loose coupling) to the provision of services seamlessly between the WLAN and the 3GPP systems (tight coupling).

We propose a WLAN/3G interworking deployment over DOCSIS CATV networks. (see Figure 3.2). Our proposal is independent of the coupling scenario chosen. WLANs have access to 3G Core Network elements through the CATV Plant. These 3G Core Network elements include SGSNs, GGSNs, AAA server, an OMC, and the CGW. A WLAN Access Point (AP) interfaces to the CATV plant using a DOCSIS CM. It uses DOCSIS MAC, and PHY protocols to send/receive data and signalling traffic to DOCSIS CMTS. The CMTS relays bidirectional traffic between WLAN APs on one side, and 3G Core Network and/or the Internet on the other side. As mentioned in Section 3.3.5, traffic to/from the 3G Core Network is routed through the Core Network HUB.

We propose a possible scenario for WLAN/3G interworking architecture, as shown in Figure 3.11. This scenario represents a compromise between loose coupling and tight coupling. We explain this WLAN/3G interworking scenario in the following sections.
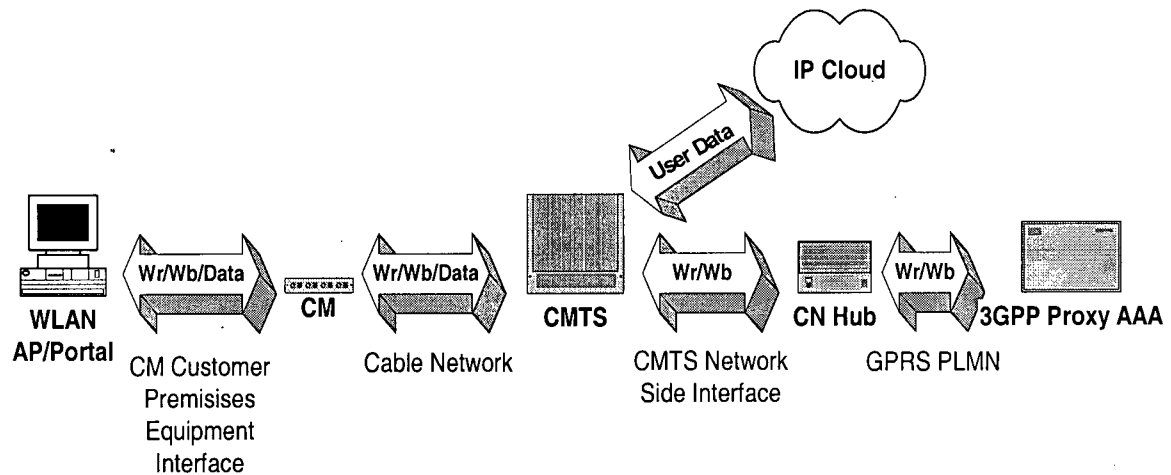
Figure 3.11 Proposed Wireless LAN over DOCSIS deployment

3.4.1 Wr and Wb Reference Points Traffic Transport

The Wr reference point connects the WLAN access network, possibly via intermediate networks, to the 3GPP proxy AAA Server [4]. The main functionality of this reference point is to transport RADIUS/Diameter frames for authentication, and authorization purposes in a secure manner.

The Wb reference point is located between the WLAN access network and 3GPP CGW. The main function of this reference point is to transport charging-related information in a secure manner. This reference point should be Diameter or RADIUS based.

We propose transporting the Wr and Wb reference points traffic over the DOCSIS CATV

plant. Figure 3.12 shows the protocol stack for transporting Wr and Wb reference points

traffic over a DOCSIS network. A WLAN AP/Portal uses a DOCSIS CM to transmit/receive

Wr and Wb packets to/from the DOCSIS CMTS. The CMTS relays bidirectional Wr and Wb

packets between the 3G Core Network (via the CN HUB) and the WLAN APs/Portals using

their IP address. It is up to the CN HUB to forward the Wr packets to the 3G AAA proxy

server, and Wb packets to the CGW, using IP addresses.

WLAN STA	WLAN AP/Portal		CMTS		CN Hub		3GPP Proxy AAA
Diameter Client Application							Diameter Server Application
TCP/SCTP							TCP/SCTP
IP	IP	IP	IP	IP	IP	IP	IP
802.2 LLC	802.2 LLC	802.2/DIX LLC	802.2/DIX LLC				
802.11 Link Security	802.11 Link Security	DOCSIS Link Security	DOCSIS Link Security	L2	L2	L2	L2
802.11 MAC	802.11 MAC	Cable MAC	Cable MAC				
802.11 Phy	802.11 Phy	DS TC Cable PMD / Up-strm Cable PMD	DS TC Cable PMD / Up-strm Cable PMD	L1	L1	L1	L1

Cable network

Figure 3.12 Proposed Wireless LAN Wb and Wr interfaces over DOCSIS protocol stack

3.4.2 WLAN Data Traffic Transport

The protocol stack for transporting WLAN data traffic over a DOCSIS CATV network is

shown in Figure 3.13. After performing Authentication and Authorization procedures, A

WLAN station can start exchanging IP packets with other users either in the same DOCSIS network or in other networks. In the upstream, A WLAN AP/Portal uses a CM to send these packets to the CMTS. Depending on the destination for these packets, the CMTS forwards them either to the Internet or to other CMs in same DOCSIS network. Downstream traffic is handled in same manner where packets destined to users within a WLAN access network are routed through the CM connected to their WLAN AP/Portal.
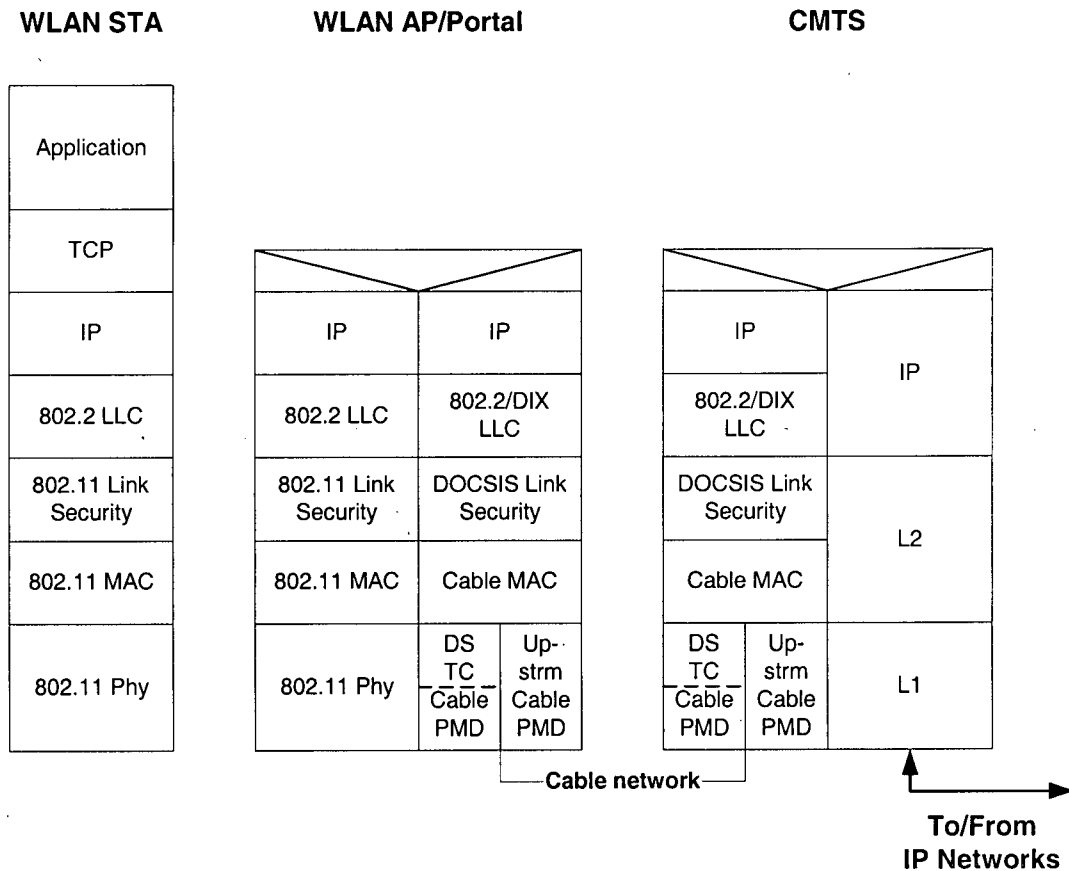
Figure 3.13 Wireless LAN data over DOCSIS protocol stack

3.5 DOCSIS QoS Framework

As presented in this chapter, we propose to use DOCSIS CATV networks for transporting 3G, and WLANs traffic. This should be performed without degrading the QoS for existing residential Internet access through the CATV plant. Although, maintaining QoS guarantees over CATV is important in both the upstream and downstream, QoS guarantees in the upstream is a more challenging task than that in the downstream. This is due to the contention based nature of the DOCSIS CATV upstream channel. Due to the challenging and urgency nature of QoS guarantees in upstream, we focus in Chapter 4, and Chapter 5 on QoS guarantees in the upstream, specifically at the MAC layer.

In the following sections, we present a QoS framework for DOCSIS networks. DOCSIS CMTS employs classifiers, policers, shapers, queue managers, and schedulers for QoS management. CMs are pretty light weight in terms of QoS as will be discussed later. These QoS mechanisms assure QoS for user data as agreed upon during connection setup.

Figure 3.14 shows an overview of the various functions employed by the DOCSIS CMTS and CMs in the upstream and downstream directions. Specific QoS functions for the CMTS and CM in the upstream and the downstream are summarized in the following sections.

3.5.1 DiffServ Support

Our QoS framework follows the DiffServ paradigm for scalability reasons. In DiffServ, some traffic is treated better than the others on an average basis. This means less delay on average, more bandwidth on average, etc. DiffServ can provide QoS guarantees for a variety of
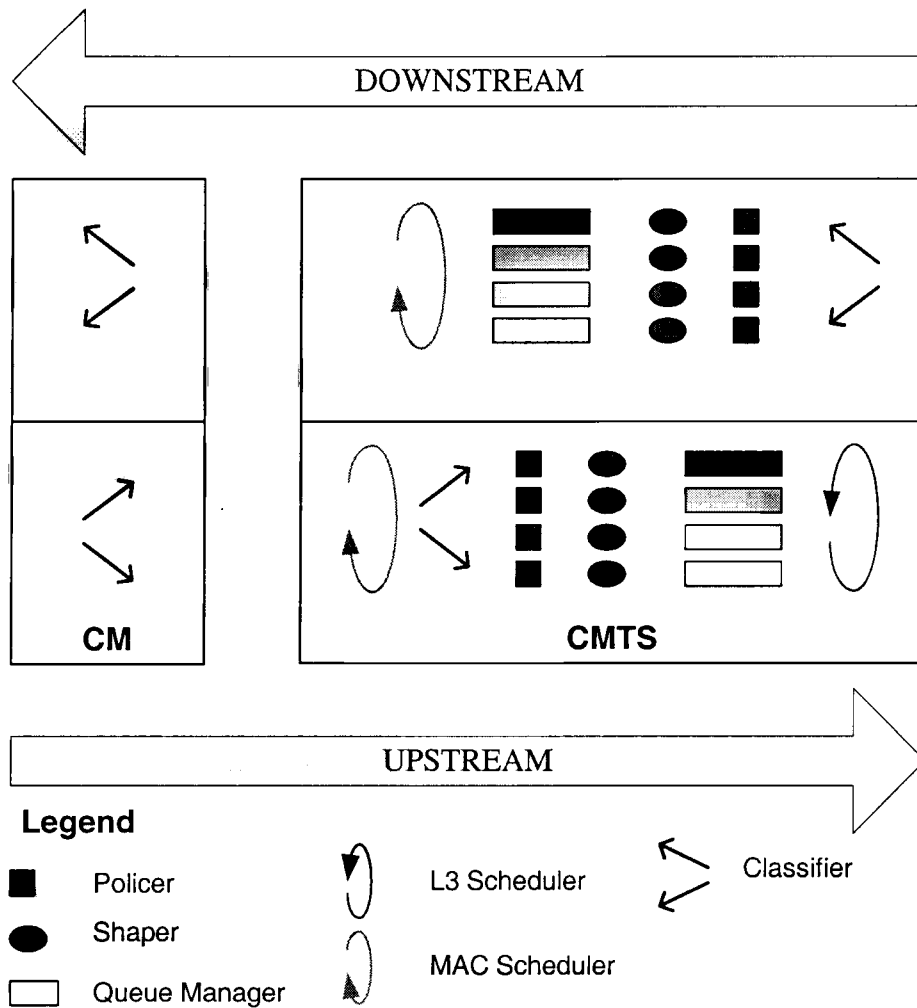
Figure 3.14 Transmission Plane QoS functions

applications with different delay, bandwidth, and packet loss requirements.

DiffServ is specified by the Internet Engineering Task Force (IETF) in RFC2475 [57]. In

DiffServ packets are classified and marked at network boundaries or hosts. A DiffServ

Codepoint (DSCP) is marked in packets which determines their Per-hop Behavior (PHB).

This is the first six bits of either the Type of Service (TOS) octet in the IPv4 header, or the

Traffic Class octet in IPv6 header. DiffServ capable network elements en route provide packets with their delay, bandwidth, and packet loss requirements according to their marked DSCP.

The IETF DiffServ Working Group has standardized several PHBs including:

- Expedited Forwarding (EF) – used to provide low delay, low jitter, and low loss services.

- Assured Forwarding (AF) – used to provide services for traffic with less stringent delay requirements than EF. There are four AF classes. Within each AF class, a packet can be assigned one of three different levels of drop precedence.

- Best-effort (BE) – a default PHB for traffic that does not require guaranteed service level.

We recommend supporting QoS in DOCSIS CMTS, and CM devices through an implementation of DiffServ that differentiates between different classes. Different traffic classes are mapped to the DiffServ Per-Hop Behaviors (PHBs). The mapping of the traffic classes to specific PHBs is configurable by an operator.

3.5.2 CMTS QoS Functions - Upstream

The CMTS is the upstream egress network element of the DOCSIS network (CM is the upstream ingress network element). It receives IP packets from CMs (residential, 3G, and WLAN CMs). These are forwarded by the CMTS to Packet Data Networks (PDNs) that include Intranets, the Internet, and 3G PLMNs. The CMTS transmission plane QoS functions

deployed in the upstream are illustrated in Figure 3.15.
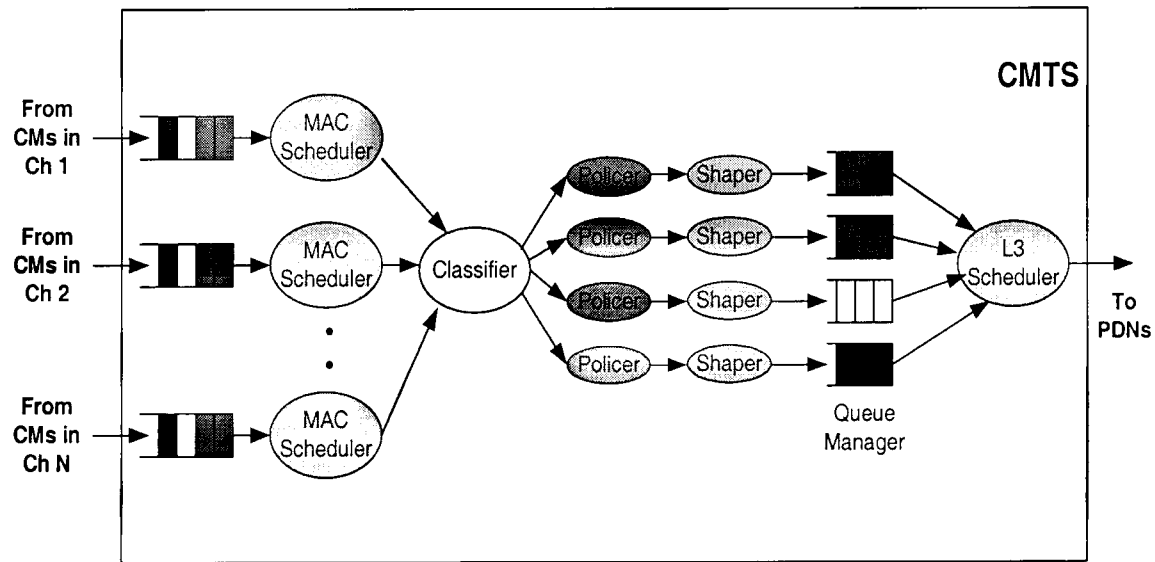
Figure 3.15 CMTS upstream QoS functions

3.5.2.1 MAC Scheduler

The MAC scheduler regulates the discipline in which upstream packets are scheduled for accessing an upstream channel. In our framework, we propose deploying two stages of scheduling in a hierarchical manner. The first stage schedules traffic in every upstream channel at the MAC layer. The second stage schedules the aggregated traffic from all upstream channels at the network layer (see Section 3.5.2.6).

We propose deploying Weighed Fair Queueing (WFQ) for scheduling at the MAC layer. WFQ mechanisms have been widely accepted for scheduling at the network layer [80]. Scheduling at the MAC layer presents more challenges than that in the network layer. In Chapter 5, we detail our WFQ based MAC scheduler. Our MAC scheduler deploys more QoS functions like

shapers and policers at the MAC layer that we don't discuss here, for clarity reasons. Rather we detail these in Chapter 5 (see Figure 5.1).

3.5.2.2 Classifier

This function classifies the packets arriving from CMs through different Channels. It classifies packets at the network layer by examining their DSCP field. This function prepare IP packets for further QoS treatment before forwarding them upstream through the ingress CMTS-NSI interface. The CMTS could either remark packets with a different DSCP or forward them using the original DSCP. Remarking the DSCP field is needed if packets are destined for a network that uses different DSCP marking for a certain QOS class. Mapping tables are needed in this case. We recommend that the CMTS forwards upstream packets which are destined to the 3G Core Network using the original DSCP field. This is to ensure consistency between DSCP marking at the 3G access network and the 3G Core Network.

3.5.2.3 Policer

The policer is deployed to ensure traffic that belong to a given connection does not exceed the QoS negotiated for that connection during admission while allowing a small degree of burstiness. A Policer deploys some metering function to monitor the behavior of a given flow and take action in case of violations. These actions could be discarding, or marking packets that belong to a certain flow.

In the upstream, policing is performed only if no policing is done at the MAC layer of the CMTS. Otherwise, if policing was performed at the MAC layer, performing this function at

the network layer becomes redundant. Deploying WFQ at the MAC layer ensures that the scheduled flows are shaped. Policed flows at the MAC layer are expected to maintain their traffic shapes as they exit the scheduler. Therefore there is no need to police those flows again.

Policing could be done per flow or per traffic class. Policing per flow is preferable since it solves the Denial of Service (DoS) problem. DoS happens when a single flow bombards the network with traffic that exceeds its negotiated contract. If policing per flow is not performed, a single user could occupy the whole bandwidth dedicated for a given traffic class, and hence can cause service denial for other users. The advantage of per flow policing comes at the expense of a high impact on the CPU resources of the CMTS. On the other hand per traffic class policing is simpler and has a lower impact on CMTS CPU resources, however, it can not avoid the DoS problem. Generic mechanisms like Token Bucket could be used for either per flow or per traffic class policing.

3.5.2.4 Queue Manager

The queue manager maintains traffic queues based on the PHB of traffic classes within the CMTS. It determines the manner in which packets are dropped during congestion. A simple queue manager could drop newly arrived packets to a full queue. We recommend using the Weighted Early Random Detection (WRED) for queue management. Adoption of WRED is recommended since it can give some preference in discarding packets during congestion. WRED has many advantages over other queue management schemes that have been widely addressed in the literature.

3.5.2.5 Shaper

A shaper prevents bursitiness by controlling the rate of traffic destined upstream. It ensures that the total amount of traffic is within the allowed average and peak throughput capabilities of the upstream schedulers. The scheduler and shaper are configured according to the SLA negotiated between the CMTS and a given PDN. A token packet shaper can be deployed for this purpose.

3.5.2.6 Network Layer Scheduler

The scheduler regulates the discipline in which packets destined upstream are scheduled for transmission. It ensures that packets with higher priorities are given preferential treatment in transmission to upstream PDNs. At this point, scheduling is performed at the network layer for packets aggregated from different upstream channels, that need to be forwarded upstream to PDNs. We recommend using WFQ for scheduling at the network layer. WFQ mechanisms have been widely accepted for scheduling at the network layer.

3.5.3 CMTS QoS Functions - Downstream

The CMTS is the downstream ingress boundary of the DOCSIS network. It receives IP packets from Packet Data Networks (PDNs) that include Intranets, the Internet, and 3G PLMNs. These are forwarded by the CMTS to CMs in the downstream at the IP layer. Also, the CMTS receives upstream packets for CMs that need to be forwarded to other CMs these are forwarded at the MAC layer by the MAC forwarder without any layer 3 intervention. Figure 3.16 shows CMTS Transmission plane QoS functions in the downstream. These are

explained in the following.
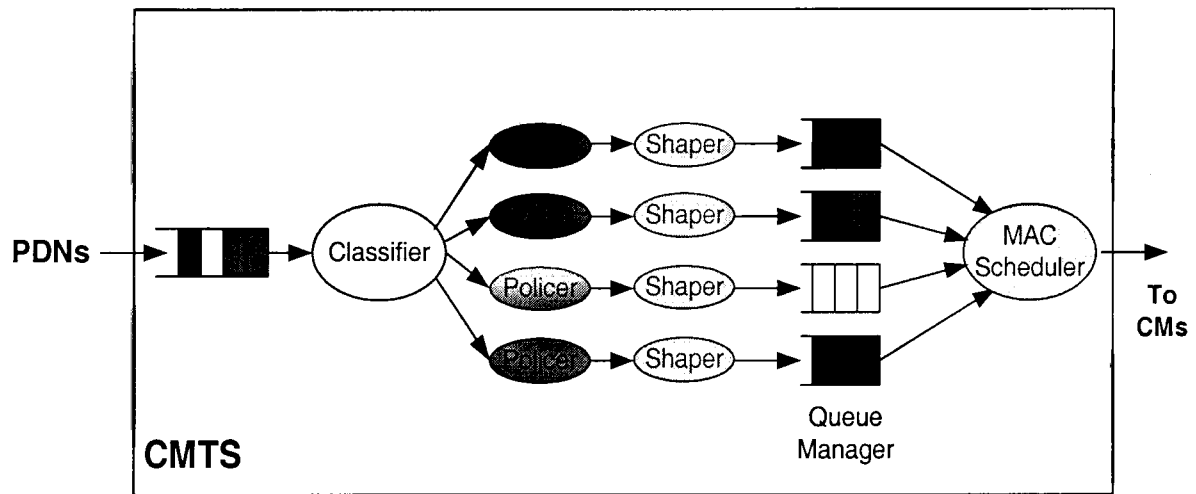
Figure 3.16 CMTS downstream QoS functions

3.5.3.1 Classifier

This function classifies the packets from PDNs by examining the DSCP in the IP header of packets arriving at the ingress interface (CMTS-NSI) to the CMTS. It forwards the received packets using the unmodified DSCP.

3.5.3.2 Policer

In the downstream, Policing is performed at CMTS-NSI to limit traffic into the DOCSIS network that arrives from any network. Alternatively, the CMTS could police traffic that arrives from "untrusted" domains only, and allow traffic that arrives from "trusted" domains to pass through without any policing. A trusted domain here refers to CMTSs that are operated by trusted operators. The ingress CMTS could assume that traffic that arrives from a trusted domain has been already policed at the egress CMTS of that domain.

60

As in the upstream, policing could be done per flow or per traffic class using a generic mechanisms like Token Bucket. Un-conforming packets could be either discarded, or marked, depending on the operator's configuration.

3.5.3.3 Queue Manager

This function has the same purpose and behavior as the upstream queue manager. WRED is used to manage queues for individual traffic classes based on PHB.

3.5.3.4 Shaper

In the downstream, a token packet shaper prevents bursitiness by controlling the rate of traffic destined downstream. It ensures that the total amount of traffic is within the allowed average and peak throughput capabilities of downstream CMs.

3.5.3.5 MAC Scheduler

In the downstream, we propose deploying a WFQ scheduler at the MAC layer. We propose eliminating network layer scheduling for redundancy reasons. If MAC layer scheduling is performed effectively, layer three scheduling becomes redundant.

3.5.4 CM QoS Functions

The CM is pretty light weight in terms of QoS functions. This is a desirable characteristic in order to keep CMs simple and affordable. Most of the intelligence is implemented in the CMTS (see Figure 3.14).

In the upstream, a classifiers marks upstream IP packets with DSCP according to the QoS

negotiated during connection setup. Upstream traffic received from 3G networks elements should be marked with a DSCP field that ensures EF PHB. Marked packets are classified into flows and queued at different queues with different SIDs at the MAC layer.

In the downstream, packets received from the CMTS are sorted into different queues according to their SID. These packets are handled to the CM network layer for forwarding to connected PCs, 3G, or WLAN network elements.

3.6 Summary

We proposed a new network architecture for deploying UMTS networks over DOCSIS CATV networks. Our proposal minimizes the need for building a backhaul network for UMTS access networks by utilizing the existing CATV plant and DOCSIS equipment and protocols. We conducted a traffic study for mapping UMTS network elements to the CATV plant. We defined the network architecture and the protocol stacks for deploying different UMTS interfaces over DOCSIS networks. In addition, we defined a network architecture and protocol stack for UMTS/WLAN interworking over DOCSIS CATV networks. Finally, we proposed a QoS framework for DOCSIS networks that facilitates downstream and upstream QoS support.

Chapter 4

Proposed DOCSIS Medium Access Control Enhancements

4.1 Introduction

This chapter proposes an efficient implementation, and enhancements to DOCSIS MAC protocol for facilitating transporting wireless data traffic over HFC CATV networks. The proposed MAC protocol can reduce the access delay for delay-sensitive traffic by 30% to 40% over existing DOCSIS MAC. This is achieved without compromising QoS guarantees for other traffic classes, or the DOCSIS channel utilization. Both wireless and wireline delay-sensitive traffic can benefit from the proposed DOCSIS MAC enhancements.

The main technical issue for supporting 3G Wireless systems over DOCSIS based CATV networks is related to Medium Access Control in the upstream (MAC). An adequate MAC

protocol for DOCSIS that efficiently supports 3G wireless traffic in addition to the existing wireline residential Internet traffic is needed.

QoS guarantees in the downstream CATV channels is not a critical task because of the broadcast nature of transmission in the downstream. It is a one-to-many scenario where the DOCSIS CMTS has sole control of the bandwidth allocation and transmission to all CMs in the downstream. There is no contention for bandwidth in the downstream which makes MAC and QoS provisioning much easier to manage than in the upstream.

On other hand, MAC and QoS guarantees in the upstream are not as simple of a task because of the many-to-one access. CMs transmissions in the upstream are subject to contention which makes QoS guarantees a difficult task to achieve. In particular, the contention nature of transmissions in the upstream makes it difficult to guarantee delay bounds for real-time traffic. There has been a significant research in the area of QoS at the MAC layer. In Appendix A, we review related previous work at the MAC layer.

Our focus is on DOCSIS MAC protocol which incorporates some of the most sophisticated MAC mechanisms reported in the literature. As summarized in Section 2.4.4, DOCSIS 1.1 MAC protocol [26] specifies a number of QoS guarantees for real-time traffic in the upstream. These include a number of upstream Service Flow scheduling services, which are designed to improve the efficiency of the poll/grant process. Among these, the Unsolicited Grant Service (UGS), Unsolicited Grant Service with Activity Detection (UGS-AD), and Real-Time Polling Service (rt-PS), are designed for real-time traffic.

Neither the UGS nor the UGS-AD are suitable for our proposed 3G wireless deployment over DOCSIS. This is due to the variable size of packets aggregated at a 3G Node to be transmitted by a CM. Both the UGS and the UGS-AD are aimed at applications that require periodic fixed size grants, where, our proposed 3G deployment requires variable size grants. The rt-PS is the only candidate for supporting our 3G traffic. However, as will be demonstrated later in this chapter, it is inefficient mainly because it restricts a CM to use only unicast polls for requesting bandwidth.

In this chapter, we propose an enhancement of DOCSIS MAC rt-PS protocol that enables it to efficiently support QoS guarantees for 3G wireless traffic as well as wireline residential Internet traffic. Also, we illustrate how 3G wireless traffic can be supported using DOCSIS MAC protocol in general. We recommend implementation options that are essential for supporting 3G wireless traffic over DOCSIS MAC protocol. Our proposed MAC protocol is designed keeping in mind the following objectives:

- Compliance with DOCSIS MAC Standard.

- QoS guarantees for different traffic classes (both wireline and wireless).

- Proposed MAC protocol changes should be transparent to existing CMs. It is not justifiable to modify millions of existing CMs.

The MAC functions consist of a master MAC located at the CMTS and a slave MAC located at CMs. The following sections describe the functional elements and the operation of the

proposed MAC protocol implementation.

4.2 CMTS MAC

The CMTS MAC acts as a master MAC. It controls the operations of the slave MACs which reside at every CM. The functions of the CMTS MAC include registration and synchronization control, bandwidth allocation, scheduling, QoS Support, and Call Admission Control (CAC).

The available frequency spectrum in the cable plant is partitioned into upstream and downstream using FDM. Upstream and downstream spectrum is further partitioned into a number of channels with data rates of 5 Mbps.

In the upstream, the CMTS MAC provides for partitioning a given upstream channel into multiple contention and reservation mini-slots, each with an implicit number that goes from 0 to a CMTS defined maximum (2^{32} - 1). A mini-slot is a power-of-two multiple of 6.25μs increments: 1, 2, 4, 8, 16, 32, 64, or 128 times 6.25μs, that can be used to transfer 16 to 48 bytes (depending on the modulation scheme). An upstream MAC frame is partitioned into maintenance mini-slots, contention mini-slots, and reservation mini-slots. Maintenance mini-slots are used for transmitting maintenance related message like synchronization and ranging. Contention mini-slots are used to transmit request mini-PDUs (mini protocol data units), while reservation mini-slots are used to transmit users' data. Integral number of mini-slots are used to transmit either variable length packets or ATM cells (note that the support for ATM is a future issue).

The CMTS MAC controls CMs registration and synchronization process. Upon receiving a registration request from a CM, the CMTS authenticates the CM, and assigns a Station ID (SID). Thereafter, the CMTS advises the CM to adjust its local time according to its distance from the CMTS (ranging). This ensures that when CMs, located at different distances from the CMTS, transmit at the same local time, their transmissions arrive at the CMTS at the same CMTS time. Also, the CMTS creates a record for the CM in active CMs lookup tables, and retrieves its subscribed QoS profile.

Upon receiving a request for establishing an upstream service flow from a CM, the CMTS creates a context and allocates upstream bandwidth for the flow. The CMTS periodically broadcasts downstream bandwidth allocation MAP messages that inform CMs of the status of their transmissions in the previous contention mini-slots (e.g. successful, collision). In addition MAP messages advice CMs of the allocation of upstream mini-slots (contention, polls, data, etc.) in a given time frame. A scheduler located at the CMTS takes care of mini-slots assignment. Figure 4.1 shows a downstream MAP message and the upstream MAC frame described by it.

4.3 CM MAC

The CM MAC located at CMs acts as a slave MAC. It is the interface between the Data Link Layer (DLL), and DOCSIS Physical Layer (PHY). The CM MAC is designed to follow the master MAC located at the CMTS. It interacts with CMTS MAC in performing registration and synchronization. Other functions include participation in contention resolution, and
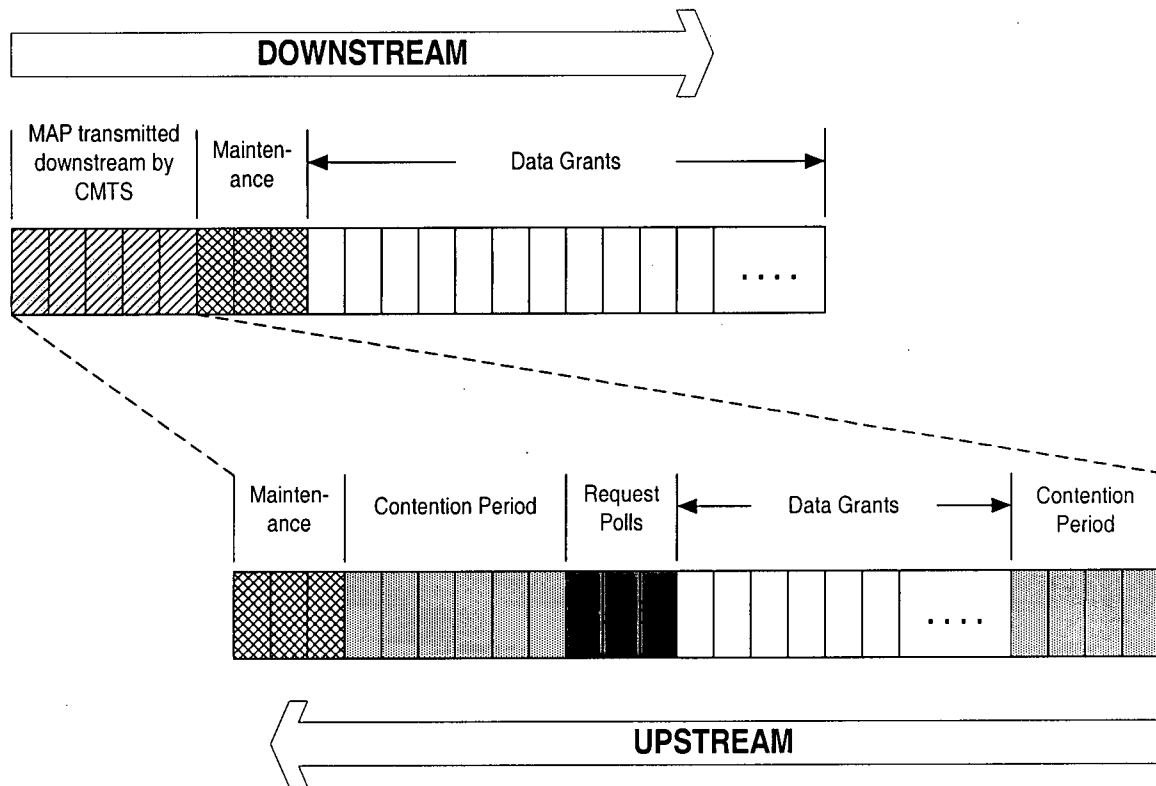
Figure 4.1 Allocation MAP

transmission of requests for bandwidth as well as upstream data.

An active CM joins the systems by undergoing a synchronization and a ranging process, where the CM is assigned a Local Identifier, and the local time for the station is adjusted according to its distance from the CMTS.

The CM MAC identifies its transmission opportunities (contention or reservation) by scanning downstream MAP messages and takes turns in transmitting requests for bandwidth and data at allocated mini-slots. Data received from upper layers is classified into a given priority classes depending on certain classifiers (e.g., DiffServ Code Point DSCP). Then the

CM transmits a request mini-PDU in the contention period of the upstream frame in order to establish a context for the flow. The CMTS uses MAP messages to indicate to various CMs the status of transmissions in the previous contention mini-slots (e.g. successful, collision). If this request is in collision, then the CM would apply the Binary Exponential Backoff Contention Resolution Algorithm (CRA) until the collisions are resolved. The CM scans future MAPs for data transmission opportunities and transmits data in assigned mini-slots. The CM which have successfully acquired reservation mini-slots can piggyback requests for extra mini-slots on their transmitted data packets using the Extended Header field.

4.4 Contention Resolution Algorithm

A simple, flexible, and efficient Contention Resolution Algorithm (CRA) is important to the operation of the DOCSIS MAC protocol. CMs send registration messages, requests for bandwidth, and some short data packets using a contention based scheme. The CRA of choice for DOCSIS MAC is based on a Truncated Binary Exponential Back-off, with the initial back-off window and the maximum back-off window controlled by the CMTS. In the following we explain how this contention based algorithm is used to send Requests for bandwidth:

- The CMTS assigns a number of upstream mini-slots for contention and creates a Request Information Element (IE) in the MAP message which is transmitted downstream to CMs. Also, the CMTS uses the MAP to inform CM about the initial back-off window and the maximum back-off window as a power-of-two value. For example, a value of 2 indicates a window between 0 and 3.

- A CM scans the MAP for contention based Request IEs. In order to reduce the probability of request collision, the CM may not start transmitting its request in the first available upstream contention mini-slot. Instead, it skips zero or more mini-slots depending on a random offset which the CM calculates based on the Data Backoff Start value in the most recent Map.

- The CMTS uses a subsequent MAP to inform CMs of the status of their transmissions (successful, collision) and bandwidth assignment or pending grants.

- If the CM receives a Data Grant (Data Grant Pending) or data Acknowledge IE from the CMTS, it recognizes that it transmission was successful, and the CRA is complete.

- If the request was in collision, the CM finds a MAP without a Data Grant (Data Grant Pending) or Data Acknowledge for it and with an ACK time more recent than the time of transmission. The CM proceeds with collision resolution by increasing its back-off window by a factor of two, as long as it is less than the maximum back-off window. Again the CM randomly selects a number within its new back-off window and repeats the deferring process described above.

- The CM keeps re-trying until either the transmission is successful or the maximum number of retries (16) has been reached, at which time the Packet is be discarded.

A key factor in reducing collisions, and hence improving the performance of the CRA is the number of mini-slots available for contention in each upstream MAC frame. It was shown in

[32] that a frame size of 2 ms with 6 contention mini-slots results in the best throughput and delay performance among all scenarios.

We adopt a 2 ms fixed length upstream MAC frame with a minimum of 6 mini-slots assigned for contention (see Figure 4.1). In addition, any unassigned mini-slots in the upstream MAC frame are assigned for contention. This has the effect of reducing collisions, stabilizing the network performance, and lowering the CMs access delay for sending Requests and short data packets. Also, Data Acknowledge, Data Grant, and Data Grant Pending Information Elements are given the highest priority in generating a MAP. This allows CMs to be informed at the earliest time possible about collision of their transmissions and expedites the contention resolution process which contributes to reducing access delay.

4.5 Cable Modem Profiles

The CMTS maintains a lookup table that acts as a database for storing CMs profiles. It stores entries for information about CMs such as subscribed, requested, and offered QoS traffic profiles (e.g., class of service, delay, and throughput). CMs could be in either of a number states (e.g., active, inactive). The CMTS updates the status of CMs including their SIDs in the lookup table to be used for different MAC operation. The CMTS maintains other temporary lookup table for some MAC operations. For instance, the CMTS maintains a lookup table for SIDs of CMs with real-time traffic which are active but temporary silent due to ON-OFF nature of some real-time traffic. Such table will be used for the Polling scheme, as explained in Section 4.7.

71

4.6 Bandwidth Allocation

A key function of the DOCSIS MAC protocol is bandwidth allocation. Bandwidth allocation is performed using the DOCSIS MAC scheduler located at the CMTS. The MAC scheduler assigns upstream mini-slots for maintenance, control, and data traffic. It constructs MAP messages that inform CMs of their assignments, and broadcasts these MAP messages to all CMs downstream.

The deployment of an efficient MAC scheduler that guarantees QoS for wireless and wireline traffic is crucial to DOCSIS MAC. The DOCSIS MAC standard does not mandate any specific scheduler, rather it is implementation dependent. In Chapter 5, we propose an efficient WFQ based MAC scheduler for DOCSIS networks.

4.7 Polling Scheme

We propose enhancements to DOCSIS MAC protocol that provide delay guarantees for real-time traffic while improving resource utilization [42]. We propose a polling scheme that can be used for providing the integrated 3G wireless traffic as well other wireline real-time traffic with QoS guarantees. Our polling scheme can be used to enhance the existing DOCSIS Real-time Polling Service (rt-PS) in the following manner:

- Our scheme allows CMs to piggyback requests while active and rely on polling only when idle, which reduces the access delay, and saves the bandwidth needed for polls. This in contrary to the existing DOCSIS rt-PS polling scheme which restricts CMs using the rt-PS to rely only on polling.

- Our scheme polls silent real-time CMs only when their delay bounds are about to be violated (a bit less than 20 ms, as well be explained later). It relies on our MAC scheduler that assigns bandwidth for requests received from polled CMs promptly. This saves the unnecessary overhead needed if all silent CMs are to be polled in every upstream MAC frame, and hence, frees some bandwidth to be effectively utilized by other traffic.

Figure 4.2, and Figure 4.3 illustrate the proposed CMTS, and CM MAC operations, respectively. Note that some MAC operations in the figures can be only understood in context of other MAC operations like concatenation and fragmentation (see Section 4.8, and Section 4.9).

Our proposed real-time polling scheme works as follows:

(1) A CM with real-time traffic sends a bandwidth request to the CMTS using contention mini-slots. If the request is in collision, the CM applies the contention resolution algorithm explained in Section 4.4 until the request eventually goes through. If the request is successfully received at the CMTS, the CMTS proceeds with step (2).

(2) The CMTS admission control checks if it can admit the new data flow without degrading the QoS for the existing connections. Depending on the result of the above check, the new connection is either admitted or rejected. An admitted CM's flow is allocated a SID, and assigned upstream mini-slots according to the MAC scheduler. Also, the CMTS creates
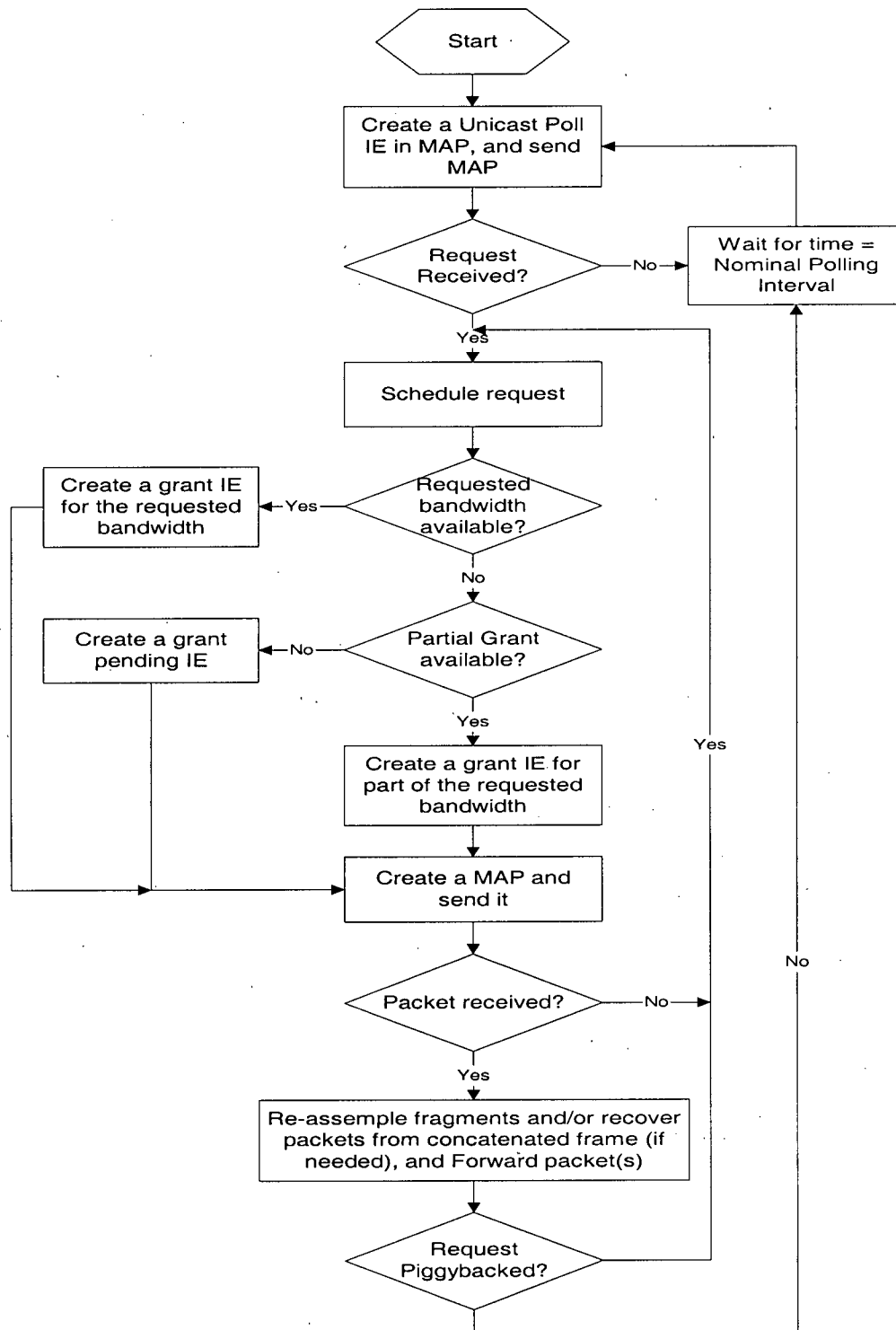
Figure 4.2 Cable Modem Termination System MAC
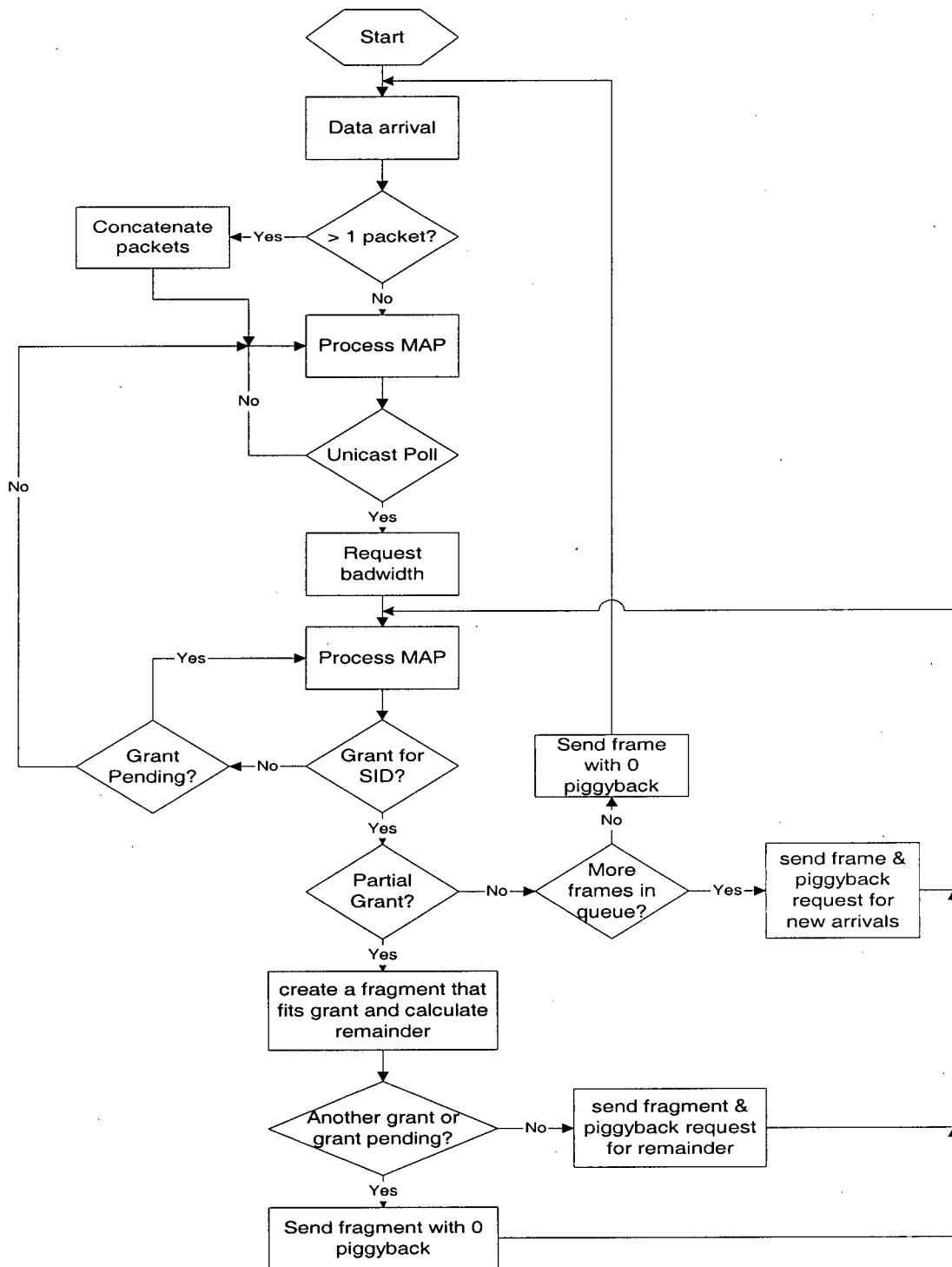
Figure 4.3 Cable Modem MAC

75

an entry for the CM in the real-time Polling Service (rt-PS) lookup table marking its status as Active/ON. The corresponding CM is informed of the outcome of the above procedure in the first subsequent downstream MAP.

(3) The CM scans MAP messages for any reserved upstream mini-slots to be used for its data. Upon allocating reserved upstream mini-slots, the CM uses the reserved mini-slots to transmit its data.

(4) If the CM receives more upstream data after sending its first request, it piggybacks a requests for extra mini-slots on the transmitted data packets using the Extended Header field.

(5) The CMTS forwards the received upstream data to its destination, and extracts the piggybacked request. The MAC scheduler assigns upstream mini-slots for the CM and informs it using the MAP message. This process continues until the CM becomes silent. This is recognized by the fact that an upstream data packet arrives at the CMTS without a piggybacked request. A CM indicates to the CMTS that the transmitted data packet is the last packet in its queue by transmitting the last packet without piggybacking a request to it.

(6) When the CMTS detects that a CM is silent as above, it stops allocating any upstream mini-slots for it. Also, the CMTS marks the status of the CM in the rt-PS lookup table as Active/OFF, and polls the CM as below.

(7) At the time of creating a MAP message, the CMTS scheduler checks the rt-PS lookup table to determine which CMs are eligible for polling. Eligibility is determined based on the CMs Nominal Polling Interval. The Nominal Polling Interval is chosen so that the access delay for a packet that arrives at a CM after a silence period does not exceed 20 ms. The CMTS scheduler creates a Unicast Poll IE in MAP for each eligible CM.

(8) If the CM is still silent, it skips the unicast polling mini-slot. An empty unicast polling mini-slot received at the CMTS indicates that the CM is still silent.

(9) The CMTS polls the silent CM every P ms, where P equals the Nominal Polling Interval. This continues until the CM uses the unicast poll mini-slot to send either a request for bandwidth when it becomes active/ON again or an explicit request to end the session.

(10) If the CM becomes active/ON, it scans MAPs for a unicast request opportunity assigned to it. Upon allocating one it sends a contention-free request to the CMTS. The CMTS halts the polling scheme, assigns the requested upstream bandwidth, and advises the CM using a MAP message. The protocol executes as in steps (3) to (10).

The above MAC protocol is expected to provide guaranteed QoS for traffic that requires tight delay bounds while having good utilization for the bandwidth of HFC networks. In Section 4.10, and Section 4.11, we evaluate the performance of the proposed MAC scheme.

4.8 Packets Concatenation

We propose aggregating traffic from a Wireless 3G Node B at a DOCSIS CM to be

transported to a 3G RNC via DOCSIS CMTS. CMs supporting traffic from a 3G Node B are expected to have higher volumes of traffic than CMs supporting residential Internet Access. As illustrated in Section 3.2, a 3G Node B can 75 wireless voice users and 50 wireless data users. In contrast to wireline residential CMs, a number of packets are expected to be queued at the 3G Node B CM most of the time. It is inefficient to request bandwidth and transmit each packet separately.

DOCSIS Packet concatenation provides the answer for the above problem [27]. Each packet is encapsulated in a single MAC frame. A CM can concatenate multiple MAC frames to be transmitted in single transmit opportunity. This Concatenated MAC frame can be transferred upstream with a single PHY overhead and Concatenation MAC Header.

A CM supporting a 3G Node B traffic encapsulates each packet it receives from the wireless side in a MAC frame. The CM calculates the bandwidth needed to transfer the concatenated MAC frame, and sends a bandwidth request to the CMTS accordingly. This request can be either piggybacked or sent in a unicast poll opportunity, as explained in Section 4.7. Upon receiving a grant, the CM transmits the whole concatenated MAC frame (or a fragment of it as explained Section 4.9). The CMTS uses the concatenation MAC header to regenerate the individual MAC frames. Finally, the CMTS removes DOCSIS MAC overhead from these individual MAC frames, and forwards individual packets to the corresponding 3G RNCs via an IP cloud.

4.9 Packet Fragmentation

A CM supporting a 3G Node B may concatenate a large MAC frame for transmission in the upstream. The CMTS may not have the available bandwidth to transmit the whole concatenated MAC frame at once. A packet may wait for a long time before it can be accommodated in a single upstream MAC time frame. If the concatenated MAC frame requires more mini-slots than the total number of mini-slots available in upstream MAC time frame, then this transmission opportunity will never come.

Fragmentation can resolve this issue in DOCSIS [27]. In contrast to concatenation, fragmentation provides the means for partitioning a packet into smaller fragments that can be transmitted individually and then re-assembled at the CMTS. Fragmentation can take place for either a single MAC frame or concatenated multiple MAC frames. It is a capability of both the CMs and the CMTS, and it is applicable only in the upstream.

When a CM request bandwidth for either a single MAC frame or a concatenated MAC frame, the CMTS could provide a Partial Grant. The CMTS may allocate a Partial Grant for the requesting CM if it can not allocate the requested mini-slots for the CM at once. The CMTS advises the CM of a Partial Grant using a MAP message. This triggers the CM to create a fragment along with fragmentation header that can fit the allocated upstream mini-slots. Additional grants that follow at later times could either trigger additional fragmentation(s) or transmitting the remainder of the MAC frame at once. The CMTS uses the fragmentation header to re-assemble the original MAC frame and forwards it to its destination. Figure 4.4

illustrates an example of fragmenting a concatenated MAC frame. Fragmentation is carried

out without regard to original packet boundaries. As mentioned earlier, fragmentation can take

place for either a concatenated MAC frame or a single MAC frame.

Orginal Concatenated MAC
Frame (208B)

6B	15B	100B	4B	6B	73B	4B
Concat HDR	MAC HDR1	PDU payload1	PCRC1	MAC HDR2	PDU payload2	PCRC2

First Grant (80B)

10B	2B	6B	15B	43B	4B
FRAG HDR	FHCS	Concat HDR	MAC HDR1	PDU payload1 (part 1)	FCRC

Second Grant (160B)

10B	2B	57B	4B	6B	73B	4B	4B
FRAG HDR	FHCS	PDU payload1 (part 2)	PCRC1	MAC HDR2	PDU payload2	PCRC2	FCRC
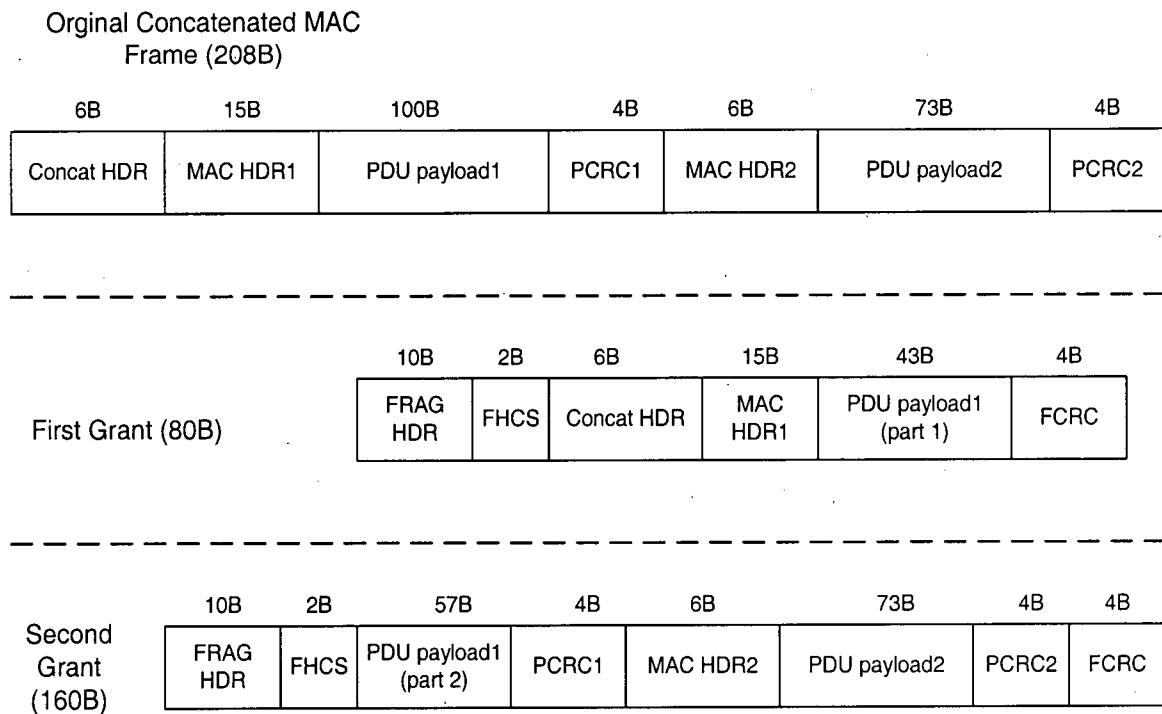
Figure 4.4 Fragmentation of a concatenated MAC frame example

There are two optional CMTS modes for performing fragmentation: The Multiple Grant

Mode, and Piggybacking Mode. It is the CMTS which determines the fragmentation mode

and CMs follow by performing a consistent set of rules. In the Piggybacking Mode, the

CMTS does not retain the state of the fragmentation. Instead, it relies on CMs to piggyback

requests for the remainder of the MAC with each fragment until the whole packet goes

through. This may require a CM to send multiple bandwidth requests before completing the

the transmission of a single MAC frame. Sending multiple requests for a single MAC frame may increase the access delay for a packet if class based queueing is used. Every time a request for a fragment arrives at the CMTS, it may have to wait at the CMTS MAC scheduler queue until other requests from other CMs in the same priority class are served. This can only be avoided if flow based queueing is used to serve request from different CMs separately.

On the other hand, in the Multiple Grant Mode, the CMTS keeps track of the state of the fragmentation which allows it to have multiple partial grants outstanding for any given CM. This mode does not require a CM to send multiple bandwidth requests for a single MAC frame. Instead, the CM makes only one successful bandwidth request for a single MAC frame, and relies on the CMTS to take care of the rest. Upon allocating any bandwidth for a request, the CMTS keeps a record of the bandwidth needed to transmit the remainder the MAC frame and allocates bandwidth for it as it becomes available. This mode helps reduce the access delay for a fragmented MAC frame.

Based on the above, we can conclude that the Multiple Grant Mode can help reduce the access delay for 3G wireless traffic and real-time traffic under any queueing discipline. The Piggybacking mode can be used for such kind of traffic only if requests are served at the CMTS in a flow based manner.

4.10 Performance Bounds

In this section, we conduct an analytical study for evaluating the delay, and throughput bounds, as detailed in the following subsections.

4.10.1 Delay Analysis

There are two types of traffic supported using DOCSIS rt-PS service. These are our proposed 3G traffic, and wireline real-time traffic. Wireline real-time traffic is an ON/OFF traffic that experiences active and silence periods. On the other hand, 3G traffic is expected to be always active since it aggregates a large number traffic sources. A SID admitted to rt-PS services can send a request using either an RTP unicast mini-slot, or piggybacking to a data packet. Note that DOCSIS standard allows only RTP unicast polls for sending request, while our proposed MAC allows both RTP unicast polls and piggybacking. In our delay analysis, we analyze both scenarios for comparison.

Proposed MAC delay bounds

In this section, we focus on deriving the delay bounds for our proposed 3G traffic deployment over DOCSIS. As explained earlier, the MAC scheduler determines the number of upstream mini-slots that need to be allocated for control (contention, polling), and data transmission. CMs are advised of this allocation using a MAP. Figure 4.5 shows a simplified upstream MAC frame for our delay analysis. A minimum of six mini-slots are allocated for contention. Unassigned mini-slots in a MAC frame are also allocated for contention. This is not the case at heavy loads where data traffic leaves no unassigned mini-slots, and restricts the number of contention mini-slots to six.

The number of RTP mini-slots is determined based on the Nominal Polling Interval for SIDs.
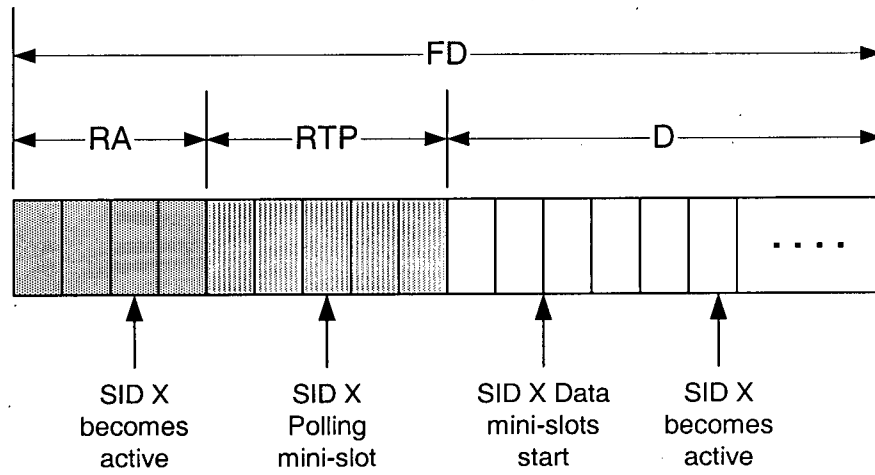
Figure 4.5 Simplified upstream MAC frame for delay analysis

Assuming that SIDs unicast RTP polls are uniformly distributed, the number of RTP polls per

MAC frame can be estimated according to the following formula:

$$\text{Number of RTP min-slots per frame} = \frac{N_{RTP} \times FD}{NPI} \tag{4.1}$$

Where N_{RTP} is the number SIDs admitted to the RTP service, FD is the upstream MAC frame

duration, and NPI is the Nominal Polling Interval.

A SID supporting 3G traffic could have a packet ready for transmission at any point in time

during a given MAC frame (see Figure 4.5). The time that an active SID waits till it sends a

request for bandwidth is a random variable. However, this delay is bounded and can be

derived using a simple mathematical analysis as in [66]. Our proposed MAC provides 3G

traffic admitted to DOCSIS RTP services with upper and lower delay bounds as analyzed

below.

The access delay (W) is composed of two parts: the detection delay (W_d), and the admission delay (W_a). The former is the waiting time till a bandwidth request by a SID is successfully received at the CMTS, while the later is the elapsed time between receiving the request till the data is received at the CMTS.

$$W = W_d + W_a \qquad (4.2)$$

For simplicity and in order to provide an insight of the proposed MAC delay bounds, we assume that detected requests sent in Frame K are allocated bandwidth in Frame K+1. Also, we assume that CMs are located 10 Km apart from the CMTS which results in a one way propagation delay (τ) of 0.05 ms (5 μs per Km). This is equivalent to 2 mini-slots duaration. Therefore, the access delay W equals the detection delay W_d plus the elapsed time till data transmission completes in Frame K+1.

The lower bounds for the detection, and total access delay can be calculated using the following formulas.

$$W_d(\text{min}) = \frac{2}{78}FD + \tau = \frac{4}{78}FD \qquad (4.3)$$

$$W(\text{min}) = \left(\frac{4}{78}FD\right)_K + \left(RA + \frac{2}{78}FD\right)_{K+1} + \tau = \frac{14}{78}FD \qquad (4.4)$$

Where FD is the MAC frame duration which is 2 ms (78 mini-slots), RA is the Random access mini-slots duration, and RTP is the RTP mini-slots duration. A minimum of 6 mini-

slots could be allocated for RA, and a minimum of 0 mini-slots can be allocated for RTP polling.

The lower delay bound represents a scenario where a short packet became ready for transmission in Frame K just before another short packet transmission in mini-slots located close to the end of Frame K. The propagation delay of 2 mini-slots time makes it impossible to piggyback a request in the last 2 mini-slots of Frame K, and get a grant in Frame k+1. The shortest packet in DOCSIS can be fit in 2 mini-slots (32 Bytes). Therefore, at best the request is piggybacked to a short packet transmitted in mini-slots number 75 and 76 of Frame K. Upstream bandwidth for the packet transmission could be allocated as early as the first Data mini-slots of Frame K+1.

The lower bound for detection time is the time required to transmit the piggybacked request (2 mini-slots) plus propagation delay. The lower bound for the total access delay is the elapsed time from the start of transmission of the piggybacked request at mini-slot number 75 till the data packet which is transmitted in Frame K+1 is received at the CTMS. That is the duration of 4 mini-slots in Frame K plus the duration of the minimum RA sub-period and 2 mini-slots for packet transmission in Frame K+1.

The upper delay bound can be estimated in the same manner. This represents a scenario where a packet became ready for transmission right after the start of transmission of an other short packet at the beginning of the data interval of Frame K, given that bandwidth assignment for data transmission was at the end of data interval of Frame K+3.

The upper bounds for the detection, and total access delay can be calculated using the following formulas.

$$W_d(\text{max}) = (D)_K + FD_{K+1} + \tau \approx 2FD \qquad (4.5)$$

$$W(\text{max}) = (D)_K + FD_{K+1} + FD_{K+2} + FD_{K+3} + \tau \approx 4FD \qquad (4.6)$$

The upper bound for detection time is the waiting time till the request is piggybacked to a data packet, which could be due for transmission at the end of Frame K+1, plus propagation delay. Piggybacking the request at the end of Frame K+1 makes it impossible to get a grant at Frame K+2 due to propagation delay. Instead, mini-slots are allocated in Frame K+3 and could be at the end of the frame. Therefore, the upper bound for the total access delay is the detection time plus elapsed time till the packet is transmitted in Frame K+3 plus the propagation delay.

Existing DOCSIS MAC delay bounds

In this section, we derive delay bound for the existing DOCSIS rt-PS MAC. As mentioned earlier, existing DOCSIS MAC restricts SIDs admitted to rt-PS to use only unicast RTP polling slots for requesting bandwidth. It does not allow piggybacking requests. We use the same methodology as above for deriving delay bounds.

We start by obtaining the lower delay bounds. The lower delay bound represents a scenario where a short packet became ready for transmission just prior to its RTP unicast polling mini-slot and that the polling mini-slots was at the end of the RTP sub-period of Frame K, given

that the assigned mini-slots for transmission was at the beginning of the Data sub-period of Frame K+1 (see Figure 4.5). This scenario also implies that there were no RTP mini-slots allocated in Frame K+1. As a result, the lower bounds for the detection, and total access delay can be calculated using the following formulas.

$$W_d(\min) = \frac{1}{78}FD + \tau = \frac{3}{78}FD \tag{4.7}$$

$$W(\min) = \left(\frac{1}{78}FD + D\right)_K + \left(RA + \frac{2}{78}FD\right)_{K+1} + \tau \approx FD \tag{4.8}$$

The lower bound for detection time is the time required to transmit the request using a unicast RTP mini-slot plus the propagation delay. The lower bound for the total access delay is the detection time, plus the elapsed time till the packet is transmitted at the beginning of the Data sub-period of Frame K+1, plus propagation delay.

The upper delay bound can be estimated in the same manner. This represents a scenario where a packet became ready for transmission shortly after the start of its RTP unicast polling mini-slot and that the polling mini-slots was at the beginning of the RTP sub-period of Frame K. The request for bandwidth was sent in the next unicast poll mini-slot which was scheduled at the end of RTP sub-period of Frame K+7, and bandwidth was granted at the end of Frame K+8.

If a SID misses a Unicast RTP polling mini-slot, the next opportunity will scheduled after the NPI time elapses. The number of frames to be skipped before scheduling the next polling

87

mini-slot for a SID is determined according to the following formula:

$$\text{Number of frames between unicast polls} = \frac{NPI}{FD} - 1 = 7 \qquad (4.9)$$

The upper bounds for the detection, and total access delay can be calculated using the following formulas.

$$W_d(\max) = (RTP + D)_K + 7FD + (RA + RTP)_{K+8} + \tau \approx 8FD \qquad (4.10)$$

$$W(\max) = (RTP + D)_K + 7FD + (RA + RTP + D)_{K+8} + FD_{K+9} + \tau \approx 10FD \qquad (4.11)$$

The upper bound for detection time is the waiting time till the request is sent at the end of the RTP sub-period of Frame K+8, plus propagation delay. The upper bound for the total access delay is the detection time, plus the elapsed time till the packet is transmitted at the end of the Data sub-period of Frame K+9, plus propagation delay.

Discussion

As seen from the above analysis, the proposed MAC can provide lower guaranteed delay bounds for the proposed 3G deployment than the exiting DOCSIS MAC. The upper bound for the access delay utilizing our proposed MAC calculated in Equation (4.6) is four times the frame size. This is 8 ms seconds in our implementation (FD = 2 ms). Whereas it is ten times the frame size (20 ms) in the case of the existing DOCSIS MAC (see Equation (4.11)). Remember this an upper bound, and access delay is expected to be less than that on average.

It is worth mentioning that the above delay bounds are obtained based on the assumption that requests for bandwidth received in Frame K are immediately assigned bandwidth in Frame K+1, as far as the MAP for Frame K+1 has not been sent already. This should be the case at least for the real-time traffic. Either a priority scheduler or a Fair Queueing scheduler with higher weights configured for real-time traffic could be employed for this purpose.

4.10.2 Throughput Analysis

The throughput of the proposed MAC depends on a number of factors that include the bandwidth allocation scheme, and the offered load traffic mix. As discussed in Section 4.10.1, the duration of Control (RA, and RTP), and data sub-periods in MAC Frame is variable, which adds to the complexity of system to be analyzed. In this section, we estimate the maximum throughput of the proposed MAC using a simple mathematical analysis as in [18].

As mentioned earlier, the data rate for a single DOCSIS upstream channel is 5 Mbps. It is partitioned into 2 ms long frames (78 mini-slots). A MAC frame is composed of three sub-periods: RA, RTP, and Data as shown in Figure 4.5. At heavy loading conditions, there are no unassigned mini-slots since all mini-slots in the data sub-period (D) will be used for data transmission. This restricts the number of contention mini-slots (RA sub-period) per frame to 6 mini-slots. The number of RTP mini-slots (RTP sub-period) per frame can be estimated using Equation (4.1).

Let β be the duration of a request mini-slot. The total duration of the request sub-period (RQ) of a frame can be estimated as:

$$RQ = RA + RTP = \left(6 + \frac{N_{RTP} \times FD}{NPI}\right)\beta \qquad (4.12)$$

Under the above circumstances, the maximum throughput (Th) of the proposed MAC can be estimated using the following equation:

$$Th = \frac{FD - RQ}{FD} = \frac{FD - \left(6 + \frac{N_{RTP} \times FD}{NPI}\right)\beta}{FD} = 1 - \left(\frac{6}{FD} + \frac{N_{RTP}}{NPI}\right)\beta \qquad (4.13)$$

Substituting the values of FD, β, N_{RTP}, and NPI used in our implementation in Equation (4.13), we obtain a maximum throughput of 78%. The average packet size in our case is 253 Bytes. This was obtained using a weighted average of three types of packets considered (3G wireless, wireline voice, and wireline data traffic). A header of 24 Bytes (18 Bytes PDU header plus 6 Bytes MAC header) is appended to each packet. This decreases the maximum effective throughput bound of the proposed MAC to 68%. This is an excellent figure for data only throughput if we consider the contention oriented nature of the upstream channel, and the QoS guarantees provided by the proposed MAC.

4.11 Performance Evaluation

In this section, we conduct a simulation study for evaluating the performance of our proposed MAC protocol. We compare it with an implementation of the DOCSIS MAC of our own, as well as an implementation by Broadcom. Due to the complexity of the system under investigation, and the large number of network parameters, most of our performance evaluation is

conducted using simulation analysis. We use OPNET simulation models for our studies. We build our simulation models based on CableLabs Opnet Common Simulation Framework.

We simulate a scenario where 149 CMs share a single 5 Mbps upstream channel. Among these CMs, there is a single CM with a 3G Node B, 74 CMs with real-time G.729 real-time voice traffic with silence deletion (8 Kbps), and 74 CMs with non-real-time data traffic (64 Kbps). A single Node B here aggregates traffic from 75 wireless voice connections, and 50 wireless data connection. The rates for wireless voice and data sources are 8 Kbps (G.729 with silence deletion), and 9.6 Kbps, respectively. These numbers are based on our traffic study presented in Section 3.2. Table 4.1 lists simulation configuration.

In our simulations, a G.729 with silence deletion voice source is modeled as an on-off source as modeled by Brady [21],[22]. A two-state Markovian model is generally used in the literature [49],[14],[71]. The silent and talkspurt periods are independent and distributed exponentially with means of 1.5 Secs and 2.25 Secs, respectively, for a voice activity cycle of 40%.

The upstream bandwidth in a single channel (5 Mbps) is divided into fixed duration 2 ms periodic time frames. A time frame is further divided into 78 time mini-slots of 16 Bytes each (0.0256 msec). A single mini-slot can be used to transmit a request, whereas an integral number of mini-slots is used to transmit one or more data packets.

The available mini-slots in an upstream time frame are assigned to different CMs SIDs (a CM could have one or more SIDs, one for each flow). Mini-slot assignment is performed by the MAC scheduler. In this study we use a priority based MAC scheduler, which gives a prefer-

Table 4.1: Simulation configuration

Parameter	Value	Comments
Upstream Channel Bandwidth	5 Mbps	
Number of CMs	149	
Number of 3G Node B CMs	1	
Number of wireline voice CMs	74	
Number of wireline data CMs	74	
Upstream MAC frame duration	2 ms	This must be distinguished from a DOCSIS MAC frame which consists of a MAC header appended to data packet.
Number of mini-slots in a frame	78	
Mini-slot duration	0.0256	
Number of contention mini-slots	6+	6 contention mini-slots are assigned in each frame. Any unused mini-slots in a frame are assigned for contention.

ence to control plane messages (e.g., requests, synchronization) in assigning upstream mini-slots. The remaining mini-slots are assigned to CMs data according to their priority. Any unassigned mini-slots in the upstream time frame are assigned for contention. Refer to Chapter 5 for details about the MAC scheduler.

We classify traffic into two priority classes with class 0 as the highest priority. Class 0 is allocated to wireless 3G Node B, as well as wireline voice traffic. Class 1 is for wireline data traffic. We show simulation results for throughput and access delay for each traffic priority class. The throughput is calculated as the number of bits per priority class received over the

simulation time, while the access delay is calculated as the elapsed time from the point when a packet belonging to a certain priority class is received at a CM until the time it leaves the DOCSIS CMTS.

We conduct a number of simulation experiments for studying the performance of our proposed MAC enhancements. We compare the performance of our proposed MAC to other existing DOCSIS MAC implementations. In the following sections, we explain the conducted simulation experiments, and discuss simulation results.

4.11.1 Dedicated DOCSIS Channel for 3G Traffic

This is a simple experiment where a complete DOCSIS upstream channel is dedicated for 3G wireless traffic. Wireline traffic is carried in different upstream channels rather multiplexed with the 3G traffic. Upstream data, and signalling traffic is framed at a 3G Node B using UMTS protocols. These frames are inserted into IP packets at a DOCSIS CM. IP packets available at the time of sending a request are concatenated in one MAC frame and have a concatenation header appended at the CM, as explained in Section 4.8. Our proposed DOCSIS MAC presented earlier in this chapter is used to transport concatenated 3G IP packet to 3G RNCs, through the DOCSIS CMTS.

Figure 4.6 shows the access delay for 3G wireless traffic in the upstream. The average access delay is also shown in the same figure. It is clear that an average access delay of about 6 ms is well below the delay bound requirement of 3G traffic (20 ms). This is desirable since 3G traffic is expected to experience additional access delay at the UMTS wireless channels. Note

that delay bound at the UMTS segments of the network are guaranteed by 3GPP QoS

mechanisms discussed in Section 2.2.2.
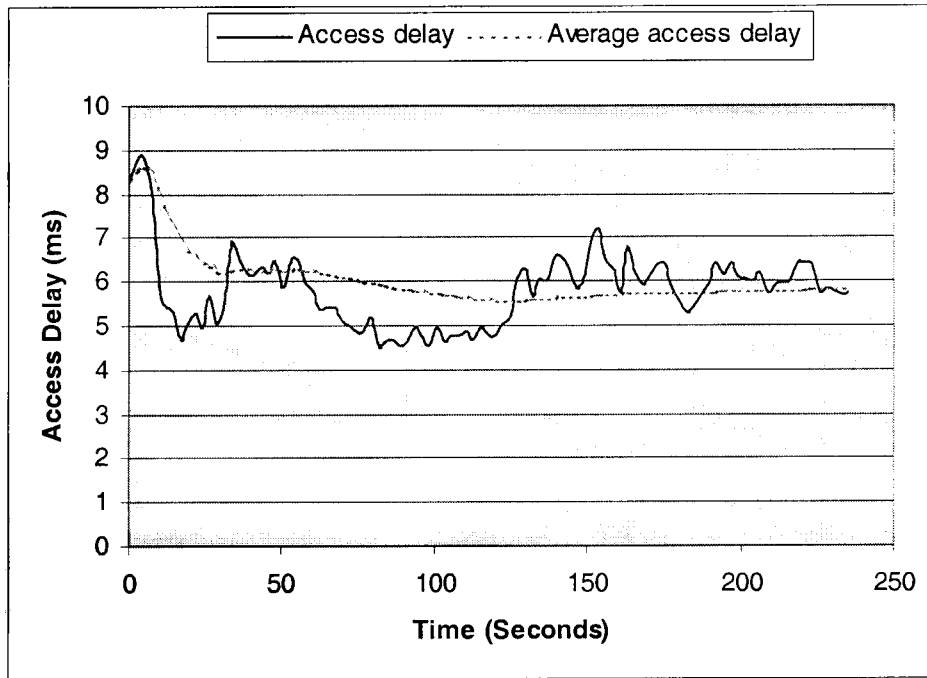
Figure 4.6 Access delay for 3G traffic using a dedicated DOCSIS channel

Figure 4.7 shows the average throughput for 3G wireless traffic. It shows 3G voice, data, and

total average throughput. As seen from this simulation experiment, our proposed MAC can

guarantee throughput requirement for 3G traffic in the DOCSIS upstream channel. However,

dedicating a whole upstream DOCSIS channel for 3G traffic is inefficient and wastes the

precious upstream bandwidth. Multiplexing 3G wireless traffic with other wireline traffic is a

desirable characteristic for achieving statistical multiplexing. In the following experiments,

we evaluate the performance of different implementations of DOCSIS MAC when 3G

wireless traffic is multiplexed with other wireline traffic in the same DOCSIS upstream
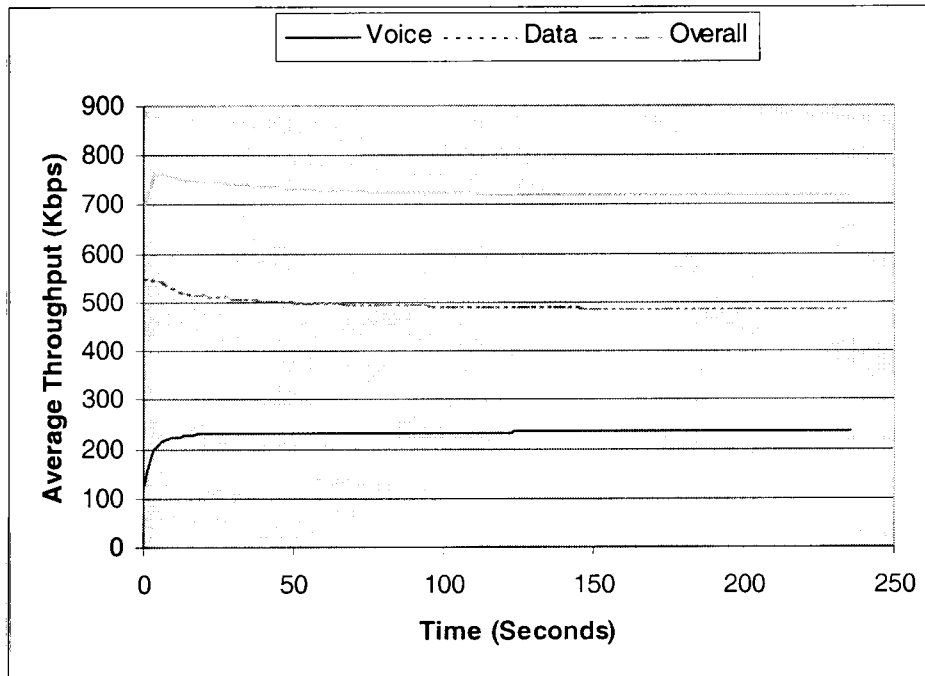
channel.

Figure 4.7 Average throughput for 3G traffic using a dedicated DOCSIS channel

4.11.2 Proposed DOCSIS MAC Enhancements

In this experiment, we evaluate the performance of our proposed DOCSIS MAC enhancements presented earlier in this chapter. In particular we focus on our proposed DOCSIS rt-PS mechanism enhancement for QoS guarantees for real-time traffic. In summary, a flow for delay-sensitive traffic establishes a service flow and get assigned a SID using DOCSIS defined procedures [27]. These SIDs are polled by the CMTS periodically. We choose a Nominal Polling Interval of 16 ms. This is to ensure a worst case access delay of less than 20 ms. In the worst case, a packet could arrive right after the SID has been polled. This SID has to wait for close to 16 ms before it is able to send a request using a unicast poll mini-slot. we

ensure that the SID is given preference in processing at the CMTS. The priority based MAC scheduler gives a high priority to delay sensitive traffic in upstream bandwidth assignment. The SID gets assigned the required upstream mini-slots, and gets informed of the assignment using a MAP message. The assigned mini-slots are used to transmit the upstream packets. We ensure that, in the worst case, the elapsed time from the point a packet arrives at a CM to the time the packet departs the CMTS is always less than the upper access delay bound of delay-sensitive traffic (20 ms).

At the time of packet transmission, the CM can piggyback requests for additional bandwidth if it has received additional packet since it transmitted its last request. The CMTS relies on these piggybacked requests to allocate more upstream mini-slots for SIDs. The CMTS is permitted to assign a unicast poll mini-slot for a SID only if the elapsed time since the last request was received (using either unicast poll mini-slot, or piggybacked) is close to 16 ms. Periodic polls every 16 ms continue until the CM uses the unicast poll mini-slot to send either a request for bandwidth when it becomes active/ON again or an explicit request to end the session.

In this experiment, we multiplex 3G wireless traffic with other wireline delay-sensitive, and best-effort traffic in the same DOCSIS upstream channel. We increase the offered load gradually. We start by having only one 3G Node active, and then we gradually increase the number of active wireline CMs until the maximum number of 149 CMs activated is reached. 3G wireless traffic (voice, and data), as well as wireline delay sensitive traffic is assigned priority 0. Wireline best-effort data traffic is assigned a priority of 1.
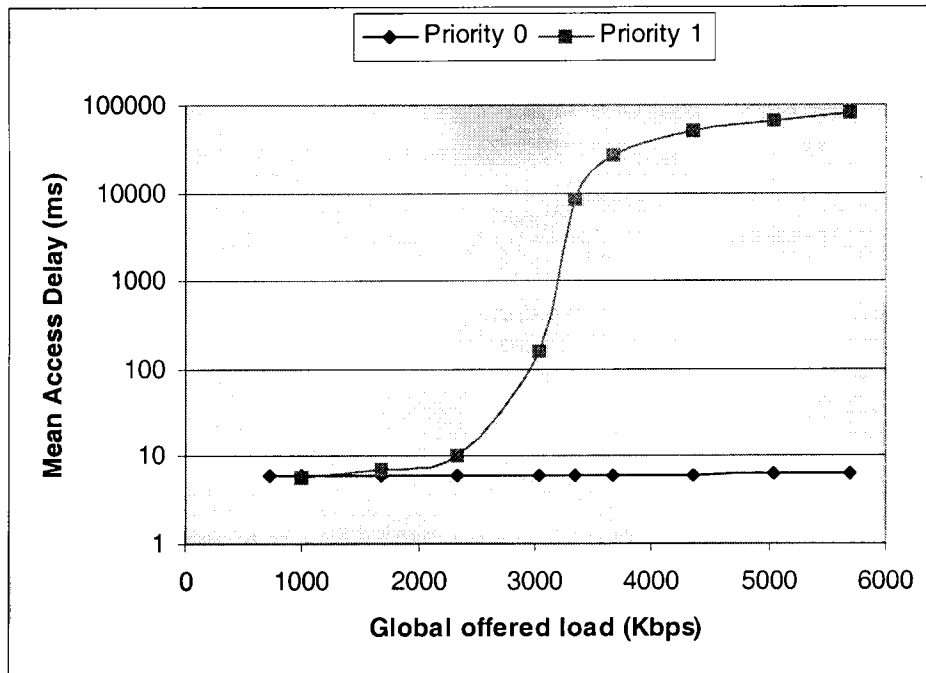
Figure 4.8 Average access delay using proposed MAC

Figure 4.8 shows the average access delay for priority 0, and priority 1 traffic classes. It is

obvious that 3G wireless traffic, and wireline delay-sensitive traffic experience guaranteed

delay bounds (about 6 ms on average) in the upstream despite the volume of the offered traffic

in the network. Figure 4.8 also shows that wireline best-effort traffic also enjoys low delay

bounds under low or moderate loading conditions. It increases exponentially as the offered

load exceeds the maximum capacity of the upstream channel.

Figure 4.9 shows the average throughput for priority 0, priority 1, as well total throughput.

Our proposed MAC offers a maximum channel efficiency of about 64%. We define channel

efficiency here as the ratio of the maximum data throughput of the channel over the channel

capacity. As seen from Figure 4.9, both priority classes experience throughput guarantees

given that the offered load is within the maximum channel efficiency. Due to the preference given to priority class 0, its throughput can still be guaranteed even if the maximum channel efficiency is exceeded because of higher offered loads. On the other hand, the throughput of priority class 1 (best-effort) start dropping as the maximum channel efficiency is exceed, which is an expected behavior.
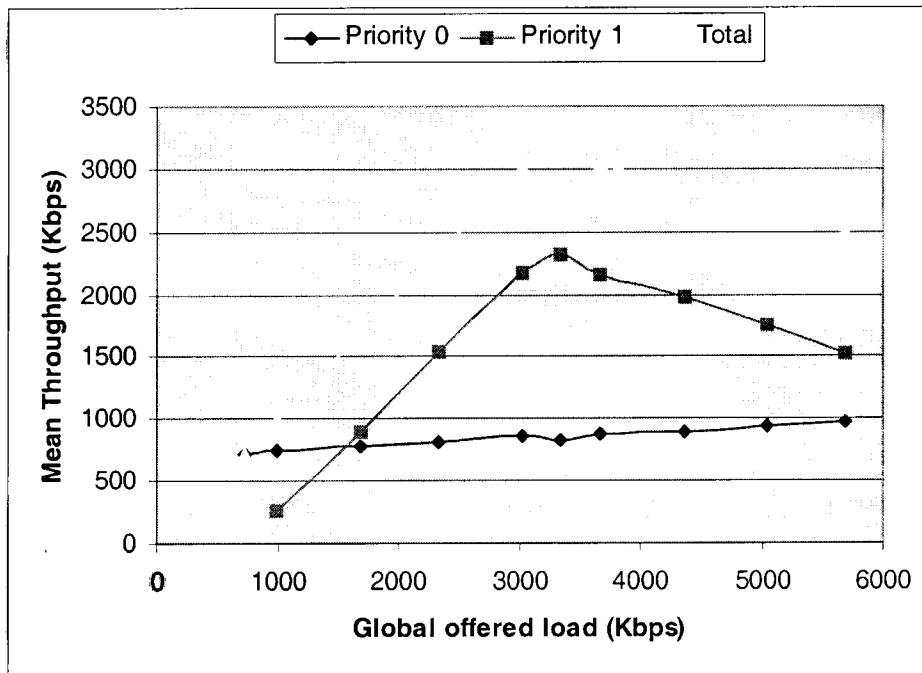
Figure 4.9 Average throughput using proposed MAC

4.11.3 DOCSIS Specifications MAC

In this experiment, we present and evaluate the performance of an implementation of DOCSIS MAC of our own. Our focus is on a specific implementation of DOCSIS rt-PS mechanisms for QoS guarantees for real-time traffic. In summary, this implementation follows DOCSIS specifications by restricting real-time traffic to use unicast poll intervals only for requesting

upstream bandwidth. Again, a service flow is established and a SID is assigned for a delay-sensitive traffic using DOCSIS defined procedures [27]. The CMTS periodically polls rt-PS SIDs in time intervals equal to the Nominal Polling Interval (16 ms in our case). These unicast, periodic polls continue for the duration of the connection. SIDs with queued traffic use the assigned unicast poll mini-slots to send requests for bandwidth. SIDs with no traffic to send skip their assigned unicast mini-slots to indicate that they have no data to send. The CMTS assigns the requested mini-slots and informs SIDs as discussed in our proposed MAC (Section 4.11.2). CMs use the assigned upstream mini-slots to transmit their data traffic, with no piggybacking allowed.

Figure 4.10 shows the average access delay for priority 0, and priority 1 traffic classes. Our implementation of DOCSIS MAC can offer delay guarantees of both priority classes. However, average access delay for priority class 0 is higher than that for our proposed MAC enmeshments as will be discussed in Section 4.12. Figure 4.10 also shows that at very low loading conditions priority class1 can experience a bit lower access delay than priority class 0. This is due to the fact that while priority class 1 could have immediate access to the channel by contending with other flows, priority 0 traffic have always to wait for an assigned unicast poll mini-slot in order to send a request. At low loads, priority class 1 has a high probability of success in sending requests using contention. As the load increases, more collisions are expected which result in higher access delay.

Figure 4.11 shows the average throughput for priority 0, priority 1, as well as the total
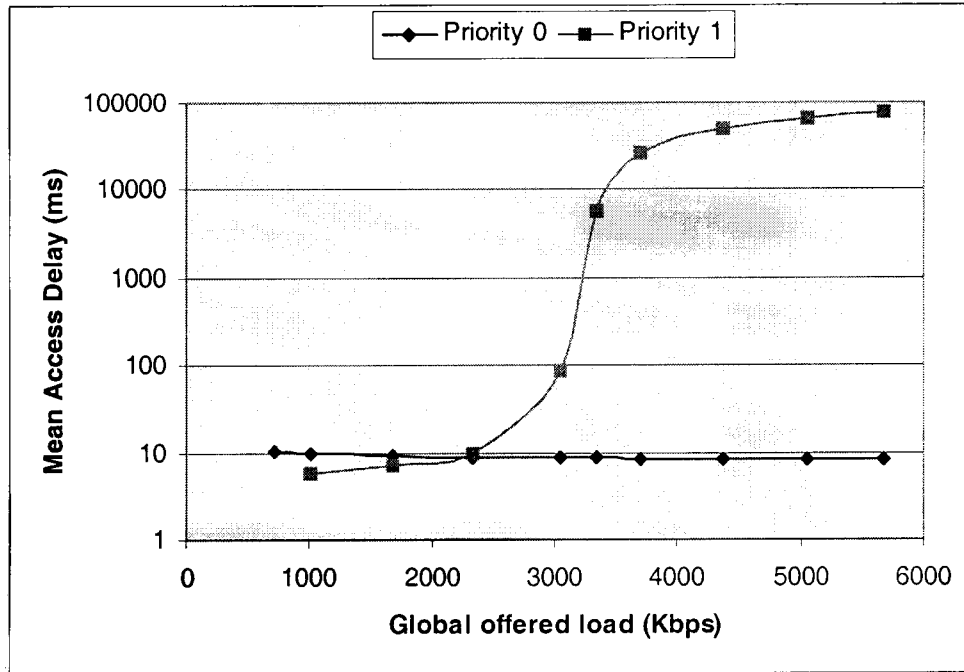
Figure 4.10 Average access delay using DOCSIS MAC

throughput. Our DOCSIS MAC implementation offers similar throughput for both classes to

that offered by our DOCSIS MAC enhancements discussed in Section 4.11.2.

4.11.4 Broadcom DOCSIS MAC Implementation

In this experiment, we evaluate the performance of another implementation of DOCSIS rt-PS

mechanism by Broadcom Corporation [23]. It has many similarities with our DOCSIS

implementation discussed in Section 4.11.3. However, it differs in the manner CMs are polled.

The CMTS polls CMs using the rt-PS as follows:

(1) The CMTS polls CMs every t ms (where t = Nominal Polling Interval = 16 ms) until a re-

quest is received.

100
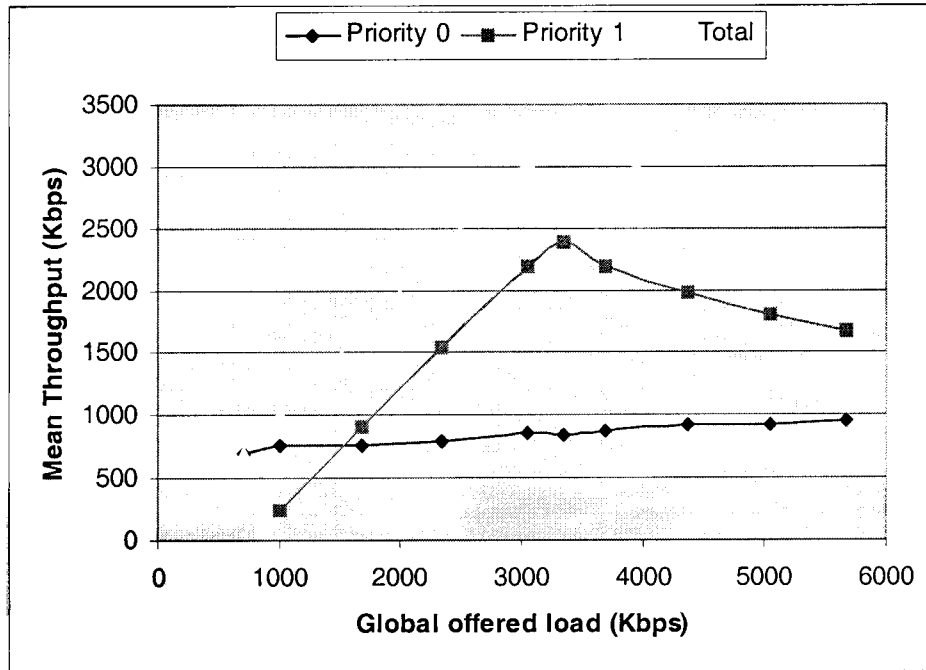
Figure 4.11 Average throughput using DOCSIS MAC

(2) Right after receiving a request from the CM, The CMTS Polls the CM every time frame

(2 ms in our case) until a second request is received, or N polls have been issued. Where

$$N = \frac{NominalPollInterval \times 2}{FrameTime}$$

(3) If a second request is received, or if no second request is received after N polls, poll as in

step (1).

Figure 4.12 shows the average access delay for priority 0, and priority 1 traffic classes. The

average throughput for priority 0, priority 1, and the total throughput is shown in Figure 4.13.

The performance of Broadcom's DOCSIS MAC implementation is comparable to that of our

proposed DOCSIS implementation discussed in Section 4.11.3. This is regardless of the

complexity of Broadcom's implementation compared to the simplicity of ours.
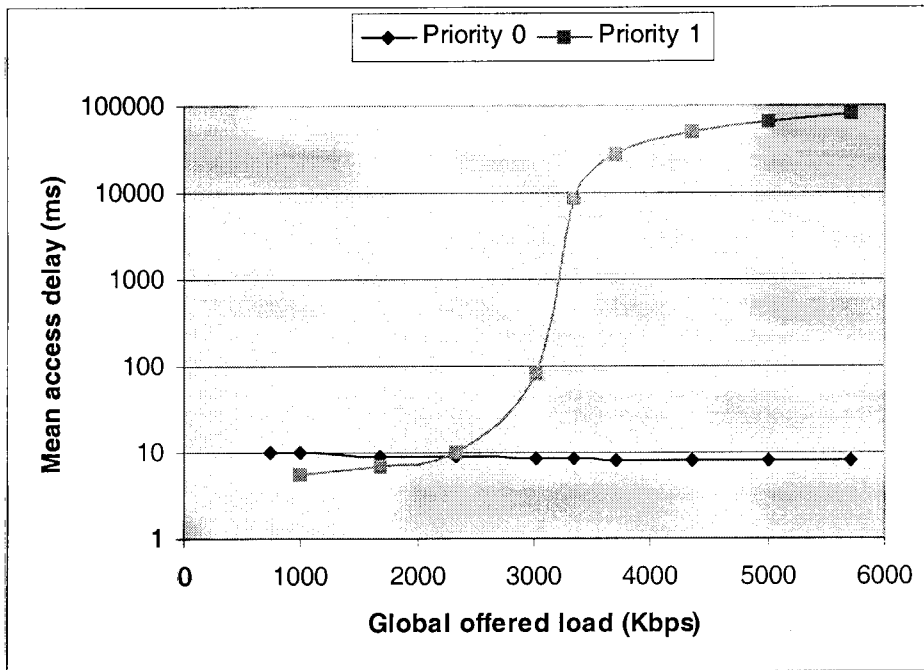
Figure 4.12 Average access delay using Broadcom DOCSIS MAC implementation

4.11.5 Fragmentation Capability

As discussed in Section 4.9, packet fragmentation is an essential feature of any DOCSIS MAC implementation, in order to be able to support 3G traffic. Packets received from a 3G Node B are concatenated at the interfaced CM for transmission upstream. Packets can buildup at the CM till the length of the concatenated packet can exceed the upstream MAC frame duration of 2 ms. A Concatenated packet from a 3G Node B may never get the chance to be transmitted upstream. Instead, it will be waiting at the head of the CM's queue, blocking other packets. Packet fragmentation can solve this problem by segmenting long packets and allocating partial grants for individual segments to be transmitted upstream, as discussed in Section

102
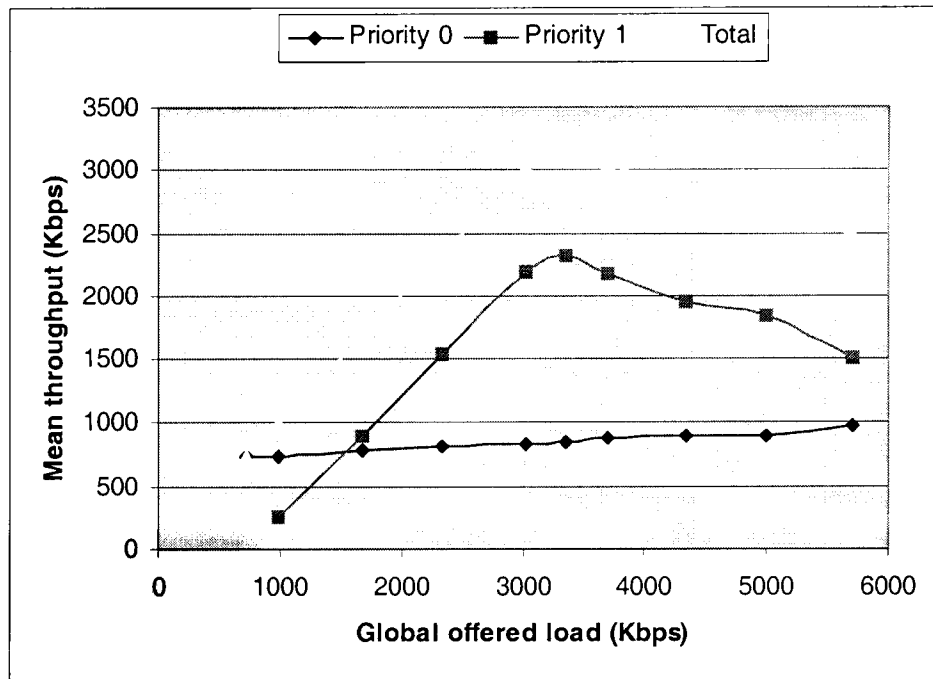
Figure 4.13 Average throughput using Broadcom DOCSIS MAC implementation
4.9.

In this experiment, we show by simulation results that fragmentation is a MUST in any

DOCSIS implementation, in order to support 3G traffic over DOCSIS. The average through-

put for priority 0, priority 1, and the total throughput is shown in Figure 4.14. As in previous

experiments, we start our simulation by having only a 3G Node B using a whole DOCSIS

upstream channel, and we increase the load gradually by having more wireline CMs active.

As can been seen from Figure 4.14, no traffic out of the 720 Kbps offered by the 3G Node B

goes through. All 3G packets wait at the CM until dropped. Priority 0 packets from other

wireline voice sources can be transmitted upstream because of their shorter packet size. This

is also evident from Figure 4.14 when priority 0 traffic throughput increases as more wireline

CM become active.
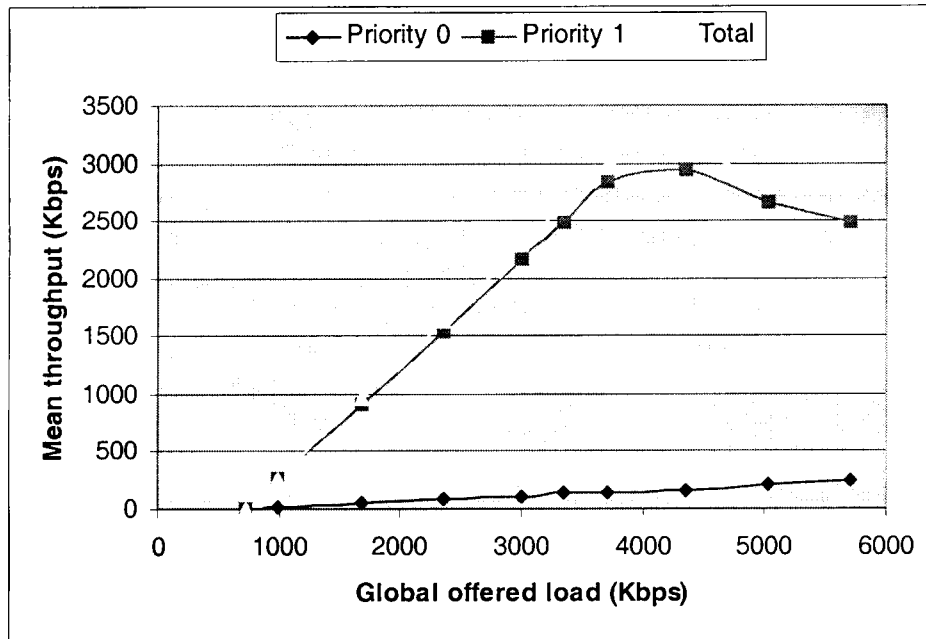
Figure 4.14 Average throughput without DOCSIS MAC fragmentation capability

The fact that 3G traffic can not be allocated bandwidth and transmitted upstream decreases the

loading conditions and gives more opportunity for other wireline traffic. It is expected that

this fact will result in low average access delay for wireline traffic. This is evident from Figure

4.14 which shows the access delay for both priority classes is lower than what is offered by

other DOCSIS MAC implementations presented earlier. This behavior is certainly not

desirable and should be avoided.

4.12 Summary

An efficient implementation of DOCSIS MAC protocol is a critical issue for supporting 3G
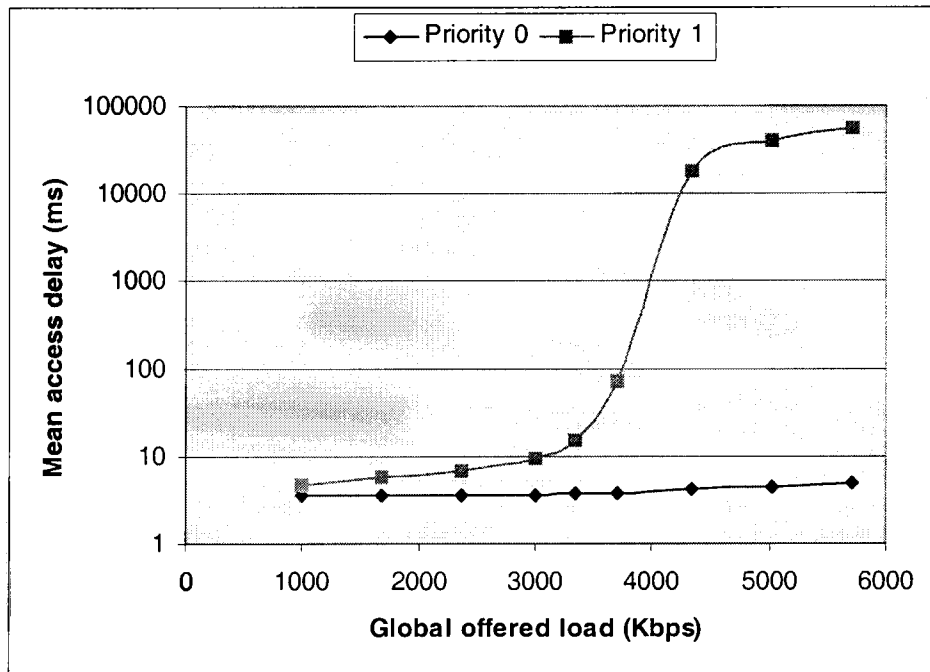
Figure 4.15 Average access delay without DOCSIS MAC fragmentation capability

wireless traffic over DOCSIS HFC CATV networks. 3G wireless traffic requires very

stringent delay, and throughput guarantees when transported over the CATV plant. DOCSIS

MAC standard specifies a framework for protocols and rules that ensure interoperability of

different CMTSs and CMs implementation. It does not impose any specific implementation.

Rather, it offers flexibility by leaving many of the MAC features to be implementation

dependent. Different vendors can have diverse interoperable DOCSIS implementations by

adhering to DOCSIS framework.

In this chapter, we demonstrated how DOCSIS MAC protocol can be used to transport 3G

wireless traffic over the CATV plant. We presented an efficient DOCSIS MAC implementa-

tion. Critical features of DOCSIS essential for the proposed 3G deployment were discussed.

We proposed enhancements to DOCSIS MAC protocol for efficient support of 3G wireless traffic over DOCSIS networks. In particular, we enhanced the DOCSIS rt-PS. The proposed enhancements can be used for 3G wireless traffic as well as wireline real-time traffic.

We conducted a study for evaluating the performance of our proposed MAC enhancements. Two other DOCSIS MAC implementations were presented and their performance was studied. Specifically, we evaluated the performance of an implementation of the existing rt-PS of our own, as well as another implementation by Broadcom. Also, the effect of packet fragmentation capability of DOCSIS MAC was studied.

Performance results show that our proposed MAC enhancements offer the lowest average delay for delay-sensitive traffic (see Figure 4.15). Depending on the loading of the DOCSIS upstream channel, our MAC enhancements can provide between 30% to 40% reduction over existing DOCSIS MAC rt-PS. This is a much desirable performance requirement for 3G wireless traffic. The reductions of access delay by our proposed MAC enhancement is achieved without compromising the delay and throughput performance requirement of other traffic classes. The total throughput of our proposed MAC is also comparable to existing DOCSIS MAC, and Broadcom DOCSIS implementation. Appendix E shows additional performance results that compare the performance of our proposed MAC to other DOCSIS MAC implementations.

Despite the adequacy of our proposed DOCSIS MAC enhancements for supporting 3G wireless traffic over DOCSIS based HFC networks, improvement is still needed in the area of
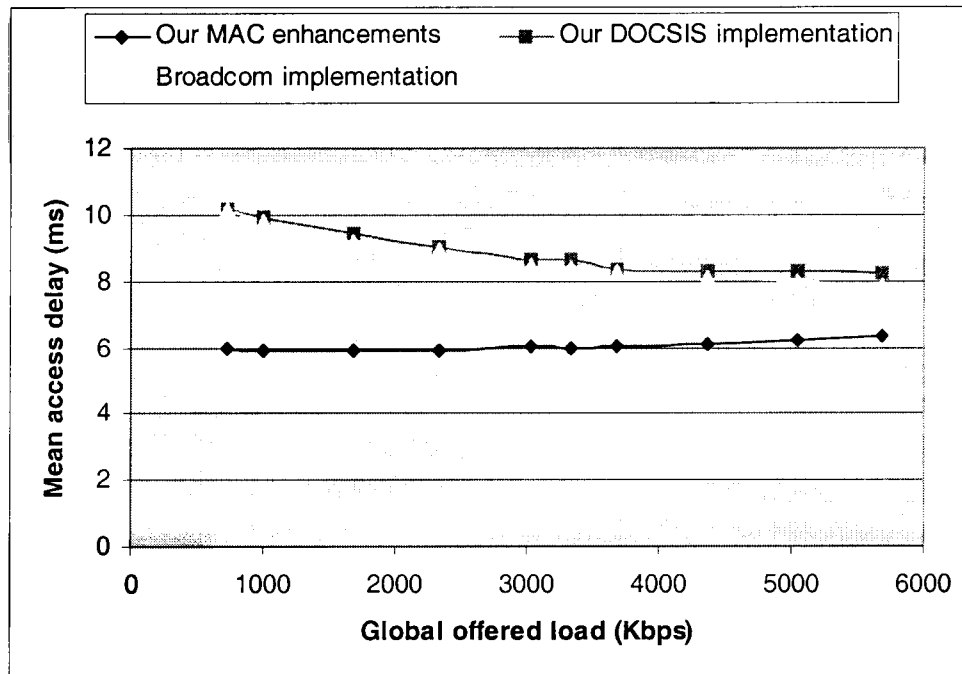
Figure 4.16 Access delay for delay-sensitive traffic using different MAC implementations

MAC scheduling. As will be shown in Chapter 5, using a priority based scheduler for allocating upstream bandwidth is favorable to high priority classes, but can be disastrous for lower priority classes. Fairness is one of the major issues in this regard. This leads us to propose a Weighted Fair Queueing based MAC scheduler, which is the subject of Chapter 5.

107

Chapter 5

Proposed DOCSIS MAC Scheduler

5.1 Introduction

This chapter proposes a superior Weighted Fair Queueing based MAC scheduler for DOCSIS

networks. The major advantage of using a fair scheduling discipline like WFQ in the MAC

layer is that the MAC can inherit the delay guarantees, and fair bandwidth allocation features

of network layer fair schedulers. The main function of the proposed scheduler is to guarantee

delay bounds, and to provide fair share of the bandwidth for different traffic flows. In

addition, it resolves any Denial of Service issues. The proposed MAC scheduler is essential

for the proposed wireless data systems deployment over CATV networks, as well as for

existing wireline users. QoS can never be assured using any sophisticated QoS mechanisms at

the upper layers unless the MAC has QoS guarantees capabilities.

A key element in DOCSIS MAC is the MAC scheduler. The MAC scheduler is located at the CMTS and takes care of upstream bandwidth allocation. Upon receiving requests for bandwidth from CMs, the MAC scheduler assigns mini-slots for different flows (contention, unsolicited grants, real-time polls, etc.). It constructs a MAP message that informs CMs of their assignments, and broadcasts this MAP message to all CMs downstream. Upon receiving a MAP message, a CM sends its data or control packets in its assigned mini-slots or contention mini-slots as defined in DOCSIS MAC protocol.

The DOCSIS standard does not mandate any specific scheduler. Instead, it specifies the protocols for requesting and granting bandwidth, and leaves the specification of candidate schedulers to be defined by implementors.

A few researchers have addressed scheduling at the control plane of HFC networks' MAC layer [39][47][53][83][97]. Most of those proposals do not target DOCSIS 1.1 QoS requirements, as specified in DOCSIS 1.1 MAC standard [26]. Note that through out this thesis, we use the term DOCSIS to refer to DOCSIS release1.1 which introduces QoS support. DOCSIS specifies a number of QoS mechanisms such as Service Flow scheduling services that need to be addressed by any potential scheduler. To the best of our knowledge, the only two proposals that try to address DOCSIS upstream QoS requirements are reported in [39], [53]. However, the work in [39] can not bee considered as a complete DOCSIS scheduling solution since it focuses only on two of DOCSIS upstream Service Flows: UGS, and BE. It does not address rt-PS, nrt-PS, or UGS-AD. The work reported in [53] addresses different DOCSIS service

flows without presenting any performance results. However, the way it schedules rt-PS is problematic. It suggest that requests for some of the rt-PS flows are aggregated in a priority queue with other nrt-PS, and BE flows. We believe that using a priority based discipline can jeopardize QoS for some of the aggregated flows. A misbehaving high priority flow can degrade the QoS offered to lower priority flows.

The targeted MAC scheduler should guarantee QoS for different flows with diverse QoS requirements. In order to accomplished that, the targeted scheduler must satisfy the following requirements [59]:

- Fairness and protection among different flows

- Performance bounds (e.g., delay, throughput)

- Ease of implementation

- Compliance with DOCSIS standard

Recognizing the above requirements, we propose to deploy a Weighted Fair Queuing (WFQ) based scheduler at the MAC layer. The guaranteed delay bounds, fairness, and flow isolation properties have established WFQ as a superior work-conserving scheduler in the network layer of packet networks [80][98]. However, our proposed WFQ MAC scheduler is fundamentally different from traditional network layer WFQ schedulers. The former schedules packets at the MAC layer based on requests for upstream bandwidth. The later

schedules the actual packets at the network layer. The end result of scheduling requests using our MAC scheduler is that the actual data packet are scheduled. In this chapter, we discuss the architecture, functionalities, and performance of the proposed MAC scheduler.

5.2 Background

WFQ, or Packet-by-packet Generalized Processor Sharing (PGPS), is an implementation of GPS scheduling that handles variable size packets instead of GPS's ideal infinitesimal size packet. WFQ transmits packets according to their finish times under GPS. In WFQ, a GPS system is simulated in parallel to the actual packet based system in order to identify the set of backlogged sessions at each instant of time and their service rates. Based on this information, a time stamp is calculated for each arriving packet, and the packets are inserted into a priority queue based on their time stamp value and transmitted in order of increasing timestamps. The time stamp specifies the finish time of the packet as if it was transmitted under GPS. For background information about WFQ refer to [59][80][98]. Our proposed WFQ based MAC scheduler is explained in Section 5.3.

5.3 Proposed WFQ MAC Scheduler

Figure 5.1 shows the architecture of the proposed WFQ based MAC scheduler for DOCSIS networks. The following sections detail the functionalities of different elements of the proposed MAC scheduler.
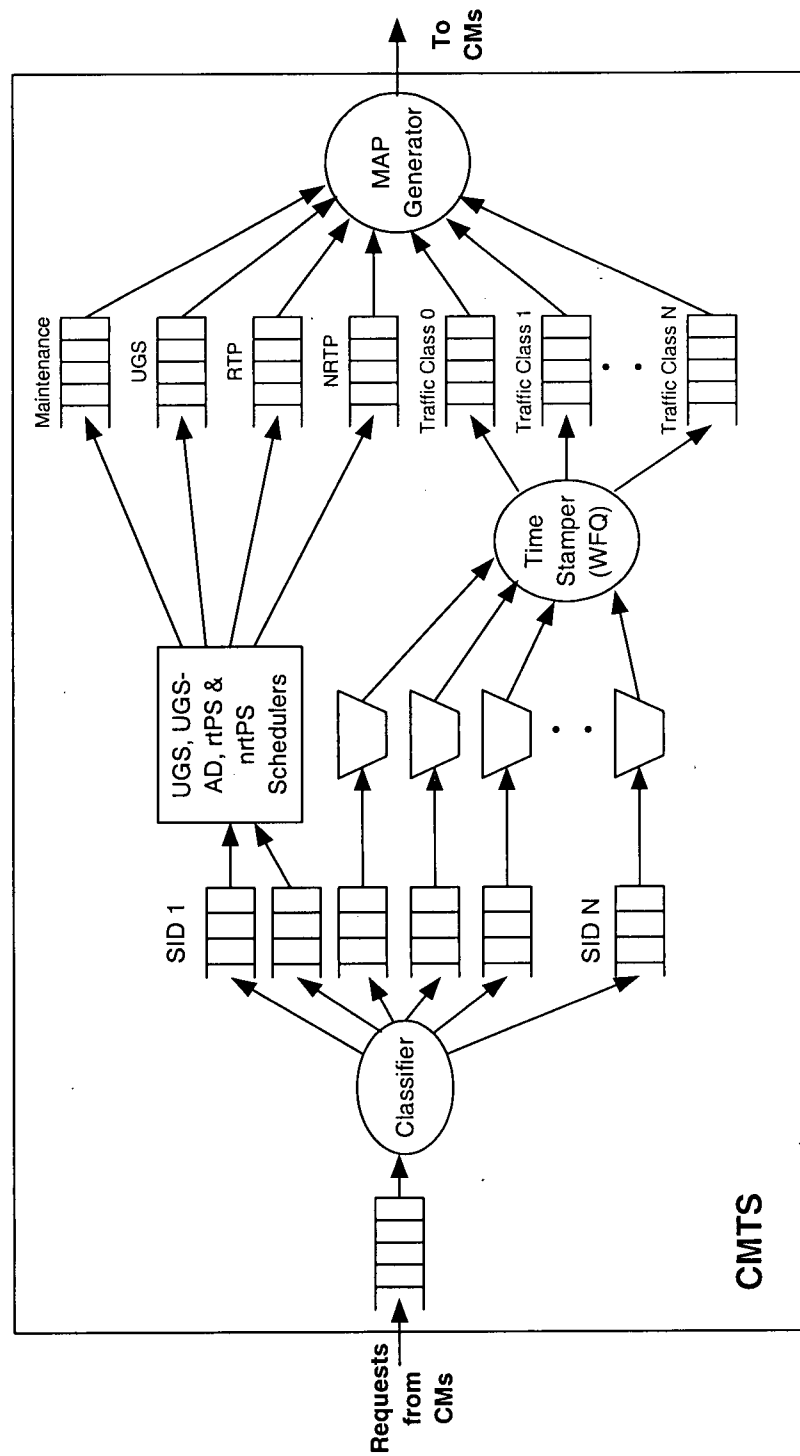
Figure 5.1 WFQ MAC scheduler - upstream

112

5.3.1 Classifier

This function classifies data packets, and control plane messages into different Service Flows based on a set of matching criteria. Classifiers are deployed in the upstream, and downstream paths of the CMs, and the CMTS. In the transmission plane, packets can be classified into one of the SIDs according to one or more of the following matching criterion [27].

- Service Flow Identifier of a specific flow to which this packet belongs.

- IP Classification Parameters including IP TOS, IP Source/Destination addresses, and TCP/UDP Source Port Start/End.

- LLC Classification Parameters that include Destination and Source MAC Address.

- IEEE 802.1P/Q Parameters such as 802.1P Priority Range.

In the control plane, MAC messages such as requests are classified into the Primary SID in both the upstream and downstream. Our focus in this section is on upstream control messages classification at the CMTS. In particular, we address classification of bandwidth request at the CMTS as it is applicable to our MAC scheduler.

Each upstream bandwidth request is classified by the CMTS classifier into one of the Service Flows. These requests can be either standalone or piggybacked to data packets. This classification is based on the SID field in the request MAC header. Classified requests are delivered to the appropriate Service Flow queues. Depending on the upstream Service Flow Scheduling

113

Service to which the SID was admitted, these queued requests can take one of two paths. Requests for UGS, and UGS-AD services are delivered to the Unsolicited Grants and Polls Module (see Section 5.3.2). On the other hand, request for rt-PS, nrt-PS, and BE services are shaped before being stamped by the WFQ time stamper, as discussed in Section 5.3.4.

5.3.2 Unsolicited Grants and Polls Module

DOCSIS specifies a number of upstream Service Flow scheduling services for improving the efficiency of the poll/grant process. Each service is tailored to a specific type of data flow. The basic services comprise: UGS, rt-PS, UGS-AD, nrt-PS, and BE service. Refer to Appendix B for background information about these services.

The Unsolicited Grants and Polls Module is responsible for generating requests for unsolicited grants and polls for these DOCSIS upstream scheduling services, as explained below. This module generates requests for unsolicited data grants and polling opportunities for sending requests, and feeds them to the MAP Generator. The MAP Generator services these requests in a priority manner, as explained Section 5.3.5. The following explains how the Unsolicited Grants and Polls Module schedules requests for different DOCSIS upstream Service Flows.

5.3.2.1 Unsolicited Grant Service (UGS)

The Unsolicited Grant Service (UGS) supports real-time Constant Bit Rate flows such as Voice over IP (VOIP). It provides fixed size data grants on a real-time periodic basis. The Unsolicited Grants and Polls Module generates requests for SID which are due to be granted

upstream bandwidth. These requests are fed to the MAP generator for data grant allocation. A generic shaper is used for generating periodic requests for SIDs admitted to use the UGS. Requests for UGS are periodically generated every t ms. The size of the requested grant, and the elapsed time between grants are controlled by the Unsolicited Grant Size, and Nominal Grant Interval parameters, respectively. Requests for additional grants for a SID can be generated based on the Queue Indicator (QI) bit of the Unsolicited Grant Synchronization Header (UGSH) (refer to [27] for information about the UGSH).

5.3.2.2 Real-Time Polling Service

The Real-Time Polling Service (rt-PS) supports real-time variable Bit Rate flows. It is aimed at supporting applications that generate variable size packets on a periodic basis. The MAC scheduler provides real-time, periodic, unicast request opportunities to CMs using this service.

The Unsolicited Grants and Polls Module generates requests for SID which are due to be polled for sending requests. The reader should distinguish between allocating unicast poll mini-slots, and allocating data grants for SIDs using the rt-PS. Unicat poll mini-slots are periodically scheduled by this module, while data grants are scheduled based on received upstream request for bandwidth. The requests for rt-PS generated by this module are fed to the MAP generator for unicast request mini-slots assignment. The RTP shaper is used for generating periodic requests for SIDs admitted to use the rt-PS. The frequency of the real-time polls is determined by the Nominal Polling Interval parameter.

5.3.2.3 Unsolicited Grant Service with Activity Detection

The Unsolicited Grant Service with Activity Detection (UGS-AD) supports flows with ON/OFF Constant Bit Rate traffic such as Voice over IP with silence suppression. USG/AD offers real-time unsolicited grants during ON periods and real-time unicast polls during OFF periods.

The Unsolicited Grants and Polls Module generates requests for data grants during active (ON) period, and switches to generating requests for unicast polls during silent (OFF) period. A combination of the mechanisms discussed in Section 5.3.2.1, and Section 5.3.2.2 are used for serving the UGS-AD. Parameters such as the Nominal Grant interval, the Unsolicited Grant Size are used to shape requests the real-time polls, and data grants.

5.3.2.4 Non-Real-Time Polling Service

The Non-Real-Time Polling Service (nrt-PS) supports non real-time variable Bit Rate flows such as FTP applications. The MAC scheduler provides periodic or non-periodic (typically in the order of a second or less) unicast request opportunities to SIDs using this service. A shaper similar to the one used for rt-PS (see Section 5.3.2.2) can be used for serving nrt-PS. The shaper generates requests for unicast polls according to the Nominal Polling Interval parameter.

5.3.3 Traffic Shapers

It has been proven that WFQ can guarantee delay bounds if the traffic is Token Bucket shaped [80]. If arrivals from flow i to a WFQ scheduler are shaped using a token bucket shaper with

116

rate of r_i and a token bucket depth of σ_i, the maximum delay experienced by a packet flow i in a WFQ scheduler is bound and equals

$$\frac{\sigma_i}{r_i} + \frac{L_i}{r_i} + \frac{L_{max}}{C} \tag{5.1}$$

Where: L_i is maximum packet size of session i, and L_{max} is the maximum packet size among all flows sharing the link.

We propose to deploy traffic shapers at CMs. These shapers ensure that traffic flows are shaped at CMs queues before requesting bandwidth from the CMTS. Alternatively, we propose to shape the requests for bandwidth at the CMTS using a token bucket shaper of parameters of r_i and σ_i. Our shaper is fundamentally different from traditional network traffic shapers. In traditional network traffic shapers the data packets are shaped, where as in our case the requests for bandwidth are shaped at the CMTS while the actual data packets are held in the CMs queues. The end result is that the actual data packet are shaped.

Our proposed CMTS shaper works as follows:

- The guaranteed link share (r_i) and maximum burst size (σ_i) of a flow are negotiated during Call Admission Control (CAC).

- The shaper generates token at a rate (r_i) and holds the tokens in bucket of size (σ_i).

- Upon an arrival of a request, the shaper scans the request for the number of mini-slots

117

requested. This number is converted to bytes and reflects the size of the actual data packet held at the CM queue.

- If the shaper has enough tokens in its bucket equivalent to the request size, the request is forwarded to the WFQ time stamper where a grant will be scheduled for it. Otherwise, the request is held in the shaper input queue.

- Whenever the token bucket shaper accumulates enough tokens equivalent to the size of the request, the request is forwarded to the WFQ time stamper.

This has the effect of shaping packets for a given flow while the actual data packets are being held at CM queue. Flows which are not conforming to their negotiated traffic profiles are penalized by not processing their requests, and therefore, holding their packets at their CMs queues. Extra packets arriving to such flows cause the concerned CMs queues to overflow and drop packets.

5.3.4 WFQ Time Stamper

Shaped requests for bandwidth received from CMs are fed to a WFQ time stamper. The WFQ time stamper simulates a GPS scheduler in the background, and stamps the requests for bandwidth with a time stamp equals to their finish time as if they were transmitted under GPS, as follows.

$$F_i^k = \frac{L_i^k}{r_i} + max(F_i^{k-1}, V(t)) \tag{5.2}$$

Where F_i^k is the time stamp for k^{th} request of the ith flow which represents the finish time under GPS, and $V(t)$ is the virtual time at time t.

According to DOCSIS specification, at most one request is allowed to be outstanding at a time. In other words, request that belong to flow i will always arrive to an empty queue. Therefore, if per flow scheduling is used F_i^{k-1} is always less than $V(t)$, and the time stamp can be calculated as follows.

$$F_i^k = \frac{L_i^k}{r_i} + V(t) \tag{5.3}$$

The virtual $V(t)$ in GPS is a piecewise linear function of time whose value at the start of a busy period is equal to zero, and whose slope changes according to the following equation.

$$\frac{\partial V(t_j + T)}{\partial T} = \frac{\sum\limits_{j=1}^{N} \bar{\phi}_j}{\sum\limits_{i \in B(t)} \bar{\phi}_i} = \frac{1}{\sum\limits_{i \in B(t)} \bar{\phi}_i} \tag{5.4}$$

Where $\bar{\phi}_i$ is SID i normalized share of bandwidth, and B(t) is the set of sessions that are backlogged at time t. At the arrival of a request to SID queue, the Virtual time $V(t)$ can be calculated as follows.

$$V(t_{j-1} + T) = V(t_{j-1}) + \frac{T}{\sum\limits_{i \in B(t)} \bar{\phi}_i} \tag{5.5}$$

The stamped requests for bandwidth are inserted into a priority queue that is sorted based on their time stamp values and serviced in order of increasing time stamps. Serviced requests for bandwidth are fed to the MAP Generator which constructs downstream MAP messages. The MAP Generator is explained in section 5.3.5.

5.3.5 MAP Generator

The MAP Generator is responsible for generating MAPs to be broadcast downstream. A MAP is a collection of Information Elements (IEs) that describe the allocation of a number of upstream mini-slots (a frame) to different CMs. Figure 4.1 shows a typical upstream MAC structure.

MAP messages must be broadcast in time for it be received and processed by the furthest CM before the start of the frame described by it. This allows all CMs to scan these MAPs for transmission opportunities (contention mini-slots, data grants, and polls for requests), and use their allocated mini-slots described in the MAP. We adopt a MAP describing a fixed frame size of 2 ms, for simplicity and performance concerns. It was shown in [32] that a frame size of 2 ms with 6 contention mini-slots results in the best throughput and delay performance among all scenarios. The MAP Generator works as follows.

• The MAP Generator starts by allocating the needed upstream mini-slots for maintenance

purposes.

- Thereafter, it allocates the first 6 mini-slots for contention.

- Then it removes the scheduled requests from the head of the priority queue and allocates upstream mini-slots for the corresponding SID until either all the requests have been satisfied, or all the mini-slots in the upstream frame being scheduled are allocated.

- If there are more requests pending grants, unsatisfied request are scheduled in subsequent MAPs.

- There might be situations where the remaining free mini-slots in an upstream frame being described by a MAP are not sufficient for satisfying the request at the head of the priority queue. In that case, partial allocation for the request is accommodated using DOCSIS MAC fragmentation feature. Additional mini-slots to satisfy the rest of request are allocated in the next upstream frame.

- If all requests are satisfied while some mini-slots are still not granted, the remaining mini-slots are allocated for contention [41]. Allocating free upstream mini-slots in a frame for contention reduces the probability of collision and therefore improves throughput and delay performance.

5.4 Performance Study

In this section, we study the performance of the proposed WFQ based MAC scheduler.

Providing throughput and delay bounds are among the main features of the proposed MAC scheduler. First, we derive throughput and delay bounds analytically in Section 5.4.1, and Section 5.4.2, respectively. Then, we conduct a simulation study to evaluate the performance of the proposed MAC scheduler, as explained in Section 5.4.3.

5.4.1 WFQ MAC Scheduler Throughput Guarantees

The proposed WFQ MAC scheduler provides throughput guarantees for different packet flows. This is accomplished by allocating bandwidth to packet flows according to their configured weights as follows:

$$r_i = \frac{\phi_i}{\phi} C \tag{5.6}$$

Where r_i is the guaranteed minimum rate of the bandwidth for session i, ϕ_i is a real number representing the share of bandwidth reserved for session i (weight), and ϕ is the sum of all configured weights.

$$\phi = \sum_{j=1}^{N} \phi_j \tag{5.7}$$

The weights for different packet flows are configured during admission according to requested, subscribed, and available bandwidth. Admission control must ensure that the sum of the guaranteed shares for all flows does not exceed the maximum throughput (C) of the upstream channel, as in Equation (5.8). The maximum throughput has been derived in Section

4.10.2 to be 78% of the available channel bandwidth. The isolation property of WFQ gives conforming flows their guaranteed share of the bandwidth despite the possibility of misbehaving flows. Misbehaving flows are the ones to be penalized for their misbehavior.

$$C = \sum_{j=1}^{N} r_j \qquad (5.8)$$

5.4.2 WFQ MAC Scheduler Delay Guarantees

In this section, we derive the upper bounds for the access delay of the proposed WFQ based MAC scheduler. The total access delay (W) is the elapsed time from the moment a packet arrives at a CM for upstream transmission till the last bit of the packet is received at the CMTS. It is composed of the detection delay (W_d), and the admission delay (W_a). The admission delay mainly consists of the queueing delay for the requests at the scheduler (W_q), and the round trip propagation delay (2τ). Therefore the total access delay for packet flow i is:

$$W_i = W_{d,i} + W_{q,i} + 2\tau \qquad (5.9)$$

The upper bound for detection delay has been obtained in Section 4.10.1. It is the time from the moment a packet arrives at a CM for upstream transmission till the request for bandwidth issued by the CM is received at the CMTS. The propagation delay depends on the distance of the CM from the CMTS. For simplicity we assume that CMs and the CMTS are 10 Km apart which results in $\tau = 0.05$ ms, as in Section 4.10.1.

The queueing delay (W_q) is the delay for queueing the request at the MAC scheduler till bandwidth is allocated to the CM. This delay is variable and depend on other requests for bandwidth in the scheduler queue. Using the proposed WFQ MAC scheduler this delay can be upper bounded for QoS guarantees. WFQ scheduler can guarantee delay bounds if the traffic is Token Bucket shaped [80]. If arrivals from flow i to a WFQ scheduler are shaped using a token bucket shaper with rate of r_i and a token bucket depth of σ_i, the maximum delay experienced by a packet flow i in a WFQ scheduler is bound and equals

$$W_{q,i} = \frac{L_{max}}{C} + \frac{\sigma_i}{r_i} + \frac{L_i}{r_i} \tag{5.10}$$

Where: L_i is maximum packet size of session i, and L_{max} is the maximum packet size among all flows sharing the link. The first term in Equation (5.10) represents a situation where an arriving request finds the scheduler busy processing a packet of length L_{max}. This forces the request to wait a maximum time of L_{max}/C.

The second term represents the time an arriving request may wait till the scheduler finishes processing a burst of length σ_i. In our case, DOCSIS does not allow more than one request per flow to be outstanding. If per flow scheduling is used, a request for flow i always finds its queue empty. Therefore, the term σ_i/r_i vanishes in our case.

The third term is the delay for processing the request itself by the scheduler. At worst, this can be L_{max}/r_i. The bandwidth share of flow i can be found as in Equation (5.6). Therefore, the

124

maximum queueing delay of request at the scheduler can be found by the following equation.

$$W_{q,i} = \frac{L_{max}}{C} + \frac{\phi L_{max}}{\phi_i C}$$

(5.11)

Substituting the queueing delay in Equation (5.9), we get the upper bound for access delay for flow i as:

$$W_i = W_{d,i} + \frac{L_{max}}{C} + \frac{\phi L_{max}}{\phi_i C} + 2\tau$$

(5.12)

In Equation (5.12), C, L_{max}, and τ are constants. The only parameters that we can control in order to reduce the access delay are $W_{d,i}$, and ϕ_i. We managed to reduce the detection time ($W_{d,i}$) as explained in Section 4.10.1. Our proposed MAC guarantees a maximum detection time of two times the frame duration (4 ms) versus eight times the frame duration (16 ms) in the case of the existing DOCSIS MAC. The weight (ϕ_i) assigned to packet flow i has a great impact on controlling the access delay. We can reduce the maximum access delay for a packet flow by assigning it a higher weight which means a higher bandwidth share.

5.4.3 Simulation Study

In this section, we study the performance of the proposed WFQ based MAC scheduler using OPNET simulation models. We compare the performance of the proposed scheduler with a Priority based MAC scheduler. We build our simulation models on CableLabs Opnet simulation framework.

We simulate a scenario where 150 CMs sharing a single 5 Mbps upstream channel. Among these CMs, there is a single CM with a 3G Node B, 16 CMs with G.711 real-time voice traffic (64 Kbps), and 8 CMs with non-real-time data traffic (64 Kbps). A single Node B here aggregates traffic from 75 wireless voice connections, and 50 wireless data connection. The rates for wireless voice and data sources are 8 (G.729) and 18 Kbps, respectively.

We classify traffic into three priority classes with class 0 as the highest priority. Class 0 is allocated to wireless 3G Node B traffic. Class 1 is for wireline voice traffic, while class 2 is for wireline data traffic. We show simulation results for throughput and access delay for each traffic priority class. The throughput is calculated as the number of bits per priority class received over the simulation time, while the access delay is calculated as the elapsed time from the point when a packet belonging to a certain priority class is received at a CM until the time it leaves the DOCSIS CMTS.

In order to be able to fairly compare the performance of the proposed WFQ based MAC scheduler with a priority based MAC scheduler, we simulate identical scenarios for both scheduling schemes. Section 5.4.4, and Section 5.4.5 present the results for the Priority, and WFQ MAC schedulers, respectively.

5.4.4 Priority Based MAC Scheduler

We configure CMs into three priority classes. Priority 0 CMs generate 1.5 Mbps, while Priority 1 and 2 generate 1.0 Mbps and 0.5 Mbps, respectively. In the first experiment, we simulate a scenario where all CMs conform to their traffic contracts by not exceeding their

126

reserved bandwidth. The Priority based scheduler manages to guarantee all three classes their

share of the bandwidth (see Figure 5.2). However, the priority scheduler fails to guarantee
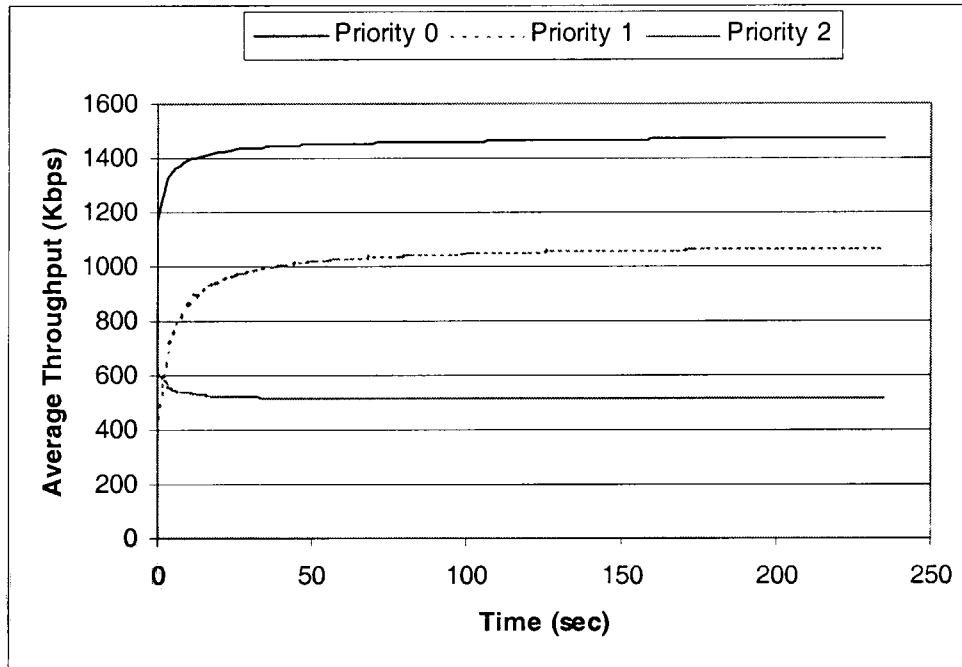
Figure 5.2 Throughput for conforming traffic using priority scheduler

delay bounds for lower priority classes. Figure 5.3, and Figure 5.4 show access delay and

average access delay for the three traffic classes, respectively. It can be seen from Figure 5.3

that even with conforming sources priority class 1 can still miss some of its deadlines. Priority

class 0 enjoys lower and bounded access delay.

In the second experiment, Priority class 0 does not conform to its traffic contract established at

connection setup (not policed). On the other hand, both priority 1, and priority 2 sources are

conforming and offer 1 Mbps and 0.5 Mbps, respectively. We increase the traffic offered by
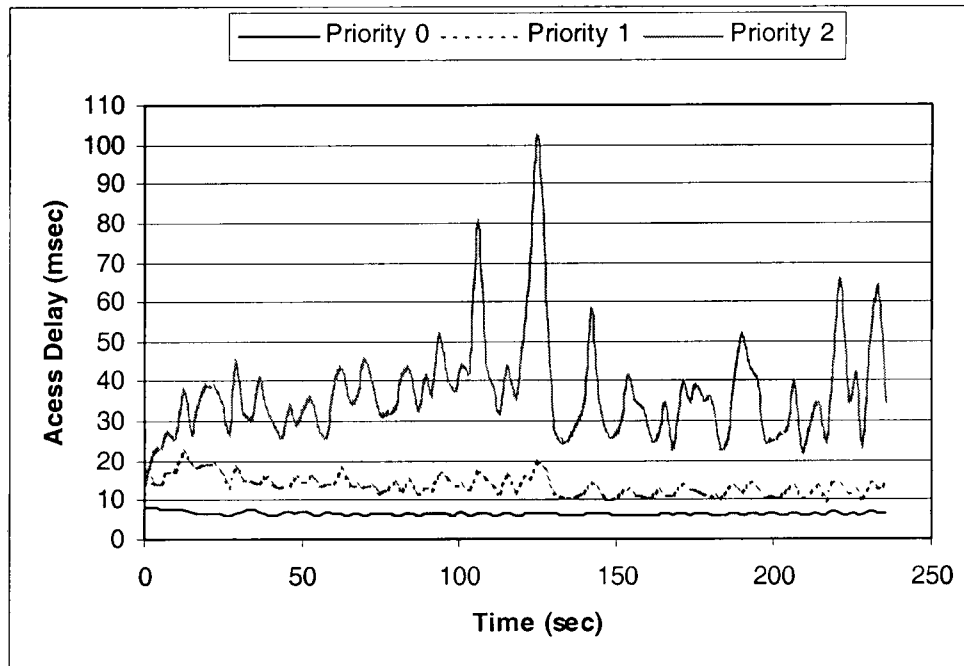
Figure 5.3 Access delay for conforming traffic using priority scheduler

priority class 0 and monitor the access delay and the throughput of the three priority classes.

Figure 5.5 shows the average throughput versus the total offered load for the three traffic classes. We gradually increase the traffic load offered by priority class 0 from 1.1 Mbps to 2 Mbps, while keeping traffic load offered by priority classes 1 and 2 constant. We can see from Figure 5.5 that the average throughput for priority 0 keeps increasing without affected the traffic from other priority classes. This remains the case until the maximum capacity of the channel is reached after which the violating priority 0 gets more than its bandwidth share at the expense of starving class 2. It can be seen from Figure 5.5 that the throughput of priority class 2 drops by about 30% while its share is taken by priority class 0.

128
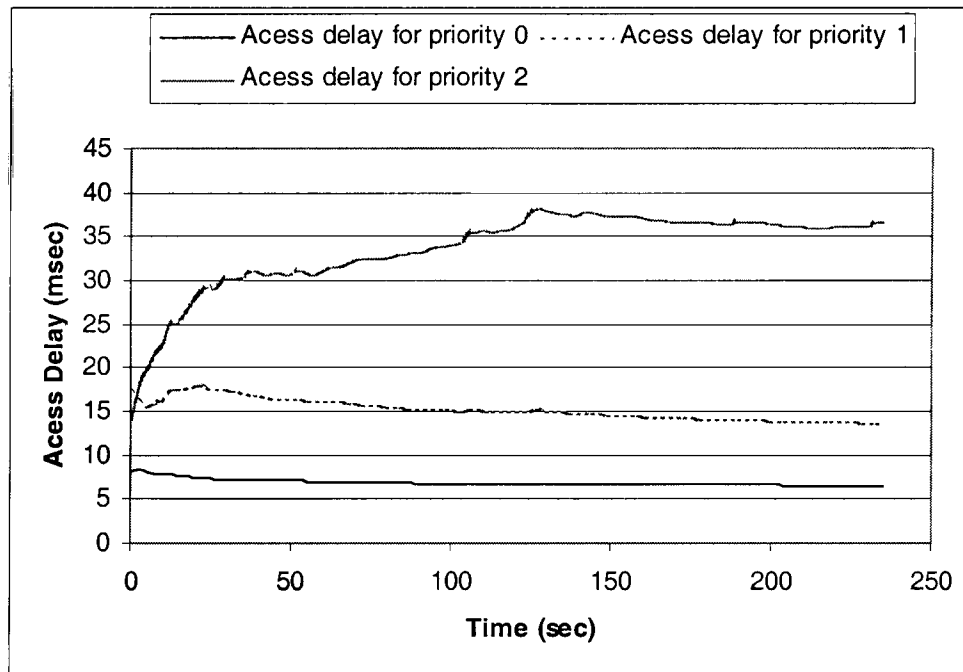
Figure 5.4 Average access delay for conforming traffic using priority scheduler

Also, using the same experiment we can see that violating priority class 0 enjoys low delay

bound while lower priority classes suffer. This can be observed from Figure 5.6. At light

traffic loads all priority classes enjoy low access delay. As the load offered by priority class 0

increases the access delay for low priority classes increase while access delay for priority   0

remains unaffected.

A Priority based MAC scheduler can not guarantee delay bounds and fair share of the

throughput for lower priority classes. This becomes obvious when a high priority class

exceeds its share the of allocated bandwidth. This motivates us to propose a WFQ based MAC

scheduler which we study its performance in the following section.
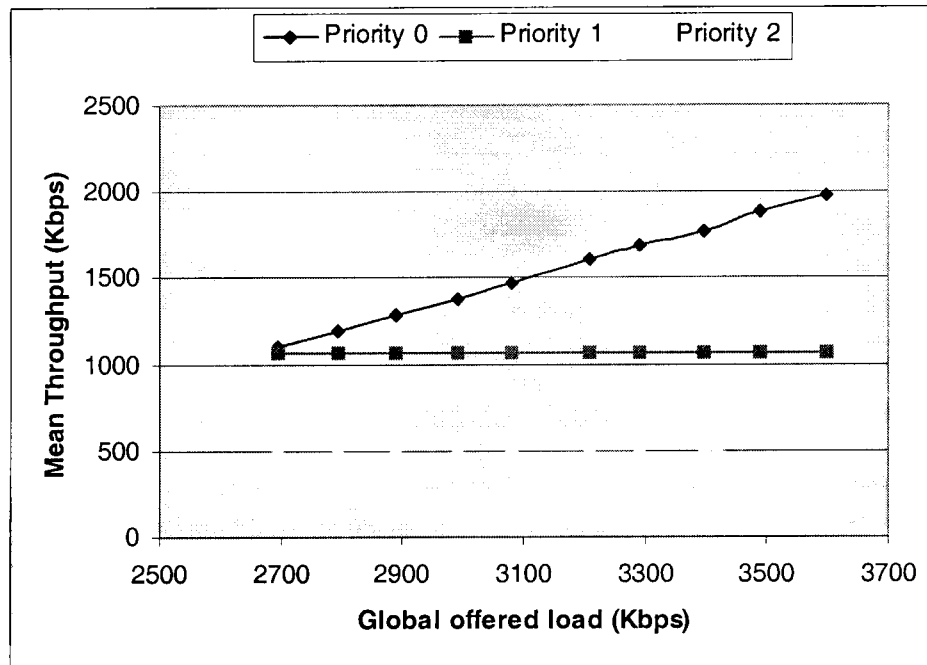
Figure 5.5 Average throughput versus offered load using priority scheduler

5.4.5 WFQ Based MAC Scheduler

In order to have a fair comparison between the proposed WFQ based MAC scheduler and a priority based scheduler we simulate similar scenarios in both cases. Here, we develop a WFQ based MAC scheduler using Opnet. We configure different weight for different priority classes. We assign Priority class 0 a weight of 3, priority class 1 a weight of 2 and priority class 2 a weigh of 1.

In the first experiment, we simulate a scenario where all CMs conform to their traffic contracts by not exceeding their reserved bandwidth. The WFQ based scheduler manages to guarantee all three classes their fair share of the bandwidth according to their assigned weight (see Figure 5.7).

130
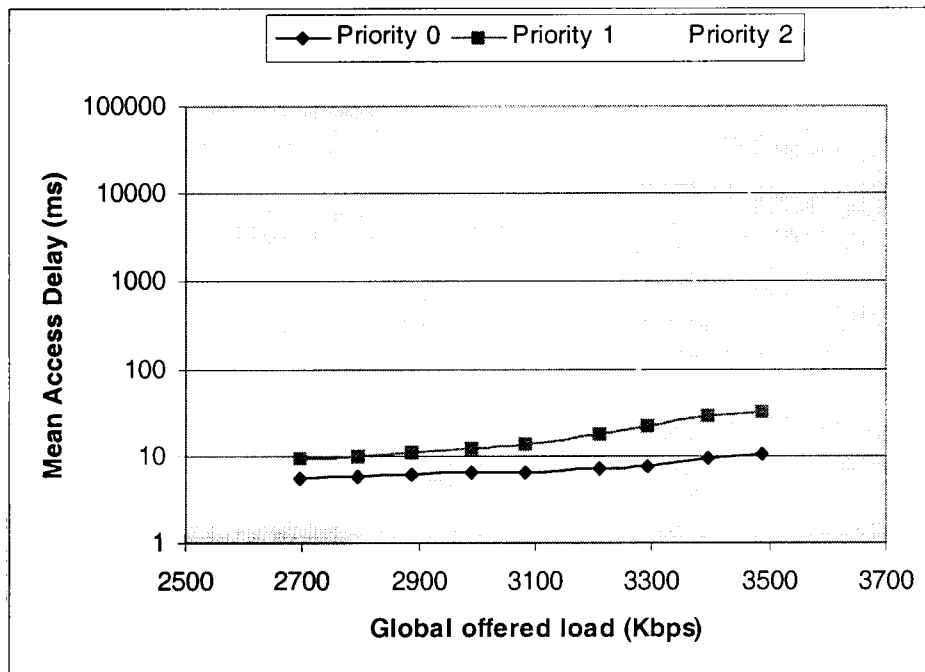
Figure 5.6 Average access delay versus offered load using priority scheduler

In addition, the WFQ scheduler can guarantee delay bounds for all traffic priority classes. Figure 5.8, and Figure 5.9 show access delay and average access delay for the three priority classes, respectively. It can be seen from Figure 5.8 that the WFQ scheduler can provide delay guarantees for real-time priority classes (priority 0 and priority 1) while granting non-real-time priority class (priority 2) acceptable delay guarantees which are lower than that provided by the priority scheduler (~40% less on average). The WFQ provides real-time traffic with the required guaranteed access delay of less than 20 ms (about 10 ms in average). This is a much desirable characteristic of the proposed scheduler.

The second experiment is similar to that for the priority scheduler where priority class 0 does not conform to its traffic contract established at connection setup (not policed). On the other
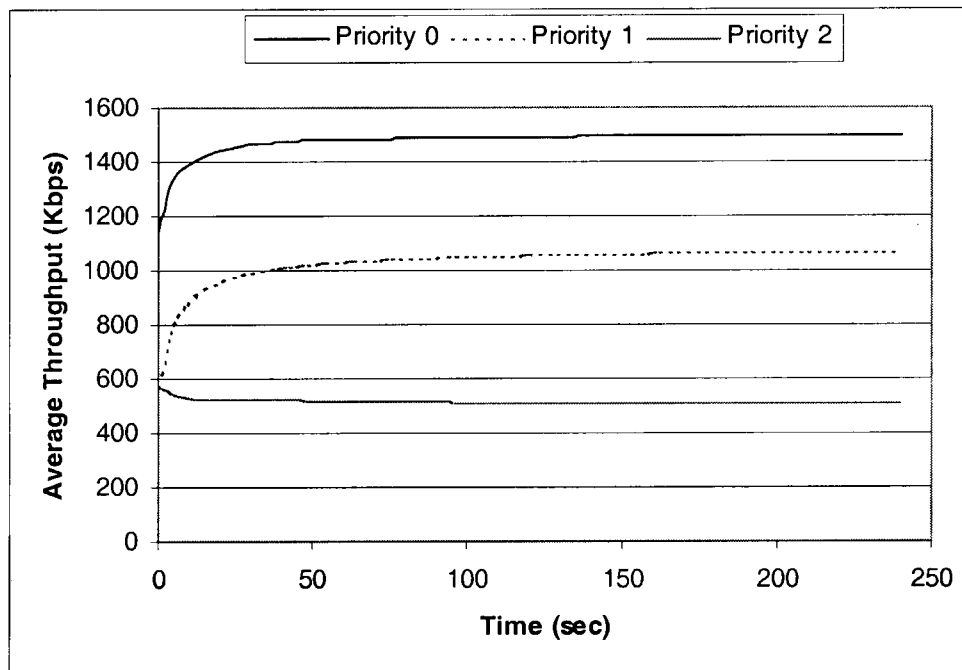
Figure 5.7 Throughput for conforming traffic using WFQ scheduler

hand, both priority 1 and priority 2 sources are conforming and offer 1 Mbps and 0.5 Mbps, respectively. We increase the traffic offered by priority class 0 and monitor the access delay and the throughput of the three priority classes.

Figure 5.10 shows the average throughput versus the total offered load for the three traffic classes. We gradually increase the traffic load offered by priority class 0 from 1.1 Mbps to 2 Mbps, while keeping traffic load offered by priority classes 1 and 2 constant. We can see from Figure 5.10 that the average throughput for priority 0 keeps increasing without affected the traffic from other priority classes. It utilizes any free bandwidth that is not used by other priority classes. This remains the case until the maximum capacity of the channel is reached after which the violating priority 0 throughput becomes constant. It can also be seen from
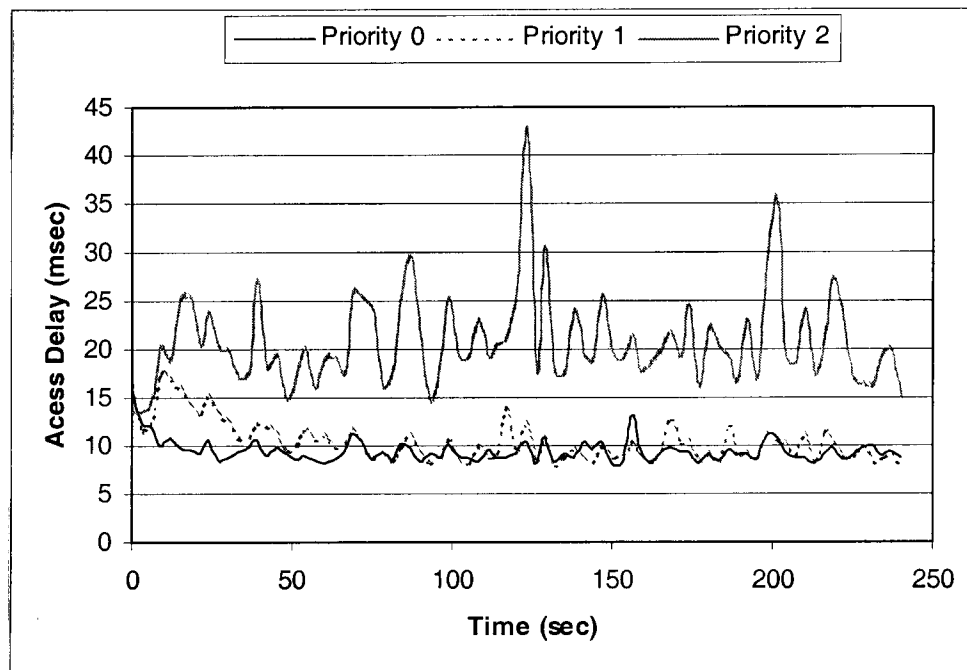
Figure 5.8 Access delay for conforming traffic using WFQ scheduler
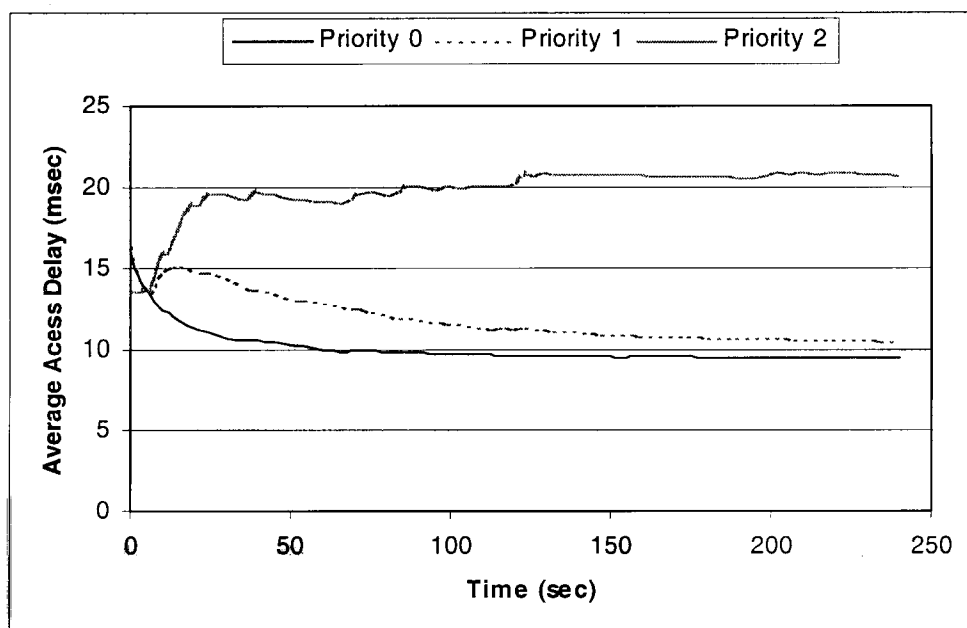
Figure 5.9 Average access delay for conforming traffic using WFQ scheduler

133

Figure 5.10 that the throughput of the lower priority classes is unaffected. This is due to excellent isolation property of WFQ schedulers.
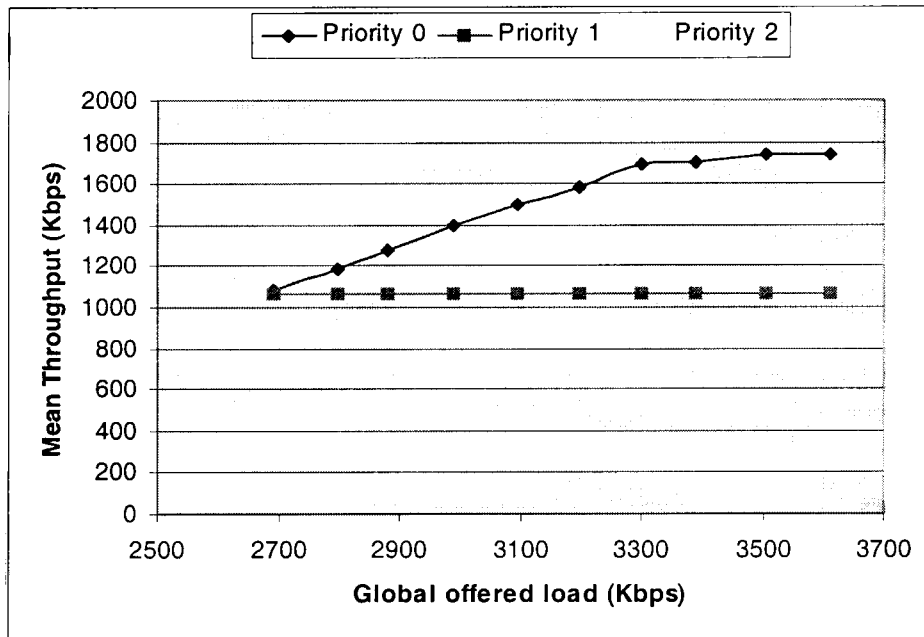
Figure 5.10 Average throughput versus offered load using WFQ scheduler

Also, using the same experiment we can see that only violating priority class 0 is punished by by experiencing high access delay when the maximum bandwidth capacity of the channel is exceeded. The access delay for lower priority classes is not affected. This can be observed from Figure 5.11. At light traffic loads all priority classes enjoy low access delay. As the load offered by violating priority class 0 increases its access delay increases while access delay for priority classes 1 and 2 remains unaffected..

It can be concluded from the above that a WFQ based MAC scheduler can guarantee delay bounds and fair share of the throughput for all conforming traffic priority classes. Only
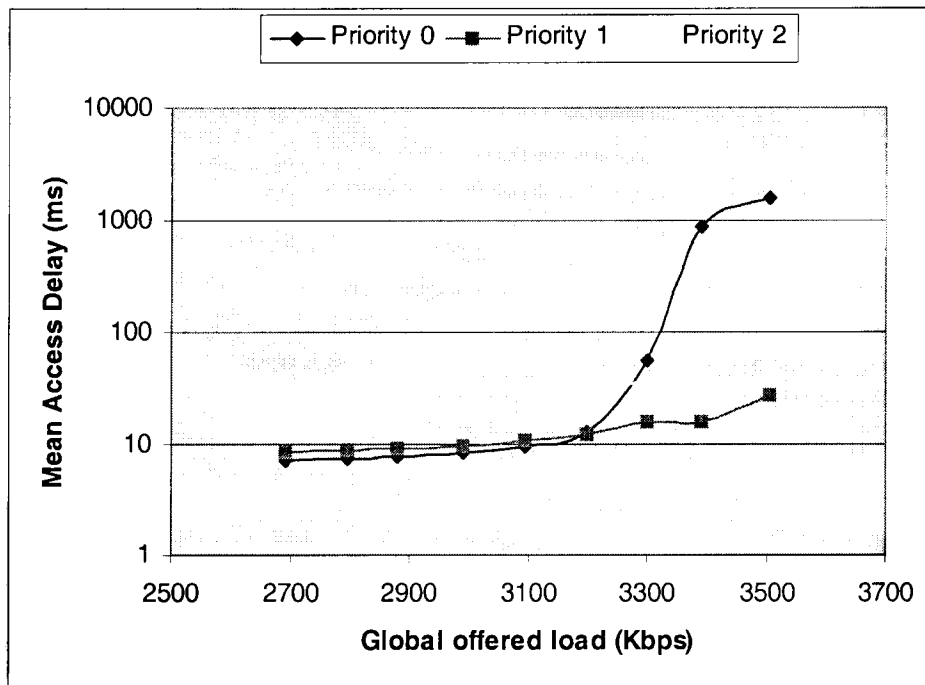
Figure 5.11 Average access delay versus offered load using WFQ scheduler

nonconforming priority classes are punished by exposure to higher access delay and limiting their throughput to at least their fair share. Traffic sources can be allocated more than their fair share only if more bandwidth is available and is not used by other conforming sources. Fairness, isolation, guaranteed fair share of the throughput, and guaranteed delay bounds characteristics of WFQ are highly desirable features of any potential scheduler. We conclude that our proposed WFQ based MAC scheduler is an excellent candidate for DOCSIS based HFC CATV network. It is important for supporting wireline traffic as well as our wireless 3G deployment over DOCSIS CATV networks.

135

5.5 Summary

In this chapter, we proposed a novel control-plane MAC scheduler based on superior Weighted Fair Queueing for DOCSIS networks. Our proposed WFQ MAC scheduler differs from traditional network layer WFQ schedulers. The former schedules packets at the MAC layer based on requests for upstream bandwidth. The later schedules the actual packets at the network layer. Also, our scheduler provides the needed functionalities for implementing different DOCSIS defined upstream service flows. Advantages of our scheduler include QoS guarantees, fairness, and flow isolation for different upstream packet flows. The proposed scheduler is essential for the proposed 3G/WLAN deployment over CATV networks, as well as for the existing CATV high-speed Internet subscribers.

Chapter 6

Conclusions

6.1 Summary and Concluding Remarks

In conclusion, this research was motivated by the need for a cost-effective infrastructure for interworking wireless data systems. The goal of this thesis is to propose a wireless data systems deployment over the HFC CATV networks.

The existing HFC CATV plant has many assets that make it an attractive candidate for providing the infrastructure needed for the wireless data systems deployment [73]. These include: excess bandwidth, real estate for electronics equipment, pole attachment rights, rights of way, and shared maintenance and customer services staff. DOCSIS is the de facto standard for supporting high-speed Internet access over the CATV plant. It is a framework for MAC, and PHY protocols for providing QoS guarantees for multimedia application over the CATV

networks.

A number of researchers have proposed wireless data systems deployment over HFC CATV networks. Most of these proposal address PCS deployment over HFC networks rather than focusing on a specific PCS/3G network. Also, previous PCS/HFC proposals do not consider using DOCSIS which is now a reality for two way communications over CATV networks. There is no previous wireless data deployment that offers a specific, complete, and standard based solution.

Instead of considering a generic wireless data systems over CATV deployment, we choose a standard based deployment. The adoption of standards represents a way for ensuring interoperability, and to put together the efforts of many vendors to bring forth the most cost effective solutions. On the wireless data systems side, we choose two dominant wireless data systems, namely: 3G UMTS, and WLAN systems. On the CATV side, we choose DOCSIS based architecture and protocols.

In summary, we propose a cost effective deployment of 3G UMTS, and WLAN systems over DOCSIS based HFC CATV networks. Cost effectiveness is provided by sharing the existing CATV network, and using the standard equipment and protocols of DOCSIS. Despite our specific deployment our proposal can be extended to deploy other wireless data systems over HFC networks.

We identified two main critical issues in accomplishing our goal. These are network architecture, and MAC related issues. Regarding network architecture issue, we started by conducting

a traffic study to verify that the CATV plant has adequate capacity for supporting wireless data users in addition to the existing services like CATV broadcasting, and high-speed Internet access. Second, we proceeded by defining a network architecture for our 3G UMTS systems deployment over DOCSIS based HFC CATV networks. We identified the required components and functionalities for our deployment. Also, we defined the required layers in the transmission, signalling, and management planes of our UMTS over HFC deployment.

Third, we proposed a network architecture for interworking WLANs with 3G systems over CATV networks. Our proposal minimizes the need for building a dedicated infrastructure for WLAN/3G interworking, specially in residential areas. Recently, WLAN/3G interworking has gained a lot of interest. This is mainly to provide true mobility and enhanced security for WLAN users. A number of scenarios for 3G and WLAN interworking have been considered. These range from common billing (loose coupling) to the provision of services seamlessly between the WLAN and the 3GPP system (tight coupling). Our proposed WLAN/3G deployment over CATV networks is independent of the interworking scenario chosen.

A final task in the architecture issues was defining a QoS framework for DOCSIS networks. This framework defined a systems level upstream, and downstream QoS architecture for DOCSIS CMTS, and CMs. Transmission plane mechanisms for QoS guarantees have been identified. QoS guarantees for both wireless, and wireline traffic were the aim of proposed QoS framework.

Regarding the MAC issues, we identified QoS guarantees at the MAC layer as the main

concern. This is mainly critical in the upstream due to the contention based nature of the DOCSIS CATV upstream channels. Wireless traffic has very stringent delay and throughput requirements when transported over DOCSIS networks.

First, we defined a DOCSIS MAC implementation for supporting QoS guarantees for delay-sensitive wireless, and wireline traffic. We identified the critical implementation options that are essential for supporting 3G wireless traffic over DOCSIS MAC protocol.

Second, we enhanced DOCSIS MAC rt-PS protocol to enable it to efficiently support QoS guarantees for 3G wireless traffic, as well as wireline residential Internet traffic. We conducted a study to evaluate the performance of the proposed MAC. Performance results show that our proposed MAC enhancements can reduce the access delay for delay-sensitive traffic by 30% to 40% over existing DOCSIS MAC rt-PS. This is a much desirable performance requirement for 3G wireless traffic. The reductions of access delay by our proposed MAC enhancement is achieved without compromising the delay and throughput performance requirement of other traffic classes. The total throughput of our proposed MAC is also comparable to an efficient implementation of existing DOCSIS MAC.

The final MAC issue is related to scheduling at the MAC layer. Despite the adequacy of our proposed DOCSIS MAC enhancements for supporting wireless traffic over DOCSIS based HFC networks, improvement was still needed in the area of MAC scheduling. Our target is to ensure that transporting wireless traffic over DOCSIS networks does not degrade QoS for existing wireline residential traffic. Upstream bandwidth allocation needs to be executed in a

140

fair manner that guarantees QoS for wireless, and wireline traffic. Using a priority based scheduler for allocating upstream bandwidth is favorable to high priority classes like wireless traffic, but can be disastrous for lower priority classes. This led us to propose a WFQ based MAC scheduler.

The proposed WFQ MAC scheduler is fundamentally different from traditional network layer WFQ schedulers. The former schedules packets at the MAC layer based on requests for upstream bandwidth. The later schedules the actual packets at the network layer. The end result of scheduling requests using our MAC scheduler is that the actual data packet are scheduled.

We conducted a performance study for our proposed WFQ MAC scheduler. We compared its performance with a priority based MAC scheduler. Performance results confirm that our WFQ MAC scheduler is superior in terms of QoS guarantees. It ensures delay bounds and fair share of bandwidth for different traffic classes. By using the proposed WFQ MAC scheduler, we ensures that while giving a high priority for the wireless traffic, QoS can still be guaranteed for lower traffic classes such as best-effort wireline traffic. This is true even if a high priority class tries to exceed its share of allocated bandwidth. Violating traffic is penalized by degrading it own QoS while other traffic can still enjoy QoS guarantees. This is superior to a priority based scheduler where a violating high priority traffic can cause DoS for lower priority traffic.

6.2 Thesis Contributions

The main contributions of this thesis can be summarized as follows:

(1) A traffic study for estimating the traffic generated by wireless data systems in different environments. The traffic study confirmed the capability of CATV networks to transport wireless data traffic on top of the existing services like CATV broadcasting, and high-speed Internet access.

(2) A cost effective network architecture for deploying 3G UMTS systems over DOCSIS based HFC CATV networks. This includes defining the required network elements, functionalities, and protocol stack.

(3) A network architecture for interworking 3G and WLAN systems over HFC CATV networks. The proposed architecture can accommodate different WLAN/3G interworking scenarios by being independent of the WLAN/3G interworking scenario chosen (e.g., loose coupling, or tight coupling).

(4) A system level QoS framework for DOCSIS networks which facilitates QoS guarantees for wireless, and wireline traffic. It defines the required upstream, and downstream QoS mechanisms in DOCSIS network elements.

(5) A DOCSIS MAC implementation for facilitating transporting wireless data traffic over HFC CATV networks. We identified the critical implementation options that are essential for supporting wireless traffic over DOCSIS MAC protocol.

(6) Enhancements for the DOCSIS MAC Real-time Polling Services that improves QoS guarantees for delay-sensitive traffic. The proposed MAC enhancements can reduce the access delay for delay-sensitive traffic by 30% to 40% over existing DOCSIS MAC rt-PS. This is achieved without compromising QoS guarantees for other traffic classes, or the DOCSIS channel utilization. Both wireless and wireline delay-sensitive traffic can benefit from the proposed DOCSIS MAC enhancements.

(7) A superior Weighted Fair Queueing based MAC scheduler for DOCSIS networks. The proposed scheduler guarantees delay bounds, and fair share of the bandwidth for different traffic classes. In addition, it resolves any DoS issues. The proposed scheduler is essential for the proposed wireless data systems deployment over CATV networks, as well as for existing wireline users.

(8) A simulation tool for evaluating the performance of the proposed architecture, and protocols. The fact that this simulation tool is developed based on CableLabs OPNET Common Simulation Framework (CSF) makes it reliable, and re-usable for comparative studies.

6.3 Future Research

There are a number of related issues that can be investigated as an extension of this research. These can be summarized as follows:

(1) Developing a suitable Call Admission Control (CAC) for DOCSIS networks. DOCSIS specifications do not impose any CAC, rather it promotes flexibility by leaving it to be implementation dependent.

(2) Developing a load balancing scheme for DOCSIS CATV networks. DOCSIS CMs can be assigned to different channels either during admission, or during a connection. DOCSIS specifications define procedures for moving CMs between channel, but does not specify the schemes for making these decisions.

(3) Investigating extension of our work to deploy other 3G systems such as CDMA 2000 over CATV networks. We expect our proposed MAC protocol and scheduler to work for the CDMA 2000 deployment as well. However, CDMA 2000 mapping to DOCSIS CATV networks need to be investigated.

(4) Extension of DOCSIS MAC to support ATM. This might be beneficial for supporting the 3G systems which use the ATM transport option. DOCSIS MAC identifies ATM support as a future issue. The operation of DOCSIS MAC with a mixture of fixed size ATM cells, and variable size IP packets needs to be investigated. Optional DOCSIS ATM related headers need to defined.

Bibliography

[1]   3GPP TS 23.060 V5.5.0, "General Packet Radio Service (GPRS); Service description; Stage 2," Release 5, Mar 2003.

[2]   3GPP TS 23.107 V5.8.0, "Quality of Service (QoS) concept and architecture," Release 5, March 2003.

[3]   3GPP TS 23.207 V5.7.0, "End-to-end Quality of Service (QoS) concept and architecture," Release 5, March 2003.

[4]   3GPP TS 23.234 V1.8.0 (draft), "3GPP system to Wireless Local Area Network (WLAN) Interworking; System Description," Release 6, April 2003.

[5]   3GPP TS 25.401 V5.5.0, "UTRAN Overall Description," Release 5, Dec. 2002.

[6]   3GPP TS 25.420 V5.1.0, "UTRAN Iur interface: general aspects and principles," Release 5, Sept. 2002.

[7]   3GPP TS 25.430 V5.2.0, "UTRAN Iub interface: general aspects and principles," Release 5, Sept. 2002.

[8]   3GPP TS 32.104 V3.5.0, "3G Performance Management (PM)," Release 1999, Dec 2001.

[9]   3GPP TS 25.442 V5.1.0, "UTRAN implementation-specific O&M transport," Release 5, sept 2002.

[10]   3GPP TR 32.800 V5.0.0, "Telecommunication Management; Management Level Procedures and Interaction with UTRAN," Release 5, Mar 2002.

[11]   N. Abramson (Ed.), "Multiple Access Communications: Foundations for Emerging Technologies," IEEE Press, 1993.

[12]   N. Amitay, "Distributed Switching and Control with Fast Resources Assignment/Handoff for Personal Communication Systems," IEEE JSAC, Vol. 11, no. 6, Aug. 1993, pp. 842-49.

[13]   ANSI/IEEE, "Std 802.11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications," 1999.

[14]   K. Apostolidis, L. F. Merakos, and X. H. Xing, "A reservation protocol for packet voice and data integration in unidirectional bus networks," IEEE Transactions on Communications, Vol. 41, No. 3, pp. 478-485, Mar. 1993.

[15] C. Aurrecoechea, A. T. Campbell, and L. Hauw, "A Survey of QoS Architectures", ACM/Springer Verlag Multimedia Systems Journal, Special Issue on QoS Architecture, Vol.6 No.3, pg. 138-151, May 1998.

[16] A. S. Beasley, "The advantages of using cable TV distribution plant for linking PCS-microcells," International Conference on Selected Areas in Wireless Communications Proceedings, pp. 292-295, 1992.

[17] E. Berruto, C. Eynard, E. Fiorina, and A. Napolitano, "Issues in protocol evaluation for personal telecommunications," Globecom '93 Proceedings, Vol. 1, pp. 555-559, Nov. 1993.

[18] D. Bertsekas, and R. Gallager, "Data Networks," 2nd Edition, Prentice Hall, 1992.

[19] C. Bisdikian, "A Review of Random Access Algorithms," IEEE 802.14-96/019, Draft Proposal.

[20] C. Bisdikian, B. McNeil, R. Norman, and R. Zeisz, "MLAP: A MAC level access protocol for the HFC 802.14 network," IEEE Communications Magazine, Vol. 34, No. 3, pp. 114-121, Mar. 1996.

[21] P. T. Brady, "A statistical analysis of on-off patterns in 16 conversations," Bell Systems Technical Journal, Vol. 47, pp. 73-91, Jan. 1968.

[22] P. T. Brady, "A model for generating on-off speech patterns in two way conversation," Bell Systems Technical Journal, Vol. 48, pp. 2445-2472, Sep. 1969.

[23] Broadcom, "OPNET/CableLabs DOCSIS Common Simulation Framework (CSF): Real-time polling overview," V1.0, Aug 1998.

[24] Broadcom, "IEEE 802.11g: The next mainstream Wireless LAN standards," White paper, http://www.commweb.com/, March 2003.

[25] CableLabs, Data-Over-Cable Service Interface Specifications, Radio Frequency Interface Specification, V1.0, SPRFI-C01-011119.

[26] CableLabs, Data-Over-Cable Service Interface Specifications, Radio Frequency Interface Specification v1.1, SP-RFIv1.1-I08-020301.

[27] CableLabs, Data-Over-Cable Service Interface Specifications, Radio Frequency Interface Specification v2.0, SP-RFIv2.0-I02-020617.

[28]   W. Carney, Texas Instruments, "IEEE 802.11g New Draft Standard Clarifies Future of Wireless LAN," White paper, http://focus.ti.com/pdfs/vf/bband/802.11g_whitepaper.pdf.

[29]   C. Carroll, "Development of integrated cable/telephony in the United Kingdom," IEEE Communication Magazine, Vol. 33, No. 8, pp. 48-60, Aug. 1995.

[30]   H. C. B. Chan, and V. C M. Leung, "Dynamic Bandwidth Assignment Multiple Access for Efficient ATM-based Service Integration over LEO Satellite Systems," Northcon'98 conference proceedings, pp. 15-20, Oct. 1998.

[31]   W. Y. Chen, and T. R. Hsing, "Architectural alternatives of wireless access through hybrid fiber/coax distribution plant," IEEE GLOBECOM '95., Vol. 2, pp. 14-16, Nov. 1995.

[32]   S. Cho, J. Kim, and S. Park, "Performance evaluation of the DOCSIS 1.1 MAC Protocol According to the structure of a MAp message," ICC 2001 proceedings, Vol. 6, pp. 1786 -1791, 2001

[33]   W. Ciciora, "An introduction to cable television in the United States," IEEE LCS Magazine, pp. 19-25, Feb. 1990.

[34]   M. D. Corner, N. Golmie, J. Liebeherr, and D. Su, "A Priority Scheme for the IEEE 802.14 MAC Protocol for Hybrid Fiber-Coax Networks,"

[35]   J. Cosmas et. al., "Overview of the mobile communications program of RACE II," IEEE Electronics ans Communication Journal, Vol. 7, No. 4, pp. 155-167, Aug. 1995.

[36]   J. E. Dail et al., "Adaptive digital access protocol: A MAC protocol for multiservice broadband access networks," IEEE Communications Magazine, Vol. 34, No. 3, pp. 104-113, Mar. 1996.

[37]   R. W. Donaldson, and A. S. Beasley, "Wireless personal communications using CATV distribution networks," CCECE Proceedings, Vol. 2, pp. 995-999, 1993.

[38]   R. W. Donaldson, and A. S. Beasley, "Wireless CATV network access for personal communications using simulcasting," IEEE Transactions on Vehicular Technology, Vol. 43, No. 3, pp. 666-671, Aug. 1994.

[39]   M. Droubi, N. Idirene, and C. Chen, "Dynamic bandwidth allocation for the HFC DOCSIS MAC protocol," 9th International Conference on Computer Communications and Networks, pp. 54-60, 2000.

[40] A. A. Elfeitori, and V. C. M. Leung, "Personal communications services over HFC CATV networks," Proceedings of CCECE, Vol. 2, pp. 462-465, May 1997.

[41] A. A. Elfeitori, and H. Alnuweiri, "Supporting Isochronous ATM Traffic over IEEE 802.14 Based HFC Access Networks," Proceedings of CCECE, 1998.

[42] A. A. Elfeitori, and H Alnuweir, "A MAC Protocol for Supporting Real-time VBR Traffic Over IEEE 802.14 Based HFC Access Networks," Proceedings of CCECE, pp. 1999.

[43] J. W. Eng, "Standards for HFC-based residential broadband: IEEE project 802.14, its mission, charter, and status," SPIE Proceedings, Vol. 2609, pp. 2-9, Oct. 1995.

[44] W. L. Er, "ATM based wireless personal communications services (PCS) over cable television (CATV) networks," M.A.Sc Thesis, Departement of Electrical and Computer Engineering, University of British Columbia, 1996.

[45] W. L. Er., T. S. Wong, and R. W. Donaldson, R.W "ATM-based wireless interworking architectures," IEEE Pacific Rim Conference 1997, Vol. 1, pp. 383 -387, Aug 1997.

[46] P. Ferguson, and G. Huston, "Quality of Service," John Wiley & Sons, Inc., 1998.

[47] N. Golmie, F. Mouveaux, and D. H. Su, "Differentiated services over cable networks," GLOBECOM '99 , Vol. 2 , pp. 1109 -1115, Dec. 1999.

[48] D. J. Goodman, "Network control for wireless Communications," IEEE Communications Magazine, Dec 1992.

[49] D. J. Goodman, R. A. Valenzuela, K. T. Gayliard, and B. Ramamurthi, "Packet reservation multiple access for local wireless communications," IEEE Transactions on Communications, Vol. 37, No. 8, pp. 478-485, Aug. 1989.

[50] G. Gruber, and N. H. Le, "Performance requirements for integrated voice/data networks," IEEE Journal on Selected Areas in Communications, Vol. sac-1, No. 6, pp. 981-1004, Dec. 1983.

[51] K. H. Han, and Y. C. Chung, "Multi-purpose fiber-optic access network (MFAN)," Optical Fiber Communication Conference and Exhibit, 2001, Vol. 3, 2001, pp. WN4-1 -WN4-4.

[52]   K. H. Han, and Y. C. Chung, "Multi-purpose fiber-optic access network (MFAN)," Optical Fiber Communication Conference and Exhibit, 2002, 17-22 Mar 2002, Page(s): 407 -408.

[53]   M. Hawa, and D. W. Petr, "Quality of service scheduling in cable and broadband wireless access systems," Tenth IEEE International Workshop on Quality of Service, pp. 247 -255, 2002.

[54]   N. Huang, C. Su, and H. Chao, "Architectures and handoff schemes for CATV-based personal communications network," IEEE INFOCOM '98, Vol. 2, pp. 748 -755.

[55]   IEEE 802.14 working group, "IEEE std. 802.14 medium access control (MAC) working draft," Draft Revision 1, Nov. 1996.

[56]   IEEE 802.14 working group, "IEEE std. 802.14 medium access control (MAC) working draft," Draft Revision 3, May 1997.

[57]   IETF, RFC 2475, "An Architecture for Differentiated Services," December 1998

[58]   M. J. Karol, Z. Liu, and K. Y. Eng, "Distributed Queueing Request Update Multiple Access (DQRUMA) for Wireless Packet (ATM) Networks," ICC '95 Proc., Seattle, WA, June 1995, pp. 1224-31.

[59]   S. Keshav, "An Engineering Approach to Computer Networking," Addison-Wesley, 1997.

[60]   F. Khan, and D. Zeghlache, "Analysis of Aggressive Reservation Multiple Access Schemes for Wireless PCS," ICC '96 Proc., June 1996, Dallas, TX, pp. 1750-55.

[61]   J. G. Kim, and I. Widjaja, "PRMA/DA: A new Media Access Control Protocol for Wireless ATM," ICC '96 Proc., June 1996, Dallas, TX, pp. 240-44.

[62]   L. Kleinrock, "Queueing Systems," Vol. 2: Computer Applications, John Willey and Sons, 1976.

[63]   L. Koh, and M. T. Liu, "A Wireless Multiple Access Control Protocol for Voice-Data Integration," Int'l. Conf. on Parallel and Dist. Sys., Tokyo, Japan, June, 1996, pp. 206-13.

[64]   F. Koperda, and B. Lin, "First phase simulation for XDQRAP," IEEE 802.14-96/062, Draft Proposal.

149

[65]   O. Kubbar, and H. T. Mouftah, "Multiple Access Control Protocols for Wireless ATM: Problems Definitions and Design Objectives," IEEE Communications Mag., pp. 93-99, Nov. 1997.

[66]   O. Kubar, "Design and Performance Evaluation of Broadband Wireless MAC Protocols," Ph.D. Thesis, Queen's University, Canada, April 2000.

[67]   A. D. Kucar, "Mobile radio: an overview," IEEE Communication Magazine, Vol. 29, No. 11, pp. 72-85, Nov. 1991.

[68]   J. Lee, "802.6 based personal communications network over CATV plant," IEEE Transactions on Broadcasting, Vol. 42, No. 1, pp. 33-41, Mar. 1996.

[69]   V. C. M. Leung, N. Qian, A.D. Malyan, and R.W. Donaldson, "Call control and traffic transport for connection-oriented high speed wireless personal communications over metropolitan area networks," IEEE J. Select. Areas Commun., Vol. SAC-12, pp. 1376-1388, October 1994.

[70]   V. O. K. Li, and X. Qiu, "Personal communication systems (PCS)," Proceedings of the IEEE, Vol. 83, No. 9, pp. 1210-1243, Sep. 1995.

[71]   O. Limb, and L. E. Flamm, "A distributed local area network packet protocol for combined voice and data transmissions," IEEE Journal on Selected Areas in Communications, Vol. sac-1, No. 5, pp. 926-934, Nov. 1983.

[72]   Y. M. Lin, and W. I. Way, "The feasibility study of transporting IS-95 CDMA signals over HFC networks," IEEE Photonics Technology Letters , Vol. 11, Issue 6, pp. 736 -738, June 1999.

[73]   S. Lipoff, "Personal communication services and cable TV," 2rd Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC) Proceedings, pp. 22-25, 1992.

[74]   H. Liu, "Architecture and signaling protocols for wireless CATM networks," ICC '99, Vol. 3, pp. 1557 -1562.

[75]   A. D. Malyan, L. J. Ng, V. C. M. Leung, and R. W. Donaldson, "Network architecture and signalling for wireless personal communications," IEEE J. Sel. Areas Commun., Vol. SAC-11, pp. 830-841, August 1993.

[76] K. S. Meier-Hellstern, G. P. Pollini, and D. J. Goodman, "Network protocols for the cellular packet switch," IEEE Transactions on Communications, Vol. COM-42, No. 2/3/4, pp.1235-1244, Feb./Mar./Apr. 1994.

[77] M. Momona, and M. Yamazaki "Framed pipeline polling for cable TV networks," IEEE 802.14-96/076, Draft Proposal.

[78] R. Mueller, "Modern cable TV networks, the ideal platform for PCS," 1993 Conference on National Telesystem Proceedings, pp. 59-64, 1993.

[79] A. Paff, "Hybrid fiber/coax in the public telecommunications infrastructure," IEEE Communications Magazine, Vol. 33, No. 4, pp. 40-45, April 1995.

[80] A. K. Parekh, and R. G. Gallager "A generalized processor sharing approach to flow control in integrated services networks: The single node case," IEEE/ACM Trans. Networking, pp. 344-357, June 1993.

[81] X. Qiu, and V. O. K. Li, "Dynamic Reservation Multiple Access (DRMA): A New Multiple Access Scheme for Personal Communication System," ACM/Baltzer Wireless Networks, Vol. 2, pp. 117-128, 1996.

[82] QUALCOMM Inc., "An overview of the application of CDMA to digital cellular systems and personal cellular networks," Research Report, May 1995.

[83] R. Rabbat, and K. Siu, "QoS support for integrated services over CATV," IEEE Communications Magazine , Vol. 37, Issue: 1 , pp. 64 -68, Jan. 1999.

[84] D. Raychaudhuri, and N. D. Wilson, "ATM-Based Transport Architecture for multiservices Wireless Personal Communication Networks," IEEE JSAC, Vol. 12, no. 8, Oct. 1994, pp. 1401-14.

[85] K. W. Richardson, "UMTS overview," Electronics & Communication Engineering Journal, Volume: 12, Issue: 3, pp. 93 -100, Jun 2000.

[86] V. Roy, and C. L. Despins, "Planning of GSM-based wireless local access via simulcast distributed antennas over hybrid fiber-coax," GLOBECOM '99, Vol. 2, pp. 1116 -1120.

[87] D. Sala, and J. O. Limb, "A protocol for efficient transfer of data over fiber/cable systems," IEEE INFOCOM Proceedings, Vol. 2, pp. 904-911, 1996.

[88] J. Sanchez, R. Martinez, and M. Marcelline, "A Survey of MAC Protocols Proposed for Wireless ATM," IEEE Network Mag., pp. 52-62, Nov. 1997.

[89] G. N. Senarath, D. Everitt, "Reduction of call drop-outs during handoffs using efficient signal strength prediction algorithms for Personal Communication Systems," Globecom '95 Proceedings, Vol. 3, pp. 2308-2312, 1995.

[90] K. Sriram, AT&T, "Adaptive digital access protocol (ADAPt): A MAC protocol for multiservice broadband access networks," IEEE 802.14-95/142, Draft Proposal.

[91] Statistics Canada, "Profile for census tracts in Matsqui and Vancouver - Part A," 1991 census report, Feb. 1993.

[92] A. Tanenbaum, "Computer Networks," 3rd edition, Prentice Hall, Inc., 1996.

[93] J. M. Ulm, LANcity Corp., "A MAC proposal for 802.14," IEEE 802.14-95/134, Draft Proposal.

[94] UMTS World, "Overview of The Universal Mobile Telecommunication System," Online tutorial, http://www.umtsworld.com/technology/overview.htm, July 2002

[95] S. Varma, "Support for ATM traffic classes in HFC networks," ICC '96 Proceedings, Vol. 1, pp. 496-502, 1996.

[96] T. S. W. Wong, "An ATM based interworking architecture for wireless personal Communications," M.A.Sc Thesis, Department of Electrical and Computer Engineering, University of British Columbia, 1996.

[97] X. Xiao, W. K. G. Seah, Y. H. Chew, and C. C. Ko, "Upstream Resource reservation and scheduling strategies for Hybrid Fiber/Coax Networks," APCC/OECC '99.

[98] H. Zhang, "Service Disciplines for Guaranteed Performance Service in Packet-switching Networks," Proc. IEEE, Vol. 83, pp. 1374–96, Oct. 1995.

Appendix A

UMTS QoS Support

The "3GPP Services and System Aspects Technical Specification Group" defined an end-to-end QoS architecture for UMTS (see Figure A.1) [2]. This QoS architecture defines a layered framework for supporting end-to-end QoS. Bearer Services are used to realize end-to-end QoS in different layers. QoS bearer services in each layer rely on services provided by lower layers. For detailed information on 3GPP QoS frame work, refer to [2].

3GPP defines the following four QoS traffic classes for UMTS.

Conversation Class: This class is aimed at supporting traffic with very strict delay bounds such as Voice over IP, and Video conferencing.

Streaming Class: Traffic with a bit less stringent delay requirement than the conversional class traffic can be handled by this class. Examples include voice, and video streaming applications.

Interactive Class: This class supports non-real-time applications such as WEB browsing. The delay requirement for this kind of traffic is loose but must be better than best-effort.

Background Class: This class targets best-effort traffic like E-mail.

In [3], a number of end-to-end QoS scenarios have been identified. Among these, scenario 2 which is based on DiffServ framework is the most supported scenario by most vendors. UMTS traffic classes can be mapped to DiffServ traffic classes at different UMTS network elements. For example, Conversational Class can be mapped to DiffServ Expedited Forwarding (EF) class. QoS mechanisms at the IP level can be used for transferring the UMTS traffic between UMTS network

Figure A.1 3GPP UMTS QoS architecture
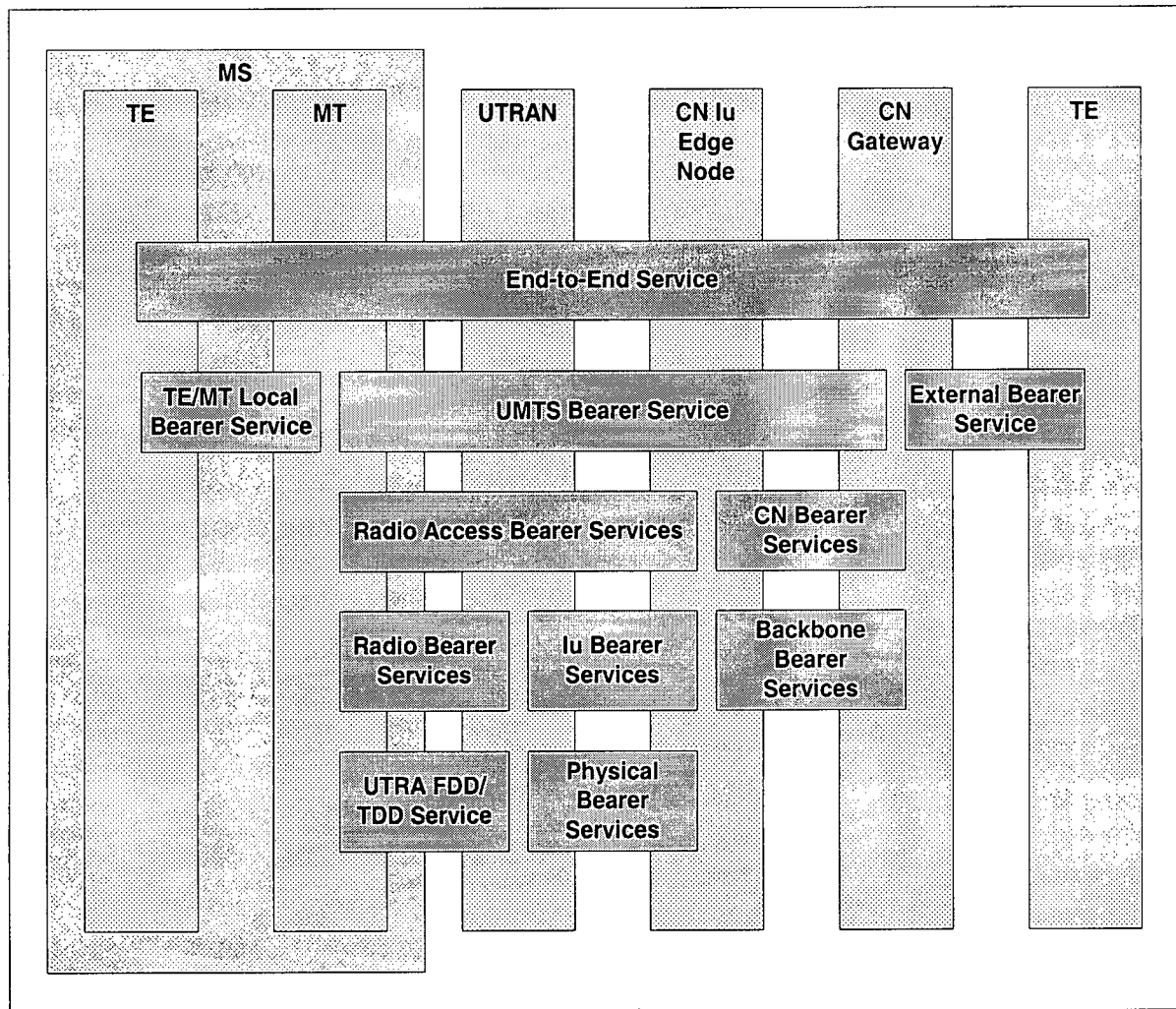
elements and inside the Internet.

3GPP standards define a QoS framework for UMTS but leaves specific implementation of QoS functions in different network elements to be implementation or vendor dependent. QoS mechanisms need to be deployed in the transmission, control, and management planes of UMTS. To ensure end-to-end capability, QoS must be deployed in the Core Network, the

UTRAN, and UE, in both the upstream and downstream. The following summarizes typical QoS functions in a UMTS network.

Transmission Plane: The transmission plane implements various traffic conditioning functions in the UTRAN, the SGSN, and GGSN. These include traffic classification, policing shaping, scheduling, and queue management. These are implemented in both the upstream and downstream. The transmission plane is responsible for maintaining QoS for local users as well as roaming users.

Control Plane: Admission control is implemented by UMTS SGSNs and GGSNs. Prior to admitting any new PDP context, the CN must determine if it can admit the requested QoS without degrading QoS for already activated PDP contexts. The QoS request can be negotiated between the CN and the UE until a compromise is reached. In addition to admission control, the UMTS network elements and underlying networks must provide the most superior QoS for signalling traffic. Dropped or delayed signalling traffic can have more serious impact on the offered QoS than that for data traffic.

Management Plane: The Management plane deploys a number of QoS mechanisms. These include QoS Configuration Management, Fault Management, Performance Management, and Policy Management. QoS Configuration Management can be performed by tweaking QoS parameters using a simple Command Line Interface (CLI). Alternatively, sophisticated interfaces can be used to transform Service Level Agreements (SLAs) to specific QoS parameters, and automatically configure UMTS network elements. Various implementation

dependent fault alarms and performance parameters can be generated and displayed to operators for fault and performance management. Policy Management is a new 3GPP release 5 issue [3]. The IETF Common Open Policy (COPS) framework has been adopted by 3GPP. The Go interface was introduced between the GGSN and a policy server for handling policy management tasks. Refer to [3] for more information on QoS Policy management.

Appendix B

DOCSIS QoS Upstream Scheduling Services

Upstream scheduling services are aimed at improving the efficiency of the poll/grant process. This enables DOCSIS MAC protocol to support latency and throughput guarantees for upstream traffic. DOCSIS MAC supports several scheduling services which are aimed at different traffic classes. The basic services supported by DOCSIS are summarized in the following [27]:

B.1 Unsolicited Grant Service (UGS)

The Unsolicited Grant Service (UGS) supports real-time Constant Bit Rate flows. It is aimed at supporting applications such as Voice over IP (VOIP) that generate fixed size packets on a periodic basis. UGS provides fixed size data grants on a real-time periodic basis. A CM uses the fixed size periodic grants for sending its data without the need for contending with other traffic. A CM using this service is prohibited from using any contention request or request/ data opportunities, and piggybacking bandwidth request. Also, the CMTS Should not provide any unicast request opportunities to a CM using this service. The key service parameters for UGS are the Nominal Grant interval, the Unsolicited Grant Size, and the Tolerated Grant Jitter (refer to [27] for more details).

B.2 Real-Time Polling Service

The Real-Time Polling Service (rt-PS) supports real-time variable Bit Rate flows. It is aimed at supporting applications that generate variable size packets on a periodic basis (e.g. MPEG video). The CMTS provides real-time, periodic, unicast request opportunities to CMs using this service. A CM can use these unicast request opportunities to send requests for bandwidth. This enables CMs with real-time variable size packet to meet their latency and throughput requirements. A CM using this service is prohibited from using any contention request or request/data opportunities, and piggybacking bandwidth request. The key service parameters for rt-PS are the Nominal Polling Interval, and the Tolerated Poll Jitter.

B.3 Unsolicited Grant Service with Activity Detection

The Unsolicited Grant Service with Activity Detection (UGS/AD) supports flows with ON/OFF Constant Bit Rate traffic such as Voice over IP with silence suppression. USG/AD is a combination of UGS and rt-PS with only one scheduling service active at a time. It offers real-time unsolicited grants during ON periods and real-time unicast polls during OFF periods. Inactivity detection by the CMTS is implementation dependent. One option is to have the CMTS detect flow inactivity by detecting unused grants. A CM using this service is prohibited from using any contention request or request/data opportunities, and piggybacking bandwidth request. The key service parameters for UGS/AD are the Nominal Grant interval, the Unsolicited Grant Size, the Tolerated Grant Jitter, Nominal Polling Interval, and the Tolerated Poll Jitter (refer to [27] for more details).

B.4 Non-Real-Time Polling Service

The Non-Real-Time Polling Service (nrt-PS) supports non real-time variable Bit Rate flows. It is aimed at supporting non real-time applications that generate variable size packets on a regular basis (e.g. high bandwidth FTP). The CMTS provides periodic or non-periodic (typically in the order of a second or less) unicast request opportunities to CMs using this service. A CM can use these unicast request opportunities to send requests for bandwidth. This enables CMs using this service to receive request opportunities even during congestion. A CM using this service is allowed to use contention request opportunities, unicast request opportunities and unsolicited data grants. The key service parameters for nrt-PS are the Nominal Polling Interval, Maximum Sustained Traffic Rate, Minimum Reserved Traffic Rate, Request/Transmission Policy and Traffic Priority.

B.5 Best Effort Service

The Best Effort (BE) service is designed to provide efficient service to best effort traffic. A CM using this service is allowed to use contention request opportunities. This will result in the CM using contention request opportunities as well as unicast request opportunities and unsolicited data grants. The key service parameters are the Maximum Sustained Traffic Rate, the Minimum Reserved Traffic Rate, and the Traffic Priority.

Appendix C

QoS Guarantees At The MAC Layer

The areas of Medium Access Control (MAC) in multiple access networks and Quality of Service (QoS) guarantees in general have been intensively investigated. However, a review study that addresses the area of MAC within the context of QoS guarantees is not currently available in the literature. This appendix discusses QoS guarantees at the MAC layers. It summarizes various MAC protocols in the literature into major categories. MAC protocols which are candidates for providing QoS guarantees are discussed. Challenges and possible solutions are presented.

The recent years have witnessed major technological advances in the area of multimedia communications. Several competing technologies have been introduced into this area. The major two competing technologies for providing multimedia communications are Asynchronous Transfer Mode (ATM), and Transmission Control Protocol/Internet Protocol (TCP/IP) suites of protocols (which is the standard de facto for the suite of protocols over the Internet). Apart from some exceptions, ATM is mainly deployed in the backbone networks. On the other hand, TCP/IP is used at the end systems and inside the Internet. The end-to-end path could be homogenous (one suite of protocols such as TCP/IP) or heterogeneous (involves different protocols such as TCP/IP and ATM). An end-to-end communication session could traverse heterogeneous networks such as fiber optic based SONET and satellite networks. ATM as a

connection oriented technology is well-engineered for guaranteed QoS in multimedia environments. Whereas TCP/IP suite of protocols have an inherited lack in this area. The problem of TCP/IP being failed to guarantee the QoS for Real Time (RT) traffic such as interactive voice and video is obvious to everyone who surfs the Internet. This fact motivated many researchers to look for solutions for guaranteeing QoS for multimedia applications over the Internet.

The problem of QoS guarantees is an end-to-end problem. In other words, QoS for multimedia applications must be guaranteed from application-to-application [15]. This implies that QoS mechanisms must be implemented at the end computers, at the access networks, and within the backbone networks, whether homogeneous or heterogeneous. These mechanisms must be implemented at different layers of a protocol stack. One such layer which is specific to multiple access networks is the MAC layer.

This appendix addresses QoS guarantees issues at the MAC layers in multiple access networks. The remainder of this appendix is organized as follows. QoS concepts, and QoS specifications are introduced in Section C.1, and Section C.2, respectively. MAC in general, and QoS guarantees at the MAC layer are discussed in Section C.3.

C.1 QoS Concepts

In the past few years, there has been much interest in the area of QoS. The term QoS is usually defined differently by different people in the computer networks community [46]. These multiple definitions often lead to multiple solutions for the problem of QoS guarantees. The

term Quality of Service consists of the two words: quality and service [46]. Within the networking community, the word quality is generally used to describe the process of delivering information with a degree or grade of excellence, whereas the word service is usually used to describe something offered to the end user of a given network (for example, video conferencing). The combination of the two words in the term Quality of Service leads to the following definition that we quote from [46]. "QoS is a measurement of how well the network behaves and an attempt to define the characteristics and properties of specific services".

C.2 QoS Specifications

QoS specifications are some parameters which are required by a certain application in order to deliver the data to the end user with a given quality. These required specifications are declared by the application and they are mapped at different layers down the protocol stack. The declaration process takes place at the beginning of a communication session and could be renegotiated during the session. This process is only of a declarative nature which isolates it from specific mechanisms for QoS assurance. QoS specifications include [15]: flow synchronization specification, flow performance specification, level of service, cost of service and QoS management. Here, we will only discuss flow performance specification as it is the most important specification at the MAC layer. For more information about other QoS specifications, please refer to [15].

Flow Performance Specification: These describe various required flow performance metrics in order to achieve the targeted quality for a certain application. These metrics include the

following [92]:

- *Throughput metrics,* which are described in terms of the required data rates metrics. These include: Peak Rate, Sustained Rate and Minimum Rate.

- *Delay metrics,* which include: Transfer Delay, Delay Variation and Delay Variation Tolerance.

- *Loss metrics,* the commonly specified metric under this category is the Loss Rate. It describes how much amount of lost data can be tolerated by the application?

C.3 QoS Guarantees At The MAC Layer

In general, computer networks can be divided into two categories: point-to-point and multiple access [92]. In point-to-point networks, there are many dedicated communication links between pairs of hosts. Communication between any pair of hosts may traverse one or more intermediate hosts. There are only two hosts connected by a dedicated link. Most backbone networks fall under this category. On the other hand, in multiple access networks, all hosts on the network share a single communication link. Data sent by any host can be received by all others. Also, two or more hosts can access the shared channel simultaneously which results in collisions. CATV and satellite networks are examples of this category.

Multiple Access networks can be classified into two classes: Centralized and Distributed. In the former, a central controller coordinates the operation of the network (e.g., CATV networks). In the later, the operation of the network is coordinated by various hosts attached to

the shared media (e.g., Ethernet LANs).

The role of a MAC protocol is to govern communications on the shared channel. It determines who has the right to access the channel and who can reveal the contents of a packet that is being receiving by all hosts. The most important issue in a multiple access networks is the bandwidth allocation. It is the job of the MAC protocol to allocate the bandwidth to different users effectively, fairly and promptly. There are many MAC protocols in the literature. These can be classified, according to their bandwidth allocation scheme, into two general categories: static and dynamic [92][65].

C.3.1 Static Allocation Protocols

In static allocation protocols, the available bandwidth is shared between users in a static fashion. Traditional MAC protocols such as Frequency Division Multiple Access (FDMA), and Time Division Multiple Access (TDMA) fall under this category. FDMA divides the available frequency spectrum between available users. On the other hand, TDMA divides the time into time slots, where users take turns in accessing these slots in a round robin fashion. Each user has the privilege to use the whole available bandwidth during its allocated time slot. Static schemes can provide guaranteed QoS for all accommodated users, but they are bandwidth inefficient [92].

C.3.2 Dynamic Allocation Protocols

Dynamic allocation protocols assign bandwidth in the shared channel to available users in a dynamic fashion. These schemes can be further divided into two subcategories: random

assignment, and demand assignment. These two subcategories are explained in the following sections.

a) Random Assignment

In random assignment schemes, all active users contend for accessing the channel according a predefined random access mechanism. The access time is either divided into time slots (slotted time systems) or left undivided (continuous time systems). In slotted time systems, a user can start transmission only at the beginning of a slot. An example of such protocol is Slotted ALOHA. On the other hand, in continuous time system, data transmission can start at any time [92]. An example of such protocol is Pure ALOHA.

Whenever a user has some data to send, it contends with other users by starting transmitting its data (in the form of variable length packets or ATM cells, however, sometimes we use the generic term packet to represent either format), either at any instant (continuous systems) or at the beginning of a slot (slotted systems). The user could also sense the channel to make sure that no other user(s) is/are using the channel (Carrier Sense Multiple Access (CSMA)). The transmissions from two or more users could overlap resulting in collisions. Collisions can be detected either by the transmitting stations themselves or by a central station that controls the access. In any case, colliding stations must retransmit their garbled packets after a random period, depending on the used Contention Resolution Algorithm (CRA). The fact that the contention resolution process could impose large delays make random assignment protocols inappropriate for QoS guarantees over multiple access networks [65]. This drawback is more

sever in networks with large propagation delays such as satellite networks (270 ms for GEO satellites).

b) Demand Assignment

Demand assignment protocols assign bandwidth to users based on their demand. Stations express their needs for bandwidth either by sending special messages called requests to a central bandwidth scheduler (centralized networks), or by announcing to other users their reservation of certain bandwidth (distributed networks). In the former, the central controller allocates bandwidth to stations depending on their requests and resources availability. In the later, all station other than the requesting station abstain from using the reserved bandwidth. Centralized demand assignment based MAC protocol are very common in the literature, and hence we focus our discussion on this category.

In general, centralized demand assignment protocols require a master clock for providing references for the start of time slots. The shared channel is divided into an upstream path (from stations to the central controller), and a downstream path (from the central controller to the stations). This is done using either Frequency Division Duplex (FDM) or Time Division Duplex (TDM) [88]. In FDM, each path is assigned a different frequency spectrum, whereas in TDM, both paths time share the whole frequency spectrum. Each path is further divided into time slots. The downstream path is a one to many communication environment, whereas the upstream is a many to one. In other words, The central controller broadcasts data in the downstream to all stations in a collision free environment, whereas in the upstream, many

166

stations contend for sending data to the central controller. QoS guarantees in the contention based upstream is a challenging task.

As mentioned above, the upstream channel is divided into time slots which are referenced by a master clock. The slot length is protocol dependent. Some proposals consider long slots that can accommodate a data packet (for example an ATM cell), others consider mini-slots (for example 8 bytes) that can only carry short messages such as requests. A group of mini-slots can be concatenated to accommodate a data packet.

A channel could be framed or frameless. In the former, a group of slots or mini-slots can form a frame of a certain duration (for example 6 ms), where the potential usage of every slot in the frame is announced by the central controller regularly. In the later, slots or mini-slots are numbered from 0 to an agreed upon maximum. The central controller regularly announces the potential usage of every slot or mini-slots within a given future time period. Figure C.1 [41] shows a typical upstream framed structure based on mini-slots. Normally, the frame is partitioned into subframes namely: a request subframe, and a data subframe. Each subframe is defined and announced by the central controller. The request subframe is used by stations for sending requests for bandwidth, whereas the data subframe is used for transporting users' data.

Examples of demand assignment based protocols include: Packet Reservation Multiple Access (PRMA) [49] and its enhancements such as Centralized (C-PRMA), Aggressive (APRMA) [60], Minipackets (MPRMA) [63], Data Steal over Voice (PRMA-DSV) [63], and

Request Subframe   **Data Subframe**

Data packet

Request mini-slot
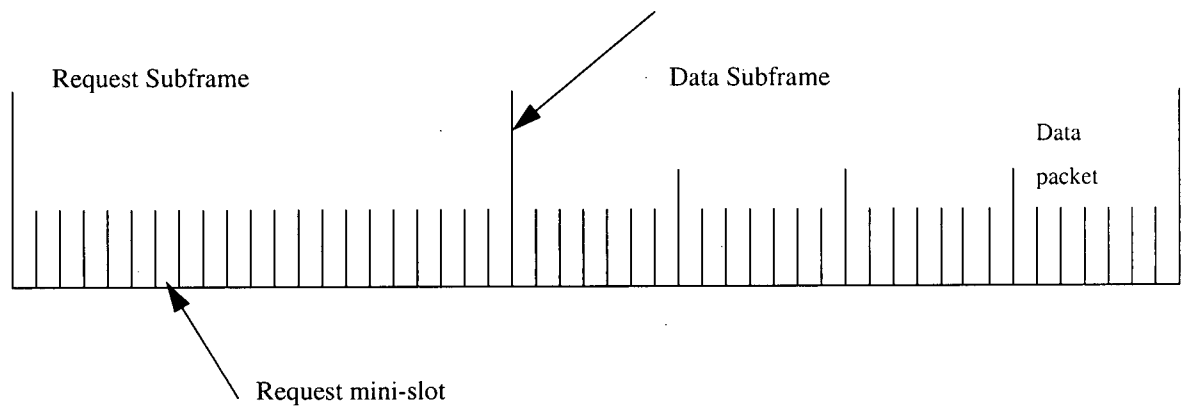
Figure C.1 MAC frame structure

PRMA/Dynamic Allocation (PRMA/DA) [61]. Other examples include Resource Auction Multiple Access (RAMA) [12], Distributed Queueing Request Update Multiple Access (DQ-RUMA) [58], Dynamic Reservation Multiple Access (DRMA) [81],and Multiservice Dynamic Reservation (MDR-TDMA) [84].

Demand assignment MAC protocols are the best candidates for providing QoS guarantees because they are based on guaranteed allocation of bandwidth to users depending on their demand. However, these protocols must be capable of providing priorities among different traffic classes. They must enable high priority applications such as real time Variable Bit Rate (rt-VBR) to communication their requests for bandwidth immediately. Also, the bandwidth scheduler must allocate the requested bandwidth for such applications and inform them via grant messages immediately.

C.3.3 QoS Guarantees at Different Phases of MAC Operation

As mentioned in Section C.3.2, demand assignment based MAC protocols are the potential candidates for providing QoS guarantees. In general, the operation of a demand assignment based MAC protocol can be divided into six phases: registration and synchronization, bandwidth request, feedback, contention resolution, bandwidth assignment, and data transfer. In order to provide QoS guarantees, different mechanisms must be implemented during some of these phases of operation. We discuss QoS guarantees within some of these phases in the following.

1) Bandwidth Request Phase

As mention earlier, stations express their needs for bandwidth using requests. Requests could be sent using contention, polling, piggybacking, or a combination of some or all of them. All of these types of communicating requests except for piggybacking are sent in the request subframe (see Figure C.1). These methods of communicating requests are explained in the following.

- **Piggybacking:** In this case, requests are sent in the data subframe using a special field in the data packets. This approach offers lower overhead and minimizes the time needed to send requests and hence provides better QoS. However, piggybacking is only possible if the station has already made reservations. Otherwise, the station needs to communicate its requests via other methods.

- **Contention:** When the method of communicating requests is contention (for example

one of the methods in 802.14 MAC protocol [56]), a stations sends its request in the predefined contention based request subframe. If this request is in collision (the station is informed by the central controller via acknowledgment messages), then the station would apply a CRA. If this request is successful, the station waits for a grant message to be sent by the central controller. The grant message informs the station of which reservation mini-slots (in the data subframe) are allocated to the station by giving these mini-slots numbers. The station can use the allocated reservation mini-slots to send its queued data and to piggyback requests. If contention is the only method for communicating request, the MAC protocol would fail to guarantee QoS for delay sensitive traffic. This is because of the unbound delay characteristics of the used CRAs.

- **Polling:** When the method of communicating requests is polling, the central controller polls stations regularly by allocating them reserved mini-slots in the request subframe. A station would use its reserved request mini-slots to send its requests. If the station has no requests, the reserved mini-slots are wasted. This approach gives guaranteed QoS at the expense of more overhead.

- **A Combination of the above:** Two or all of the above methods could be used to communicate requests. Piggybacking is always used with one or both of the other two methods. A Combination of contention and polling is also used. For example, contention could be used under light loads while polling could be used under heavy loads. Other combinations are also possible.

2) Feedback Phase

The central controller must provide feedback regarding the status of transmissions in the previous slots. The feedback could be binary (collision, success) as in binary tree mechanisms, or ternary (collision, success, empty) as in ternary tree mechanisms [19]. Also, the central controller should process successfully received requests, allocate bandwidth to stations (this process will be explained later), and inform stations of these allocations via grant messages. In order to provide QoS guarantees, both types of feedback should be sent by the central controller at the earliest time possible. Fast feedback regarding collisions allow stations to start the contention resolution process earlier which shortens the contention resolution time. On the other hand, early received grant messages will enable stations to start their data transmissions at the earliest possible time. The time saved when using fast feedback contributes to the overall guarantees for QoS.

3) Contention Resolution Phase

As mentioned earlier, the request messages from two or stations could collide. In this case, a CRA must be applied to resolve the contention. The CRA algorithm is part of the Random Access Algorithm (RAA). In order to provide QoS guarantees, the RAA must be stable, fast in resolving collisions, bandwidth efficient, and priority based. These characteristics are discussed in the following:

- **Stability:** It implies that for a network operating under a RAA, if the new arrival rate (λ) is less than or equal to the maximum throughput of the RAA (Λ), λ must be equal

to the departure rate (μ) [19]. A RAA is called unstable if the underlaying network has $\lambda \leq \Lambda$ while $\lambda > \mu$. This is the result of many unresolved collisions that if persist result in a deadlock. Unstable RAA are not suitable to work in networks which are aimed at providing QoS guarantees. The purely acknowledgment based ALOHA family of algorithms have been proved to be unstable under heavy loads [11][18][19][62]. However, ALOHA can be stabilized by dynamically changing the retransmission probability as the backlogged stations change [19]. On the other hand, splitting CRA such as ternary algorithms are stable and, therefore, they are the candidate algorithms for QoS guarantees.

- **Speed in resolving collisions:** This means that the CRA must be very prompt in resolving collisions. This characteristic is important to all types of traffic. However, delay sensitive traffic is the major beneficent. One way of expediting the CRA is to use ternary feedback based mechanisms. These tend to resolve contention faster than the binary feedback based ones. This is because a feedback regarding an empty slot indicates that stations allocated to that slot are idle, and hence, allows the CRA to skip sure collisions between the remaining stations by splitting them immediately [19]. A second method of expediting the CRA is to dynamically adjust the Capturing probability. This adjustment could be made based on the number of backlogged stations, their activity factor, the number of contention slots, or a combination of them similar to that in [30]. A third method of expediting the CRA is to assign many slots as contention slots. One way of doing this is to convert all slots that are not reserved for data into

contention slots [41][81].

- **Bandwidth efficiency:** This means that the RAA must have a high throughput with a minimal bandwidth usage. A CRA could assign a great number of contention slots for colliding station so that contention can be resolved promptly. However, this approach could waste bandwidth. The number of contention slots assigned for new arrivals as well as the ones used by the CRA must be optimized while maintaining high throughput and minimum delay.

- **Priority handling:** A CRA aimed at proving QoS guarantees must deploy a priority based mechanism for contention resolution. In resolving contentions, delay sensitive traffic must be given higher priority than delay insensitive traffic. One way of achieving this is to partition contention slots into a number of clusters. Each class of applications is assigned a specific cluster. This allows the applications to be given preference during the initial contention period as well as after collisions. High priority traffic could be assigned more contention slots than other traffic. Also, the CRA could start resolving collisions among high priority traffic before other types of traffic as in [34].

4) Bandwidth Assignment Phase

When request are successfully received at the central bandwidth scheduler, the bandwidth assignment phase starts. The bandwidth scheduler receives request for bandwidth from stations using any of the methods mentioned earlier in this section. Then, it sorts the requests into different queues according to their priorities. Finally, it checks the available bandwidth

and assigns bandwidth to stations based on their requested amount of traffic and its class, and the available bandwidth.

In order to provide QoS guarantees, bandwidth assignment must be done quickly, fairly, and on a priority basis. In a network of a great number users, each with different traffic classes, the bandwidth scheduler should perform its task on the fly using very fast parallel processing architectures. In assigning bandwidth, priority should be given to delay sensitive traffic while being fair to delay tolerant traffic. An example of a priority based bandwidth scheduler that has been proposed for use in IEEE 802.14 based CATV networks is detailed in [36].

Appendix D

UTRAN Interfaces Protocol Stacks

In this appendix, we present UTRAN protocol stacks related to this thesis. Specifically, we present simplified protocol stacks for Iu-b, Iu-r interfaces. These are essential for understanding the proposed UMTS over DOCSIS CATV network architecture presented in Chapter 3. For detailed information on these interface, please refer to [6], [7].

D.1 Iu-b Interface

The Iu-b interface interconnects a Node B to an RNC within a UTRAN. Figure D.1 shows a simplified version Iu-b Interface protocol stack [7].

D.2 Iu-r Interface

Within a UTRAN, the Iu-r interface interconnects RNCs of different Radio Network Subsystems. A simplified protocol stack for the Iu-r Interface is shown in Figure D.2 [6].

Node B   **RNC**

Frame Protocol for RACH, CPCH [FDD], FACH, DCH, DSCH, USCH [TDD]
UDP
IP
Data Link
PHY

Frame Protocol for RACH, CPCH [FDD], FACH, DCH, DSCH, USCH [TDD]
UDP
IP
Data Link
PHY

-----------------Iu-b-----------------
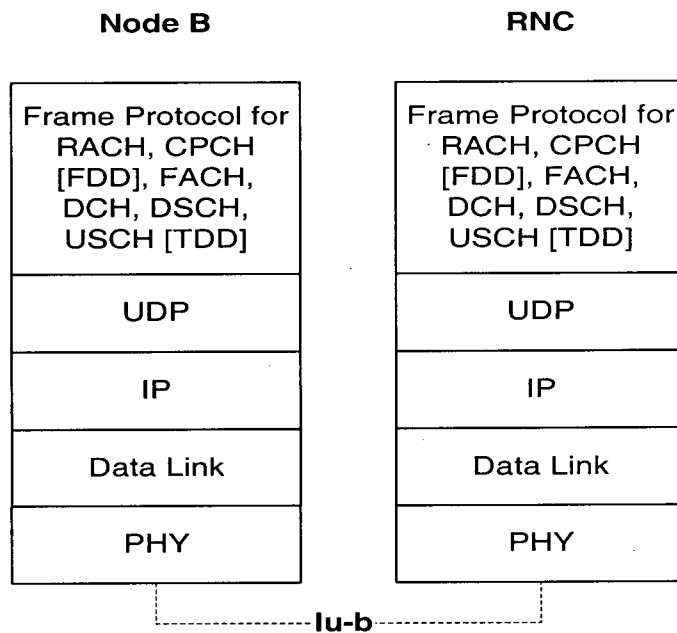
Figure D.1 A simplified Iu-b Interface protocol stack

RNC   **RNC**

Frame Protocol for RACH, CPCH [FDD], FACH, DCH, DSCH, USCH [TDD]
UDP
IP
Data Link
PHY

Frame Protocol for RACH, CPCH [FDD], FACH, DCH, DSCH, USCH [TDD]
UDP
IP
Data Link
PHY

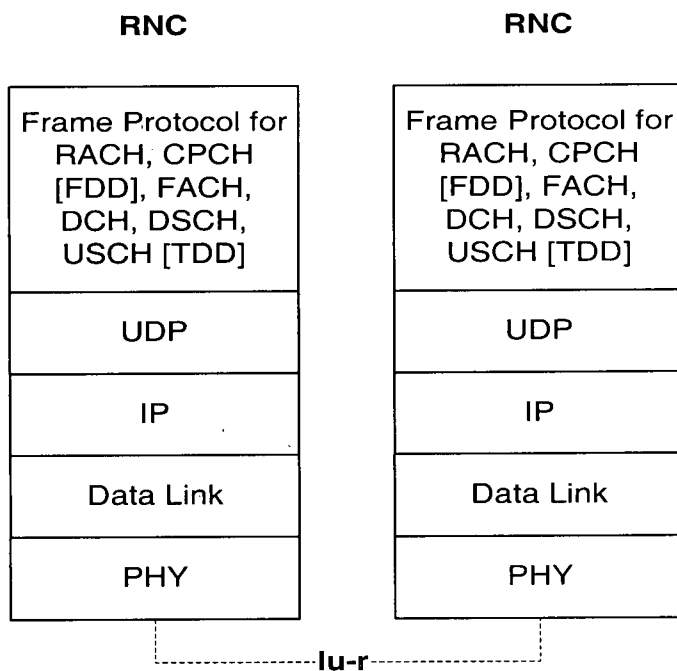-----------------Iu-r-----------------

Figure D.2 A simplified Iu-r Interface protocol stack

176

Appendix E

Proposed MAC Additional Performance Results

As mentioned in Section 4.12 our proposed DOCSIS MAC enhancements offers the lowest average delay for delay-sensitive traffic without compromising the performance requirements of other traffic classes, or the total throughput of DOCSIS upstream channels.

In this appendix, we show addition performance results that compare the delay, and throughput offered by our proposed MAC to that offered by existing DOCSIS MAC, as well as an implementation by Broadcom Inc. The performance of the above three DOCSIS MAC alternatives has already been discussed in Section 4.11. This appendix only shows performance results for comparison reasons.

Figure E.1 compares the average access delay for Best-effort traffic offered by our proposed MAC with that offered by other DOCSIS MAC implementations. The throughput for delay-sensitive traffic, Best-effort traffic, and total throughput is compared in Figure E.2, Figure E.3, and Figure E.4, respectively.
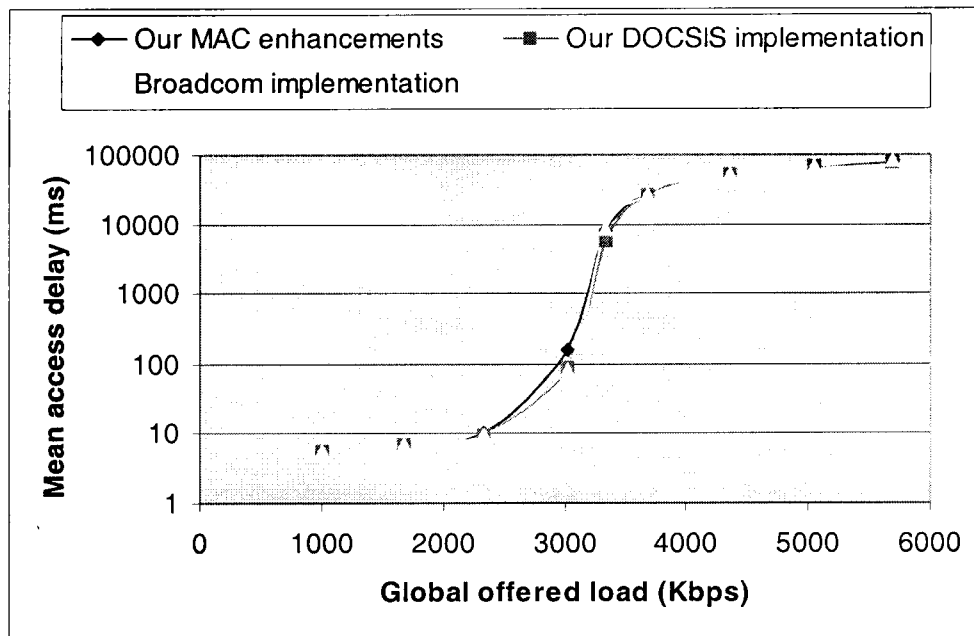
Figure E.1 Access delay for best-effort traffic using different MAC implementations
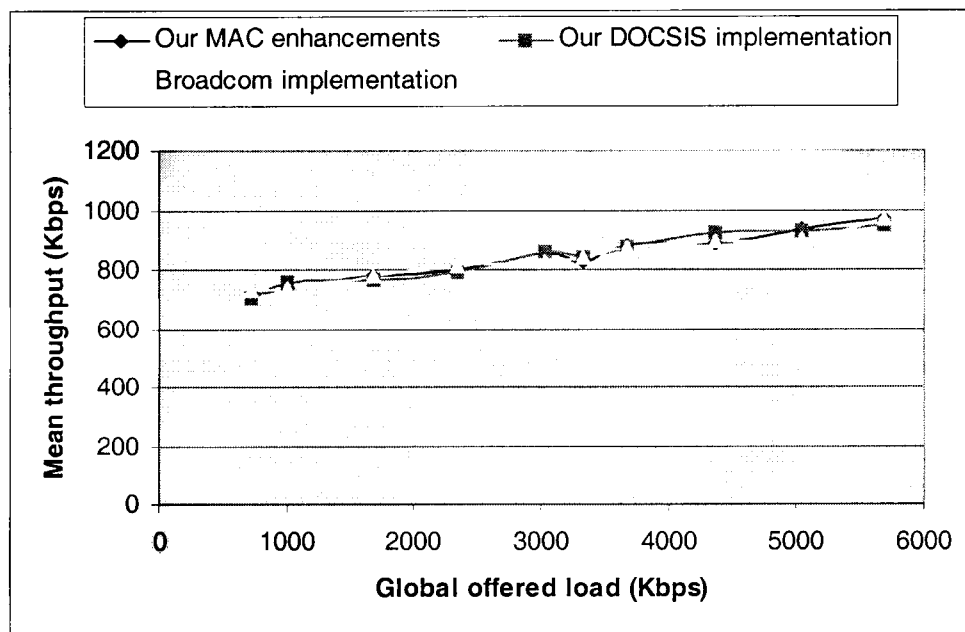

Figure E.2 Throughput for delay-sensitive traffic using different MAC implementations
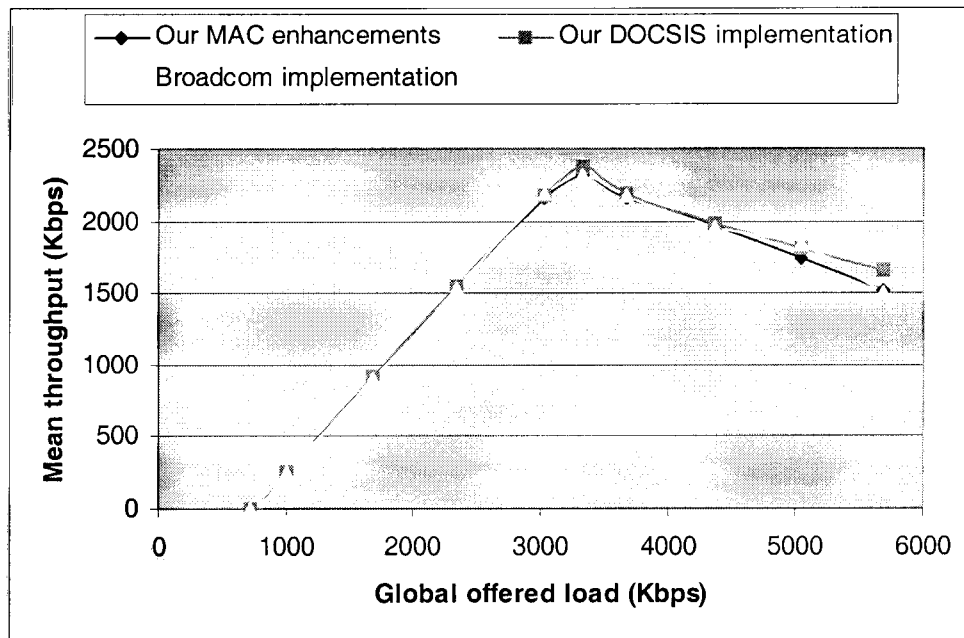
178

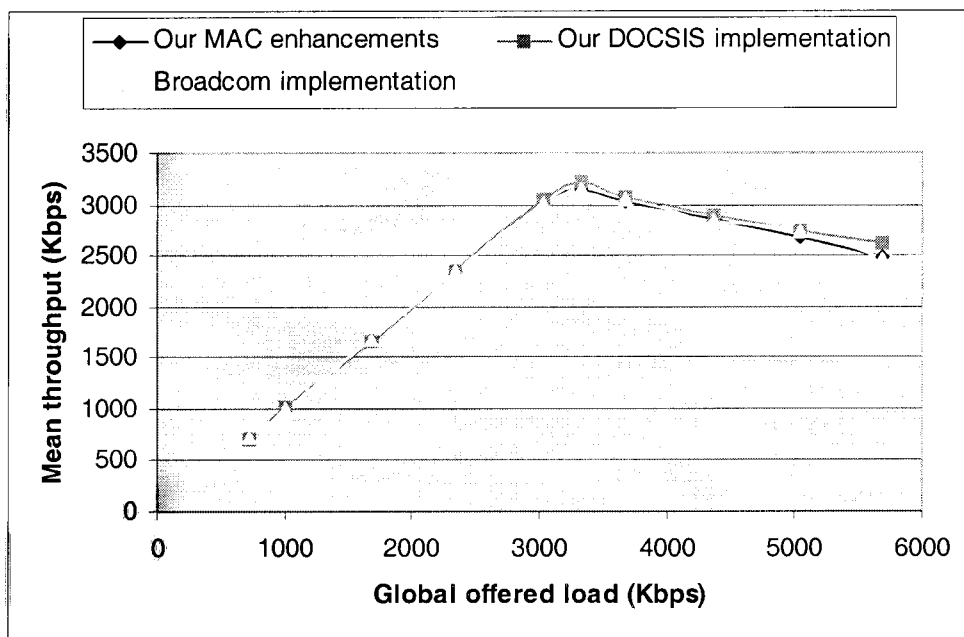Figure E.3 Average throughput for best-effort traffic using different MAC implementations

Figure E.4 Average total throughput using different MAC implementations

179

Glossary

2G	Second Generation
3G	Third Generation
3GPP	Third Generation Partnership Project
AAA	Authentication, an Authorization and Accounting
AAL	ATM Adaptation Layer
AAL5	ATM Adaptation Layer type 5
ACK	Acknowledgement
AMPS	Advanced Mobile Phone System
AP	Access Point
APN	Access Point Name
ATM	Asynchronous Transfer Mode
BE	Best Effort
BS	Base Station
BSS	Basic Service Set
CAC	Call Admission Control
CATV	Cable Television
CBR	Constant Bit Rate
CDMA	Code Division Multiple Access
CDR	Call Detail Record
CGF	Charging Gateway Functionality
CM	Cable Modem
CMTS	Cable Modem Termination System
CN	Core Network
C-PRMA	Centralized Packet Reservation Multiple Access
CRC	Cyclic Redundancy Code
CS	Circuit Switched
CSF	Common Simulation Framework

CSMA	Carrier Sense Multiple Access
CSMA/CD	Carrier Sense Multiple Access with Collision Detection
DHCP	Dynamic Host Configuration Protocol
DLL	Data Link Layer
DNS	Domain Name System
DOCSIS	Data Over Cable Systems Interface Specifications
DoS	Denial of Service
DS	Distribution System
DSCP	DiffServ Codepoint
D-TDMA	Bandwidth on Demand-Time Division Multiplexing
EDGE	Enhanced Data Rates for Global Evolution
ESS	Extended Service Set
ETSI	European Telecommunications Standards Institute
FCFS	First Come First Serve
FDD	Frequency Division Duplex
FDM	Frequency Division Multiplexing
FDMA	Frequency Division Multiple Access
FN	Fiber Node
GGSN	Gateway GPRS Support Node
GMM/SM	GPRS Mobility Management and Session Management
GPRS	General Packet Radio Service
GPS	Generalized Processor Sharing
GSM	Group Special Mobile/Global System of Mobility
GSN	GPRS Support Node
GTP	GPRS Tunnelling Protocol
GTP-C	GTP Control Plane
GTP-U	GTP User Plane

HE	Headend
HFC	Hybrid Fiber Coaxial
HLR	Home Location Register
IBSS	Independent Basic Service Set
IE	Information Element
IETF	Internet Engineering Task Force
IMSI	International Mobile Subscriber Identity
IP	Internet Protocol
IPv4	Internet Protocol version 4
IPv6	Internet Protocol version 6
Kbps	Kilobits per second.
LAN	Local Area Networks
LLC	Logical Link Control
MAC	Medium Access Control
Mbps	Megabits per second
MCNS	Multimedia Cable Network System
M-MAC	Master-Medium Access Control
MPRMA	Mini Packets Reservation Multiple Access
N-PDU	Network Protocol Data Unit
nrt-PS	Non-real-time Polling Services
OF	Optical Fiber
OMC	Operation and Management Center
PCN	Personal Communication Networks
PCS	Personal Communication Services
PCU	Packet Control Unit
PDCH	Packet Data CHannel
PDCP	Packet Data Convergence Protocol

PDN	Packet Data Network
PDP	Packet Data Protocol, e.g. IP
PDU	Protocol Data Unit
PGPS	Packet-by-packet Generalized Processor Sharing
PHB	Per-hop Behavior
PHY	Physical Layer
PRMA	Packet Reservation Multiple Access
PSTN	Public Switched Telephone Network
PTN	Personal Telecommunication Number
QAM	Quadrature Amplitude Modulation
QoS	Quality of Service
QPSK	Quadrature Phase Shift Keying
RA	Request Access
RNC	Radio Network Controller
RNS	Radio Network Subsystems
RTP	Real-time Polling
rt-PS	Real-time Polling Services
S-Aloha	Slotted-Aloha
SGSN	Serving GPRS Support Node
SID	Station ID
SM	Short Message
S-MAC	Slave-Medium Access Control
SMS	Short Messaging Service
SRNC	Serving RNC
SRNS	Serving RNS
STA	Station
TCP	Transmission Control Protocol

TDD	Time Division Duplex
TDM	Time Division Multiplexing
TDMA	Time Division Multiple Access
TLLI	Temporary Logical Link Identity
TMSI	Temporary Mobile Subscriber Identity
TOS	Type of Service
UDP	User Datagram Protocol
UE	User Equipment
UGS	Unsolicited Grant Service
UGS-AD	Unsolicited Grant Service with Activity Detection
UMTS	Universal Mobile Telecommunications System
UTRAN	UMTS Terrestrial Radio Access Network
VLR	Visitor Location Register
VOIP	Voice over IP
WAN	Wide Area Network
WFQ	Weighted Fair Queueing
WLAN	Wireless Local Area Network
WRED	Weighted Early Random Detection

Symbols

Ga	Charging data collection interface between a CDR transmitting unit (e.g. an SGSN or a GGSN) and a CDR receiving functionality (a CGF).
Gb	Interface between an SGSN and a BSS.
Gc	Interface between a GGSN and an HLR.
Gd	Interface between an SMS-GMSC and an SGSN, and between an SMS-IWMSC and an SGSN.
Gf	Interface between an SGSN and an EIR.
Gi	Reference point between GPRS and a packet data network.
Gn	Interface between two GSNs within the same PLMN.
Gp	Interface between two GSNs in different PLMNs. The Gp interface allows support of GPRS network services across areas served by the co-operating GPRS PLMNs.
Gr	Interface between an SGSN and an HLR.
Gs	Interface between an SGSN and an MSC/VLR.
Iu	Interface between the RNS and the core network. It is also considered as a reference point.
Uu	Interface between the mobile station (MS) and the Iu mode network. The Uu interface is the Iu mode network interface for providing GPRS services over the radio to the MS.