

**Unified Multiservice All-Packet-Based Broadband Access
Architecture: An Evolutionary Path Towards a Fixed Mobile
Convergence Solution**

by

Sherif R. Sherif

A dissertation submitted to the Graduate Faculty in Engineering in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2008

UMI Number: 3303792

Copyright 2008 by
Sherif, Sherif R.

All rights reserved.

UMI[®]

UMI Microform 3303792

Copyright 2008 by ProQuest Information and Learning Company.
All rights reserved. This microform edition is protected against
unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

© 2008

SHERIF R. SHERIF

All Rights Reserved

This manuscript has been read and accepted for the Graduate Faculty in Engineering in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

_____	Prof. Mohamed Ali
Date	Chair of Examining Committee
_____	Dean. Mumtaz Kassir
Date	Executive Officer

Prof. Samir Ahmed (EE Dept, CCNY, CUNY) _____

Prof. Roger Dorsinville (EE Dept, CCNY, CUNY) _____

Prof. Neophytos Antoniadis (Engineering Dept, College of Staten Island, CUNY) _____

Dr. Rakesh Bakshi (Alcatel-Lucent, RF System Performance) _____
Supervision Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

Unified Multiservice All-Packet-Based Broadband Access Architecture: An Evolutionary Path towards a Fixed Mobile Convergence

by

Sherif R. Sherif

Adviser: **Professor Mohamed Ali**

Mobile networks and services have gone beyond voice-only communication services and are rapidly evolving towards data-centric services. Emerging mobile data services are expected to see the same explosive growth in demand that internet and wireless voice services have seen in recent years. Demand for data-centric mobile services including web surfing, IPTV, music and video downloads, and other streaming services, will emerge as an equally significant market driver as that of the traditional mobile voice services. With this trend, mobile networks will experience the same paradigm shift that has already occurred in wireline networks, namely the shift from voice-centric to data-centric networks.

In anticipation of this paradigm shift, mobile operators are currently planning to migrate their existing TDM-based mobile network infrastructure to a fully packet-based network. This migration is mainly to support the emerging 3G/4G data-centric services along with the diverse QoS and rate requirements set by these services faces several technical as well as economical hurdles.

To address some of these key challenges, this thesis argues for an immediate migration from current TDM-based mobile environment to fully packet-based mobile networks with a fiber based access infrastructure. Specifically, this work proposes and devises cost-effective multiservice Ethernet-based mobile backhaul networking architecture with a fiber based access infrastructure. The proposed architecture extends the reach of Ethernet access to base stations and provides a scalable mobile backhaul infrastructure that can accommodate emerging 3G/4G and legacy traffic as well as subscriber growth.

Rather than deploying a totally new packet-based mobile infrastructure, the proposed architecture utilizes the already existing EPON access infrastructure, this enables a seamless migration path to fixed/mobile convergence and creates the potential of supporting several powerful and cost-effective packet-based fixed/mobile access architectural models.

This work will be divided into two main sequential phases. In the first phase, we propose a novel EPON architecture that uses a fully distributed arbitration scheme that supports differentiated QoS. In the second phase, based on the developed decentralized EPON infrastructure, several powerful and cost-effective packet-based fixed/mobile access architectural models are presented. The performances of each model are compared with each other and with that of traditional RAN model.

Acknowledgements

I would like first to express sincere admiration and respect to my supervisor Professor Mohamed A. Ali who has provided me with invaluable academic advice. He devoted much of his time to discussing details of this research with me and helping me in so many other ways. I have been greatly influenced by his methodology in the field of scientific research, and have benefited from his detailed comments on every aspect in this thesis.

I would also like to thank the Electrical Engineering faculty and specially the examining committee for their support and for giving me the chance to achieve my goals.

I would like to thank my family for their continuous support. They were always there and gave me the love and support when I needed them the most. Special thanks to my father Professor Ragaei Sherif for being my inspiring figure.

Finally, Very special thanks for my wife and my lovely kids who helped me a lot in so many ways. Without my wife support, this work would not have been possible.

This dissertation is dedicated to my late mother, my father, and my loving wife.

TABLE OF CONTENTS

1	CHAPTER 1: INTRODUCTION	2
1.1	Introduction	2
1.2	Thesis Motivation	6
1.3	Thesis Statement.....	10
2	CHAPTER 2: OVERVIEW OF ETHERNET PASSIVE OPTICAL NETWORKS (EPON) AND ARCHITECTURES	17
2.1	Introduction	17
2.2	Overview of Ethernet-over-PON Technologies and Architectures	21
3	CHAPTER 3: PROPOSED DISTRIBUTED EPON ARCHITECTURE	25
3.1	The Proposed Distributed Architecture	25
3.2	Detailed Proposed Cycle	28
4	CHAPTER 4: PROPOSED INTEGRATED SCHEDULING AT ONU AND DYNAMIC BANDWIDTH ALLOCATION SCHEMES	32
4.1	Introduction	32
4.2	Queue Management with Priority Queuing (Intra-ONU Scheduling).....	33
4.2.1	Strict Priority Scheduling	34
4.2.2	Fair Non-Strict Priority Scheduling.....	36
4.3	Dynamic Bandwidth Allocation at the ONU (Intra-ONU Scheduling).....	36
4.3.1	Overview of Previous/Current DBA Schemes	36
4.3.2	Decentralized Dynamic Bandwidth Allocation Schemes at the ONU.....	38
5	CHAPTER 5: DECENTRALIZED EPON PERFORMANCE EVALUATION	44
5.1	Introduction	44
5.2	Numerical Results without CoS Support.....	45
5.3	Numerical Results for Differentiated Classes of Service.....	49
5.3.1	Traffic Model for Differentiated CoS	49
5.3.2	Detailed Numerical Results for Differentiated CoS	51
5.4	Conclusion.....	62
6	CHAPTER 6: NOVEL WIRELESS BACKHAUL ARCHITECTURE	64

6.1	Introduction	64
6.2	Background and Motivation.....	65
6.3	Overview of Radio Access Networks and Traditional Wireless Backhaul architecture	66
6.4	Proposed Wireless Backhaul Architectures	68
6.4.1	Proposed Standalone Wireless Backhaul Architectures	71
6.4.1.1	Smooth Transitional Wireless Backhaul Standalone Architecture for 2G/3G and 3G+	71
6.4.1.2	Fully Packet-Based Standalone Wireless Backhaul Architecture.....	74
6.4.2	Proposed Integrated Wired/Wireless Backhaul Architecture	76
6.5	Bandwidth Allocation Schemes for Proposed Architecture	78
6.5.1	Decentralized Architecture Bandwidth Allocation Schemes	78
6.5.1.1	Decentralized Uplink DBA Scheme	79
6.5.1.2	Decentralized Downlink DBA Scheme	81
6.5.2	Centralized Architecture Bandwidth Allocation Schemes.....	81
6.5.2.1	Centralized Uplink DBA Scheme.....	82
6.5.2.2	Centralized Downlink DBA Scheme.....	82
6.6	Wireless Classes of Services (CoS) and Traffic Characteristics.....	83
6.7	Wireless Backhaul QoS Traffic Model	86
7	CHAPTER 7: PROPOSED WIRELESS BACKHAUL PERFORMANCE EVALUATION.....	90
7.1	Introduction	90
7.2	Numerical Results for the Evenly Loaded Base Stations Scenarios.....	92
7.3	Numerical Results for the Unevenly Loaded Base Stations Scenarios.....	95
7.3.1	Uplink Numerical Results	96
7.3.1.1	Uplink: Centralized EPON Wireless BH.....	96
7.3.1.2	Uplink: Decentralized EPON Wireless BH	101
7.3.2	Downlink Traffic Numerical Results.....	106
7.3.2.1	Downlink: Centralized EPON Wireless BH.....	106
7.3.2.2	Downlink: Decentralized EPON Wireless BH	110
8	CHAPTER 8: CONCLUSION.....	113
9	REFERENCES.....	116
10	PUBLICATIONS	120

LIST OF FIGURES

Figure 2.1: EPON Downstream overview.....	23
Figure 2.2: EPON Upstream overview	23
Figure 3.1: Proposed Distributed EPON Architecture.....	26
Figure 3.2: (a) Cycle updating process (b) Transmission process.....	28
Figure 4.1: Illustrative example of a scheduler’s operation.....	35
Figure 5.1: Average Packet Delay for Centralized and Decentralized architectures with various DBAs	46
Figure 5.2: Maximum Packet Delay for Centralized and Decentralized architectures with various DBAs.	47
Figure 5.3: Improvement in packets dropped using Remainder DBA.	49
Figure 5.4: Average packet delay for various classes of services.....	52
Figure 5.5: Maximum packet delay for various classes of services as a function of network Load for the IPACT Limited DBA and the Decentralized Limited DBA	53
Figure 5.6: Class P0 average packet delay as a function of total network load for the centralized IPACT and the decentralized architecture under various DBAs.	55
Figure 5.7: Class P1 average packet delay as a function of total network load for the centralized IPACT and the decentralized architecture under various DBAs.	56
Figure 5.8: Class P2 average packet delay as a function of total network load for the centralized IPACT and the decentralized architecture under various DBAs.	57
Figure 5.9: Class P0 Maximum packet delay as a function of total network load for the centralized IPACT and the decentralized architecture under various DBAs.	58
Figure 5.10: Class P1 Maximum packet delay as a function of total network load for the centralized IPACT and the decentralized architecture under various DBAs.	59
Figure 5.11: Class P2 Maximum packet delay as a function of total network load for the centralized IPACT and the decentralized architecture under various DBAs.	60
Figure 5.12: Improvement in the number of Class P2 packets dropped using Remainder DBA.....	61
Figure 6.1: Traditional Wireless Backhaul Architecture	68
Figure 6.2: Proposed EPON Wireless Backhaul Architecture	71
Figure 6.3: Proposed 2G/3G/3G+ Smooth Transitional Wireless Backhaul Architecture	73
Figure 6.4: Base Station Admission Control	74
Figure 6.5: Proposed Standalone Fully Packet-Based Wireless Backhaul Architecture	75
Figure 6.6: Proposed RAN Architecture OLT/RNC non co-located.....	76

Figure 6.7: Proposed Integrated Wired/Wireless Backhaul Architecture.....	78
Figure 6.8: Uplink RLP Mechanism.....	85
Figure 6.9: Downlink RLP Mechanism	86
Figure 7.1: Evenly Loaded BSs Network Utilization EPON BH vs. Traditional BH.....	93
Figure 7.2: Evenly Loaded BSs Network Utilization at 0.83 Network Load.....	94
Figure 7.3: Evenly Loaded BSs Network Utilization statistics at Network Load=0.83 ..	94
Figure 7.4: Uplink Utilization: Centralized EPON Wireless BH vs. Traditional BH	97
Figure 7.5: Uplink Utilization statistics at Network Load=0.24	97
Figure 7.6: Uplink Utilization Statistics at Network Load=0.83 (Centralized EPON BH vs. Traditional BH)	98
Figure 7.7: Uplink Utilization Time Series at Network Load=0.83 (Centralized EPON BH vs. Traditional BH).....	98
Figure 7.8: Uplink Packet Drop Rate: Centralized EPON Wireless BH vs. Traditional BH	100
Figure 7.9: Uplink Utilization: Decentralized EPON Wireless BH vs. Centralized EPON and Traditional BH	103
Figure 7.10: Uplink Utilization Statistics at Network Load=0.83 (Decentralized EPON BH vs. Centralized EPON and Traditional BH)	104
Figure 7.11: Uplink Utilization Time Series at Network Load=0.83 (Decentralized EPON BH vs. Centralized EPON and Traditional BH)	104
Figure 7.12: Uplink Utilization Statistics at Network Load=1.13 (Decentralized EPON BH vs. Centralized EPON and Traditional BH)	105
Figure 7.13: Uplink Utilization Time Series at Network Load=1.13 (Decentralized EPON BH vs. Centralized EPON and Traditional BH)	105
Figure 7.14: Downlink Utilization: Centralized EPON Wireless BH vs. Traditional BH.	107
Figure 7.15: Downlink Utilization statistics at Network Load=0.83 (Centralized EPON BH vs. Traditional BH).....	108
Figure 7.16: Downlink Utilization statistics at Network Load=0.98 (Centralized EPON BH vs. Traditional BH).....	108
Figure 7.17: Downlink Utilization Time Series at Network Load=0.83 (Centralized EPON BH vs. Traditional BH).....	109
Figure 7.18: Downlink Utilization Time Series at Network Load=0.98 (Centralized EPON BH vs. Traditional BH).....	109
Figure 7.19: Downlink Utilization: Decentralized EPON Wireless BH vs. Centralized EPON BH and Traditional BH	111
Figure 7.20: Downlink Utilization Statistics at Network Load=0.98 (Decentralized EPON BH vs. Centralized EPON and Traditional BH)	111

LIST OF TABLES

Table 7.1 Unevenly Loaded Base Stations Scenarios.....	91
Table 7.2 Packet Drop per Class of Service at Network Load = 0.83.....	100

Chapter 1

INTRODUCTION

1 Chapter 1: Introduction

1.1 Introduction

Recent rapid advances in broadband access technologies in wireline and wireless mobile networks promises seamless, all-over access to unlimited information and entertainment to consumers and business users. Mobile and fixed-line service providers (SPs) worldwide are being challenged to deliver new and innovative offerings beyond voice and basic data services to their customers. To meet these challenges, SPs are introducing dramatic changes in both their fixed and mobile broadband access networking infrastructures.

On the fixed broadband access side, there is a growing perception that copper access networks will soon no longer be able to meet the ever-growing consumer demand for bandwidth. This, along with a combination of regulatory and competitive forces, as well as recent rapid advances and standardizations of Passive Optical Network (PON) technology, are finally prompting carriers around the world to consider PON-based Fiber-To-The-Home (FTTH) systems as possible successor to current copper-based access solutions. FTTH is the ultimate level of access, allowing end users to access the backbone networks through the gigabit capacity of a fiber optic cable.

A PON connects a group of Optical Network Units (ONUs) located at the subscriber premises to an Optical Line Terminal (OLT) located at the service provider's facility. It consists of a single, shared optical fiber (trunk) connecting an OLT to a passive star coupler (SC), which splits the downstream signal to multiple ONUs over dedicated short

optical fiber. Traffic from an OLT to an ONU is called “downstream” (point-to-multipoint), and traffic from an ONU to OLT is called ‘upstream’ (multipoint-to-point). Among the various PON-based FTTH solutions, single channel Time-Division Multiplexed PON (TDM-PON) architecture is currently the most viable solution [1-4].

Due to their reduced operational and equipment costs, TDM-PONs have been widely accepted as a viable technology for the implementations of FTTH solutions, and are being deployed in the field in several places around the world [1-4]. TDM-PON-based FTTH access solutions including broadband PON (BPON), gigabit PON (GPON), and Ethernet PON (EPON) are finally emerging into the mainstream and are set to revolutionize the access infrastructure worldwide. It is widely anticipated that, over the next decade, copper access networks worldwide will be largely replaced by fiber access networks, marking the beginnings of a new era of a mass migration to PON-based FTTH solutions.

On the wireless side, the mobile telecommunications industry is experiencing rapid growth both in terms of subscribers as well as high-speed mobile data services. Emerging mobile data services are expected to see the same explosive growth in demand that Internet and wireless voice services have seen in recent years. It is widely anticipated that demand for data-centric mobile services including web surfing, IPTV, music and video downloads, mobile TV, and other streaming services, will emerge as an equally significant market driver as that of the traditional mobile voice services. To meet the growing demands of both existing and forthcoming high-speed data services, service providers continue to evolve their mobile networks and services from 2G (second generation) to 2.5G to 3G and to beyond 3G (4G).

The first mobile services, which are commonly referred to as First Generation (1G) wireless, were analog. Analog systems were primarily based on circuit-switched technology and designed for voice, not data. In the 1990s, mobile services based on digital mobile technologies signaled the creation of the Second Generation (2G) of wireless services. 2G used several technologies including GSM (Global System for Mobile Communications), TDMA (Time Division Multiple Access), and CDMA (Code Division Multiple Access). CDMA and TDMA were deployed in the various parts of the U.S., while GSM was deployed as the common standard in Europe. 2G is also mostly based on circuit-switched technology. Extension of 2G system is introduced in 2.5 G systems such as General Packet Radio Service (GPRS) and Enhanced Data Rates for Global Evolution (EDGE). GPRS and EDGE are two interim technologies, which supported higher data rates than that of 2G networks and provided a migration path to the Third Generation (3G) of wireless services. Current 2G networks (e. g. GSM and CDMAone) are designed primarily for voice traffic with support of low speed data via GPRS or at enhanced rates using EDGE or 1x-CDMA2000. Neither the first nor the second generations of wireless technologies were designed for multi-media services, such as the Internet. [20-23]

The need to support a wider range of advanced services along with the diverse quality of service (QoS) and rate requirements set by these services have led to the development of 3G mobile telecommunications systems. The concept of 3G wireless technology represents a shift from voice-centric services to multimedia-oriented (voice, data, video, gaming and other real-time) services. 3G wireless technology represents the convergence of various 2G wireless telecommunications systems into a single global system that

includes both terrestrial and satellite components. 3G networks can use a variety of present and future wireless network technologies, including GSM, CDMA, TDMA, WCDMA, CDMA2000, UMTS and EDGE. Third generation networks such as Universal Mobile Telecommunications System (UMTS) and CDMA2000 1x EV-DO Rev0 and RevA support higher data rates (up to 3Mb/s) [20-21] and are currently well into deployment offering a variety of data services including mobile video and TV.

In anticipation of an increased demand for more advanced multimedia applications which require data rates higher than those supported by 3G, High-Speed Packet Access (HSPA) standards, which represent an ongoing evolution from 3G to 3.5G services, including High-Speed Downlink Packet Access (HSDPA) and High-Speed Uplink Packet Access (HSUPA) have been introduced in enhanced releases 6 and 7 of 3G UMTS. HSDPA delivers speeds comparable to or better than current fixed-line broadband access systems up to 14.4 Mb/s peak air throughput per user, while HSUPA provides improved uplink performance of up to 5.76 Mb/s theoretically. Since 2006, UMTS networks in many countries have been or are in the process of being upgraded with HSDPA, with some networks offering full HSDPA 14.4 Mb/s downlink capacity and full HSUPA 5.76 Mb/s uplink capacity.

The ultimate vision is the evolution to a fully IP-based mobile infrastructure capable of providing 100 Mb/s and 1 Gb/s speeds both indoors and outdoors, with premium quality and high reliability. This is the envisioned Fourth-Generation wireless systems (4G), also known as beyond 3G. A 4G system will be able to provide a comprehensive IP solution where voice, data and streamed multimedia can be given to users on an "Anytime, Anywhere" basis, and at higher data rates than previous generations. The introduction of

High Speed OFDM Packet Access (HSOPA) and Ultra Mobile Broadband (UMB) technologies are paving a smooth evolutionary path towards 4G and are often referred to as pre-4G.

High Speed OFDM Packet Access (HSOPA) is a Long Term Evolution (LTE) upgrade path for UMTS systems, which is also often referred to as Super 3G. HSOPA is currently under development, aiming for maximum transfer rates of 100 Mb/s for down-link and 50 Mb/s for uplink. UMB is part of a project within 3GPP2 to improve the CDMA2000 mobile phone standard for next generation applications and requirements. The system is based upon IP networking technologies running over a next generation radio system, with peak rates of up to 280 Mb/s down-link and 75 Mb/s uplink. The UMB standardization is expected to be completed soon, with commercialization taking place around mid-2009.

1.2 Thesis Motivation

It is thus quite clear from the above discussion that the technology is already in place to support quite advanced mobile data services, specifically IP/Ethernet-based services. It is also clear that mobile networks and services have gone beyond voice-only communication services and are rapidly evolving towards data-centric services. HSPA and CDMA 1x Evolution-Data Optimized (EV-DO) devices now on the market support peak download rates of several Mbit/s, which has led to a significant increase in data traffic volumes. With this trend, mobile networks will certainly experience the same paradigm shift that has already occurred in wireline networks - namely the shift from voice-centric to data-centric networks.

In anticipation of this paradigm shift, mobile operators are currently planning to migrate their existing TDM-based mobile network infrastructure to a fully packet-based network. However, current mobile backhaul infrastructure remains the main hurdle to such a transition. The problem is further exacerbated by the fact that mobile backhaul accounts for up to 40% of total operational cost in 2G networks, and 60% or more in 3G. The migration to a packet-based mobile backhaul infrastructure that can support the emerging 3G/4G data-centric services along with the diverse QoS and rate requirements set by these services, that is the focus of this thesis, faces several technical as well as economical hurdles including:

First, as outlined above, each new wireless generation provides wider range of services as well as new capabilities. Such high performance, however, comes at a price: an exponential increase in the bandwidth required to backhaul cellular traffic across the Radio Access Network (RAN) from individual Base Stations (BSs) to the Base station Controller (BSC) or Radio Network Controller (RNC). The BSC is used in 2G networks, while the RNC is used in 3G and later generation networks. Radio Access Network (RAN) provides a connection between the end-user mobile station over air interface and the core network over a landline transmission network usually called – mobile backhaul.

The fundamental problem is that the majority of today's RANs are built on legacy TDM-based circuit switched infrastructures, where each BS is connected directly to the BSC/RNC (in "star topology" architecture) over dedicated point-to-point T1/E1 lines. Though acceptable for voice and low data rate applications (simple messaging services and email), T1/E1 capacity is inadequate for higher mobile data rates. With early deployment of HSDPA and HSUPA under way and HSOPA as well as UMB, the

exponential spike in capacity required to transport these data-intensive applications with peak air throughput per user reaching 14-200 Mb/s, is far beyond what can realistically be achieved using today's T1/E1 leased lines. Assuming that the requisite additional T1/E1 lines were readily available (few tens to more than a hundred T1/E1 lines per BS) from the landline operator, they would be cost-prohibitive.

Another major limitation facing current traditional RAN architectures is the fundamental problem of inefficient utilization of limited available network resources. Because of the static-mapping nature of TDM transport, for each and every channel on the air interface, regardless if carrying data or idle, the appropriate channel resource must be allocated to that channel on the outgoing T1/E1 link. Furthermore, the TDM bandwidth is dedicated on a point-to-point basis from each base station all the way back to the BSC/RNC. This problem is more pronounced under typical non-uniform network traffic loads scenario, i. e., when some BSs, for a given interval, are heavily loaded while others are underutilized or are totally idle. In this case, the unused dedicated channel capacities of those lightly loaded/idle BSs cannot be shared by any of the other heavily loaded BSs attached to the BSC/RNC, leading to the waste of scarce network resources. There is no way to dynamically move capacity from a heavily loaded BS to a lightly loaded/idle BS even within the same cell site. In other words, additional backhaul expense results from unused bandwidth that is often stranded in the "wrong place at the wrong time." Therefore, it is essential that future RAN architectures must support dynamic bandwidth allocation (DBA) and sharing.

In addition to the aforementioned technical hurdles, which can be addressed, the most serious challenge to this migration scenario, however, is an economical one, namely, the

decoupling of operating costs and revenues. It is well known that, as was the case in wireline, “the revenue per bit for data services is significantly lower than that for voice services” and, thus, this anticipated increase in data traffic will probably not be associated with an equal increase in revenue. It is therefore imperative to drastically reduce the cost per bit of transporting data traffic over tomorrow’s packet-based mobile backhaul infrastructure and into the core network. Furthermore, it is equally important that this infrastructure must have the capability of efficiently supporting both existing services (which are paying for operating costs), and emerging 3G/4G data-intensive services, minimizing the bandwidth required and thereby the cost of supporting these services.

In view of the above discussion and reasoning, the following summarizes the status of current mobile networks: First, optimized for slow growing, narrowband, circuit-switched voice traffic, current mobile backhaul networks can’t cope with the dynamic and bursty traffic pattern of the emerging 3G/4G data-centric multimedia services. Second, these T1/E1- based access transport networks lack the dynamic functionality and scalability needed to keep pace with the increasing volumes and unpredictability of data traffic. Third, the transition from current mobile backhaul infrastructure that is built primarily for voice traffic to a network supporting increasingly large volumes of data and ultimately multimedia traffic is not feasible via conventional approaches that apply point-solutions and workarounds to the currently existing infrastructure.

Taking into account the current status of mobile backhaul infrastructure along with the fact that the ongoing evolution of mobile wireless services is dominated by technologies that support advanced data-intensive mobile services and are inherently packet-oriented,

such as 1x EV-DO RevA, HSDPA and HSUPA, migration to a fully-packet-based RAN infrastructure is inevitable.

1.3 Thesis Statement

To address some of these key challenges, this thesis argues for an immediate migration from current TDM-based mobile environment to fully packet-based mobile networks with a fiber based access infrastructure. Specifically, this work proposes and devises cost-effective multiservice Ethernet-based mobile backhaul networking architecture with a fiber based access infrastructure. The proposed architecture extends the reach of Ethernet access to base stations and provides a scalable mobile backhaul infrastructure that can accommodate emerging 3G/4G and legacy traffic and services as well as subscriber growth.

The key for a successful migration strategy rests entirely upon the cost-effectiveness of the proposed architecture. To achieve the most cost-effective migration strategy, rather than deploying a totally new packet-based mobile infrastructure, the proposed architecture utilizes the already existing EPON access infrastructure such that both RNC and OLT are collocated at the central office, while the BS is collocated with the ONU. The RNC along with each BS is equipped with an 802.3ah interface. Utilizing the existing wireline EPON infrastructure as a single packet-based radio access transport network enables a seamless migration path to fixed/mobile convergence and creates the potential of supporting several powerful and cost-effective packet-based fixed/mobile access architectural models including:

1. An independent stand alone Ethernet-based RAN architecture model that enables the backhaul of next generation (3G/4G) voice and data traffic and services over a single fully packet-based RAN infrastructure. It is shown that this architecture utilizes available network resources (bandwidth) in both upstream (BSs to RNC) and downstream (RNC to BSs) directions much more efficiently compared to that of a typical TDM-based circuit switched RAN architecture. This reduces the aggregate bandwidth consumed per BS freeing up capacity in the network to support heavily loaded BSs and/or new mobile data services. This is achieved via utilizing efficient DBA and sharing schemes, which is independently implemented by each BS in the upstream direction, as well as statistical multiplexing of voice and data traffic.
2. A near-term stand alone Ethernet-based RAN transitional model that enables the backhaul of current and next generations (2G/3G/4G) of voice and data traffic and services over a single packet-based RAN infrastructure. This model consolidates 2G/3G/4G access networks and provides a basis for graceful transition to a Next-Generation (NG) fully-packet-based network. Since 2G BSs have E1/T1 interfaces, a TDM-over-packet converter is required at each 2G BS to map T1/E1 circuits into Ethernet frames.
3. A hybrid model that support EPON's traditional fixed users as well as the packet-based RAN's mobile users. In this architecture, an ONU and a BS can be integrated into a single module in terms of both software and hardware. This model provides the best overall system performance in terms of cost-effectiveness, bandwidth utilization, and QoS. This is because the integrated

control module housed at each ONU/BS has global information about the entire fixed/mobile network status including the aggregate bandwidth requirements of both fixed and mobile users. Thus, the processes of bandwidth allocation and packet scheduling as well as prioritizing different class of services (for either fixed or mobile users) are globally optimized.

There are several technical and economical drivers behind selecting native Ethernet as a viable alternative to TDM leased lines or ATM for transporting cellular traffic including: 1) the inherent benefits of simplicity, flexibility, and low cost have uniquely positioned Ethernet as the leader for the inexpensive transport of packet-based technologies; 2) Ethernet naturally support IP services, which are expected to dominate NG 4G mobile services; 3) Ethernet offers more capacity at finer granularity; 4) some radio equipment manufacturers have already introduced Ethernet ports in their EV-DO BSs and their UMTS equipment; and 5) Ethernet services are provided over a standard, widely available and well-understood Ethernet interface.

Despite these advantages, Ethernet presents a major technical challenge when used as a transport technology in the RAN. Traditionally, Ethernet provides best effort traffic delivery and doesn't ensure end-to-end QoS guarantees. Although IEEE 802.1Q specifies three priority bits, Ethernet has no true class of service provision, such as DiffServ, and therefore cannot mark packets for prioritization, scheduling, and policing. QoS support is missing due to the lack of connection admission control and/or packet scheduling mechanisms, which could introduce packet loss, jitter, and delay. To address this problem, the first phase of this work will focus on developing novel fully distributed DBA, packet scheduling, and queue management schemes to ensure that Ethernet can

support differentiated class of services including voice, video, and best effort data, at least in a typical wireline EPON access environment.

As this work will show, integration of both inter-ONU/BS scheduling (upstream bandwidth allocation) and intra-ONU/BS scheduling (queue management and priority queuing at the ONU/BS) mechanisms facilitates the support of differentiated QoS across the proposed RAN. In contrast to traditional centralized EPON architectures, where the OLT arbitrates upstream transmissions by allocating an appropriate timeslot to each ONU, the integration process requires each ONU/BS to independently perform the process of upstream bandwidth allocation. It also requires direct intercommunication among ONUs/BSs. This in turn requires a fully distributed control plane among the ONUs/BSs for ONU/BS-ONU/BS communication as well as upstream access to the OLT/RNC.

Note that the processes of moving the functionality of the packet scheduler and the upstream DBA module from the RNC to the BSs as well as achieving direct intercommunication among BSs are in full compliance with 3G standards and the vision of 4G. This also leads to efficient radio resource scheduling, optimum upstream bandwidth allocation, and faster re-transmission. Another important advantage is that some of the burden has been off loaded from the RNC to the BSs. This means that the conventional centralized EPON infrastructure where the process of upstream bandwidth provisioning is located at the distant OLT/RNC, and where there is no inter-ONU/BS communications at all, is not adequate for supporting the proposed RAN architecture. The challenge is to introduce a novel fully distributed EPON architecture, which forms the corner stone for the proposed RAN architecture, with as minor changes as possible to

the conventional centralized architecture, while maintaining the IEEE 802.3ah traditional EPON standards.

To achieve our overall objectives, this work will be divided into two main sequential phases. In the first phase, we propose and devise a novel Ethernet-based PON architecture that uses a fully distributed time division multiple access arbitration scheme. A fully distributed control plane, which achieves intercommunication among the ONUs/BSs is devised and presented. To implement a distributed control plane, direct connectivity (communicability) between the ONUs/ BSs should be in place without imposing any constraint on the PON topology. Supported by the distributed control plane, we develop several decentralized QoS-based DBA algorithms that dynamically allocate upstream network resources among ONUs/BSs fairly and efficiently, in which the OLT is excluded from bandwidth arbitration process.

The proposed distributed EPON architecture supports differentiated services through the integration of both scheduling mechanisms at the ONU/BS (intra-ONU/BS scheduling). The introduction of this integration feature that can only be supported by a decentralized architecture provides better QoS guarantees. It is shown that these distributed networking architectures and the associated bandwidth allocation algorithms and protocols have characteristics that make them far better suited for provisioning Quality of Service (QoS) schemes necessary for properly handling data, voice, video, and advanced mobile data services over a single line.

In the second phase, based on the developed decentralized EPON infrastructure, several powerful and cost-effective packet-based fixed/mobile access architectural models are presented. The performance as well as the pros and cons of each model are compared

with each other and with that of traditional TDM-based RAN model. The DBA and packet scheduling algorithms developed in the first phase are then used to assess the overall performance of the Ethernet-based mobile backhaul infrastructure in both upstream and downstream directions. The performance metrics used are network utilization, packet drop rate, and Peak to Average network utilization, which is a performance metric widely used by network planners to assess network architecture stability and predictability.

Chapter 2

OVERVIEW OF ETHERNET PASSIVE OPTICAL NETWORKS (EPON) AND ARCHITECTURES

2 Chapter 2: Overview of Ethernet Passive Optical Networks (EPON) and Architectures

2.1 Introduction

Recent advances in optical networking technologies have fueled tremendous growth in both backbone and metropolitan access network (MAN) capacity. At the same time, the performance of end-users computing equipment has reached gigahertz speeds. The conduits linking the high-speed end-user equipment to the high capacity backbone networks, however, remain a bottleneck. These connections are commonly referred to as the “Last Mile”.

Because end users are becoming more sophisticated and rich multimedia and real-time services are becoming more popular, the current Last Mile capacity is rapidly becoming unacceptable. While recent advances in Last Mile technology have increased capacity from the range of 56kb/s for a dial-up modem to a few Mb/s for a cable modem or digital subscriber line (DSL) connection, this is still far short of the gigabit line speed necessary to support rich multimedia and real-time services. Increasing this capacity to support these advanced services is one of the most significant problems facing providers and local carriers today.

Fiber-To-The-Home (FTTH) is the ultimate level of access, allowing end users to access the backbone networks through the gigabit capacity of a fiber optic cable. Unfortunately current systems have proven too complex and expensive to be commercially viable. To

lower the cost and expedite the implementation of FTTH, Passive Optical Network (PON) based-solutions have been proposed [1-2]. PONs are point-to-multipoint fiber optical networks with no active elements in the signal's path. It is likely that the reduced equipment costs and the reduced operational costs of PONs will enable carriers to justify FTTH, thus solving the Last Mile bottleneck. PONs are classified into two main networking architectures, namely, ATM-based PON (APON) and Ethernet-based PON (EPON) [1-2].

APONs were standardized and developed around 1995 through the work of the Full Services Access Network (FSAN) initiative [1]. The FSAN recommendation (ITU G.983) defines a PON-based optical access network that uses Asynchronous transfer mode (ATM) as its layer 2 data link protocol. At that time, ATM was viewed by many as the technology that will dominate the Local Area Network (LAN), MAN, and backbone. Since that time, ATM (and APONs) has lost favor and Ethernet has emerged as the frontrunner technology for transporting data, video, and voice services over a single platform. Ethernet currently accounts for more than 85 percent of all installed connections and more than 95 percent of all LANs. Because the length structure of the Ethernet frame is variable (up to 1518 bytes), Ethernet is tailor-made for transporting IP traffic and dramatically reduces the overhead incurred by ATM. With IP firmly entrenched as the dominant internetworking protocol, Ethernet is gaining new grounds in the access, MAN and Wide Area Network (WAN).

With the rapid decline in the cost of fiber optics and Ethernet equipment, EPON is poised to emerge as the technology of choice for the next-generation broadband access network [3-7]. EPON is one of the access solutions considered by the new IEEE 802.3ah Ethernet

in the First Mile (EFM) Task Force [2], focusing on direct support of Ethernet services. Since most of the key technologies are borrowed from APON, EPON is very similar to APON in its basic operations. Like APON, EPON uses Coarse Wavelength Division Multiplexing (CWDM) and Time-Division Multiplexing (TDM) to provide bi-directional, point-to-point communications over a fiber and maintains frame structure for both downstream and upstream communications. One significant distinction is that EPON uses variable-length IEEE 802.3 frames, whereas in APON, data is transmitted in fixed-length 53-byte cells (48-byte payload and five-byte overhead). While EPONs have recently been getting more and more attention from both the research community and the telecommunication industry [3-8], its deployment in access networks is still facing several obstacles and challenges.

To date, the mainstream EPONs bandwidth allocation schemes as well as the new IEEE 802.3ah EFM Task Force specifications [2] have been centralized, relying on a component in the central office (Optical Line Terminal (OLT)) to provision upstream traffic. Hence, the OLT is the only device that can arbitrate time-division access to the shared channel. Since the OLT has global knowledge about the state of the entire network, this is a centralized control plane in which the OLT has centralized intelligence. It is the purpose of this work to propose distributed solutions to this problem, and in the process to prove that these distributed networking architectures and the associated bandwidth allocation algorithms and protocols have characteristics that make them far better suited for provisioning Quality of Service (QoS) schemes necessary for properly handling data, voice, video, and other real-time streaming advanced multimedia services over a single line.

Specifically, this work proposes a novel Ethernet over Star Coupler-based PON architecture that uses a fully distributed time division multiple access arbitration scheme. Supported by the decentralized scheme, we develop several QoS-based dynamic bandwidth allocation (DBA) algorithms in which the OLT is excluded from the implementation of the time slot assignment. In contrast to the mainstream centralized EPON architectures that combine priority queuing (intra-ONU scheduling) with DBA schemes (inter-ONU scheduling) [7-11], the proposed distributed EPON architecture supports differentiated services through the integration of both scheduling mechanisms at the ONU (intra-ONU scheduling). The introduction of this integration feature that can only be supported by a decentralized architecture provides better QoS guarantees. Furthermore, in addition to the added flexibility and reliability, the overall performance of the proposed decentralized EPON architecture and the associated bandwidth allocation algorithms are shown to be at least as efficient as their centralized counterparts.

To implement a distributed control plane, direct connectivity (communicability) between the ONUs should be in place without imposing any constraint on the PON topology. In the proposed architecture, the ONUs exchange signaling and control information concerning their queue status and their transmission needs amongst themselves. Then, the ONUs simultaneously and independently run instances of the same DBA algorithm outputting identical bandwidth allocation results. Once the algorithm is run, the ONUs sequentially and orderly transmit their data without any collisions, eliminating the OLT's centralized task of processing requests and generating grants for bandwidth assignment.

2.2 Overview of Ethernet-over-PON Technologies and Architectures

A PON is a point-to-multipoint fiber optical network with no active elements in the signal's path. It consists of a single, shared optical fiber connecting a service provider's central office (head end) to a passive star coupler (SC)/optical splitter/combiner, which is located near residential customers. The SC is intentionally positioned a substantial distance away from the central office (CO), but close enough to the customers in order to save fiber. Each customer receives a dedicated short optical fiber but shares the long distribution trunk fiber. All transmissions in a PON are performed between an Optical Line Terminal (OLT) and Optical Network Units (ONUs) (Fig. 1). Traffic from an OLT to an ONU is called 'downstream' (point-to-multipoint), and traffic from an ONU to the OLT is called 'upstream' (multipoint-to-point) [3-7]. Two wavelengths are used: typically 1310 nm (λ_{up}) for the upstream transmission and 1490 nm (λ_d) for the downstream transmission. The OLT resides in the central office, connecting the optical access network to the metro or backbone network, where the ONU is located at either the curb (Fiber To The Curb; FTTC solution) or the end-user location (Fiber To The Building and Fiber To The Home; FTTB and FTTH respectively). A single PON typically serves from 16-64 customers. PONs can be deployed in a 1:N tree, tree-and-branch, ring, or bus topology.

In the downstream direction, Figure 2.1 shows an overview simplified illustration of EPON downstream, an EPON operates as a broadcast and select network. The OLT has the entire bandwidth of the channel to broadcast standard formatted 802.3 Ethernet frames to all ONUs. Each ONU extracts those packets that contain the ONU's unique Media Access Control (MAC) address. In the upstream direction, Figure 2.2 shows an

overview simplified illustration of EPON upstream, multiple ONUs share the transmission channel. Thus, the ONUs need to employ some arbitration mechanism to avoid collisions. In that case, each ONU transmits within a dedicated time slot and the OLT receives a continuous stream of collision-free frames from multiple ONUs.

The IEEE 802.3ah task force is actively standardizing the control and management messages used to control the data exchange between the OLT and the ONUs as well as the processing of these messages through the development of Multi-Point Control Protocol (MPCP). Note that MPCP is not concerned with any particular bandwidth allocation; it is merely a supporting protocol that facilitates the implementation of various bandwidth allocation algorithms in EPON. The protocol relies on two Ethernet control messages (GATE and REPORT) in its regular operation. The OLT assigns the Transmission Windows (TWs) via the GATE messages.

In general, the OLT arbitrates the upstream transmissions by allocating an appropriate timeslot/transmission window to each ONU. An ONU is only allowed to transmit during the TW allocated to it by the OLT. Each ONU uses a set of queues to store its Ethernet frames and starts transmitting them as soon as its TW starts. An ONU can support up to 8 priority queues as defined in 802.1Q [12]. Within each cycle, in order to inform the OLT about its bandwidth requirements, ONUs use REPORT Messages that are also transmitted along with the data in the TW. The ONU should also account for additional overhead when requesting the next time slot; this includes 8 bytes frame preamble and 12 bytes Inter-Frame Gap (IFG) between two consecutive frames. Between the TW of two ONUs there is a certain guard time “g” needed to account for the laser on and off times, receiver recovery times, round trip delay (which relates to the physical distance between

communicating ONUs) and other optic related issues. Upon receiving a REPORT, the OLT passes the message to a Dynamic Bandwidth Allocation (DBA) module, which performs the bandwidth allocation computation.

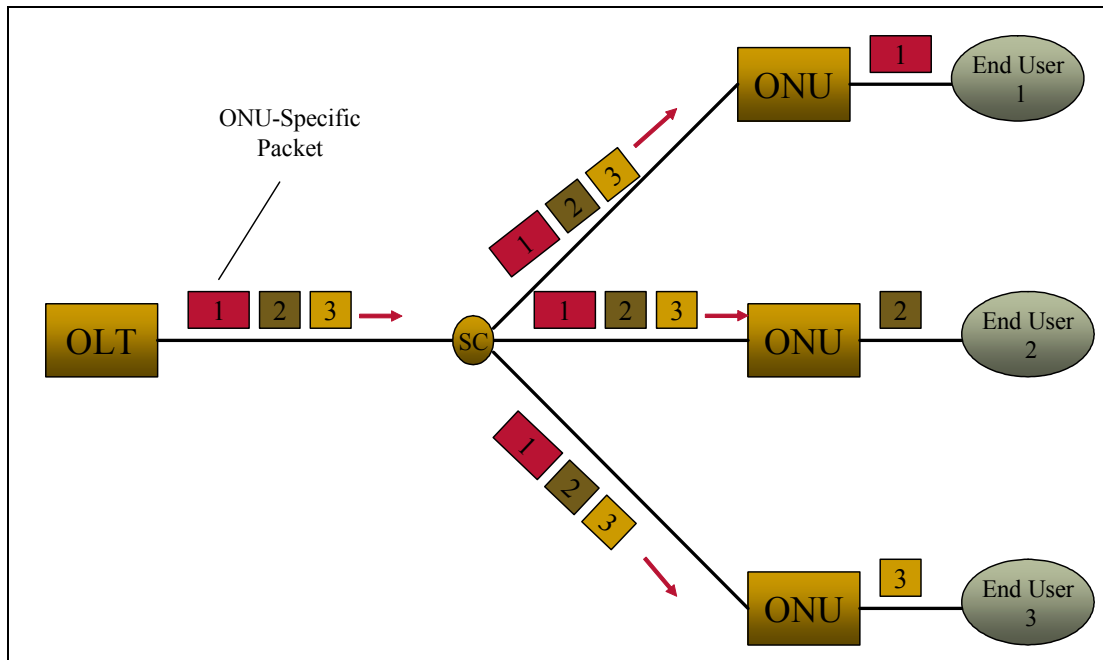


Figure 2.1: EPON Downstream overview

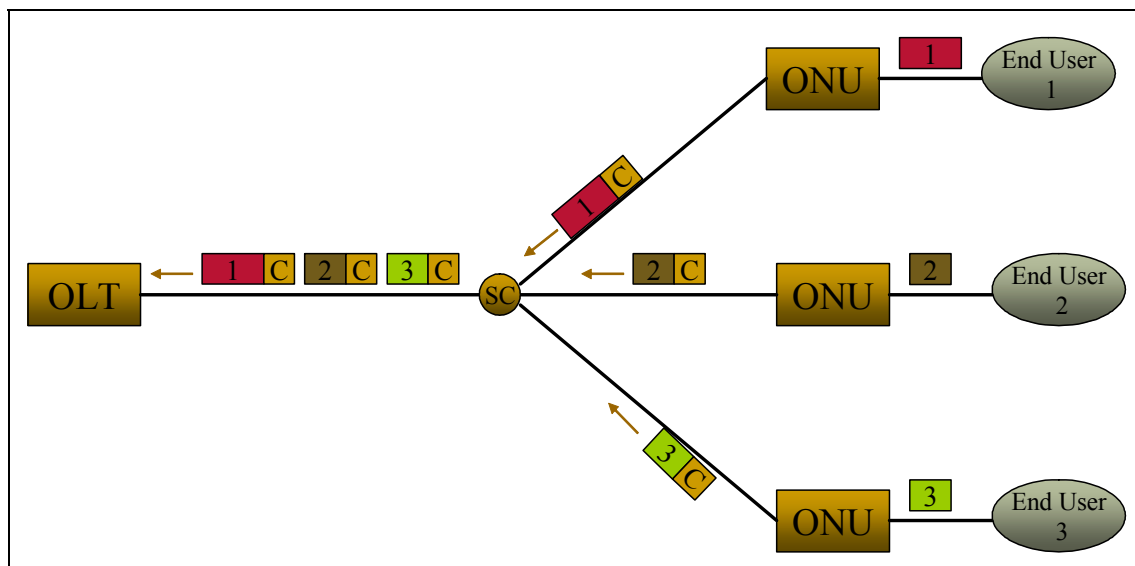


Figure 2.2: EPON Upstream overview

Chapter 3

PROPOSED DISTRIBUTED EPON ARCHITECTURE

3 Chapter 3: Proposed Distributed EPON Architecture

3.1 The Proposed Distributed Architecture

Figure 3.1 shows the general architecture of this approach [13]. As can be seen from Figure 3.1, a portion of the optical signal power transmitted by an upstream transmitter (λ_{up}) toward the OLT will be redirected back and broadcasted to all ONUs. This can be achieved by connecting two ports of a 3xN SC with each other through an optical isolator as shown in Figure 3.1 [14]. Note that in addition to the conventional transceiver maintained at each ONU (a λ_{up} transmitter and a λ_d receiver), this approach requires an extra receiver tuned at λ_{up} . A baseband direct detection circuit is needed to detect the redirected control channel (λ_{up}) in order to recover the control update information.

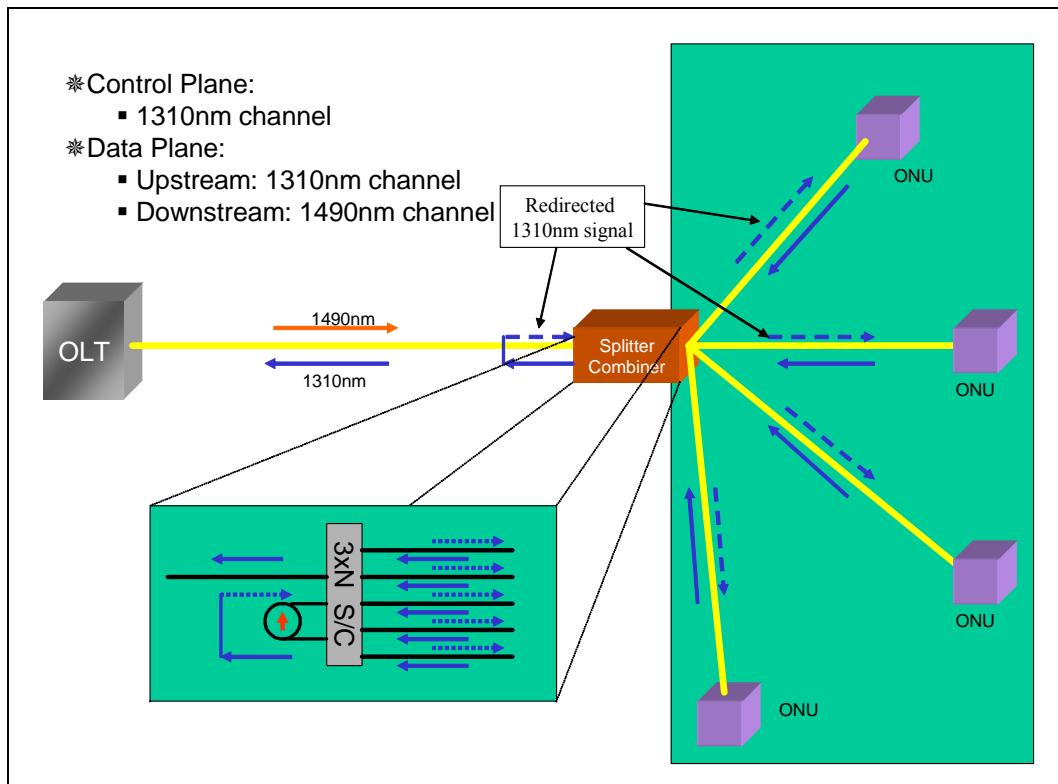


Figure 3.1: Proposed Distributed EPON Architecture

This architecture assumes a cycle-based upstream link, where a cycle is defined as the time that elapses between two executions of the scheduling algorithm. The cycle size can either have fixed, or variable length confined within certain lower and upper bounds to accommodate the dynamic upstream traffic conditions. The cycle is divided into three periods: a static *update period* (control plane), a fixed *waiting period* (processing control messages and running the algorithm) and a dynamic *transmission period* (data plane).

The proposed cycle, along with the details of how the control plane performs the updating process is shown in Figure 3.2a in three phases. Each ONU transmits its update

control message in its own assigned fixed time slot (first phase). These messages are then combined at the SC and a multiplexed update message is created (second phase). In the third phase, a fraction of the multiplexed control signal is transmitted through the first output port of the SC and propagates to the OLT (which could discard it, make use of it as a synchronization message, and/or process the control information). Another fraction of the multiplexed control signal is redirected back and broadcasted to all ONUs (through the isolator). A baseband direct detection circuit located at each ONU is used to detect the redirected control channel (λ_{update}). The detected signal is then processed in order to recover the control data information belonging to each of the other $(N-1)$ ONUs. Since there are only two operating communication wavelengths (λ_{up} and λ_{d}), signaling and upstream transmission take place on the same communication channel (λ_{up}) and the periods will appear sequentially as on the top of Figure 3.2.

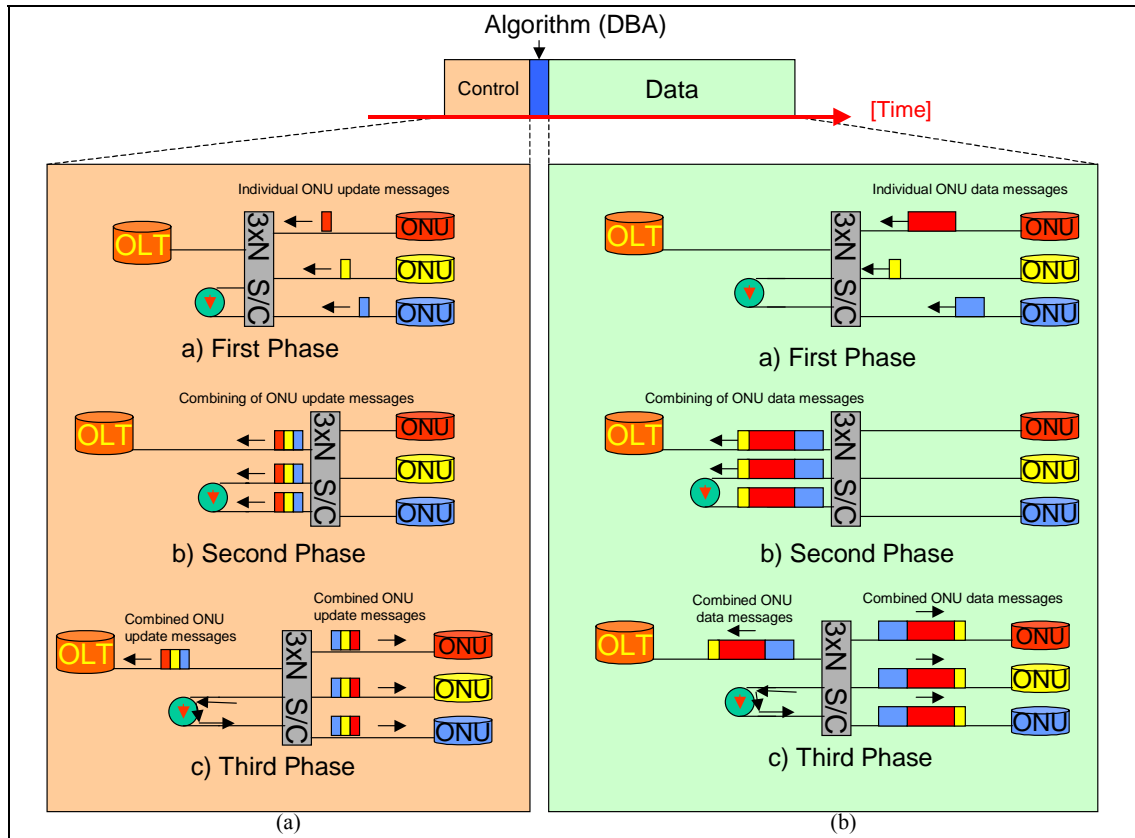


Figure 3.2: (a) Cycle updating process (b) Transmission process.

3.2 Detailed Proposed Cycle

1. The First Period (Control Plane): The update period is divided into N equal fixed time slots where N is the maximum number of users the network is designed to accommodate. The update period is used for the ONUs to communicate their status and to exchange signaling and control message information with one another. Each ONU uses its own fixed time slot within the update period to transmit its control message. For simplicity, and to avoid collisions, the assignment of these N timeslots follows a fixed Time-Division Multiple Access (TDMA) assignment since control messages are fixed in size. Note that the control slots in the proposed distributed scheme are all transmitted

sequentially in one period (update period). This is in contrast to the centralized schemes reported above [3-11], where the control slot (REPORT Message) of each ONU is transmitted along with the data in the TW allocated to it by the OLT. All control update messages are transmitted as Ethernet frames. Because the signaling information is segregated from the upstream traffic, signaling information can be timelier and complete thus increasing the efficiency of the dynamic bandwidth allocation algorithm. These enhanced DBA algorithms would have the ability to support better QoS characteristics because transmission of the signaling information is not constrained by the shared data/control upstream channel associated with the centralized schemes.

2. The Second Period (Algorithm Execution): The second period of fixed length is a **waiting period** (no upstream transmissions are allowed during this period) and is used for allowing the ONUs to process the information gathered from the multiplexed control message. Each ONU maintains a table with information about the state of the queues at each other ONU. This information is updated each cycle whenever the ONU receives a new multiplexed control message from all other ONUs. The DBA algorithm module uses the table maintained at each ONU. Note that instances of the same DBA algorithm are executed simultaneously and independently at each ONU. An execution of the algorithm produces a *unique* and identical set of ONU assignments w_i ; where w_i is the amount of bytes that an ONU is allowed to transmit in its TW during the cycle. It is critical that the algorithm produces a unique outcome for any arbitrary set of inputs. In other words, it should not incorporate any assumptions or randomness to handle exceptions. This is because several instances of it will run locally and independently at each ONU.

3. The Third Period (Data Plane): The third period or (**transmission period**) is essentially a giant slot used for actual upstream data transmission. During the transmission period, the ONUs follow exactly the allocation scheme the algorithm produced (i.e., their transmissions start at specific times and last for specific bytes) as shown in Figure 3.2b. Note that the order of ONUs transmission may be different in each cycle and need not be fixed. It is rather a function of the ONU's traffic demand. This is a major advantage compared to the fixed transmission order proposed in [8].

Note that all ONUs are synchronized to a common reference clock extracted from the OLT downstream traffic. Clocking information, in the form of a synchronization marker, is included at the beginning of each downstream frame cycle. The synchronization marker is a one-byte code that is transmitted every 2 ms to synchronize the ONUs with the OLT [1-2]. The TDM controller for each ONU, in conjunction with timing information from the OLT, controls the upstream transmission of the variable-length packets within the dedicated time slots. Maintaining proper time sync between different channels is required for the appropriate operation of the DBA algorithm.

Chapter 4

PROPOSED INTEGRATED SCHEDULING AT ONU AND DYNAMIC BANDWIDTH ALLOCATION SCHEMES

4 Chapter 4: Proposed Integrated Scheduling at ONU and Dynamic Bandwidth Allocation Schemes

4.1 Introduction

In order for the mainstream centralized EPON architectures to support differentiated QoS, two independent scheduling mechanisms are required:

a) ***Scheduling at the OLT (inter-ONU scheduling)***: The OLT is the only device that can arbitrate the upstream transmissions by allocating an appropriate TW to each ONU. In this case, the OLT passes the request messages to a dynamic bandwidth allocation module (co-located with the OLT) that performs the bandwidth allocation computation and generates grant messages.

b) ***Scheduling at the ONU (intra-ONU scheduling)***: In this case, queue management and priority queuing are used to divide the bandwidth allocated by the OLT to a given ONU among the different class of services (based on their priorities) supported by that ONU.

In general, combining both scheduling mechanisms is required to enable centralized EPON networking architectures to support differentiated QoS. However, since the two scheduling schemes are independent of each other, the final bandwidth allocated to a particular class of service for a given ONU may not be the optimum choice. In the case of the proposed decentralized architecture described in previous chapters, however, instances of the same DBA algorithm are executed simultaneously at each ONU. Thus,

both scheduling tasks (DBA scheme and priority queuing) are performed at the ONU leading to the notion of integrating both scheduling mechanisms at the ONU.

4.2 Queue Management with Priority Queuing (Intra-ONU Scheduling)

Bandwidth management and fair scheduling of different traffic classes will play an important role in supporting QoS in the emerging EPON-based differentiated services (Diff-serv¹)-capable access network [11]. Priority queuing is considered a useful and relatively simple method for supporting differentiated service classes.

Each ONU maintains three separate priority queues that share the same buffering space. We consider three priority classes P0, P1, and P2, with P0 being the highest priority (constant-bit-rate) and P2 being the lowest (non-real time data transfer). These classes are used for delivering voice (constant-bit-rate or CBR), video stream (variable-bit-rate or VBR), and best-effort data and they allow easy mapping of DiffServ's Expedited Forwarding (EF), Assured Forwarding (AF), and Best effort (BE) classes into 802.1D classes. Packets are first segregated and classified (packets classification is performed by checking the type-of-service (TOS) field of each IP packet encapsulated in the Ethernet frame) and then placed into their corresponding appropriate priority queue. The queues in each ONU share common memory of fixed size.

¹ Diffserv is an IETF framework for classifying network traffic into classes, with different service level for each class.

4.2.1 Strict Priority Scheduling

A priority-based scheduler is required for scheduling packet transmission. Strict priority scheduling mechanism (defined in P802.1D) schedules packets from the head of a given queue only if all higher priority queues are empty. For instance, if an arriving packet with priority P1 finds the shared buffer full in the ONU, it can displace one or more P2 packets from the P2 queue, such that the arriving P1 packet can be placed into the P1 queue. Similarly, arriving P0 packet can displace P1 and P2 packets from their queues.

Figure 4.1 will be used as an illustrative example to illustrate the operation of such a scheduler, where ONU_i is requesting transmission. At time t_0 , ONU_i sends a control (report) message during the update period requesting bandwidth based on its buffer occupancy. The report message will first reach the SC (Star Coupler) where it is then combined (multiplexed) with all other ONU report messages forming an update-multiplexed message. The update message is then broadcasted (reflected back) to all ONUs, arriving at ONU_i at time t_1 . Upon receiving the multiplexed update message, the ONU passes the message to the DBA module which, in turn, allocates a timeslot to the ONU. Let's assume that the algorithm processing ends at time t_2 , and that the transmission is scheduled to start at time t_3 . During the waiting time $(t_3 - t_0)$, more packets may contend for the buffer. Thus, in strict priority scheduling, the high priority packets arriving during the waiting period will be scheduled ahead of the already reported (before t_0) lower priority packets.

This will penalize low priority traffic with indefinite increase in packet delay and higher packet loss, as low priority queues will have to wait till all higher priority queues are empty. It has been shown in [9-11] that using strict priority scheduling would generate what was termed “*light-load penalty*” in which delay for some traffic classes increases when network load decreases. To alleviate this problem, this work only considers a modified version of the *fair non-strict priority scheduling* (reported in [11]) which is described below.

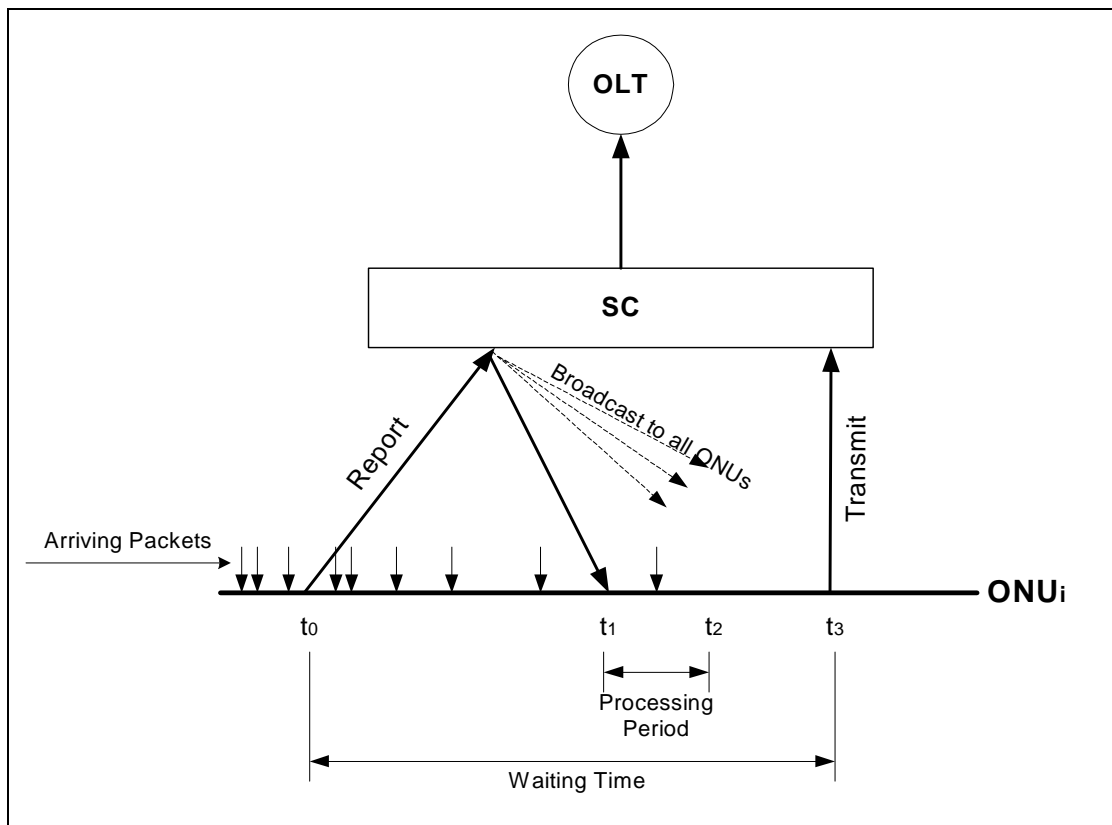


Figure 4.1: Illustrative example of a scheduler's operation

4.2.2 Fair Non-Strict Priority Scheduling

In this scheduling scheme, the ONU keeps track of the total number of bytes (per queue) that it has reported in the last update message (i.e., the report transmitted during the last cycle), and it transmits the reported data first, regardless of their priorities [11]. Only those packets that arrived before t_0 will be transmitted first provided that the bandwidth of the allocated timeslot allows for the transmission. Thus, if higher priority packets arrived after t_0 , they have to wait until the reported (arrived before t_0) lower priority traffic is transmitted. If packets arriving before t_0 are all scheduled, and if the current timeslot can still accommodate more traffic, it will be allocated for higher priority traffic. This scheme will ensure fairness in scheduling packets.

4.3 Dynamic Bandwidth Allocation at the ONU (Intra-ONU Scheduling)

4.3.1 Overview of Previous/Current DBA Schemes

Several centralized allocation schemes have recently been reported in the literature ranging from a static allocation scheme (fixed TDMA) to a dynamically adapting scheme based on instantaneous queue size in every ONU (statistical multiplexing) [3-7, 15-17]. The simplest is the static TDMA scheme in which every ONU gets a fixed timeslot [3]. While this scheme is very simple, it results in inefficient upstream channel utilization, since statistical multiplexing between the ONUs is not possible. An OLT-based polling scheme, called interleaved polling with adaptive cycle time (IPACT) based on *Grant* and *Request* messages, has been presented in [8]. This scheme uses an interleaved polling

approach where the next ONU is polled before the transmission from the previous one has arrived. This scheme provides statistical multiplexing for ONUs and results in efficient upstream channel utilization.

Using IPACT, several bandwidth allocation algorithms were studied in [7-8], namely fixed, limited, gated, constant credit, and linear credit. Amongst these algorithms, the limited (where the OLT grants the requested number of bytes, but no more than a given predetermined maximum B_{\max}) was shown to exhibit the best performance. The basic limited service scheme as well as several variants of it were also combined with both strict priority and priority queuing (Intra-ONU scheduling) in order to support differentiated services [9-11].

All of the above referenced DBA schemes are OLT-based, that is the OLT has centralized intelligence. Two major limitations may render the performance of the centralized DBA schemes sub-optimal: 1) The bandwidth granted by the OLT, during cycle n , to ONU_i is only determined by the content of the single report message transmitted in the previous cycle $n-1$ by ONU_i (i.e., the OLT does not take into account all other ONU requests). Thus, the process of bandwidth allocation is not globally optimized. 2) Due to the bursty nature of Ethernet traffic, some ONUs might have less traffic to transmit while other ONUs may require more bandwidth than B_{\max} . For instance, assume that ONU_i requests an amount of bandwidth $R_i < B_{\max}$, while ONU_j requests an amount of bandwidth $R_j > B_{\max}$. Although there is an excess amount of bandwidth ($B_{\max} - R_i$) that can be used by ONU_j due to limitation # 1 cited above, ONU_j

can only be granted bandwidth B_{\max} . This may lead to inefficient utilization of bandwidth resources.

4.3.2 Decentralized Dynamic Bandwidth Allocation Schemes at the ONU

The proposed decentralized EPON architecture can address some of the limitations of the centralized DBA schemes cited above and can further provide several advantages as follows:

- Since the bandwidth allocation computation is performed after receiving and processing all ONUs requests (processing period) (i.e., the computation takes into account the entire network status), the bandwidth allocation process now reflects the entire network information collectively, leading to a globally optimized decision.
- In contrast to the centralized architectures where the order of ONUs transmission is fixed in each cycle (sequential), the decentralized architecture has the added flexibility of varying the order of ONUs transmission according to the ONUs traffic demands and priority. Thus, the order of ONUs transmission may be different in each cycle and need not be fixed.
- Since the DBA computation is based on the global network information, the heavily loaded ONUs may be allocated the remaining excessive bandwidth that is not utilized by the lightly loaded ONUs.
- Given that DBA and priority queuing scheduling tasks are both executed at the ONU, the DBA module can integrate both scheduling information to yield a

globally optimized bandwidth allocation to a particular class of service in a given ONU.

Because the centralized limited DBA scheme was shown to exhibit the best performance in [7-8], we will consider this scheme as a reference model for comparing the performance of our distributed architecture versus that of the centralized scheme reported therein. In the following section, we will take advantages of the proposed decentralized architecture and present two enhanced variants of the centralized limited DBA scheme. The first approach varies the order of ONUs transmission while the second exploits the fact that the DBA computation is based on global network information.

Overview of Limited DBA Scheme:

This algorithm is cycle-based, where a cycle is defined as the time that elapses between two executions of the scheduling algorithm. A cycle has a variable length size confined within certain lower and upper bounds, which we denote as T_{MIN} and T_{MAX} (sec). Thus, the algorithm schedules between B_{MIN} and B_{MAX} (bytes) at a time, where B_i is found by multiplying T_i with the line rate. In this scheme, the ONU will be granted the requested number of bytes, but no more than a given predetermined maximum W_{MAX} (maximum transmission window). If Req_i is the requested bandwidth of ONU_{*i*} and $Grant_i$ is the granted bandwidth, $Grant_i$ is then equal to

$$Grant_i = \begin{cases} Req_i & \text{if } Req_i < W_{MAX} \\ W_{MAX} & \text{if } Req_i \geq W_{MAX} \end{cases}$$

W_{MAX} is determined by the maximum cycle time T_{MAX} [7-8]

$$T_{MAX} = N \left(Guard_{total} + \frac{W_{MAX}}{Transmission_Speed} \right) \quad N: \text{number of ONUs.}$$

In the decentralized architecture, $Guard_{total}$ is the total guard time added per cycle, and is equal to

$$Guard_{total} = Guard_{transmission} + Guard_{report}$$

$Guard_{transmission}$ is the guard time added between two consecutive transmission windows (TWs) during the transmission period (second period) and $Guard_{report}$ is the guard time added between two consecutive report messages during the update period (first period).

In the centralized architecture, $Guard_{report} = 0$, as the TW contains both report message and data packets to be transmitted.

Proposed Decentralized Collective DBA Algorithms:

All proposed DBA schemes are Collective DBAs in which the DBA module processes the information gathered from all ONUs. In this approach, the DBA scheme will reflect the entire network status collectively as opposed to IPACT, where the DBA module at the OLT processes just a single ONU request message at a time (in a sequential order) and independently for each ONU [8].

A. Transmission Order-Based Enhanced DBA algorithms (TOEDBA)

Three variants of the TOEDBA algorithms will be evaluated:

1. **Sequential:** This DBA is implemented for comparison purposes only; it uses the same approach used in IPACT [8], where the ONUs transmission order during the transmission period (third period) is fixed (sequential).
2. **Shortest Request First (SRF):** The ONUs transmission order will start at the ONU with the shortest bandwidth request in ascending order. The ONU with the shortest bandwidth request will transmit first.
3. **Longest Request First (LRF):** The ONUs transmission order will start at the ONU with the longest bandwidth request in descending order. The ONU with the longest bandwidth request will transmit first.

B. Efficient Bandwidth Utilization-Based Enhanced DBA Algorithm (Remainder DBA)

As mentioned above, due to the bursty nature of Ethernet traffic, some ONUs might have less traffic to transmit while other ONUs may require more than W_{MAX} . Consequently, ONUs can be classified into two groups based on their requested bandwidth, namely: lightly loaded ONUs which might have requests less than bandwidth W_{MAX} ; and heavily loaded ONUs which may have requests more than bandwidth W_{MAX} . The lightly loaded ONUs with $Req_i < W_{MAX}$ will contribute a total remainder bandwidth:

$$Remainder_{total} = \sum_i^L (W_{MAX} - Req_i) \quad L: \text{Number of lightly loaded ONUs}$$

The heavily loaded ONUs with $Req_i \geq W_{MAX}$ will require a total over the limit bandwidth:

$$OverLimit_{total} = \sum_i^H (Req_i - W_{MAX}) \quad H: \text{Number of heavily loaded ONUs}$$

The total remainder bandwidth can be fairly distributed amongst the heavily loaded ONUs to expand their maximum transmission window as follows [11]:

$$\Delta W_i = \text{Remainder}_{total} \times \frac{\text{Req}_i}{\text{OverLimit}_{total}}$$

Where ΔW_i is the extra bandwidth allocated to ONU_i.

$$BW_{Granted} = \begin{cases} \text{Req}_i & \text{If } \text{Req}_i \leq W_{MAX} \\ \text{Req}_i & \text{If } \text{Req}_i > W_{MAX} \ \& \ \text{Remainder}_{total} \geq \text{OverLimit}_{total} \\ \text{Req}_i + \Delta W_i & \text{If } \text{Req}_i > W_{MAX} \ \& \ \text{Remainder}_{total} < \text{OverLimit}_{total} \end{cases}$$

To support differentiated classes of service, the remainder bandwidth can be distributed per priority queue for a given ONU using a weight function P_{Q_i} . In this case Q is the priority queue number (class of service, i.e., P₀, P₁, and P₂) and P_{Q_i} is the weight allocated to a given priority queue (ranging from 0 to 1) for ONU_i, where $\sum_{Q=0}^k P_{Q_i} = 1$. The total remainder bandwidth can then be distributed amongst the queues of the heavily loaded ONUs to expand their maximum transmission window as follows:

$$\Delta WP_{Q_i} = \text{Remainder}_{total} \times \frac{\text{Req}_i}{\text{OverLimit}_{total}} \times P_{Q_i}$$

Chapter 5

DECENTRALIZED EPON PERFORMANCE EVALUATION

5 Chapter 5: Decentralized EPON Performance Evaluation

5.1 Introduction

In this section, we compare the overall performance of the proposed decentralized architecture using the two enhanced DBA algorithms presented in the previous section versus the centralized IPACT architecture using the limited DBA scheme. An event-driven packet-based simulation model was developed using C++. Two simulation programs were developed, one for the centralized IPACT architecture and the other for the decentralized architecture, with identical network parameters. The performance metrics used here are average packet delay, maximum packet delay, and the number of packets dropped.

The traffic model used for the performance evaluation is the same as that reported in [8] where each ONU has a number of ON/OFF sources, each with a Pareto distribution governing the lengths of the ON/OFF periods, in order to capture the self-similar nature of Ethernet traffic. To compare the performance results of the proposed decentralized model with that of the centralized scheme (IPACT) of [8], we used a system with 16 ONUs, access link data rate from users to an ONU of 100 Mb/s, and a 1 Gb/s upstream link data rate (from an ONU to the OLT). The distance between the OLT and the ONUs is 20 km. Maximum cycle time $T_{MAX} = 2\text{ms}$. The guard time separating two consecutive transmission windows is set to $5\ \mu\text{s}$ which are reasonable choice to account for the laser on and off times, receiver recovery times, round trip delay (which depends

on the physical distance between communicating ONUs) and other optic related issues. [2]. For the decentralized architecture, the guard time separating two consecutive control messages is set to $0.1 \mu\text{s}$. Buffer size in each ONU is 8 Mbytes. Each point on the following plots corresponds to a sample of 50 million packets, averaged over four different runs.

The Total Network Load used in the figures below is calculated by the following formula:

$$\text{Total Network Load} = \text{Number of ONUs} \times (\text{ONU Load}) \times \frac{\text{Line rate User - to - ONU}}{\text{EPON Line rate}}$$

For example; If: ONU load = 0.7,

Then: Total Network Load = $16 \times 0.7 \times (100\text{Mbps}/1\text{Gbps}) = 1.12$.

5.2 Numerical Results without CoS Support

Figure 5.1 compares the average packet delay as a function of total network load, for centralized IPACT architecture using limited DBA scheme versus the decentralized architecture using the two enhanced DBA schemes (TOEDBA scheme with its three variants, Sequential, Shortest Request First (SRF), and Longest Request First (LRF) and Remainder DBA algorithm). Figure 5.2 presents the same comparison as Figure 5.1 for the maximum packet delay in order to show that the maximum packet delay exhibited by the network in the decentralized approach is similar to the maximum packet delay of the centralized scheme.

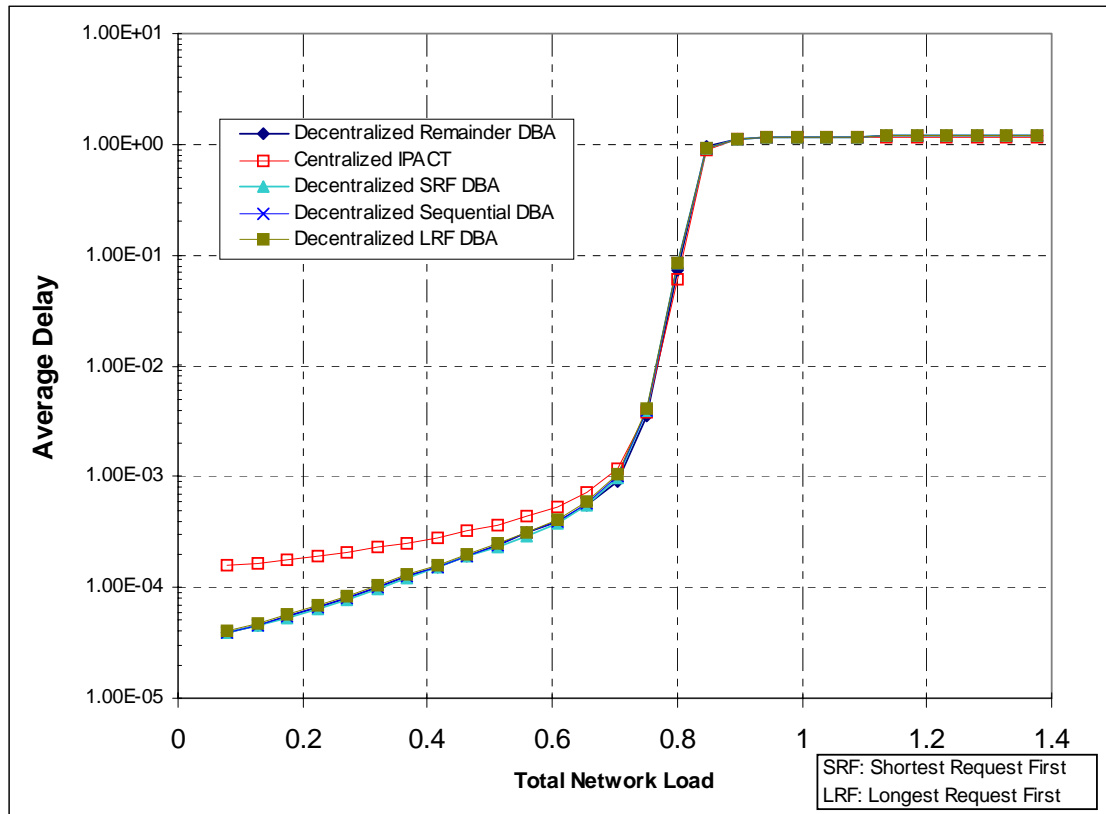


Figure 5.1: Average Packet Delay for Centralized and Decentralized architectures with various DBAs

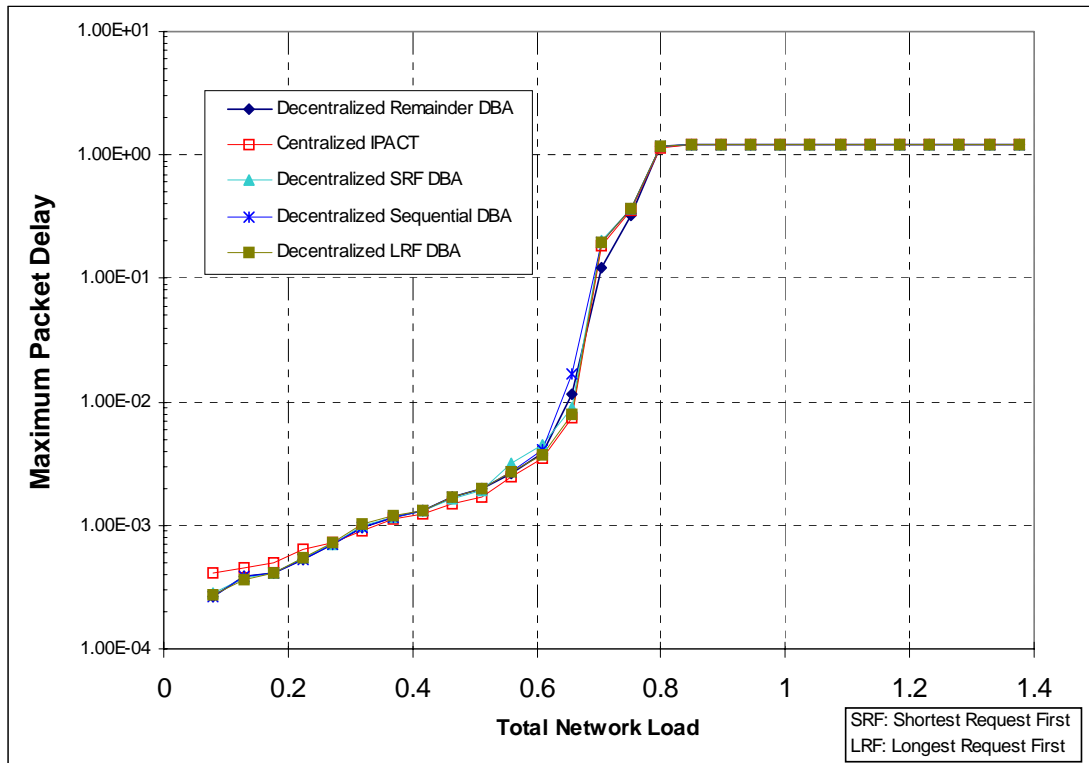


Figure 5.2: Maximum Packet Delay for Centralized and Decentralized architectures with various DBAs.

As can be seen from Figure 5.1, the decentralized architecture demonstrates an improvement over the centralized IPACT approach in terms of average packet delay, as long as the total network load is less than 0.8. This is mainly due to two reasons: (a) By interchanging the order of transmission, a given ONU's update message is closer in time to its corresponding transmission. Thus, a more current depiction of its buffer status is governing the transmission. As the load increases, more ONUs request more than the maximum allowed window, and thus more get the same allocation (maximum window). This, in turn, eliminates the advantage of the interchanged order at higher loads (b) In the case of IPACT, if an ONU has no packets to send, the algorithm stops at this ONU in order to transmit its report message (polling table) which consequently adds an associated

guard time. In the decentralized architecture, however, due to the fact that the DBA is a collective algorithm, those ONUs with no data to send are not granted a transmission window. At light loads, because the probability of having an ONU with no packets to transfer is high (due to the bursty nature of self-similar traffic), this will result in less guard times added between ONU transmissions, which in turn translates to better performance in terms of average delay.

Due to the ONUs limited buffer size (8 Mbytes) it was observed that packets start to drop when the total network load exceeds 0.8. Figure 5.3 shows the difference in number of packets dropped for both the decentralized *Limited DBA* scheme and the decentralized *Remainder DBA* algorithm in order to demonstrate the advantage of using the Remainder DBA scheme. In both algorithms, the fixed sequential ONUs transmission order is used. When a positive value is depicted in the plot, that means that the number of packets dropped using the remainder DBA is less than the number of packets dropped using the Limited DBA and vice versa. It can be seen from the figure that using the Remainder DBA algorithm minimizes the number of packets dropped.

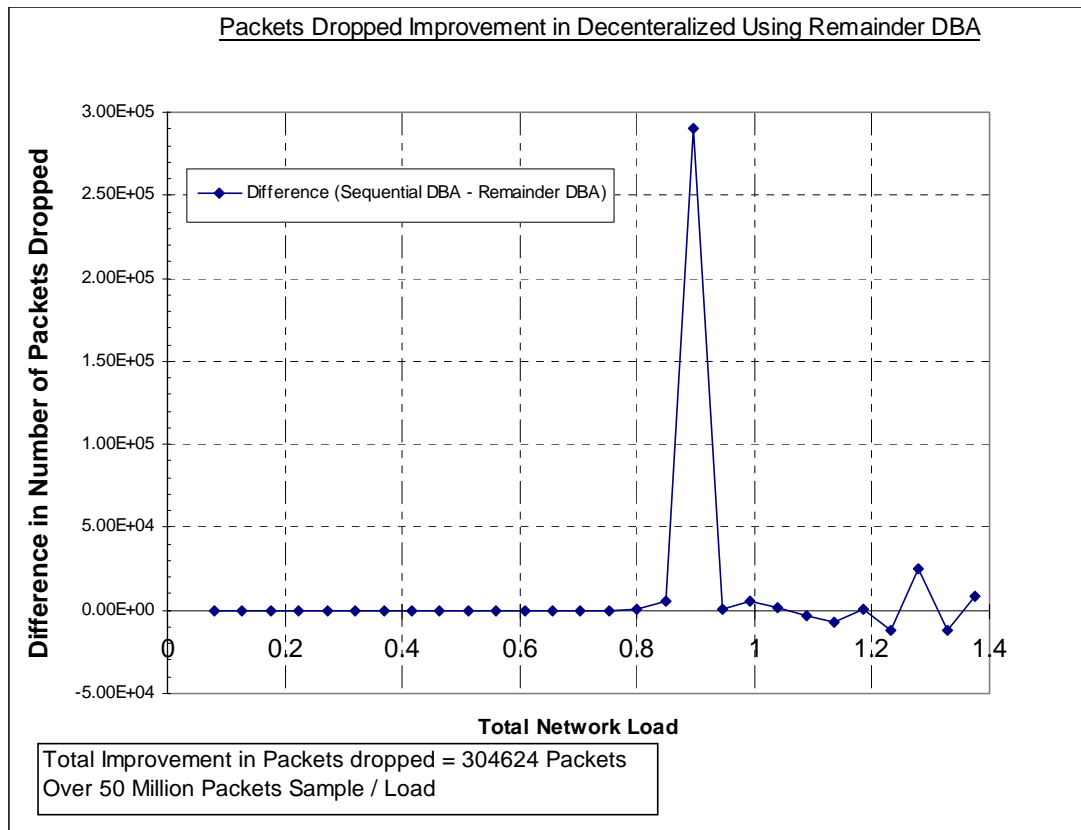


Figure 5.3: Improvement in packets dropped using Remainder DBA.

5.3 Numerical Results for Differentiated Classes of Service

5.3.1 Traffic Model for Differentiated CoS

We consider three priority classes P0, P1, and P2, with P0 being the highest priority (constant-bit-rate) and P2 being the lowest (non-real time data transfer). These classes are used to for delivering voice (constant-bit-rate or CBR), video stream (variable-bit-rate or VBR), and best-efforts data. Each ONU maintains three separate priority queues that share the same buffering space.

- **Class P0** is used to emulate a circuit over packet connection. P0 traffic has constant-bit rate (CBR). In our model, we chose to emulate a T1 connection. The T1 data arriving from the user is packetized in the ONU by placing 24 bytes of data in a packet. Including Ethernet and UDP/IP headers this results in a 70-byte frame generated every $125 \mu\text{s}$. Hence, the P0 data consumed 4.48 Mbps of bandwidth. This is the highest priority traffic.
- **Class P1** consisted of VBR video streams that exhibit properties of self-similarity [18-19] and long-range dependence. Packet sizes in P1 streams ranged from 64 bytes to 1518 bytes.
- **Class P2** is also used to emulate self-similar and long-range dependant traffic. This class also has variable-size packets ranging from 64 bytes to 1518 bytes. This class has the lowest priority. This priority level is used for non-real time data transfer. There is no delivery or delay guarantees in this service.

In our simulation, when we varied the load, P0 was always kept constant (4.48 Mbps).

The remaining load was split equally between P1 and P2.

For example, if we assume an ONU with 0.7 load then:

The CBR P0 load will be: $70 \text{ bytes every } 125 \mu\text{s} = (4.48\text{Mbps}/100\text{Mbps}) = 0.0448$;

$P1 \text{ Load} = (0.7 - 0.0448)/2 = 0.3276$;

$P2 \text{ Load} = (0.7 - 0.0448)/2 = 0.3276$

$\text{Total ONU Load} = 0.0448 + 0.3276 + 0.3276 = 0.7$

Total Network Load = $16 * 0.7 * 100\text{Mbps}/1\text{Gbps} = 1.12$

5.3.2 Detailed Numerical Results for Differentiated CoS

In this section, we compare the overall performance of the proposed decentralized architecture when differentiated classes of services are supported. We compare the two enhanced DBA algorithms versus the centralized IPACT architecture using the limited DBA scheme. The same event-driven packet-based simulation model that was discussed in previous sections was modified to handle differentiated classes of services with the same network parameters. *Fair Non-Strict Priority Scheduling* is used for both architectures.

Figure 5.4 and Figure 5.5 compare the average and maximum packet delay as a function of total network load for various classes of services for the decentralized architecture (using limited DBA scheme with sequential ONUs transmission order) versus the centralized IPACT architecture (using limited DBA). As it can be seen from Figure 5.4, the average packet delay of P0 and P1 of the decentralized architecture is always less than those of IPACT for all network loads. It can be also seen that the average packet delay of P2 of the decentralized architecture is always less than that of P2 of IPACT as long as the load is less than 0.7. For loads higher than 0.7, the average packet delay is almost the same for both architectures. This is again due to the same two reasons mentioned in the discussion of Figure 5.1 above. Figure 5.5 presents the same comparison as Figure 5.4 for the maximum packet delay. This is again used to show that

the maximum packet delay exhibited by the network in the decentralized approach is similar to the maximum packet delay of the centralized scheme.

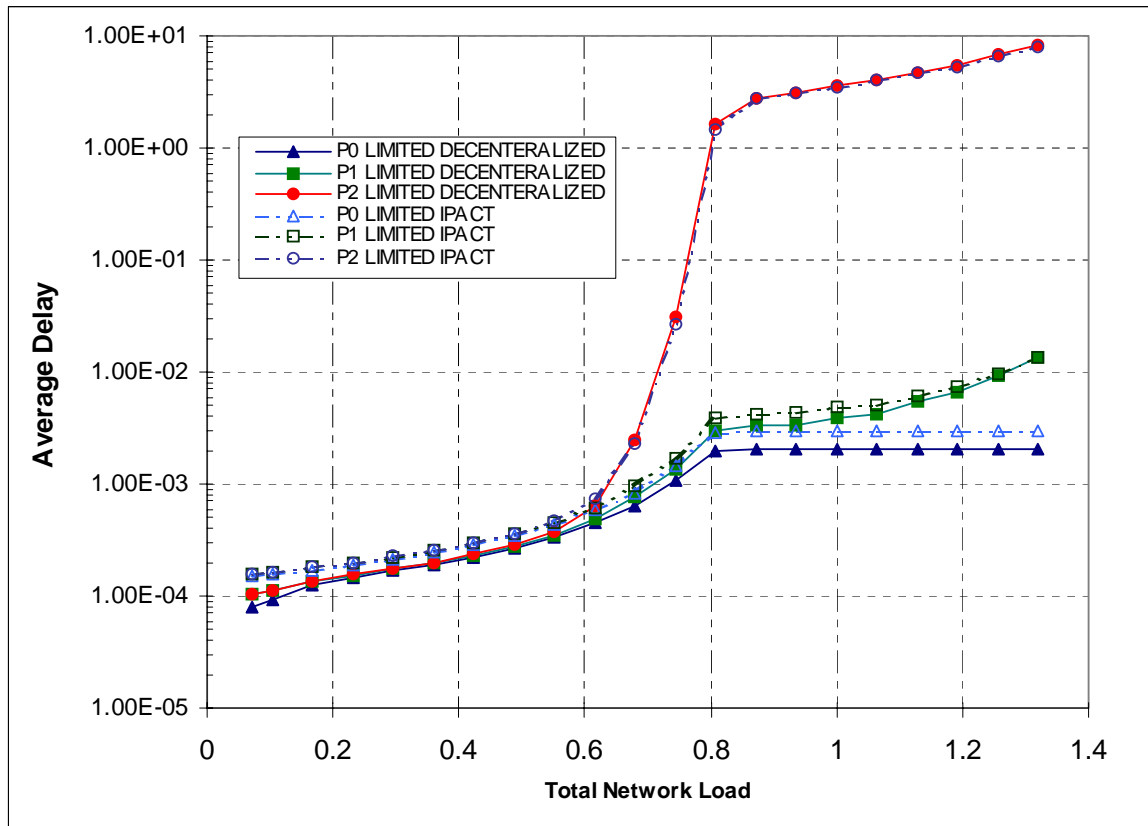


Figure 5.4: Average packet delay for various classes of services comparing IPACT Limited DBA and Decentralized Limited DBA.

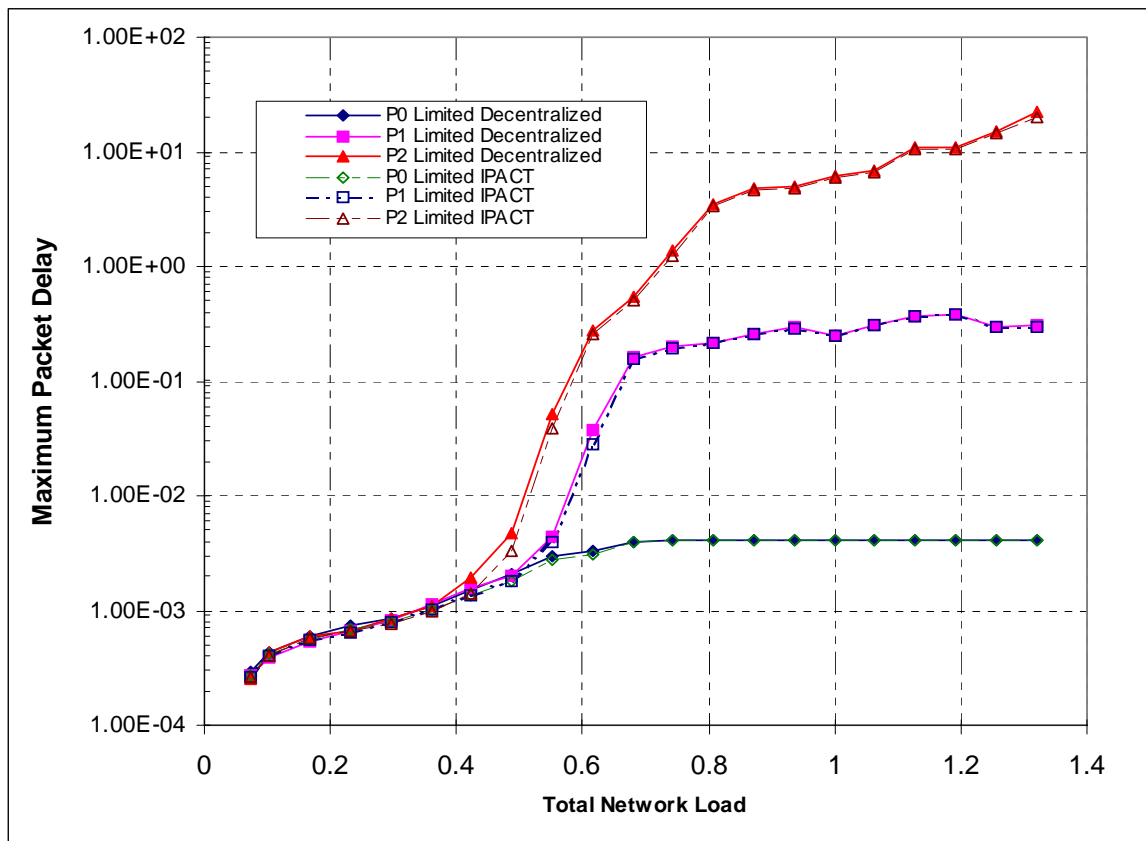


Figure 5.5: Maximum packet delay for various classes of services as a function of network Load for the IPACT Limited DBA and the Decentralized Limited DBA

Figure 5.6, Figure 5.7 and Figure 5.8 compare the average packet delay as a function of total network load, for a given class of service (Figure 5.6 represents P0, Figure 5.7 represents P1, and Figure 5.8 represents P2), for centralized IPACT architecture using limited DBA scheme versus the decentralized architecture using the enhanced DBA schemes (Limited Sequential, SRF, LRF and Remainder).

In general, and for all three classes (P0, P1, and P2), the centralized IPACT architecture always exhibits higher packet delays. It can be seen from the figures that the delay exhibits a clear knee at total network loads around 0.8. At this load, the network begins to exhibit signs of saturation: buffers are full and a large amount of packets are dropped for P2. For IPACT, the average delay for P0 at saturation loads was about 2ms, but for all decentralized DBAs it saturated at 1ms. At loads less than 0.8, P0 of the decentralized SRF (Shortest Request First) algorithm is shown to exhibit the smallest packet delay. This is because in SRF, ONUs with short requests will jump ahead in the transmission leading to better overall packet delay performance. On the other hand, the LRF DBA packet delay performance was the worst amongst all the decentralized DBA schemes. However, LRF had a better P2 packet loss compared to the SRF.

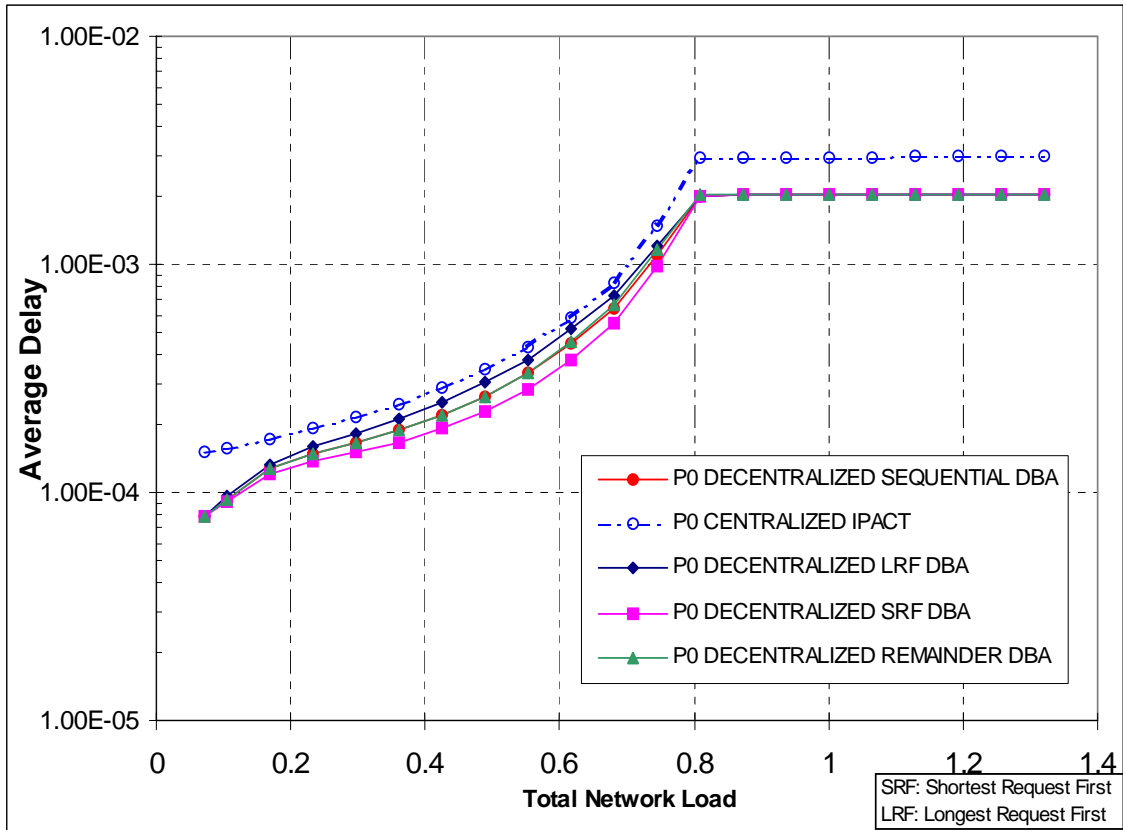


Figure 5.6: Class P0 average packet delay as a function of total network load for the centralized IPACT and the decentralized architecture under various DBAs.

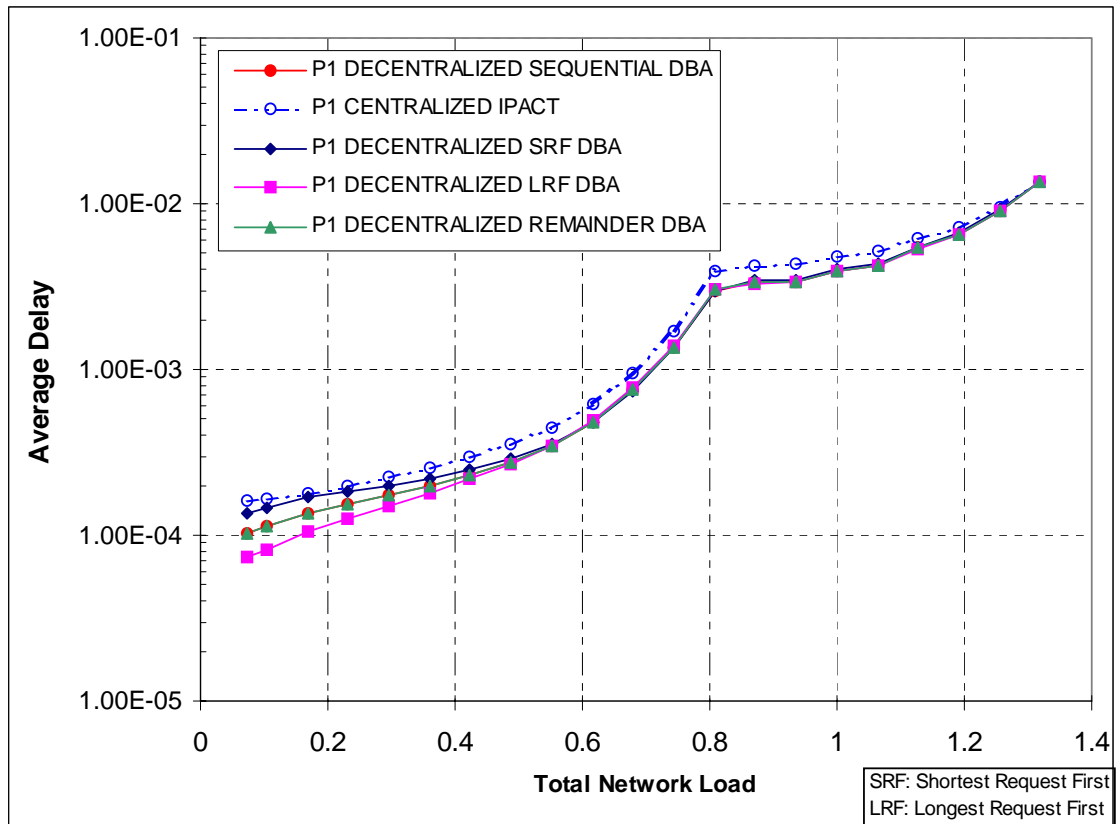


Figure 5.7: Class P1 average packet delay as a function of total network load for the centralized IPACT and the decentralized architecture under various DBAs.

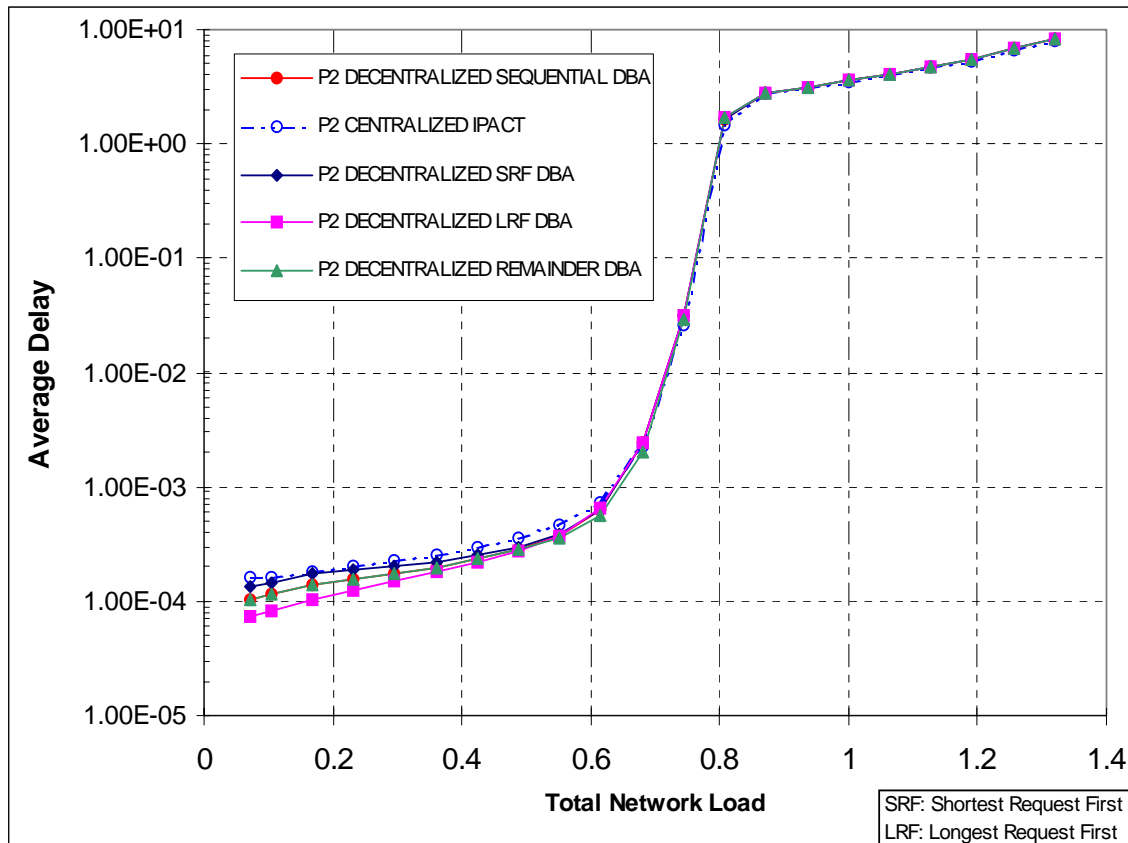


Figure 5.8: Class P2 average packet delay as a function of total network load for the centralized IPACT and the decentralized architecture under various DBAs.

Figure 5.9, Figure 5.10 and Figure 5.11 present the maximum packet delay comparison between the centralized IPACT architecture using limited DBA and the decentralized architecture using various DBA algorithms. Each plot represents one class (P0, P1 and P2 respectively). Clearly, the maximum packet delay exhibited by the network is similar for all classes for both the centralized and decentralized approaches.

From our simulations, it was observed that all packets dropped were from priority class P2, starting at network loads of 0.8. In order to improve the number of packets dropped

we used the Remainder Collective DBA algorithm discussed in previous sections. Figure 5.12 demonstrates the improvement in number of packets dropped using the Remainder DBA. The plot shows the difference in packets dropped (Sequential DBA – Remainder DBA).

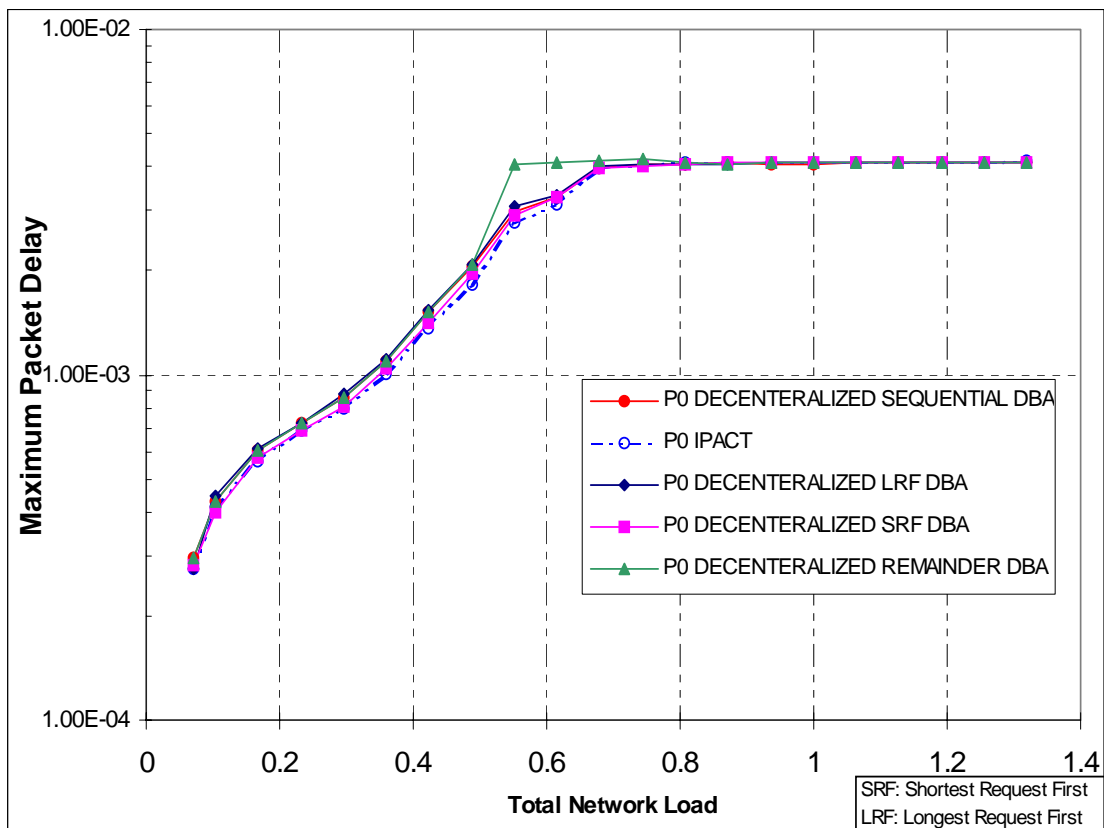


Figure 5.9: Class P0 Maximum packet delay as a function of total network load for the centralized IPACT and the decentralized architecture under various DBAs.

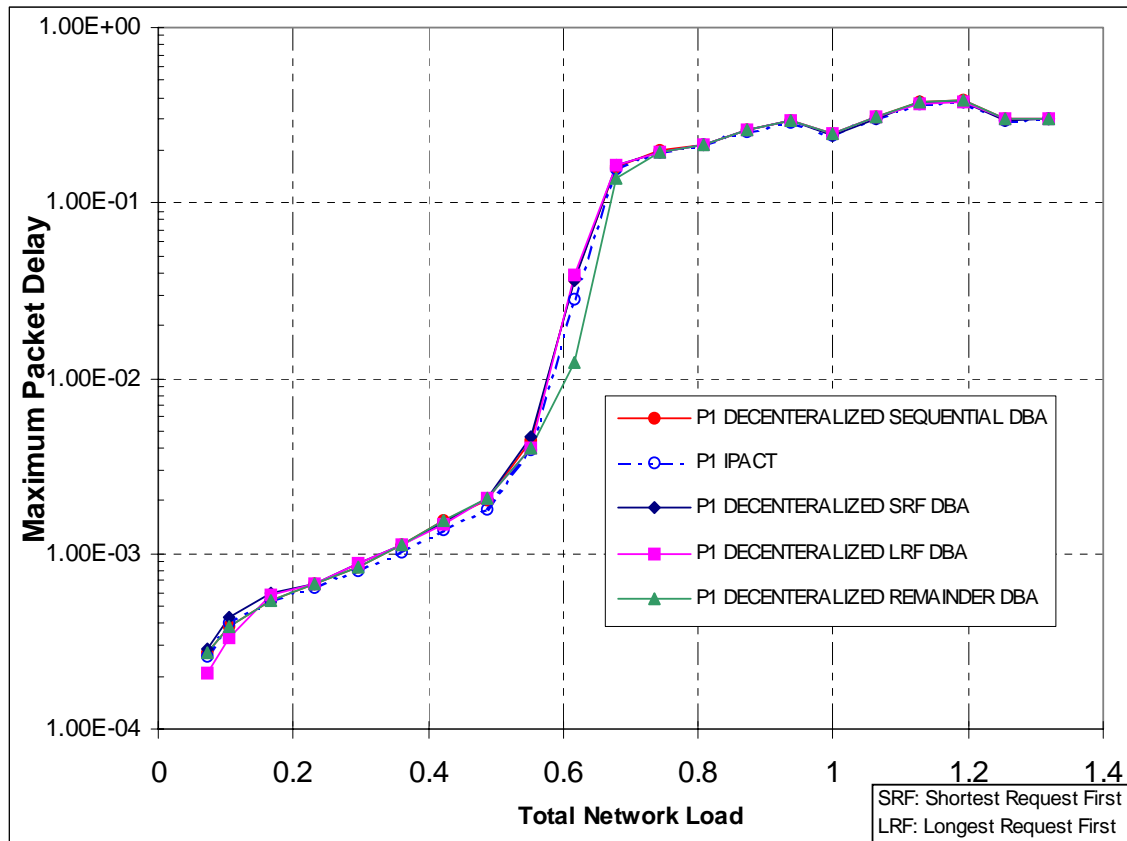


Figure 5.10: Class P1 Maximum packet delay as a function of total network load for the centralized IPACT and the decentralized architecture under various DBAs.

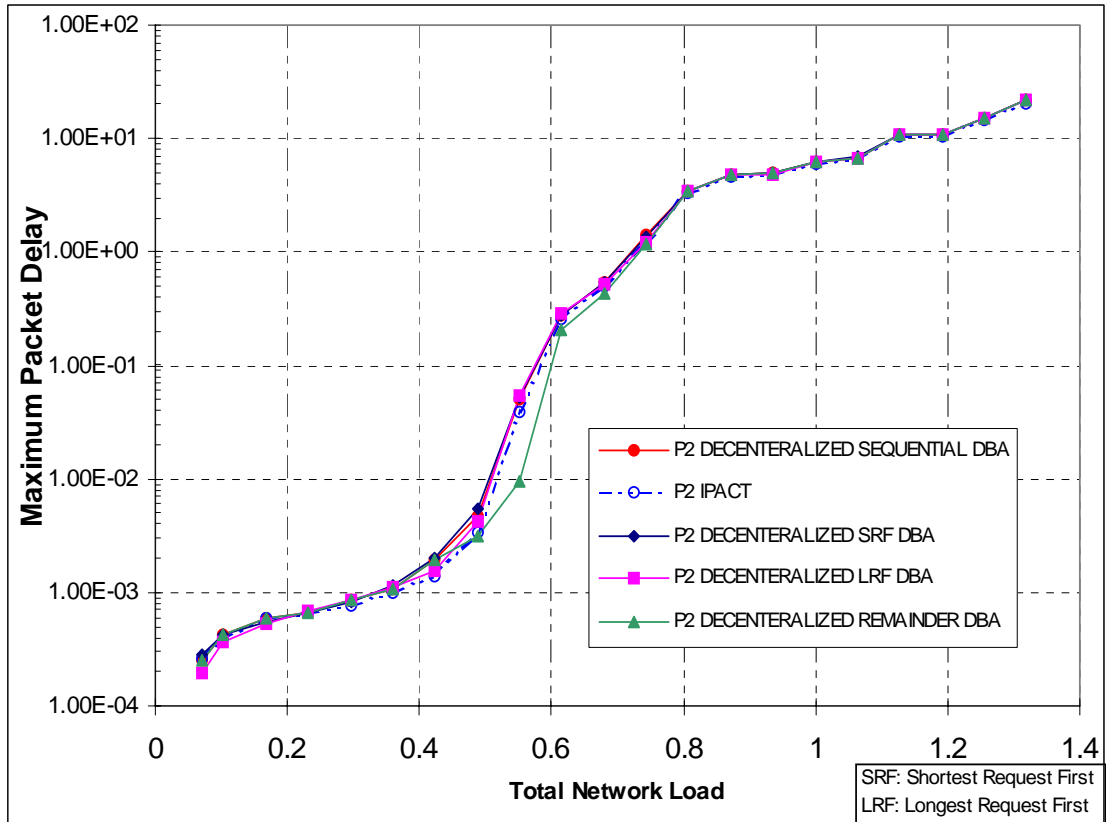


Figure 5.11: Class P2 Maximum packet delay as a function of total network load for the centralized IPACT and the decentralized architecture under various DBAs.

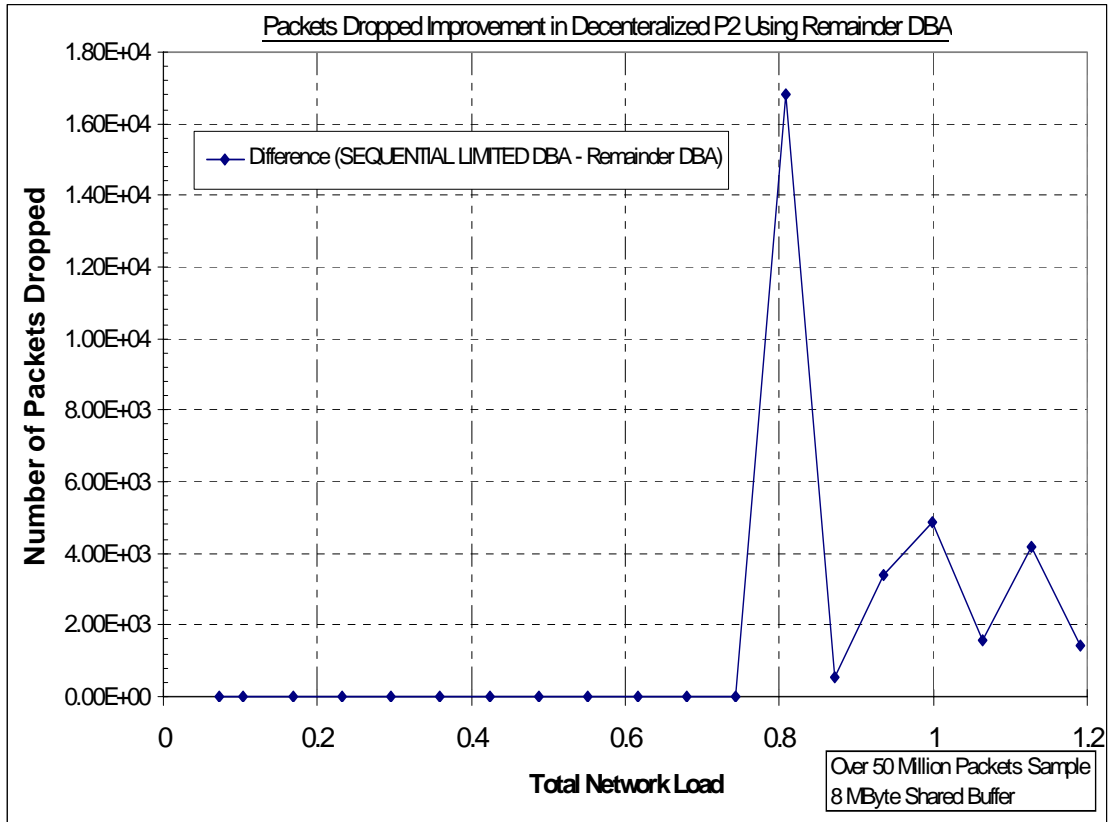


Figure 5.12: Improvement in the number of Class P2 packets dropped using Remainder DBA

5.4 Conclusion

This work has proposed a novel Ethernet over Star Coupler-based PON architecture that uses a fully distributed time division multiple access arbitration scheme. Supported by the decentralized scheme, we have developed several QoS-based dynamic bandwidth allocation (DBA) algorithms in which the OLT is excluded from the implementation of the time slot assignment. In contrast to the mainstream centralized EPON architectures that combine priority queuing (Intra-ONU scheduling) with DBA schemes (inter-ONU scheduling), the proposed distributed EPON architecture supports differentiated services through the integration of both scheduling mechanisms at the ONU (Intra-ONU scheduling). It has been shown that these distributed EPON networking architectures and the associated bandwidth allocation algorithms and protocols have characteristics that made them far better suited for provisioning Quality of Service (QoS) schemes necessary for properly handling data, voice, video, and other real-time streaming advanced multimedia services over a single line. Furthermore, in addition to the added flexibility and reliability, the overall performance of the proposed decentralized EPON architecture and the associated bandwidth allocation algorithms are shown to be at least as efficient as their centralized counterparts.

Chapter 6

NOVEL WIRELESS BACKHAUL ARCHITECTURE

6 Chapter 6: Novel Wireless Backhaul Architecture

6.1 Introduction

The majority of mobile networks still use "star topology" architecture in their backhaul where each Base Station (BS) is connected directly to the Radio Network Controller (RNC) over dedicated T1/E1 lines. Since each T1/E1 line is dedicated to a particular BS, each Transmitter/Receiver (TxRx) of that BS must be assigned to dedicated channels on a given T1/E1. The Star topology offers the easiest deployment and maintenance. This is specially true in the initial stages of network deployment (GreenField) as well as new wireless technology introductions. But later on and as the network traffic starts to ramp up (through additional subscribers or increased network usage due to bandwidth-demanding applications or a combination of both) it becomes very costly to deliver broadband mobile services due to the need of the extra T1/E1 lines.

This work proposes using Ethernet Passive Optical Network (EPON) as the Wireless Backhaul. A comparison between the performances using traditional backhaul network versus EPON as wireless backhaul is presented. Even and un-even BS (Base Station) loads were used. The loading is represented using wireless/mobile traffic characteristics with different CoS (Class of Service). We proved that using EPON architecture would improve network utilization and packet drop rate in both uplink and downlink, in addition to the significant cost reduction. The EPON architectures used were both the centralized in addition to our proposed decentralized architecture.

6.2 Background and Motivation

Vast majority of mobile networks are designed and dimensioned to comply with the requirements of worst-case conditions in which busy hour or peak traffic demand can arise simultaneously in all of the backhaul network elements. To meet capacity demands, the TxRxs are assigned to each BS according to busy hour call usage projections. The busier an area is, the more TxRxs need to be installed in the BSs to provide sufficient capacity, and the more backhaul bandwidth is required to transport increased traffic. There is also a coverage demand, so the mobile service providers will add additional BSs to insure full service coverage for all areas (urban, sub-urban and rural), regardless of usage remaining low in some areas.

This will lead to situation where the traffic channels in one system are rarely simultaneously busy and the result is - inefficient allocation of limited network resources.

Another issue that must be taken into account is inherent inefficiency. The service providers have no way of using extra capacity on seldom used lines to carry overflow traffic from busier lines. Consequently, while service providers can lease additional T1/E1 lines for BSs in high-traffic areas, they must absorb the cost of providing spare backhaul bandwidth for BSs in low-use areas

6.3 Overview of Radio Access Networks and Traditional Wireless Backhaul architecture

Ethernet-based Radio Access Networks (RAN) are expected to be the next generation access networks in UMTS and CDMA networks. There are several benefits of Ethernet-based RAN including lower costs, flexibility of merging wired and wireless networks, network scalability and the capability of delivering End-to-end Quality of Service (QoS).

Currently, third-generation wide-area wireless networks based on the CDMA2000 [20-22] and UMTS [23] are being deployed throughout the world. These networks provide both voice and high-speed data services to the mobile subscriber. As the cost of these services is being reduced to attract more subscribers, it becomes important for the network operators to reduce their capital and operating expenses. In wireless access networks today, the base stations and the radio network controllers are connected by point-to-point T1/E1 links. These backhaul links are expensive and add to operating costs.

The wireless access network consists of a set of base stations (BS) that are managed by a Radio Network Controller (RNC). A Radio Access Network (RAN) connects the BSs to the RNCs. The RNC performs a number of functions, including soft-handoffs, reverse outer loop power control, and termination of the Radio Link Protocol (RLP) for data users.

Because end users are becoming more sophisticated and rich multimedia and real-time services are becoming more popular, the current capacity is rapidly becoming

unacceptable and the throughput bottleneck is at the T1/E1 links. In addition the QoS is becoming an essential need for delivering Next Generation Network (NGN) applications.

Majority of mobile networks still use "star topology" architecture in their backhaul, where each BS is connected directly to the RNC over dedicated T1/E1 lines. Since each T1/E1 line is dedicated to a particular BS, each Transmitter/Receiver (T_xR_x) of that BS must be assigned to dedicated channels on a given T1/E1. The Star topology offers easiest deployment and maintenance, especially in initial stages of network development, but in case of extensive growth of subscribers and installed network base this solution, however, increases costs of delivering mobile services.

Figure 6.1 shows the traditional wireless backhaul architecture. As can be seen from Figure 6.1, in 3G systems each BS is connected to the RNC by a dedicated point-to-point link, this link is widely deployed as Ethernet/IP over N E1/T1. Similarly, the 2G BSs are connected to the Base Station Controller (BSC) by a dedicated point-to-point link with E1/T1 TDM links.

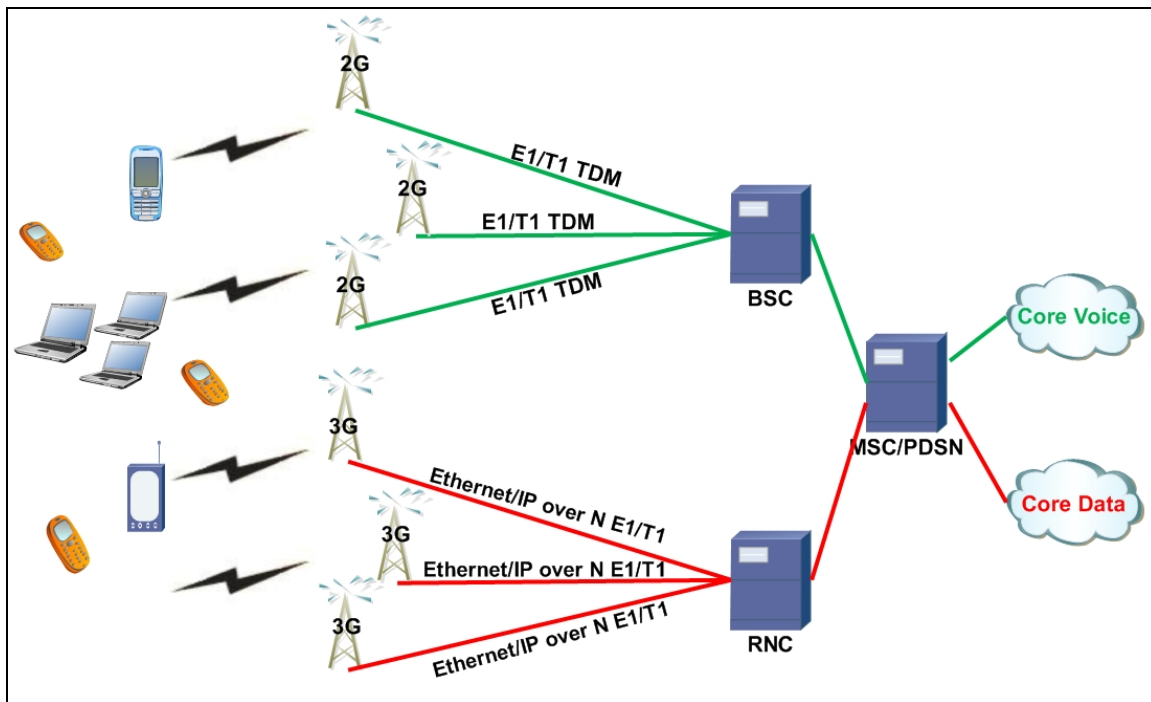


Figure 6.1: Traditional Wireless Backhaul Architecture

6.4 Proposed Wireless Backhaul Architectures

The proposed architecture will utilize the existing PON infrastructure which is currently being deployed worldwide. Analogous to PON systems, each base station receives a dedicated short optical fiber but shares the long distribution trunk fiber. Traffic from an RNC to a BS is called ‘downstream’ (point-to-multipoint), and traffic from a BS to the RNC is called ‘upstream’ (multipoint-to-point) [3-7]. Two wavelengths are used: typically 1310 nm (λ_{up}) for the upstream transmission and 1490 nm (λ_d) for the downstream transmission. A single cluster can typically serve from 16-64 base stations.

The proposed architecture offers the following significant advantages:

- a) Proposed architecture can support triple play services as being proposed in 3rd generation packet based wireless networks (3G) and the emerging 4G networks with full end-to-end differentiated QoS support.
- b) Proposed distributed architecture enables the base stations to communicate with each other, this is in compliance with the emerging 4G standards, whereas will be shown in upcoming sections, BSs can perform some of the functionality typically executed by the RNC thus offloading some of the RNC complexity.

Figure 6.2 shows the general architecture of this approach. As can be seen from Figure 6.2, the ONU will be co-located with the BS, for every cluster of BSs there will be one SC (Passive star coupler (SC)/optical splitter/combiner). All base stations of the same cluster will be connected to the star coupler co-located with one of the base stations in the cluster. The SC is connected to the OLT; the OLT in this architecture is part of the Radio Network Controller (RNC). The SC is intentionally positioned a substantial distance away from RNC, but close enough to the base stations in order to save fiber.

All BSs will utilize shared upstream link, and the RNC will utilize shared downstream link, this requires arbitration mechanisms and the utilization of DBA schemes at both the RNC and BSs (if fully distributed), as one of the architectures proposed here is a fully distributed architecture, where the BSs will have to communicate with each other and each must independently run uplink DBA.

This work utilizes the 802.3ah standard [2] as the BS will have an 802.3ah WAN interface. To control the Point-to-Multipoint fiber network, we will use the Multi-Point Control Protocol (MPCP). The MPCP specifies a control mechanism between a Master and Slaves units connected to a Point-to-Multi-Point (P2MP) segment to allow efficient

transmission of data, MPCP is defined within the MAC Control layer, introducing new 64-byte control messages:

- GATE and REPORT are used to assign and request bandwidth
- REGISTER is used to control the auto-discovery process

MPCP provides hooks for network resource optimization. Bandwidth reporting satisfies requirements by BSs for DBA. Optical parameters are negotiated to optimize performance.

The BS/ONU performs an auto-discovery process which includes ranging and the assignment of both Logical Link IDs and bandwidth. Using timestamps on the downstream GATE MAC Control Message, the BS/ONU synchronizes to the RNC/OLT timing. It receives the GATE message and transmits within the permitted time period. The RNC/OLT generates time stamped messages to be used as global time reference. It generates discovery windows for new BSs, and controls the registration process.

Downstream: The PON handles the physical broadcast of 802.3 frames. Broadcast frames are extracted by the logical link ID in the preamble. 64-byte GATE messages are sent downstream to assign upstream bandwidth.

Upstream: The MPCP uses time slots containing multiple 802.3 frames, where each BS sends a 64-byte REPORT message, with BS state information to the OLT. There are no collisions and no packet fragmentation in this architecture.

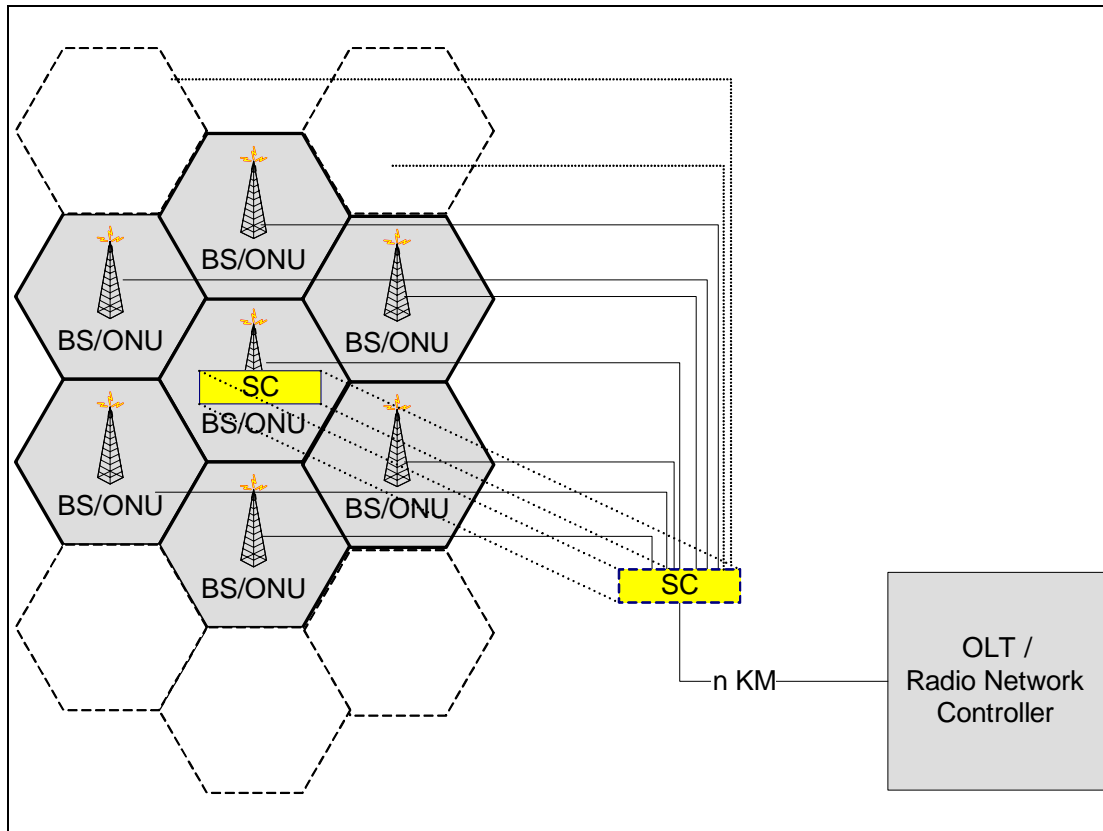


Figure 6.2: Proposed EPON Wireless Backhaul Architecture

6.4.1 Proposed Standalone Wireless Backhaul Architectures

6.4.1.1 Smooth Transitional Wireless Backhaul Standalone Architecture for 2G/3G and 3G+

In this proposed architecture shown in Figure 6.3, we are offering a smooth transitional scenario of 2G, 3G and 3G+ systems with only one aggregated backhaul network. By utilizing the existing PON infrastructure, the 3G/3G+ BS will have a 100 Base T Ethernet interface, this interface will be connected directly to the ONU (ONU output supports the 802.3ah standard). The ONU can be co-located with the 3G/3G+ BS or serving as an

interface card inside the BS platform. The 2G BS has TDM N E1/T1 interface, by using a packet mapper, the circuit switch BS output will be carried over packetized 100 Base T Ethernet interface, which is then connected directly to the ONU.

This architecture will utilize the 802.3ah standard, and to control the Point-to-Multipoint fiber network, we will use the Multi-Point Control Protocol (MPCP). The RNC will utilize shared downstream link, this requires arbitration mechanisms and the utilization of DBA schemes at both the OLT/RNC and ONUs.

In distributed architecture, since the integrated scheduling and DBA algorithms will be conducted at each ONU, then the ONU can feedback the information in the control plane to the BSs, eventually, the BSs can make use of the ONU bandwidth allocation information for admission control. Figure 6.4 demonstrates the congestion control operation, when congestion bottleneck is at the backhaul and not on the radio interface, the ONU will report that the BS requested bandwidth cannot be achieved and that the backhaul is congested, this will trigger the admission control module at the base station to block future users till congestion is released. When backhaul congestion is released the ONU will send another control message to inform the base station admission control module.

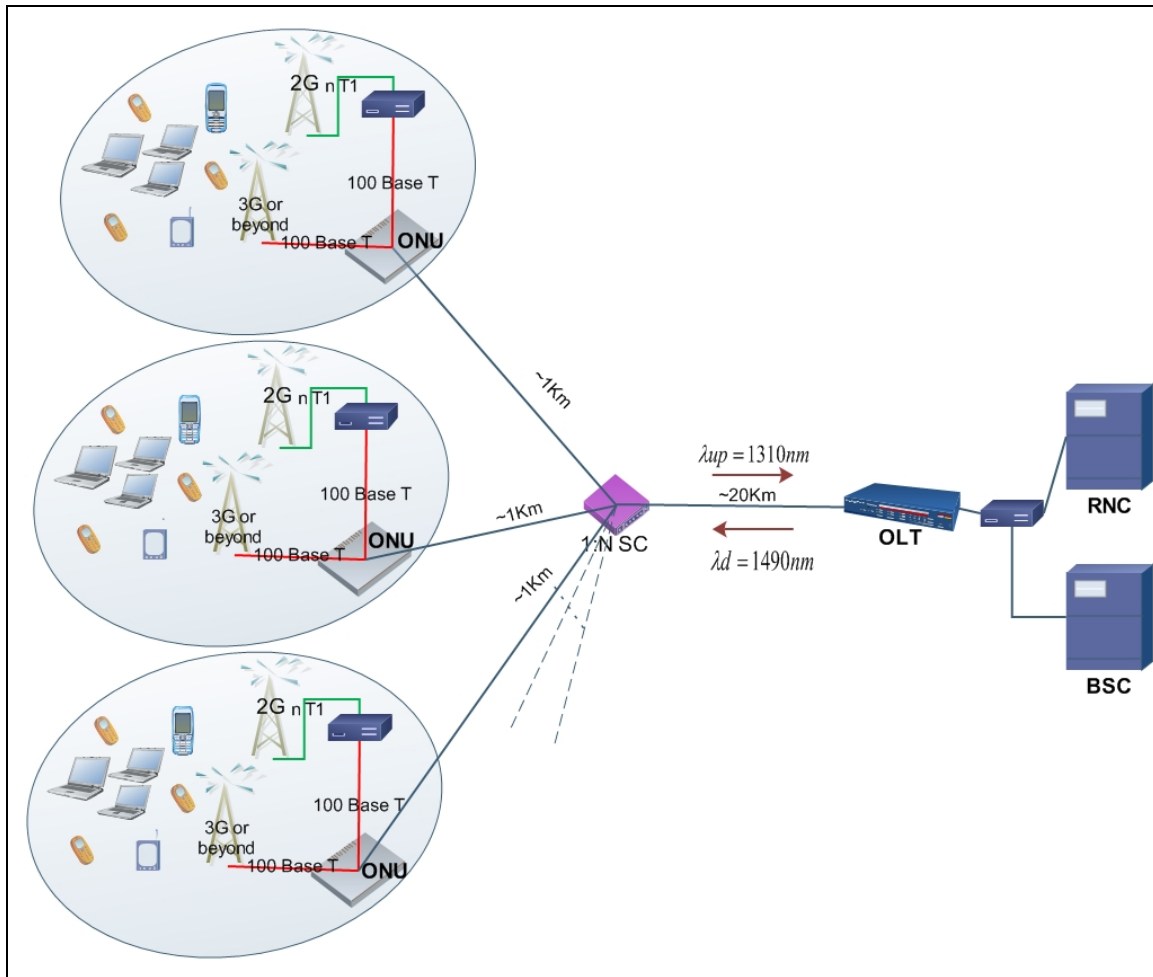


Figure 6.3: Proposed 2G/3G/3G+ Smooth Transitional Wireless Backhaul Architecture

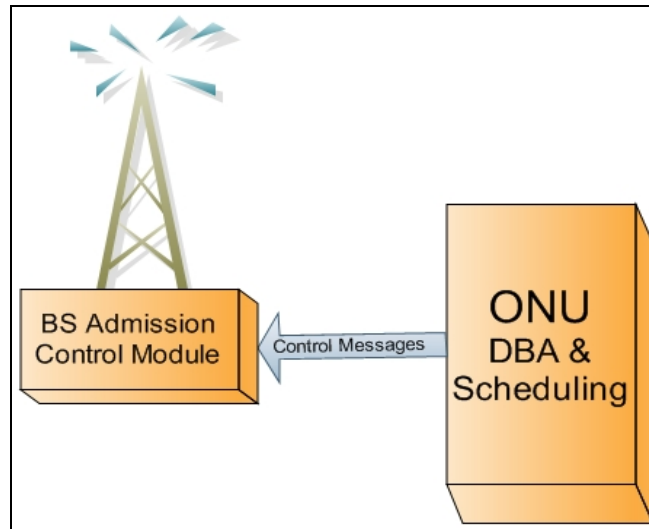


Figure 6.4: Base Station Admission Control

6.4.1.2 Fully Packet-Based Standalone Wireless Backhaul Architecture

This proposed architecture is targeting fully packet-based wireless systems. Figure 6.5 illustrates the architecture diagram, in which each BS and the RNC will be equipped with an 802.3ah interface. As described previously, the backhaul will utilize the existing PON infrastructure, in which the SC is intentionally positioned a substantial distance away from RNC, but close enough to the base stations in order to save fiber. This architecture similarly will utilize the 802.3ah standard, and to control the Point-to-Multipoint fiber network, we will use the Multi-Point Control Protocol (MPCP). The RNC will utilize shared downstream link, this requires arbitration mechanisms and the utilization of DBA schemes at both the RNC and BSs.

In distributed architecture, since the integrated scheduling and DBA algorithms will be conducted at each BS, then in case of backhaul congestion the DBA and scheduling module can send internal control messages to the BS admission control module to block

future users till congestion is released and another control message will be sent from the DBA and scheduling module to the admission control module.

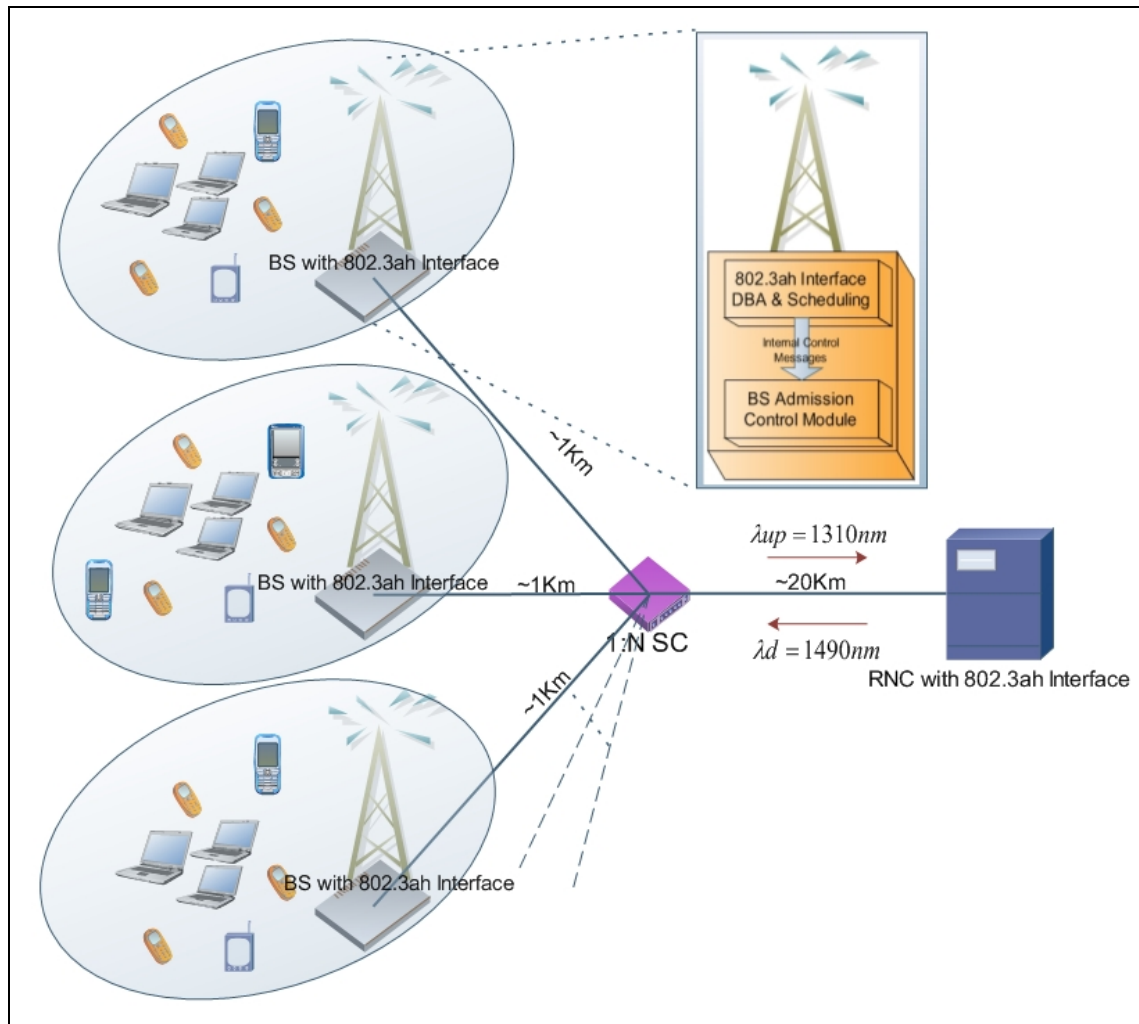


Figure 6.5: Proposed Standalone Fully Packet-Based Wireless Backhaul Architecture

Figure 6.6 shows another way to implement the architecture where multiple OLT will serve single RNC, this architecture will be suitable for suburban and less populated coverage where single RNC will serve several long-reach clusters of BSs. For this

architecture the OLT and the RNC will be non co-located, each cluster of BSs will be connected to a single OLT, to achieve connectivity with the single RNC, the RNC and each OLT will be connected to a WAN where full connectivity is achieved.

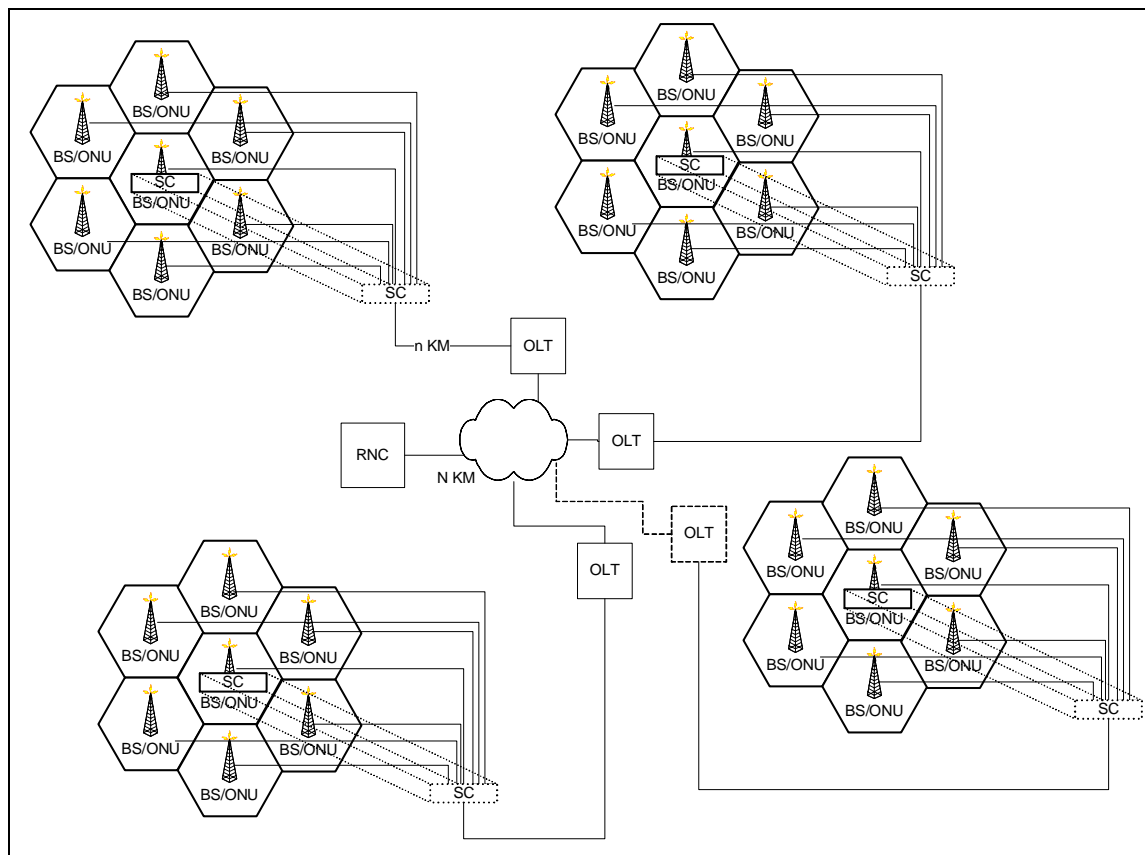


Figure 6.6: Proposed RAN Architecture OLT/RNC non co-located

6.4.2 Proposed Integrated Wired/Wireless Backhaul Architecture

The proposed integrated architecture will support both fixed (wired) users and wireless base station mobile users. The two devices will support common Ethernet standard (100 Base T). The ONU can have two different port ranges, the first port range will support

fixed users, and the second port range will support wireless mobile users. The port ranges will be used by the ONU to identify and differentiate between mobile users versus fixed users. This will help achieve the following functions:

- a) **Wireless/Mobile admission control:** Analogous to previous architectures, in case of backhaul congestion, after identifying wireless mobile traffic, the ONU can send control messages to the BS to trigger wireless admission control.
- b) **Service Level Agreements:** If carriers request higher priority for wireless/mobile traffic, then this can be achieved by identifying the wireless/mobile traffic and assigning it in higher priority queues than fixed users (in distributed architecture DBA and scheduling is done at the same ONU box).

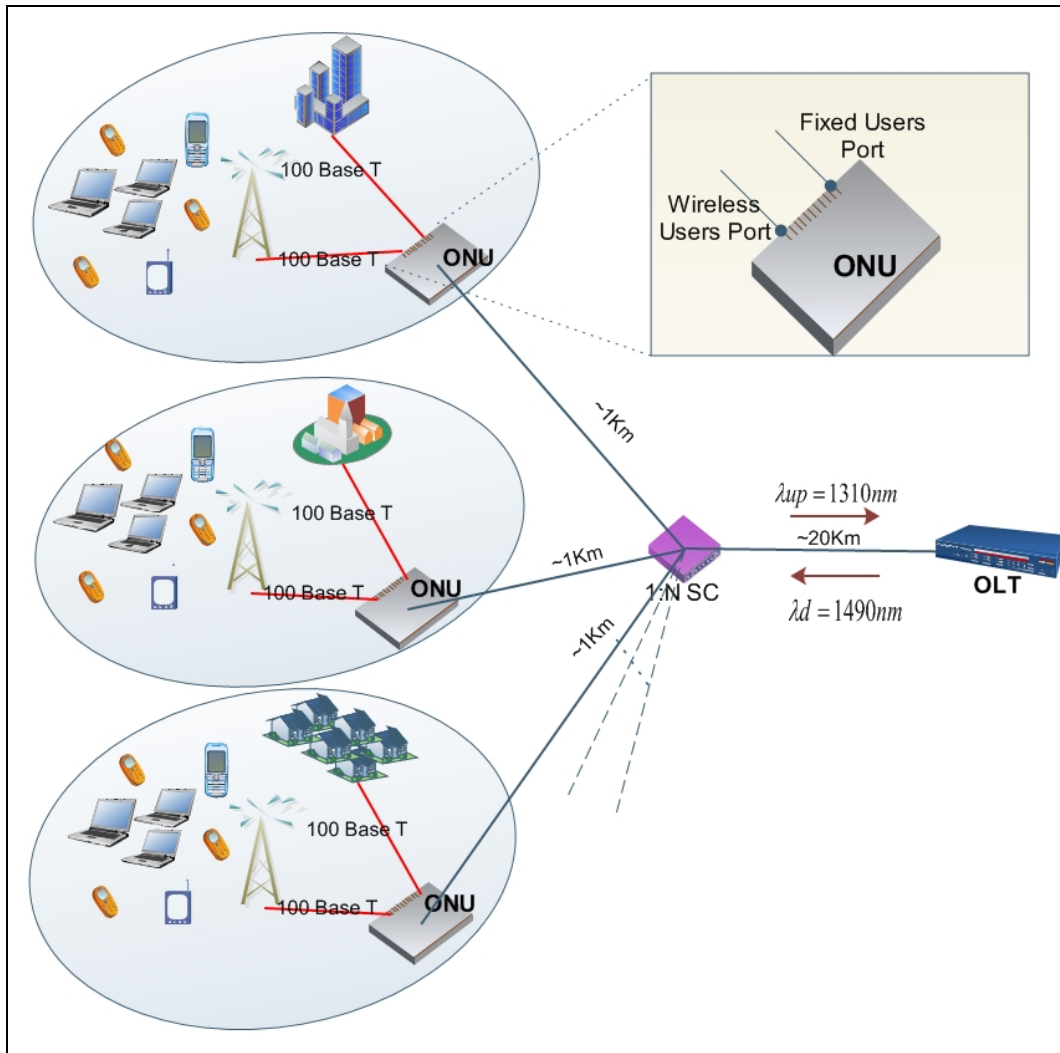


Figure 6.7: Proposed Integrated Wired/Wireless Backhaul Architecture

6.5 Bandwidth Allocation Schemes for Proposed Architecture

6.5.1 Decentralized Architecture Bandwidth Allocation Schemes

Mainstream EPONs bandwidth allocation schemes as well as the new IEEE 802.3ah EFM Task Force specifications [2] have been centralized, relying on a component in the

central office (RNC or OLT) to provision upstream traffic. Hence, the RNC/OLT is the only device that can arbitrate time-division access to the shared channel. Since the RNC/OLT has global knowledge about the state of the entire network, this is a centralized control plane in which the RNC/OLT has centralized intelligence. It is the purpose of this work to propose distributed solutions to this problem, and in the process to prove that these distributed networking architectures and the associated bandwidth allocation algorithms and protocols have characteristics that make them far better suited for provisioning Quality of Service (QoS) schemes necessary for properly handling data, voice, video, and other real-time streaming advanced multimedia services over a single line. The *decentralized scheme* used is the one described in chapters 2, 3, 4 and 5.

6.5.1.1 Decentralized Uplink DBA Scheme

For the *decentralized scheme*, the *uplink* instances of the same DBA algorithm are executed simultaneously at each BS. Thus, both scheduling tasks (DBA scheme and priority queuing) are performed at the BS leading to the notion of integrating both scheduling mechanisms at the BS. Also, each BS will be of complete knowledge of the queue status of all other BSs in the same cluster, this will lead to the ability of executing a better efficient DBA and can apply fairness among BSs. For *decentralized uplink* traffic, we used the following proposed DBA

Remainder DBA:

The *Remainder DBA* is an Efficient Bandwidth Utilization-Based Enhanced DBA Algorithm. Due to the bursty nature of Ethernet traffic, some BSs might have less traffic to transmit while other BSs may require more than W_{MAX} . Consequently, BSs can be classified into two groups based on their requested bandwidth, namely: *lightly loaded BSs*

which might have requests less than bandwidth W_{MAX} ; and *heavily loaded BSs* which may have requests more than bandwidth W_{MAX} . The lightly loaded BSs with $Req_i < W_{MAX}$ will contribute a total remainder bandwidth:

$$Remainder_{total} = \sum_i^L (W_{MAX} - Req_i) \quad L: \text{Number of lightly loaded BSs}$$

The heavily loaded BSs with $Req_i \geq W_{MAX}$ will require a total over the limit bandwidth:

$$OverLimit_{total} = \sum_i^H (Req_i - W_{MAX}) \quad H: \text{Number of heavily loaded BSs}$$

The total remainder bandwidth can be fairly distributed amongst the heavily loaded BSs to expand their maximum transmission window as follows [11]:

$$\Delta W_i = Remainder_{total} \times \frac{Req_i}{OverLimit_{total}}$$

where ΔW_i is the extra bandwidth allocated to BS_i.

$$BW_{Granted} = \begin{cases} Req_i & \text{If } Req_i \leq W_{MAX} \\ Req_i & \text{If } Req_i > W_{MAX} \text{ \& } Remainder_{total} \geq OverLimit_{total} \\ Req_i + \Delta W_i & \text{If } Req_i > W_{MAX} \text{ \& } Remainder_{total} < OverLimit_{total} \end{cases}$$

To support differentiated classes of service, the remainder bandwidth can be distributed per priority queue for a given BS using a weight function P_{Q_i} . In this case Q is the priority queue number (class of service, i.e., P₀, P₁, and P₂) and P_{Q_i} is the weight allocated to a given priority queue (ranging from 0 to 1) for BS_i, where $\sum_{Q=0}^k P_{Q_i} = 1$. The total

remainder bandwidth can then be distributed amongst the queues of the heavily loaded BSs to expand their maximum transmission window as follows:

$$\Delta WP_{Q_i} = Remainder_{total} \times \frac{Req_i}{OverLimit_{total}} \times P_{Q_i}$$

6.5.1.2 Decentralized Downlink DBA Scheme

For the *downlink* traffic of the *decentralized scheme*, the OLT/RNC has the queue information of each BS traffic. Therefore, the OLT/RNC will be able to make a collective decision based upon all queues information. The DBA enforced will be the *Remainder DBA* described in the previous section, as the remainder DBA will guarantee efficient usage of all bandwidth available per cycle and at the same time will guarantee fairness among BSs. Note that for the *decentralized scheme* the *downlink* transmissions from the OLT/RNC to each BS will not have any GATE message (64 Byte) added at the end of each transmission, this will enhance the *decentralized scheme* bandwidth utilization for the downlink traffic.

6.5.2 Centralized Architecture Bandwidth Allocation Schemes

In the centralized architecture the OLT is the only device that can arbitrate time-division access to the shared uplink channel. The OLT provision the upstream traffic. Since the OLT has global knowledge about the state of the entire network, this is a centralized control plane in which the OLT has centralized intelligence.

6.5.2.1 Centralized Uplink DBA Scheme

For the *centralized scheme*, the *uplink* will be using a cycle-based DBA algorithm, where a cycle is defined as the time that elapses between two executions of the scheduling algorithm. A cycle has a variable length size confined within certain lower and upper bounds, which we denote as T_{MIN} and T_{MAX} (sec). Thus, the algorithm schedules between B_{MIN} and B_{MAX} (bytes) at a time, where B_i is found by multiplying T_i with the line rate. In this scheme, the BS will be granted the requested number of bytes, but no more than a given predetermined maximum W_{MAX} (maximum transmission window). If Req_i is the requested bandwidth of ONU_i and $Grant_i$ is the granted bandwidth, $Grant_i$ is then equal to

$$Grant_i = \begin{cases} Req_i & \text{if } Req_i < W_{MAX} \\ W_{MAX} & \text{if } Req_i \geq W_{MAX} \end{cases}$$

W_{MAX} is determined by the maximum cycle time T_{MAX} [7-8]

$$T_{MAX} = N \left(Guard_{total} + \frac{W_{MAX}}{Transmission_Speed} \right) \quad N: \text{number of BSs.}$$

$Guard_{total}$ is the guard time added between two consecutive transmission windows (TWs) during the transmission period.

6.5.2.2 Centralized Downlink DBA Scheme

Analogous to the *downlink* traffic of the *decentralized scheme*, we will use the *Remainder DBA* for the *downlink* traffic of the *centralized scheme* as the OLT/RNC has the queue

information of each BS traffic. Therefore, the OLT/RNC will be able to make a collective decision based upon all queues information.

For the uplink *Fair Non-Strict Priority Scheduling* is used for both architectures.

Fair Non-Strict Priority Scheduling:

In this scheduling scheme, the BS keeps track of the total number of bytes (per queue) that it has reported in the last update message (i.e., the report transmitted during the last cycle), and it transmits the reported data first, regardless of their priorities [11]. Only those packets that arrived before t_0 will be transmitted first provided that the bandwidth of the allocated timeslot allows for the transmission. Thus, if higher priority packets arrived after t_0 , they have to wait until the reported (arrived before t_0) lower priority traffic is transmitted. If packets arriving before t_0 are all scheduled, and if the current timeslot can still accommodate more traffic, it will be allocated for higher priority traffic. This scheme will ensure fairness in scheduling packets.

6.6 Wireless Classes of Services (CoS) and Traffic Characteristics

In a typical wired system the classes of services usually use three classes (BE, Video and Voice). In the wireless systems the traffic characteristics leads to the concept of introducing an additional CoS which arises due to lower layer retransmission of lost packets from BS to RNC and vice versa.

Providing Internet services in third generation wireless networks is a very important and challenging issue. TCP is the transport layer protocol used in many Internet applications,

such as HTTP, FTP, TELNET, etc. However, it is well known that TCP will experience significant performance degradation when applied to links with high loss rate (e.g., 5 percent or higher), which is very likely in wireless links. This is mainly caused by TCP's congestion control mechanism which reduces window size at the occurrence of packet loss [24].

Many mechanisms have been proposed to solve the problem of TCP performance degradation over wireless networks. For example, lower layer retransmissions have been implemented in various 3G Code Division Multiple Access (CDMA) systems to provide reliable wireless links for TCP. They include Radio Link Protocol (RLP) in cdma2000 standard [25] and Radio Link Control (RLC) in Universal Mobile Telecommunication System (UMTS) [23].

The IS-707 RLP mechanism in cdma2000 is a pure Negative Acknowledgment (NAK)-based selective repeat protocol [26], i.e. the receiver only requests retransmission for the incorrectly packets. RLP performs a partial link recovery through a limited number of packet retransmission. Figure 6.8 shows the uplink RLP mechanism, Frame 1 and Frame 2 were sent from the mobile and received with no error at the RNC, frame 3 and frame 4 are either lost or received but with errors, when frame 5 is received by RNC, the RNC will detect RLP holes (from the jumping of 2 RLP sequence numbers). RNC will send one NAK request message for frames ranges from 3 to 4, once received by the mobile; the mobile will retransmit frames 3 and 4. Figure 6.9 shows the Downlink RLP mechanism, where Frame 1 and Frame 2 were forwarded from the RNC and received with no error at the mobile, frame 3 and frame 4 are either lost or received but with errors, when frame 5 is received by the mobile, the mobile will detect RLP holes (from

the jumping of 2 RLP sequence numbers). Mobile will send a one NAK request message for frames ranges from 3 to 4, once received by the RNC, the RNC will retransmit frames 3 and 4.

Practically, the retransmission is permitted for non-real time application (Best-Effort data). When the retransmitted packets arrive at ONU/BS, the ONU/BS will mark the packet with its appropriate Diff-Serv class. The next section will include the details of traffic model used in our study.

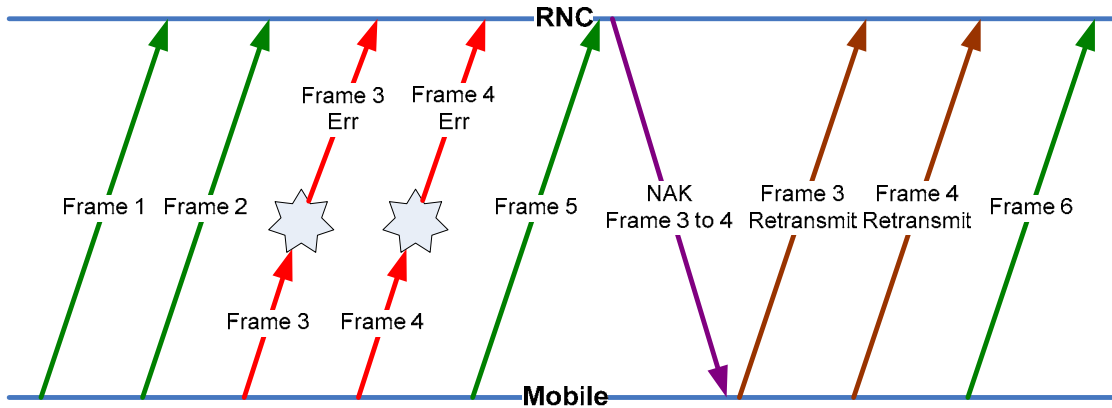


Figure 6.8: Uplink RLP Mechanism

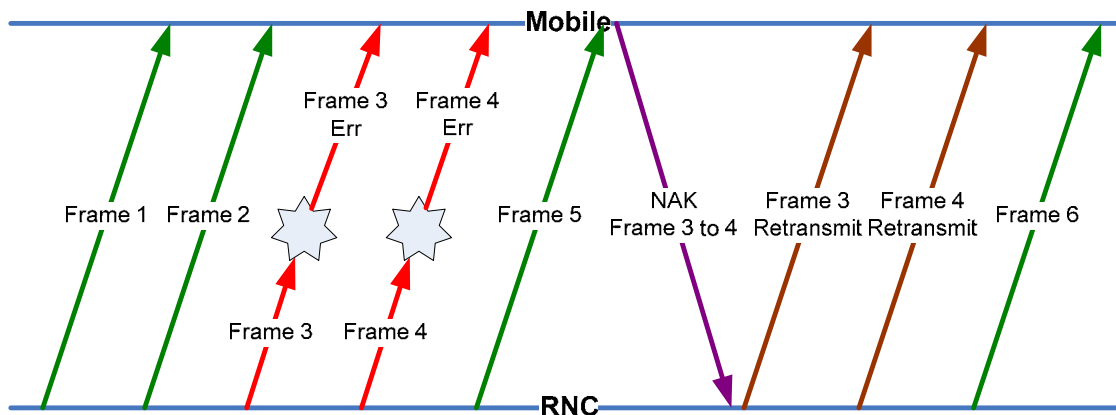


Figure 6.9: Downlink RLP Mechanism

6.7 Wireless Backhaul QoS Traffic Model

We consider four priority classes P0, P1, P2, and P3 with P0 being the highest priority (constant-bit-rate) and P3 being the lowest (non-real time data transfer). These classes are used to for delivering voice (constant-bit-rate or CBR), Wireless Retransmitted Packets (retransmission is only permitted for Best-Effort traffic), video stream (variable-bit-rate or VBR), and best-effort data. Each BS maintains four separate priority queues that share the same buffering space.

- **Class P3** is used to emulate self-similar [18-19] and long-range dependant traffic. This class has variable-size packets ranging from 64 bytes to 1518 bytes. This class has the lowest priority. This priority level is used for non-real time Best-Effort (BE) data transfer. There is no delivery or delay guarantees in this service.

- **Class P2** consisted of VBR video streams that exhibit properties of self-similarity [18-19] and long-range dependence. Packet sizes in P2 streams ranged from 64 bytes to 1518 bytes.
- **Class P1** The wireless network is using The IS-707 RLP mechanism in cdma2000. The receiver (RNC) only requests retransmission for the incorrectly packets. The retransmission is permitted only for Best-Effort data, assuming retransmission rate of 1% of BE data.

The selection of 1% retransmission rate was proven by [27] in a TCP Performance over cdma2000 RLP study to give the best overall throughput. The retransmitted packets will be marked with class P1 priority marking.

- **Class P0** is used to emulate a circuit over packet connection. P0 traffic has constant-bit rate (CBR). In our model, we chose to emulate a voice T1 connection. The voice data arriving from the user is packetized by placing 24 bytes of data in a packet. Including Ethernet and UDP/IP headers this results in a 70-byte frame generated every $125 \mu s$. Hence, the P0 data consumed 4.48 Mbps of bandwidth.

This is the highest priority traffic.

In our simulation, when we varied the load, P0 was always kept constant (4.48 Mbps).

The remaining load was divided between P1, P2 and P3 according to the following rule:

$$P2 \text{ Load} = P3 \text{ Load}$$

$$P1 \text{ Load} = (P3 \text{ Load}) * 0.01$$

For example, if we assume a BE load of 0.32 then:

$$\text{The CBR P0 load will be: } 70 \text{ bytes every } 125 \mu s = (4.48\text{Mbps}/100\text{Mbps}) = 0.0448;$$

$$P3 \text{ Load} = 0.32$$

$$P2 \text{ Load} = P3 \text{ Load} = 0.32$$

$$P1 \text{ Load} = (P3 \text{ Load}) * 0.01 = 0.0032$$

$$\text{Total ONU Load} = P0 \text{ Load} + P1 \text{ Load} + P2 \text{ Load} + P3 \text{ Load}$$

$$\text{Total ONU Load} = 0.0448 + 0.0032 + 0.32 + 0.32 = 0.688$$

For Class P3 (Best Effort Traffic) The traffic model used for the performance evaluation is the same as that reported in [8] where each BS has a number of ON/OFF sources, each with a Pareto distribution governing the lengths of the ON/OFF periods, in order to capture the self-similar nature of Ethernet traffic.

Chapter 7

PROPOSED WIRELESS BACKHAUL PERFORMANCE EVALUATION

7 Chapter 7: Proposed Wireless Backhaul Performance Evaluation

7.1 Introduction

In this chapter, we compare the overall performance of the proposed EPON Wireless Backhaul with traditional wireless backhaul defined in previous section. An event-driven packet-based simulation model was developed using C++. Three simulation programs were developed:

- a) Traditional wireless backhaul
- b) Centralized EPON wireless backhaul
- c) Decentralized EPON wireless backhaul

Two different scenarios of traffic loading are considered:

1. In the first scenario we will consider the widely used assumption that all BSs has uniform traffic, this will be used as a reference model and for illustration purpose. We simulated the network load with an evenly distributed load over the 16 BSs. Although this scenario is rarely occurring in real life scenario, but in most of the previous studies [7-11] it has been the only loading scenario used.
2. In the second scenario we consider the most realistic and practical case in which some BSs might be almost idle during given period and other BSs might be heavily loaded. To simulate real life scenarios and to demonstrate uneven loading scenarios, we used uneven BSs loads for the 16 BSs under different total network loading conditions, we divided the BSs into “Heavily Loaded”, “Super Heavily

Loaded”, “Moderately Loaded” and “Lightly Loaded”, Table 7.1 shows the details for different loading scenarios.

Heavily Loaded BSs		Lightly Loaded BSs		Total Network
# BSs	BS Load	# BSs	BS Load	Load
1	0.8488	15	0.1051	0.24
3	0.8488	13	0.1051	0.39
5	0.8488	11	0.1051	0.54
Super Heavily Loaded BSs		Moderately Loaded BSs		Total Network
# BSs	BS Load	# BSs	BS Load	Load
1	1.85	15	0.3262	0.68
2	1.85	14	0.3262	0.83
3	1.85	13	0.3262	0.98
4	1.85	12	0.3262	1.13

Table 7.1 Unevenly Loaded Base Stations Scenarios

The performance metrics used here are network utilization, packet drop rate and we will consider another important performance metric that is widely used by network planners which is called utilization *Peak to Average*, this performance metric indicates the network architecture stability and predictability.

$$Peak_to_Average(Utilization) = Max(Utilization) / Average(Utilization)$$

To compare the performance results of the proposed EPON wireless backhaul model with that of the traditional wireless backhaul, we used a system with 16 BSs, for the proposed EPON as wireless backhaul architecture, upstream link data rate (from an BS to the RNC) is 1 Gb/s. Each BS will have an ONU interface (802.3ah WAN interface), and the RNC will have an OLT function.

The distance between the RNC and the BSs is 20 km. Maximum cycle time $T_{MAX} = 2\text{ms}$. The guard time separating two consecutive transmission windows is set to $5 \mu\text{s}$ which are reasonable choice to account for the laser on and off times, receiver recovery times, round trip delay (which depends on the physical distance between communicating BSs) and other optic related issues. Buffer size in each BS is 1 Mbytes. Each point on the following plots corresponds to a sample of ~ 250 million packets for a simulated time of 600 sec, averaged over four different runs. For the traditional wireless backhaul all network parameters (buffer size, distances, loading scenarios, simulation time) are identical to the proposed architecture parameters except to be fair in the comparative study, each BS will have a maximum transmission speed of 62.5 Mbps (to give a total of 1Gbps of total aggregated network speed) $1\text{Gbps} = 62.5\text{Mbps} * 16$

7.2 Numerical Results for the Evenly Loaded Base Stations Scenarios

Figure 7.1 shows uplink utilization under various network loads (evenly loaded), for all the uplink loading scenarios with total network load less than or equal 0.83 the network utilization for the proposed EPON wireless backhaul architecture was almost identical to the traditional wireless backhaul architecture. For network loads above 0.83, the traditional backhaul recorded slightly higher utilization than the EPON backhaul.

Figure 7.2 shows the network utilization over time at a single network load of 0.83, the EPON wireless backhaul architecture demonstrated network utilization stability over time whereas the traditional wireless backhaul demonstrated fluctuation and high variation in utilization over time. Figure 7.3 shows the network utilization statistics at the same network load of 0.83, although the average utilization is almost identical but EPON wireless backhaul proved to have much lower standard deviation.

This proves that with using evenly loaded BSs, which is a model widely used in previous studies, the numerical results for the EPON Wireless Backhaul utilization were comparable to the Traditional Wireless Backhaul at almost all network loads. Although the EPON wireless backhaul utilization average for network loads higher than 0.83 were slightly less than the traditional wireless backhaul utilization, but EPON wireless backhaul proved to have greater utilization stability with less variations in utilization, and this strengthen the network stability and predictability.

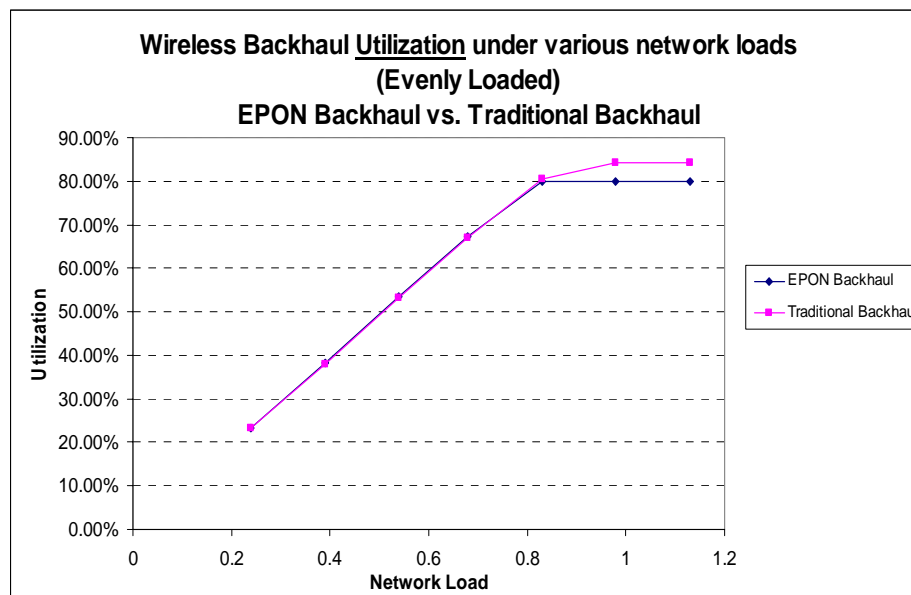


Figure 7.1: Evenly Loaded BSs Network Utilization EPON BH vs. Traditional BH

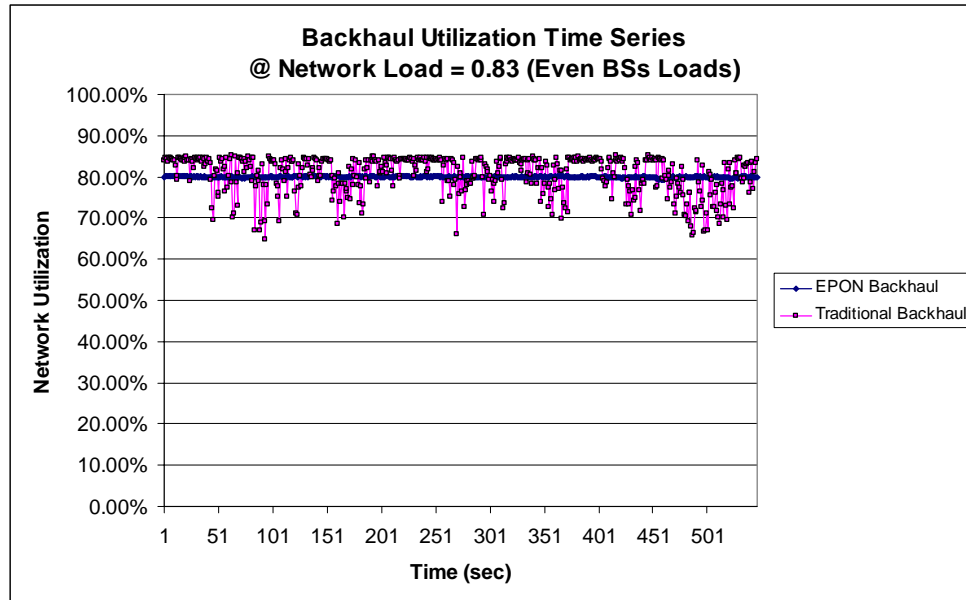


Figure 7.2: Evenly Loaded BSs Network Utilization at 0.83 Network Load

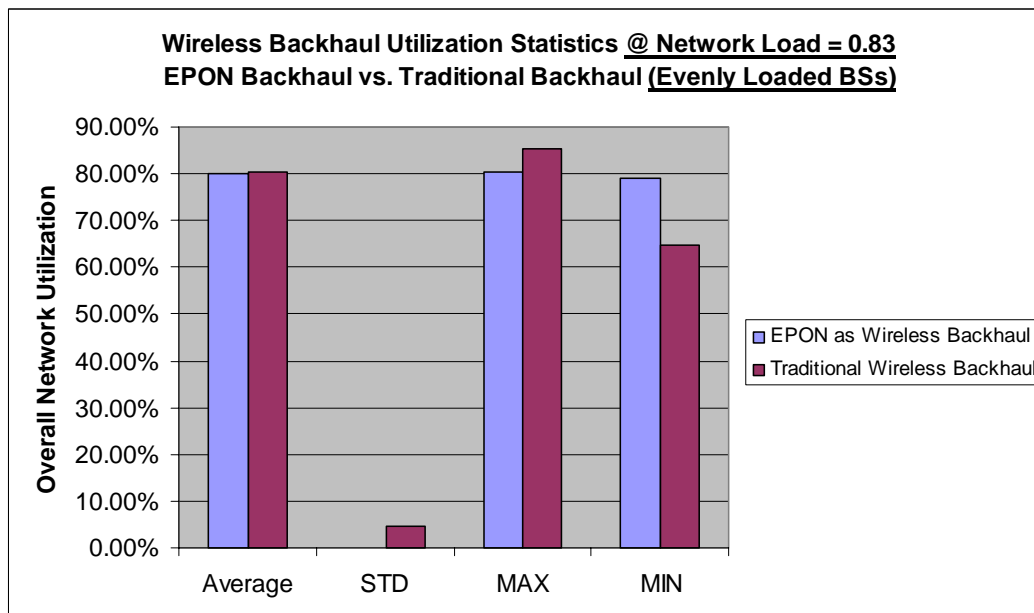


Figure 7.3: Evenly Loaded BSs Network Utilization statistics at Network Load=0.83

7.3 Numerical Results for the Unevenly Loaded Base Stations Scenarios

Since using the unevenly loaded BSs scenarios are more realistic, so we will use these scenarios in the entire comparative study and numerical results for the rest of this chapter.

The comparison will cover the following:

- Uplink Traffic
 - Traditional Wireless Backhaul
 - Centralized EPON Wireless Backhaul
 - DBA used: Limited DBA described in section 6.5
 - Priority scheduling used: Fair Non-Strict Priority Scheduling described in section 6.5
 - Decentralized EPON Wireless Backhaul
 - DBA used: Remainder DBA described in section 6.5
 - Priority scheduling used: Fair Non-Strict Priority Scheduling described in section 6.5
- Downlink Traffic
 - Traditional Wireless Backhaul
 - Centralized EPON Wireless Backhaul
 - DBA used: Remainder DBA described in section 6.5
 - Decentralized EPON Wireless Backhaul
 - DBA used: Remainder DBA described in section 6.5

- Note: No GATE message is used in decentralized architecture.

7.3.1 Uplink Numerical Results

7.3.1.1 Uplink: Centralized EPON Wireless BH

Figure 7.4 shows uplink utilization under various network loads comparing Centralized EPON Wireless BH versus Traditional Wireless BH. Through all the loading conditions the Centralized EPON wireless backhaul showed much higher utilization. High utilization peak to average observed in traditional backhaul even at low network loads (Network load = 0.24), the traditional backhaul recorded 1.548 utilization peak to average. Whereas for Centralized EPON Wireless Backhaul the utilization peak to average was at a low level of 1.195.

Figure 7.5 shows the network utilization statistics at a single network load of 0.24, Traditional Wireless Backhaul showed higher standard deviation for network utilization on the contrary Centralized EPON Wireless Backhaul showed very stable utilization with minimal standard deviation. Figure 7.6 shows the network utilization statistics at a single network load of 0.83, similarly, Traditional Wireless Backhaul showed higher standard deviation for network utilization, as for the Centralized EPON Wireless Backhaul it recorded almost zero standard deviation. Figure 7.7 shows comparison for the utilization time series at a single network load of 0.83, from this figure we can observe the network utilization stability for the Centralized EPON Wireless Backhaul architecture versus the high variance and fluctuation in utilization for the traditional backhaul.

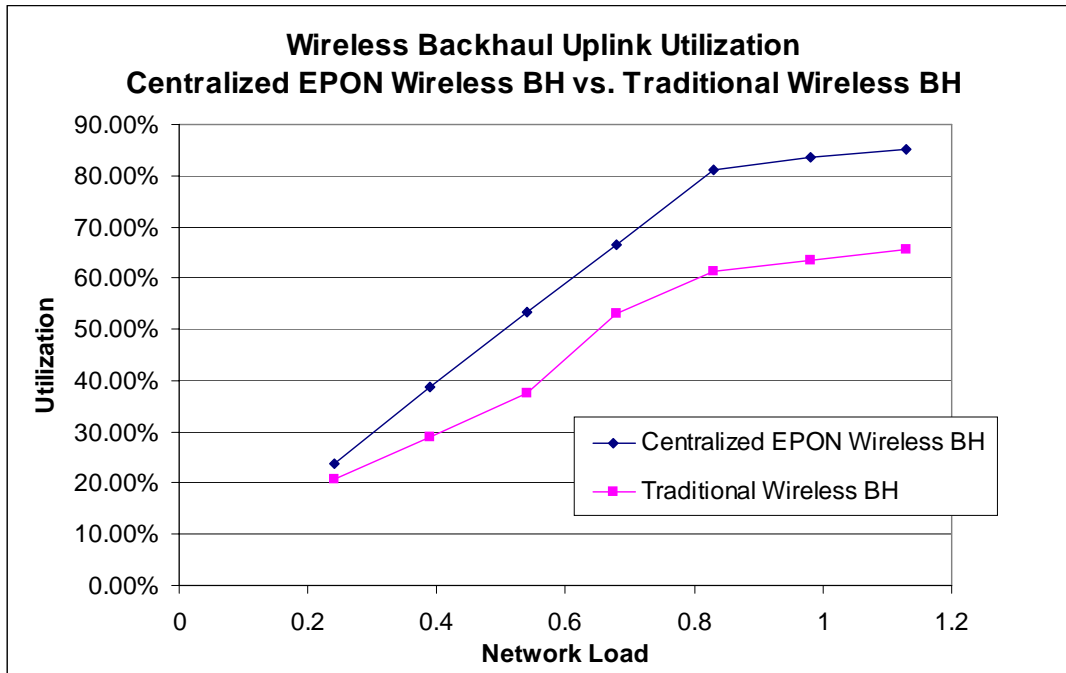


Figure 7.4: Uplink Utilization: Centralized EPON Wireless BH vs. Traditional BH

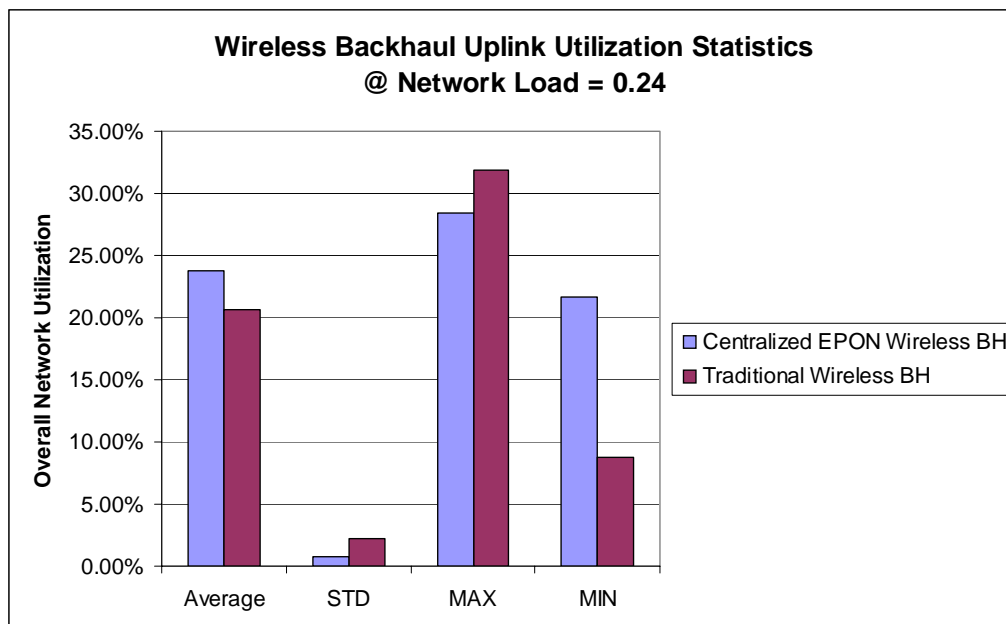


Figure 7.5: Uplink Utilization statistics at Network Load=0.24

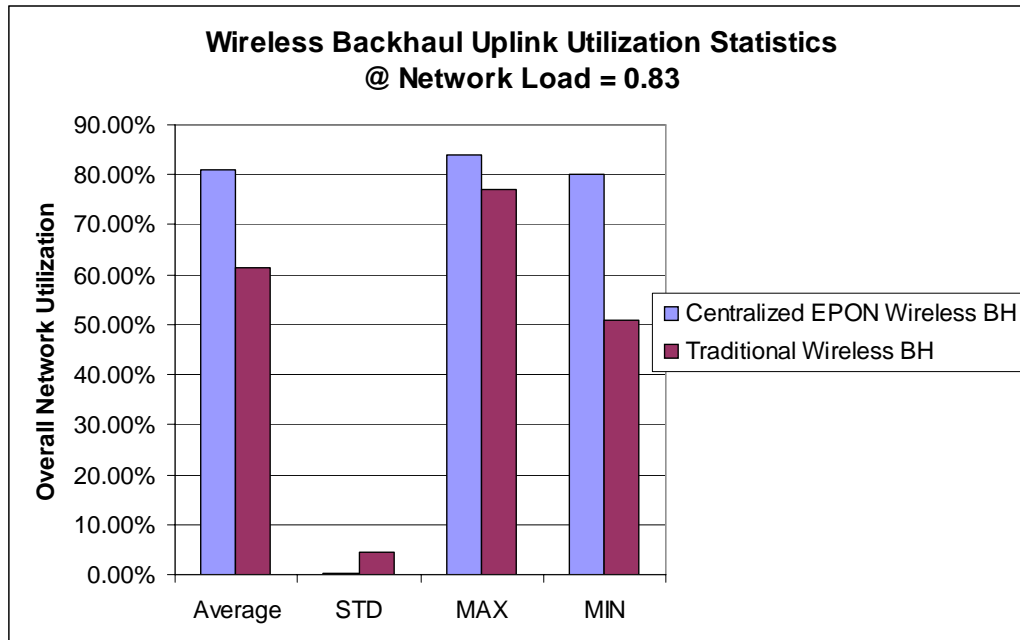


Figure 7.6: Uplink Utilization Statistics at Network Load=0.83 (Centralized EPON BH vs. Traditional BH)

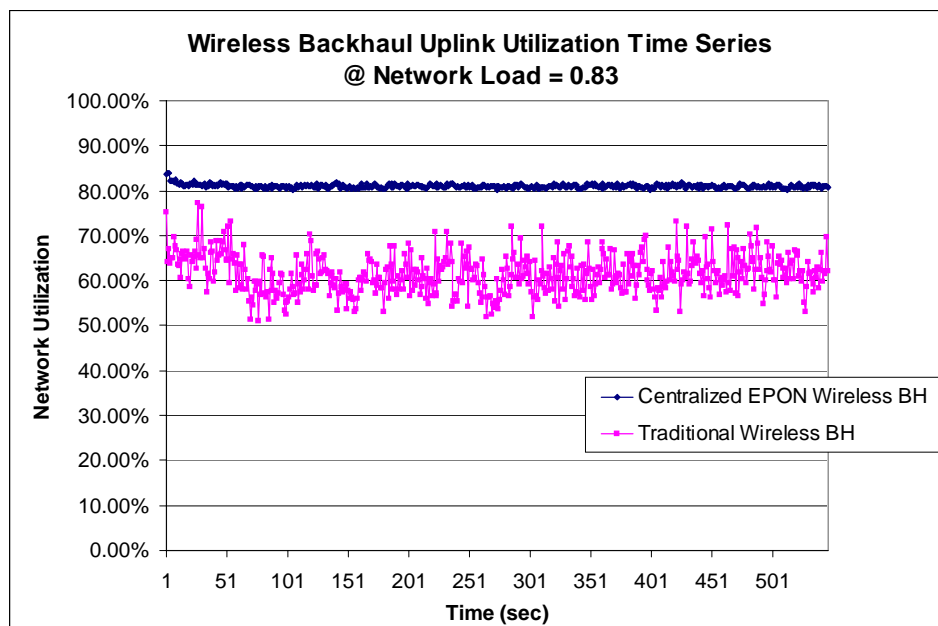


Figure 7.7: Uplink Utilization Time Series at Network Load=0.83 (Centralized EPON BH vs. Traditional BH)

Packet Drop Rate:

Figure 7.8 shows the packet drop rate under various network loads comparing Centralized EPON Wireless BH versus Traditional Wireless BH. In the figure, the packet drop rate is calculated by dividing the total number of packets dropped from all queues by all the packets generated. Centralized EPON Wireless Backhaul showed much lower packet drop rate under all the loading conditions ranging from total network load of 0.24 up-to 1.13. For Centralized EPON Wireless Backhaul, no packet drop observed at all loads less than or equal 0.68, whereas for the traditional backhaul even for the lowest network load (0.24) packet drop was observed (0.23%). Although the percentage of packet drop at that load is not significant, but at this very low network load we should expect 0% packet drop rate. Due to the bursty nature of the BE and Video traffic, even under low loads the traditional backhaul showed packet drop, as for the EPON Wireless Backhaul the aggregation of the traffic into single line allowed the network to absorb the burstiness into a more constant overall traffic.

Additionally, for low BS loads (BS load = 0.1051), the traditional backhaul showed some packet drop percentage due to the burstiness in traffic, whereas for the proposed centralized architecture the burstiness phenomenon was absorbed by using the uplink DBA algorithm and by aggregating the traffic into a single line connection.

Table 7.2 shows the packet drop rates per Class of Service at a single network load of 0.83. For Centralized EPON Wireless Backhaul all the packets dropped were only from the BE Class P3, as for the Traditional Backhaul class P1 dropped 0.17% of all P1 traffic, and class P2 dropped 13.78% of all P2 traffic, and also class P3 dropped 46.86% out of all P3 traffic. This proves that the proposed centralized EPON wireless architecture

performed far better than the traditional wireless backhaul, and also it signifies that using the proposed architecture will benefit the enforcement of the differentiated QoS, classes P0, P1 and P2 showed 0% packet drop rate. Even for the total packets dropped, the proposed architecture experienced only 8.31% versus 21.18% packet drop for the traditional backhaul. This proves that on all packet drop rate aspects under this high network load of 0.83, the proposed centralized EPON wireless architecture performed much better than the traditional wireless backhaul.

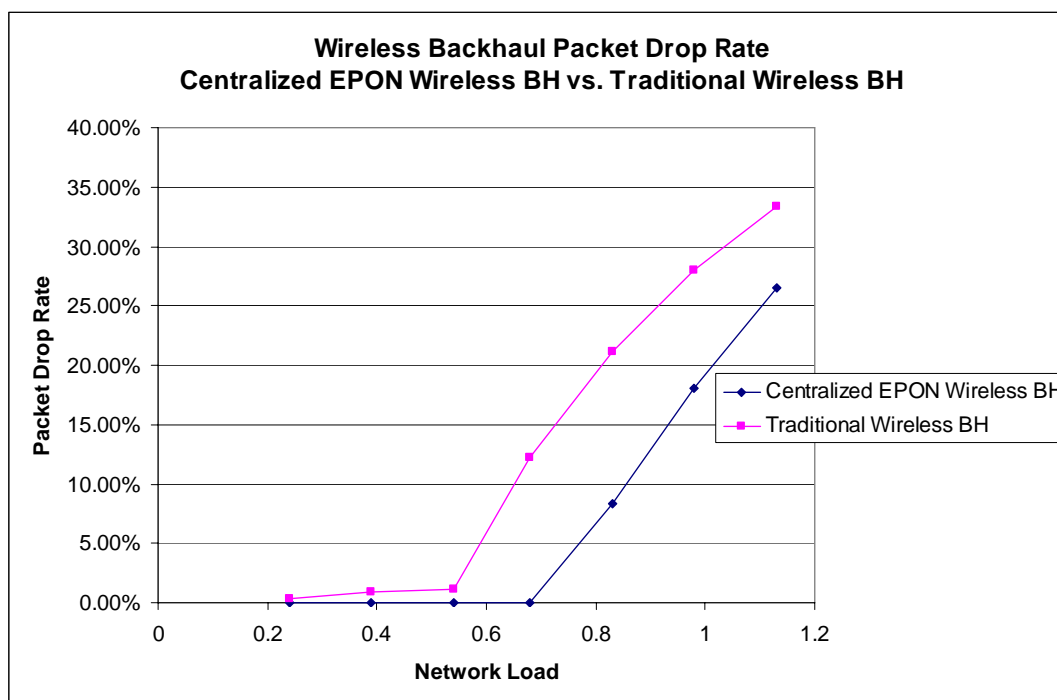


Figure 7.8: Uplink Packet Drop Rate: Centralized EPON Wireless BH vs. Traditional BH

	Audio %Drp_P0	Re-Transmit %Drp_P1	Video %Drp_P2	Best-Effort %Drp_P3	%TotalPktsDropped
Centralized EPON Wireless BH	0.00%	0.00%	0.00%	23.62%	8.31%
Traditional Wireless BH	0.00%	0.17%	13.78%	46.86%	21.18%

Table 7.2 Packet Drop per Class of Service at Network Load = 0.83

7.3.1.2 Uplink: Decentralized EPON Wireless BH

Figure 7.9 shows uplink utilization under various network loads comparing Decentralized EPON wireless backhaul versus both Centralized EPON Wireless BH and Traditional Wireless BH. We observed that through all the loading conditions the Decentralized EPON wireless backhaul showed much higher utilization than Traditional Wireless Backhaul. As for Decentralized EPON Wireless Backhaul versus Centralized EPON Backhaul, the Decentralized EPON Wireless Backhaul showed higher utilization than the Centralized on all network loads above 0.64, and for network loads less than or equal 0.64 it showed almost identical network utilization as the Centralized EPON Wireless Backhaul.

Figure 7.10 shows the uplink network utilization statistics at a single network load of 0.83, the figure illustrates the comparison between Decentralized EPON Wireless Backhaul versus both Centralized Wireless Backhaul and Traditional Wireless Backhaul. From the figure we can deduce that at network load 0.83, Decentralized EPON Wireless Backhaul had the highest average utilization than both Centralized EPON and Traditional Wireless Backhaul, also it recorded lower standard deviation than traditional backhaul, and identical standard deviation as the centralized backhaul. Figure 7.11 shows the uplink network utilization time series comparison at the same single network load of 0.83, from this figure we can observe the network utilization stability for the Decentralized and Centralized EPON Wireless Backhaul architecture versus the high variance and fluctuation in utilization for the traditional backhaul, also, it shows that through all instances the decentralized recorded higher utilization than centralized architecture.

Figure 7.12 shows the uplink network utilization statistics at a single network load of 1.13, it shows the comparison between Decentralized EPON Wireless Backhaul versus both Centralized Wireless Backhaul and Traditional Wireless Backhaul. From the figure we can deduce that at network load 1.13, the Decentralized EPON Wireless Backhaul had the highest average utilization than both Centralized EPON and Traditional Wireless Backhaul, also it recorded lower standard deviation than traditional backhaul and identical standard deviation as the centralized backhaul. Figure 7.13 shows the uplink network utilization time series comparison at a single network load of 1.13, from this figure we can observe the network utilization stability for the Decentralized and Centralized EPON Wireless Backhaul architecture versus the high variance and fluctuation in utilization for the traditional backhaul, also, it shows that through all instances the decentralized recorded higher utilization and less fluctuation than centralized architecture.

Decentralized EPON Wireless Backhaul lead to a better uplink network utilization for the following reasons

- Bandwidth allocation computation is performed after receiving and processing all BSs requests (processing period) (i.e., the computation takes into account the entire network status), the bandwidth allocation process now reflects the entire network information collectively, leading to a globally optimized decision.
- In contrast to the centralized architectures where the order of BSs transmission is fixed in each cycle (sequential), the decentralized architecture has the added flexibility of varying the order of BSs transmission according to the BSs traffic

- demands and priority. Thus, the order of BSs transmission may be different in each cycle and need not be fixed.
- Since the DBA computation is based on the global network information, the heavily loaded BSs are allocated the remaining excessive bandwidth that is not utilized by the lightly loaded BSs.
 - Given that DBA and priority queuing scheduling tasks are both executed at the BS, the DBA module integrates both scheduling information to yield a globally optimized bandwidth allocation to a particular class of service in a given BS

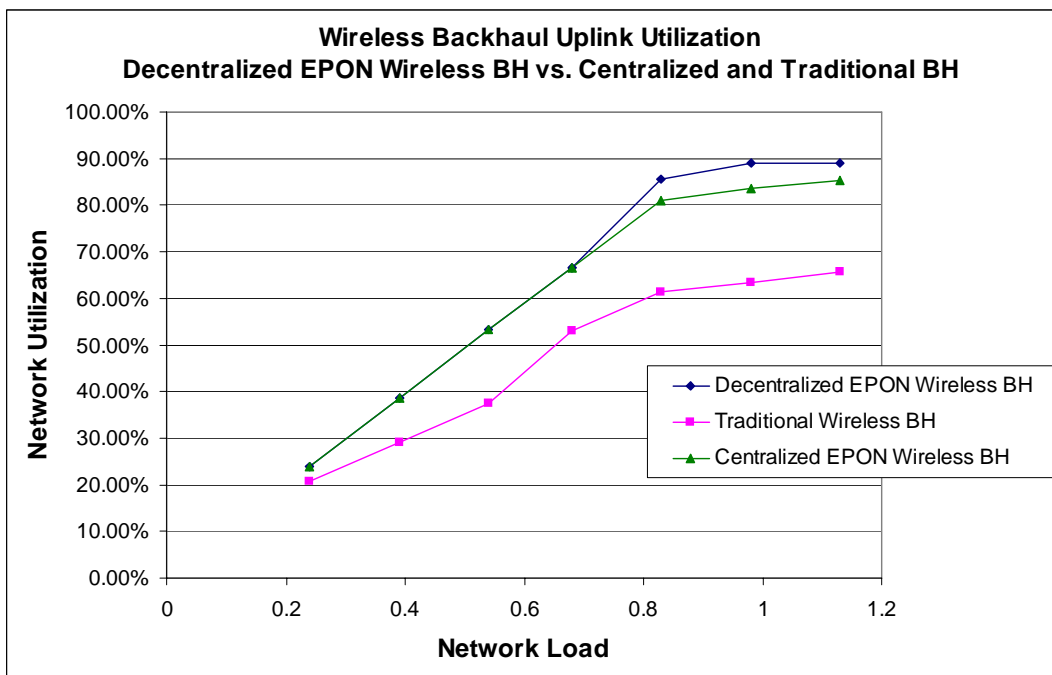


Figure 7.9: Uplink Utilization: Decentralized EPON Wireless BH vs. Centralized EPON and Traditional BH

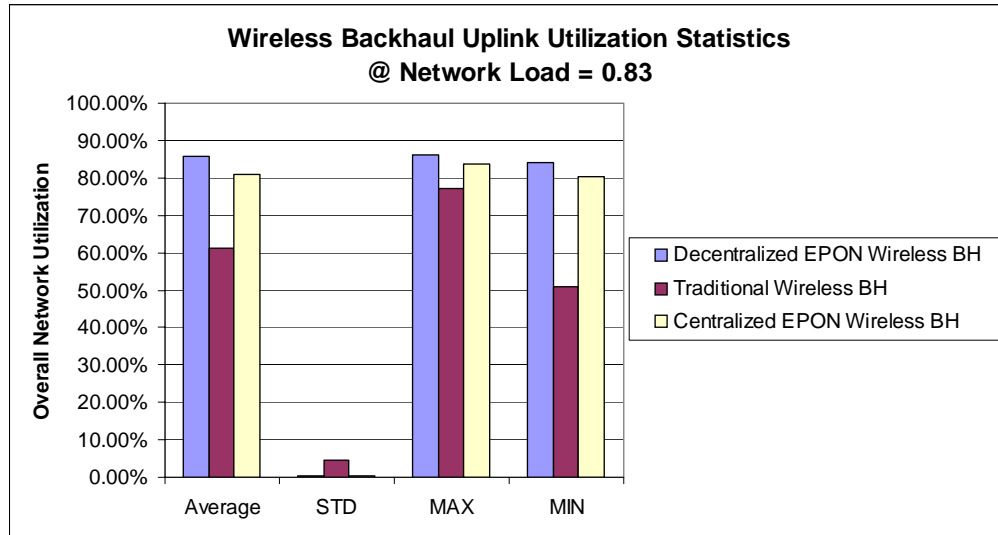


Figure 7.10: Uplink Utilization Statistics at Network Load=0.83 (Decentralized EPON BH vs. Centralized EPON and Traditional BH)

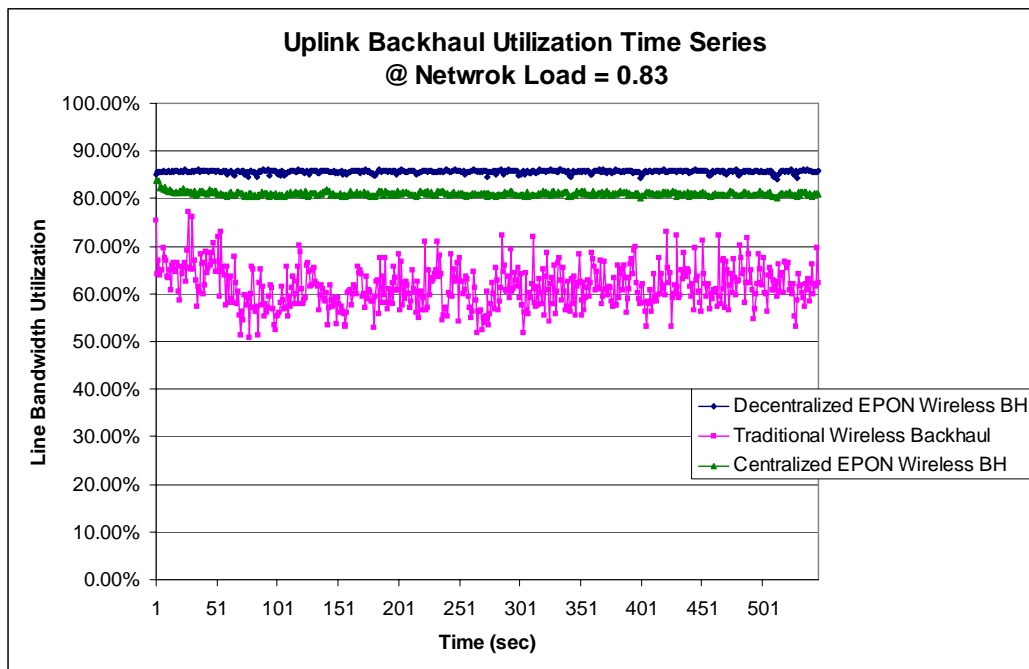


Figure 7.11: Uplink Utilization Time Series at Network Load=0.83 (Decentralized EPON BH vs. Centralized EPON and Traditional BH)

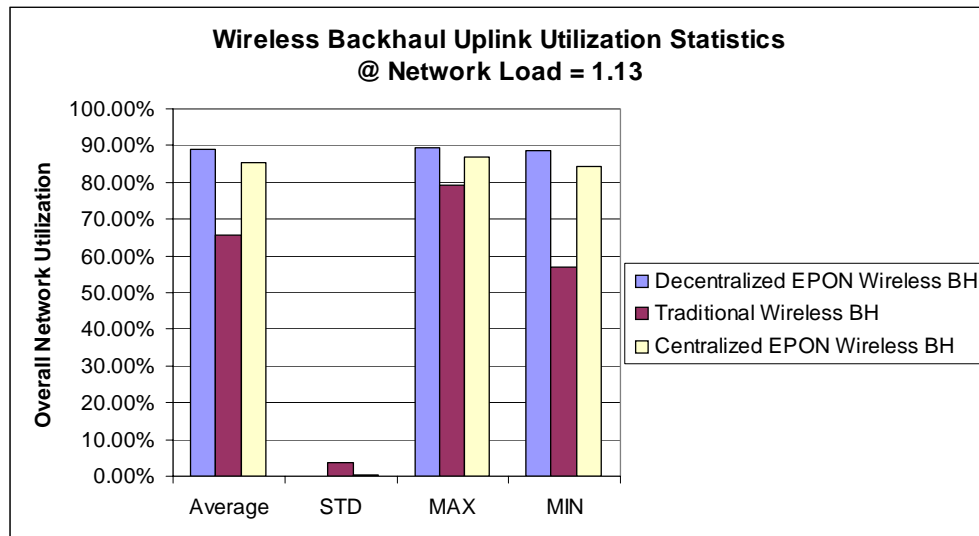


Figure 7.12: Uplink Utilization Statistics at Network Load=1.13 (Decentralized EPON BH vs. Centralized EPON and Traditional BH)

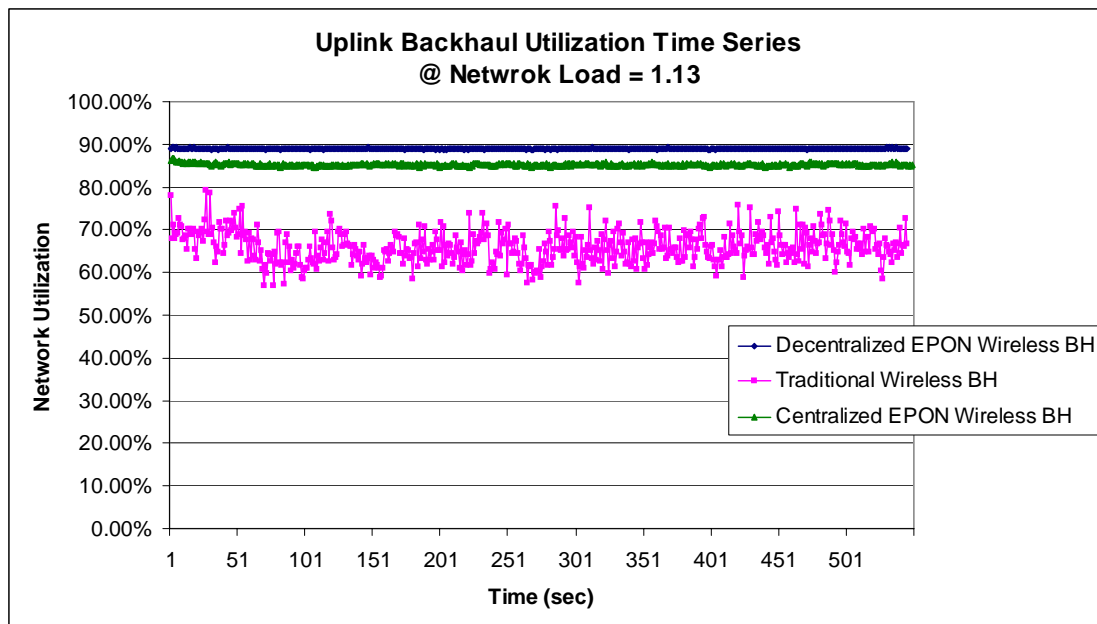


Figure 7.13: Uplink Utilization Time Series at Network Load=1.13 (Decentralized EPON BH vs. Centralized EPON and Traditional BH)

7.3.2 Downlink Traffic Numerical Results

7.3.2.1 Downlink: Centralized EPON Wireless BH

As mentioned in section 6.5.2.2, for the downlink, the Centralized EPON Wireless Backhaul is using the proposed remainder DBA.

Figure 7.14 shows downlink utilization under various network loads comparing Centralized EPON Wireless BH versus Traditional Wireless BH, the Centralized EPON Wireless Backhaul showed much higher downlink utilization than the traditional backhaul. The downlink utilization for Centralized EPON Wireless Backhaul at network loads of 0.68 and higher recorded higher network utilization than that of the uplink utilization shown in Figure 7.4. This is due to applying the proposed enhanced centralized DBA (remainder DBA), in high network loads the benefit of the algorithm will help achieve better packet drop rate and eventually helped achieve higher utilization. Another reason contributed to the utilization difference for uplink versus downlink, in the uplink there is Guard Time needed to account for the laser on and off times in between BS/ONU transmission, whereas for the downlink there is no Guard Time needed.

Figure 7.15 shows the downlink utilization statistics at a single network load of 0.83, and Figure 7.16 shows the downlink utilization statistics at a single network load of 0.98. From both figures it is obvious that the Centralized EPON wireless backhaul proved to have higher average utilization and much lower standard deviation. Figure 7.17 shows comparison for the utilization time series at network load 0.83, and Figure 7.18 shows

comparison for the utilization time series at network load 0.98, both figures prove the network utilization stability for the Centralized EPON wireless backhaul architecture versus the fluctuation and high variance in utilization observed in traditional backhaul.

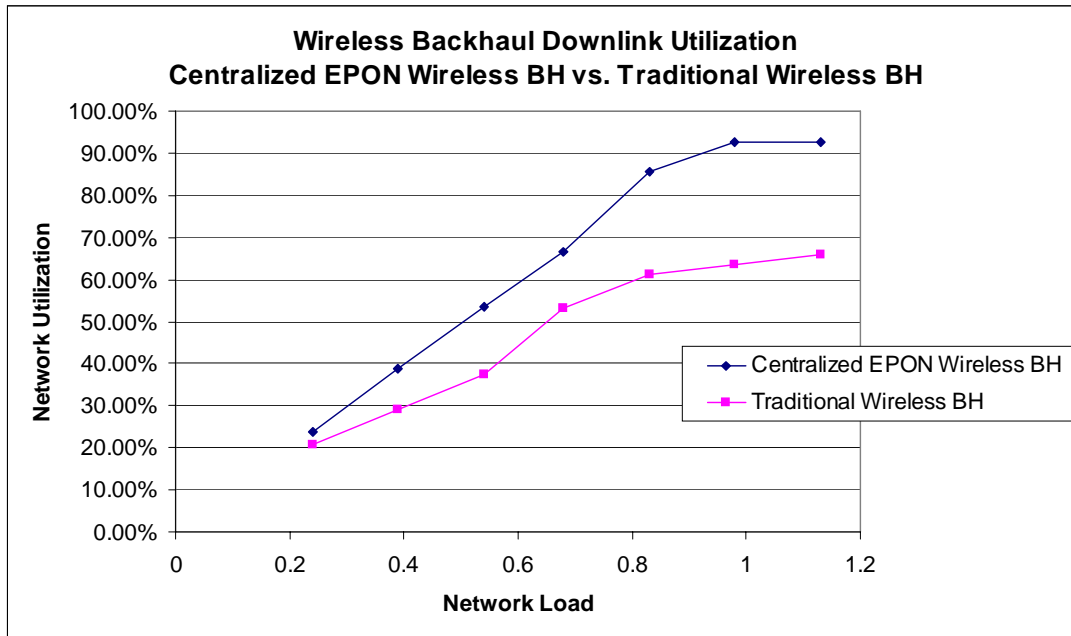


Figure 7.14: Downlink Utilization: Centralized EPON Wireless BH vs. Traditional BH

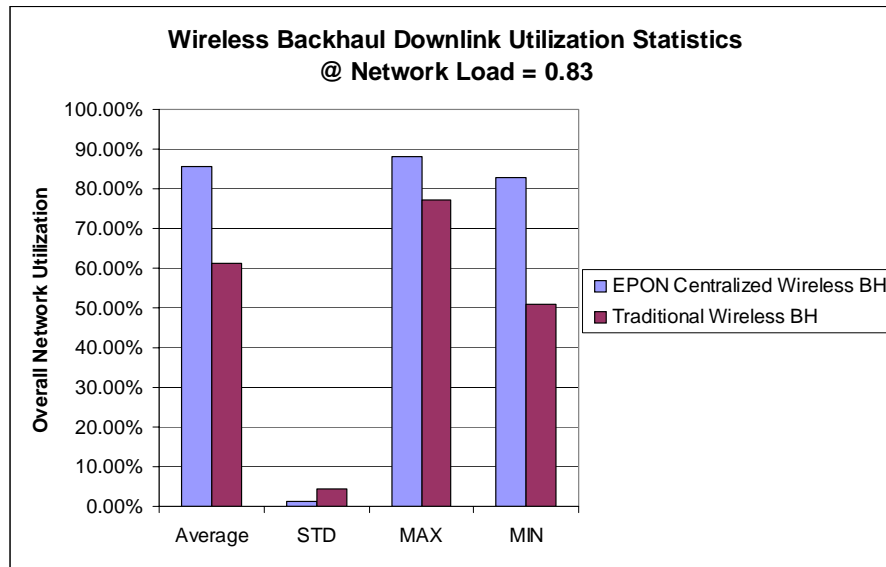


Figure 7.15: Downlink Utilization statistics at Network Load=0.83 (Centralized EPON BH vs. Traditional BH)

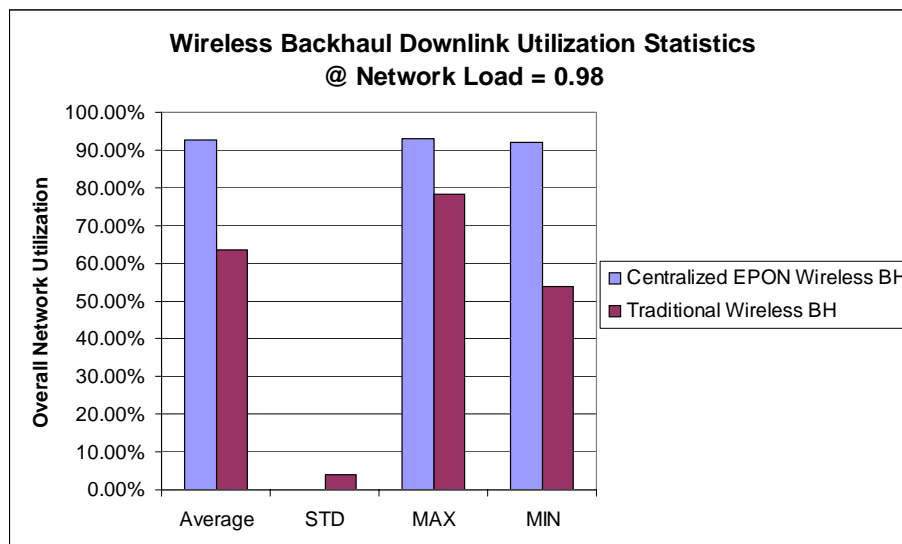


Figure 7.16: Downlink Utilization statistics at Network Load=0.98 (Centralized EPON BH vs. Traditional BH)

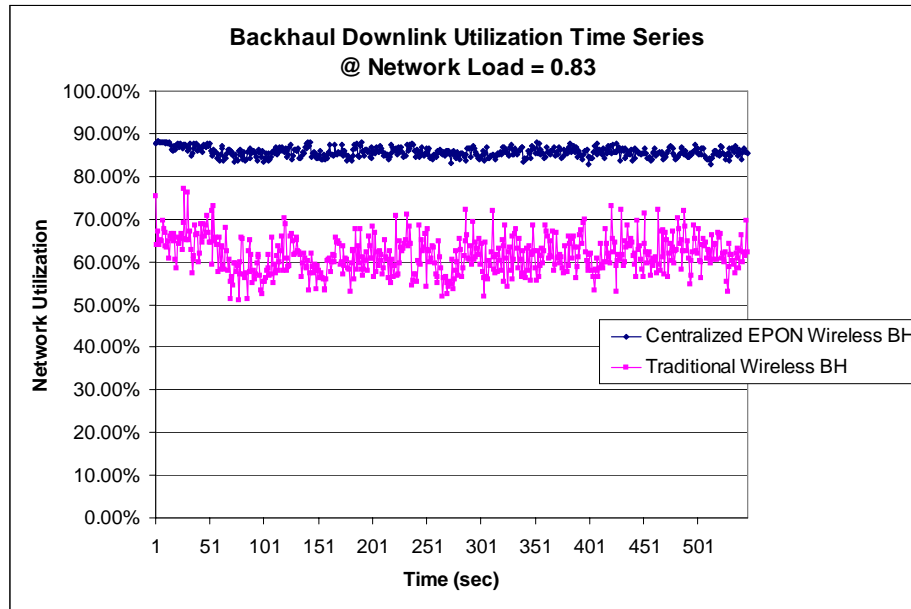


Figure 7.17: Downlink Utilization Time Series at Network Load=0.83 (Centralized EPON BH vs. Traditional BH)

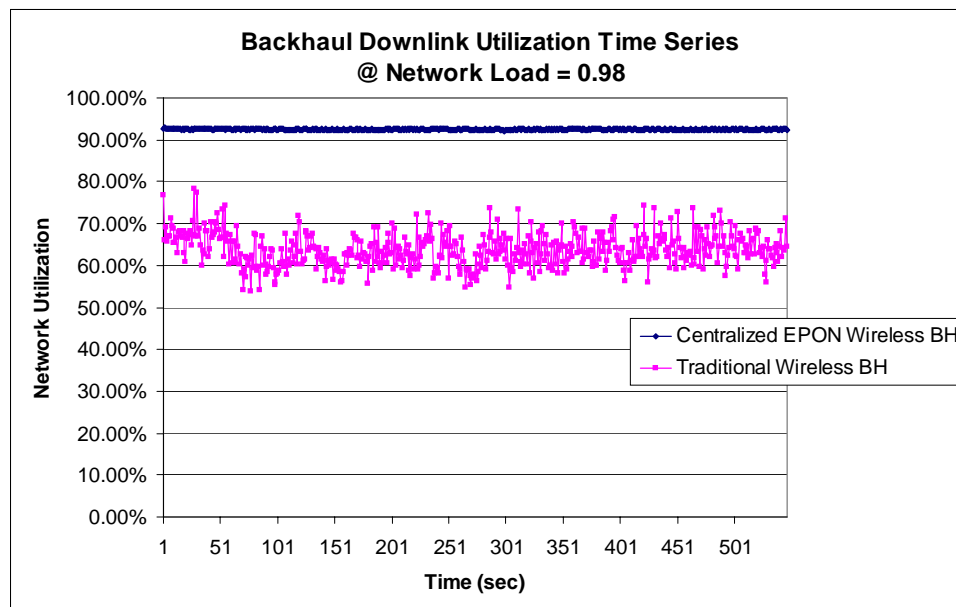


Figure 7.18: Downlink Utilization Time Series at Network Load=0.98 (Centralized EPON BH vs. Traditional BH)

7.3.2.2 Downlink: Decentralized EPON Wireless BH

Figure 7.19 shows downlink utilization under various network loads comparing Decentralized EPON wireless backhaul versus both Centralized EPON Wireless BH and Traditional Wireless BH, it shows almost identical downlink utilization for Decentralized EPON Wireless Backhaul compared to Centralized Wireless Backhaul, the decentralized recorded slightly better downlink utilization under network load of 0.83 and above. Whereas by comparing decentralized EPON wireless backhaul versus the traditional backhaul, the decentralized showed much higher downlink utilization than the traditional backhaul.

Figure 7.20 shows the downlink utilization statistics at a single network load of 0.98, comparing Decentralized EPON wireless backhaul versus both Centralized EPON Wireless BH and Traditional Wireless BH, the decentralized EPON wireless backhaul had an average utilization of 93%, which proved to be slightly higher than centralized wireless backhaul which had an average utilization of 92.5%. On the other hand both proposed architectures had significantly higher average utilization than the traditional wireless backhaul which recorded only 63.5%. Identical standard deviation was observed for both decentralized and centralized EPON wireless backhaul architectures 0.10%, as for the traditional backhaul it recorded higher utilization standard deviation of 4.09%.

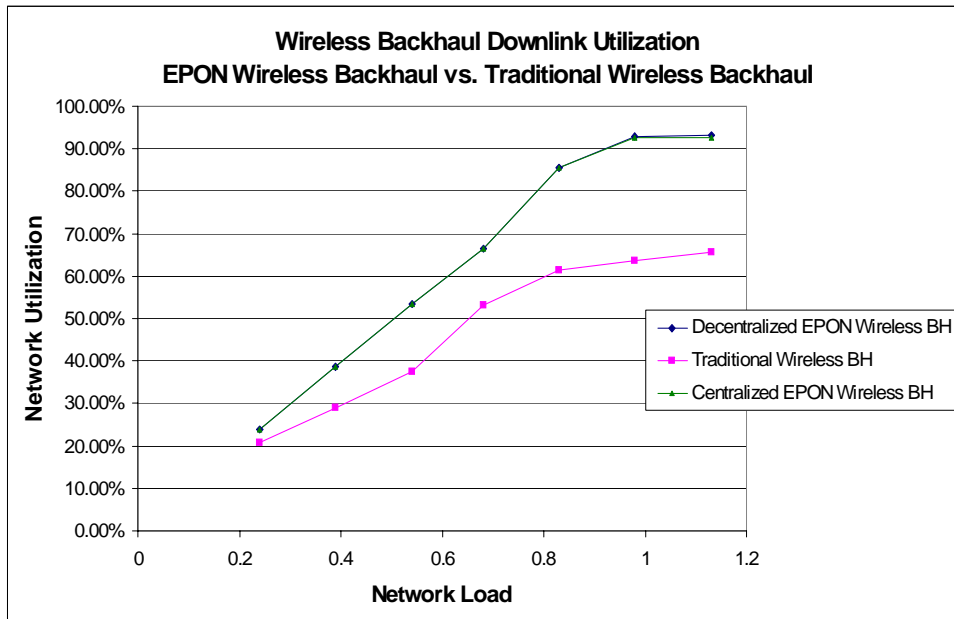


Figure 7.19: Downlink Utilization: Decentralized EPON Wireless BH vs. Centralized EPON BH and Traditional BH

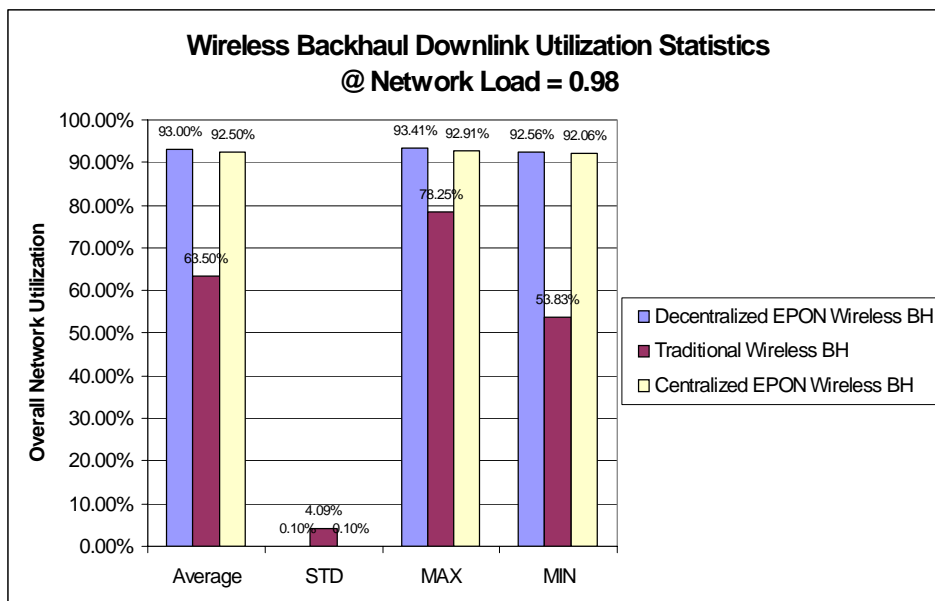


Figure 7.20: Downlink Utilization Statistics at Network Load=0.98 (Decentralized EPON BH vs. Centralized EPON and Traditional BH)

Chapter 8

CONCLUSION

8 Chapter 8: Conclusion

This thesis proposed and devised cost-effective multiservice Ethernet-based mobile backhaul networking architecture with a fiber based access infrastructure. The proposed architecture extended the reach of Ethernet access to base stations and provided a scalable mobile backhaul infrastructure that can accommodate emerging 3G/4G and legacy traffic and services as well as subscriber growth. The architecture utilized the already existing EPON access infrastructure

This work showed that the integration of both inter-ONU/BS scheduling (upstream bandwidth allocation) and intra-ONU/BS scheduling (queue management and priority queuing at the ONU/BS) mechanisms facilitated the support of differentiated QoS across the proposed RAN. In contrast to traditional centralized EPON architectures, where the OLT arbitrated upstream transmissions by allocating an appropriate timeslot to each ONU, the integration process required each ONU/BS to independently perform the process of upstream bandwidth allocation. It also required direct intercommunication among ONUs/BSs. This in turn required a fully distributed control plane among the ONUs/BSs for ONU/BS-ONU/BS communication as well as upstream access to the OLT/RNC.

The processes of moving the functionality of the packet scheduler and the upstream DBA module from the RNC to the BSs as well as achieving direct intercommunication among BSs are in full compliance with 3G standards and the vision of 4G. This also led to efficient radio resource scheduling, optimum upstream bandwidth allocation, and faster

re-transmission. Another important advantage is that some of the burden had been off loaded from the RNC to the BSs. This means that the conventional centralized EPON infrastructure where the process of upstream bandwidth provisioning is located at the distant OLT/RNC, and where there is no inter-ONU/BS communications at all, is not adequate for supporting the proposed RAN architecture. This work introduced a novel fully distributed EPON architecture, which formed the corner stone for the proposed RAN architecture, with as minor changes as possible to the conventional centralized architecture, while maintaining the IEEE 802.3ah traditional EPON standards.

In the first phase, we proposed and devised a novel Ethernet-based PON architecture that used a fully distributed time division multiple access arbitration scheme. The fully distributed control plane achieved intercommunication among the ONUs/BSs was devised and presented. Supported by the distributed control plane, we developed several decentralized QoS-based DBA algorithms that dynamically allocated upstream network resources among ONUs/BSs fairly and efficiently, in which the OLT was excluded from bandwidth arbitration process.

The proposed distributed EPON architecture supported differentiated services through the integration of both scheduling mechanisms at the ONU/BS (intra-ONU/BS scheduling). The introduction of this integration provided better QoS guarantees. It was shown that these distributed networking architectures and the associated bandwidth allocation algorithms and protocols had characteristics that made them far better suited for provisioning Quality of Service (QoS) schemes necessary for properly handling data, voice, video, and advanced mobile data services over a single line.

In the second phase, based on the developed decentralized EPON infrastructure, several powerful and cost-effective packet-based fixed/mobile access architectural models were presented. The performance as well as the pros and cons of each model were compared with each other and with that of traditional RAN model. The DBA and packet scheduling algorithms developed in the first phase were used to assess the overall performance of the Ethernet-based mobile backhaul infrastructure in both upstream and downstream directions.

9 References

- [1] Full Services Access Networks, <http://www.fsanet.net/>.
- [2] IEEE 802.3 Ethernet in the First Mile Study Group,
<http://www.ieee802.org/3/efm/public/index.html>
- [3] B. Lung, "PON Architecture 'Future proofs' FTTH," *Lightwave Magazine*, vol. 16, no. 10, Sept. 1999, pp. 104-7
- [4] D. Sala and A. Gummalla, "PON Functional Requirements: Services and Performance", Presented at IEEE 802.3ah meeting in Portland OR, July 2001.
- [5] G. Kramer and G. Pesavento, "Ethernet Passive Optical Network (EPON): Building a Next Generation Optical Access Network," *IEEE Com. Mag.*, pp. 66-73, Feb. 2002.
- [6] Alloptic, "Ethernet Passive Optical Networks," The International engineering Consortium, <http://www.iec.org> .
- [7] G. Kramer, et al. "Ethernet PON: Design and Analysis of an Optical Access Network," *Photonic Network Communications Journal*, vol. 3, No. 3, July 2001.
- [8] G. Kramer, B. Mukherjee, and G. Pesavento, "IPACT: A dynamic Protocol for an Ethernet PON (EPON)," *IEEE Comm. Mag.*, pp. 74-80, Feb. 2002.

- [9] G. Cramer, B. Mukherjee and A. Maislos, "Ethernet Passive Optical Network (EPON): A Missing Link in an End-to-End Optical Internet" in S. Dixit "Multiprotocol over WDM: Building the Next Generation Internet" John Wiley and Sons, March 2003.
- [10] G. Cramer, et al., "On Supporting Differentiated Classes of Service in EPON-Based access network" Journal of Optical Networks, 2002.
- [11] C. Assi, Y. Ye, S. Dixit, and M. A. Ali, "Dynamic Bandwidth Allocation for Quality of Service Over Ethernet PONs", IEEE Journal of Selected Areas in Communications, Dec. 2003
- [12] Virtual Bridged Local Area Networks, IEEE Std. 802. Iq., 1998.
- [13] A. Hadjiantonis, S. Sherif, A. Khalil, T. Rahman, G. Ellinas, M. Arend, M. Ali, and A. Shami, "A Novel Decentralized Ethernet-Based PON Architecture", the proceedings of IEEE ICC2004, Paris, France, 20-24 June 2004, paper ON06-8.
- [14] Chae Chang-Joon, Elaine Wong, Rodney S. Tucher, "Optical CSMA/CD Media Access Scheme for Ethernet over Passive Optical Network," IEEE Photonics Technology Letters, vol. 14, no.5, May 2002.
- [15] D. Nikolova, B. Van Houdt, and C. Blondia, "Dynamic bandwidth allocation algorithm in EPON: A simulation study," Proceedings Opticomm'03, Dallas, TX, October 2003.
- [16] M. Ma, Y. Zhu, and T. H. Cheng, "A bandwidth guaranteed polling MAC protocol for Ethernet passive optical networks," in Proc. IEEE INFOCOM'03, San Francisco, CA, Mar.–Apr. 2003, pp. 22–31.

- [17] K. Rege et al., “QoS management in trunk-and-branch switched Ethernet networks,” *IEEE Communication Magazine*, vol. 40, pp. 30–36, Dec. 2002.
- [18] Vern Paxson and Sally Floyd, “Wide Area Traffic: The Failure of Poisson Modeling”, *IEEE/ACM Transactions on Networking*, vol. 3, no. 3, June 1995, pp. 226-244.
- [19] W. Willinger, M. S. Taqqu, and A. Erramilli, “A bibliographical guide to self-similar traffic and performance modeling for modern high-speed networks,” in *Stochastic Networks*. Oxford, U.K.: Oxford Univ. Press, 1996, pp. 339–366.
- [20] 3GPP2 C.S002-A, “Physical Layer Standard for cdma2000 Spread Spectrum Systems”
- [21] 3GPP2 C.S0024 Version 2.0, “cdma2000 High Rate Packet Data Air Interface Specification”
- [22] TIA/EIA/cdma2000, *Mobile Station - Base Station Compatibility Standard for Dual-Mode Wideband Spread Spectrum Cellular Systems*, Washington: Telecommunication Industry Association, 1999.
- [23] 3G Partnership Project, “Release 99”.
- [24] Haitao Lin, and Sajal K. Das, “Performance Study of Link Layer and MAC Layer Protocols to Support TCP in 3G CDMA Systems” *IEEE Transactions on Mobile Computing*, vol. 4, no. 5, September/October 2005
- [25] *Data Service Options for Spread Spectrum Systems: cdma2000 High Speed Packet Data Service Option 34*, TIA/EIA/IS-707-A-1.11 (PN-4541.11), 2004.
- [26] TIA-707.10-B,”*Data Service Options for Spread Spectrum System: Radio Link Protocol Type 3*”, July, 2004

- [27] Farooq Khan, Sarath Kumar, Kamesh Medepalli and Sanjiv Nanda, "TCP Performance over cdma2000 RLP" Vehicular Technology Conference Proceedings, May 2000, Page(s): 41 - 45 vol.1

10 Publications

- [1] S. Sherif, A. Hadjiantonis, G. Ellinas, C. Assi and M. Ali, “A Novel Distributed Ethernet-Based PON Access Architecture for Provisioning Differentiated QoS”, Proc. IEEE/ISO Journal of Lightwave Technology, Volume 22, Issue 11, Nov. 2004, Pages 2483 – 2497
- [2] A. Hadjiantonis, S. Sherif, A. Khalil, T. Rahman, G. Ellinas, M. Arend, M. Ali, and A. Shami, “A Novel Decentralized Ethernet-Based PON Architecture”, the proceedings of IEEE ICC2004, Paris, France, 20-24 June 2004, paper ON06-8