Next-Generation Self Healing Broadband Access PON Architectures for Supporting Triple Play Services & Private Networking Capability

by

A S M Delowar Hossain

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

2007

UMI Number: 3283591

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

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 A. 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 Antoniades, Engr. Dept., College of Staten Island, CUNY

Dr. Andrew Wallace, AT &T

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

Next-Generation Self Healing Broadband Access PON Architectures for Supporting Triple Play Services & Private Networking Capability

by

A S M Delowar Hossain

Adviser: Professor M.A. Ali

This thesis proposes and devises a simple and cost effective packet-based broadband access solution for next-generation access networks that is scalable, reliable, and capable of delivering to end-users emerging triple play services over a single converged access network. Specifically, this work examines the technological requirements and assesses the performance analysis and feasibility for implementing a novel self healing ring-based local access PON architecture that addresses the limitations of current tree-based PON architectures including supporting private networking capability as well as providing a simple and cost-effective fully distributed resilience capabilities against any and all types of networking failures.

The main characteristic of the proposed architecture is that it supports a fully distributed control plane among ONUs for ONU-ONU communication (private networking capability). The control plane supports fully distributed fault detection and recovery mechanisms as well as a decentralized DBA scheme in which the OLT is excluded from both arbitration and fault detection/recovery processes.

We show that the proposed distributed solutions can overcome the problems associated with current centralized ones, and in the process prove that these distributed networking architectures and the associated fault detection and recovery schemes as well as bandwidth allocation algorithms and protocols have characteristics that make them far better suited for supporting cost effective resilience capabilities and for provisioning QoS schemes necessary for properly handling data, voice, video, and other real-time streaming advanced multimedia services over a single line. The objective of the proposed distributed solutions is to enhance the Ethernet’s capabilities to ensure that real-time voice and IP video services can be delivered reliably over a single platform with the same QoS and ease of management as ATM or SONET.

Supported by the distributed architecture, this work develops several decentralized QoS-based DBA algorithms in which the OLT is excluded from the arbitration process. The proposed distributed architecture guarantees the delivery of delay and jitter-sensitive real-time services through the integration of both scheduling mechanisms at the ONU (inter and intra-ONU scheduling). The introduction of this integration feature can only be supported by a distributed architecture.

The proposed decentralized fault and protection schemes are capable of protecting against both node and distribution/trunk fiber failures. These schemes enable the restoration of all network traffic including upstream, downstream, and LAN data. In addition, these schemes can also protect against any combination of concurrent double failures including trunk/distribution fiber breaks and node failures. Furthermore, the recovery time associated with any and all different distribution network/trunk failures is still within the delay-bound limit required for delivering guaranteed triple play services.

Acknowledgements

My deepest gratitude goes to my adviser and committee chair Prof. M.A. Ali. His extensive knowledge, experience and exceptional ability to find new approaches for difficult problems were pivotal in this work. I am deeply indebted to Prof. Ali for his extensive discussions and exhaustive editing of my publications and thesis. This work would not have been completed without his encouragement and patience.

I am also grateful to Prof. Samir Ahmed, Prof. Roger Dorsinville, Prof. Neophytos Antoniades and Dr. Andrew Wallace for serving on my supervisory committee. Their insightful recommendations have improved the quality of this work.

I would like to thank my friends Dr. Abdullah Shami, Dr. Chadi Assi and Hasan Erkan for their insightful comments in many issues in the research.

Finally, and most importantly, I am grateful to my parents for cultivating in me the passion for learning. I am also grateful to my wife and children for their continuous patience and support.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

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 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" or "Access Networks". The explosive growth of Internet traffic has further underscored the limitations of access network capacity.

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 advancement 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 (CM) 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.

While DSL and/or CMs seem to offer a satisfactory solution for current bandwidth demands, however, as the number of end users attached to a given access network grows each year, and as their usage patterns evolve to include more bandwidth intensive applications such as video and image-based services, there emerges an acute need for a next-generation packet-based access solution, that is the focus of this thesis; one that is simple, cost effective, scalable, survivable, and capable of providing each end-user with at least a 50-100 Mb/s of bundled voice, data, and video services over a single converged access network.

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 [1-10]. 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.

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 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). In the downstream direction, a PON operates as a broadcast and select network. In the upstream direction, multiple ONUs share the transmission channel. Thus, ONUs need to employ some arbitration mechanism to avoid collisions. In general, OLT arbitrates upstream transmissions via a Dynamic Bandwidth Allocation (DBA) module that performs the bandwidth allocation computation.

PON-based FTTH solutions have been classified into four different architectures based on layer 2 data link protocol technology being used. These include Broadband PON (BPON), Gigabit PON (GPON), ATM-PON, and Ethernet PON (EPON). 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 LAN, MAN, and backbone. Since that time, ATM (and APONs) has lost favor.

Since that time, Ethernet has emerged as the frontrunner technology for transporting data, video, and voice services over a single platform. 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 [6-10]. EPON is one of the access solutions considered by the new IEEE 802.3ah Ethernet in the First Mile (EFM) Task Force [6], focusing on direct support of Ethernet services.

1.2 Thesis Motivation

Current PON-based FTTH architectures including BPON, GPON, and EPON enjoy low installation and maintenance cost and are being deployed in the field in several places around the world. These different PON architectures are traditionally deployed as tree topologies, which provide added flexibility and simplicity in the deployment of access networks. However, tree-based topologies have several inherent limitations including:

1. To date, mainstream PON DBA schemes have been centralized – relying on a component at the distant OLT to arbitrate upstream transmission. In addition to the conventional "single-point of failure" problem (failure of OLT software will bring down the whole access network), centralized DBA schemes can't support triple play services efficiently and the process of bandwidth allocation is suboptimal.
2. Lack of a simple and cost-effective protection and/or restoration capabilities against failures in the distribution network. ITU-T G.983.1 recommended four possible protection schemes, which duplicate fibers and equipment at the ONUs and OLT. These schemes are all centralized and can significantly alter the cost-

effectiveness of the PONs since they require many redundant components as well as many spare fibers connections to each ONU. Furthermore, the centralized processes of detecting and restoring distribution fiber breaks at the distant OLT are lengthy and complex processes and require many changes at each ONU.

3. Inability to support a truly shared Local Area Network (LAN) capability among end users.
4. Furthermore, the above limitations lead to inefficient utilization of network resources (both upstream and downstream).

These limitations combined have placed several technical hurdles to the evolution of current centralized PON-based FTTH architectures to an envisioned cost effective, scalable and reliable next-generation broadband access network, including:

1. Scalability problem: as the number of users grows, the OLT must support more than just a single PON; this means adding more hardware and software complexity at the OLT. As more and more users are added up, the OLT will no longer be able to scale with the added complexity.
2. Quality of Service (QoS) problem: Currently Ethernet provides best effort traffic delivery. 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 and scheduling. In addition, centralized DBA schemes can’t guarantee the delivery of delay and jitter-sensitive real-time services that have traditionally not been the focus of Ethernet.

1.3 Thesis Statement

It is the purpose of this thesis to propose and devise a simple and cost effective packet-based broadband access solution for next-generation access networks that is scalable, reliable, and capable of delivering to end-users emerging triple play services over a single converged access network. Specifically, this work examines the technological requirements and assesses the performance analysis and feasibility for implementing a novel self healing ring-based local access PON architecture that addresses the limitations of current tree-based PON architectures including supporting private networking capability as well as providing a simple and cost-effective fully distributed resilience capabilities against any and all types of networking failures.

The fact that well over 95% of all data traffic either originates or terminates as Ethernet has prompted us to select native Ethernet as the potential convergence solution for the proposed access network. Because the length structure of the Ethernet frame is variable (up to 1518 bytes), Ethernet is tailor-made for transporting IP traffic. With IP firmly entrenched as the dominant internetworking protocol, Ethernet is gaining new grounds in the access, MAN, and WAN. With its simplicity, scalability, ubiquity, and natural support for IP services, Ethernet provides a compelling case.

The main characteristic of the proposed architecture is that it supports a fully distributed control plane among ONUs for ONU-ONU communication. The control plane supports fully distributed fault detection and recovery mechanisms as well as a decentralized DBA scheme in which the OLT is excluded from both arbitration and fault detection/recovery processes.

We show that the proposed distributed solutions can overcome the problems associated with current centralized ones, and in the process prove that these distributed networking architectures and the associated fault detection and recovery schemes as well as bandwidth allocation algorithms and protocols have characteristics that make them far better suited for supporting cost effective resilience capabilities and for provisioning QoS schemes necessary for properly handling data, voice, video, and other real-time streaming advanced multimedia services over a single line. The objective of the proposed distributed solutions is to enhance the Ethernet's capabilities to ensure that real-time voice and IP video services can be delivered reliably over a single platform with the same QoS and ease of management as ATM or SONET.

Supported by the distributed architecture, this work develops several decentralized QoS-based DBA algorithms in which the OLT is excluded from the arbitration process. In contrast to centralized architectures where the order of ONUs transmission is fixed in each cycle, the distributed architecture has the added flexibility of varying the order of ONUs transmission according to ONUs traffic demands and priority. Most importantly, the proposed distributed architecture guarantees the delivery of delay and jitter-sensitive real-time services through the integration of both scheduling mechanisms at the ONU (inter and intra-ONU scheduling). The introduction of this integration feature can only be supported by a distributed architecture.

The proposed decentralized fault and protection schemes are capable of protecting against both node and distribution/trunk fiber failures. These schemes enable the restoration of all network traffic including upstream, downstream, and LAN data. In addition, these schemes can also protect against any combination of concurrent double failures including trunk/distribution fiber breaks and node failures. Furthermore, the recovery time associated with any and all different distribution network/trunk failures is still within the delay-bound limit required for delivering guaranteed triple play services.

In addition to the added flexibility and reliability of a distributed scheme, the proposed architecture eliminates the OLT's centralized task of failure detection and subsequent recovery scenarios as well as the task of processing requests and generating grants (GATE messages) for bandwidth allocations. This reduces the additional processing complexities and delays at the OLT as well as maximizes downstream bandwidth utilization.

1.4 Organization of the Thesis

In this thesis a number of new contributions are proposed that could eradicate the existing shortcomings of PON and in turn make broadband access network more efficient. The thesis outline is as follows:

Chapter 2 discusses the current access solutions and their shortcomings. It also discusses the promising solution of fiber optical technology, leading to the discussion of PON and eventually EPON. An overview of EPON is also presented.

Chapter 3 proposes the ring-based distributed EPON architecture with LAN capability. The details of distributed architecture along with its operational principles are described. A distributed DBA scheme is also proposed, where ONUs exchange control information among them to perform DBA (OLT is relieved from that burden). The distributed DBA globally optimizes decision making process. The current LAN emulation problems and increased end-to-end delay of LAN traffic are also addressed. In addition to the added flexibility and reliability of a distributed scheme, this architecture demonstrates the following several advantages over a typical star-based centralized PON architecture: (i) eliminates the typical utilization of an upstream burst-mode receiver and associated design challenges at the OLT; (ii) increases the available downstream and upstream channel bandwidth; (iii) improves LAN and WAN traffic performance significantly; (iv) supports more efficient upstream channel utilization.

Chapter 4 extends the globally optimized distributed DBA of the proposed architecture to support QoS. Several decentralized QoS-based DBA algorithms are proposed in which the OLT is excluded from the arbitration process. In contrast to centralized architectures where the order of ONUs transmission is fixed in each cycle, the proposed distributed scheme has the added flexibility of varying the order of ONUs transmission according to ONUs traffic demands and priority. Most importantly, the proposed scheme guarantees the delivery of delay and jitter-sensitive real-time services through the integration of both scheduling mechanisms at the ONU (inter and intra-ONU scheduling).

Chapter 5 analyzes the downstream bandwidth utilization problem of centralized architecture due to the LAN emulation and GRANT messages. A solution is proposed which demonstrates notable improvement in downstream performance.

Chapter 6 discusses the importance of survivability of ring-based PON architecture. A distributed failure detection scheme is proposed for the ring-based distributed EPON architecture. An APS protection scheme is introduced, whereby in the event of a failure in an ONU, trunk fiber, or distribution fiber, its detection and subsequent protection switching is efficiently initiated by the affected ONU. Significant performance improvement over the centralized counterpart is demonstrated.

Chapter 7 gives concluding remarks and future challenges in this work.

Chapter 2

Overview of Access Solutions

Due to the importance of the access network, the networking community has renamed the access segment of the network from “last mile” to the “first mile”. The first mile connects the service provider central offices to businesses and residential subscribers. Also referred to as the subscriber access network, or the local loop, it is the network infrastructure at the neighborhood level (Fig. 2.1).

Since 1990, the data traffic is growing enormously. In certain years (1995/96), this growth peaked at 1000% [1]. Data traffic has already surpassed the voice traffic. This trend is likely to continue in the future. More and more subscribers require the same network performance as they see on corporate LANs. In a nutshell, subscribers demand access solutions that are broadband, offer Internet media-rich services, and are reasonably priced.

To face the end users demands, service providers came up with Digital Subscriber Line (DSL) technology and cable service. DSL uses the same twisted pair as telephone lines and requires a DSL modem at the customer premises and Digital Subscriber Line Access

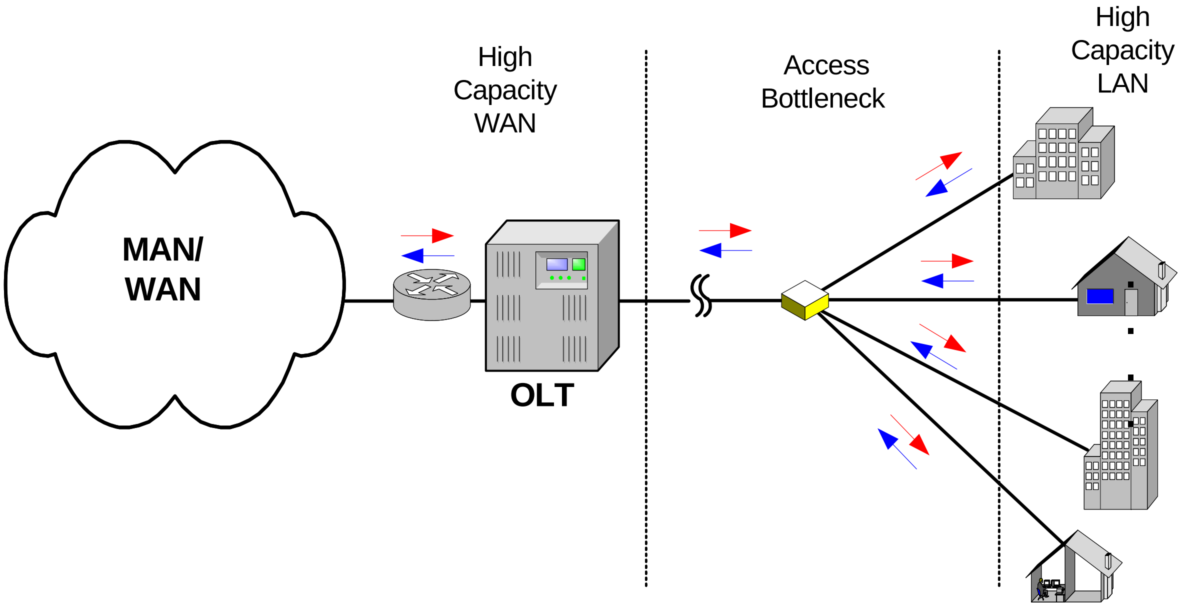


Figure. 2.1 : Access Network

Multiplexer (DSLAM) in the central office (CO). The data rate provided by DSL is typically offered in a range from 128 kbps to 1.5 Mbps. Even though it is an improvement, it cannot support emerging Internet media-rich services. In addition, the physical area that one CO can cover with DSL is limited to distances less than 5.5 km, which covers approximately 60% of potential subscribers. Constrained by signal distortion and crosstalk, the service providers generally do not provide DSL services to subscribers located more than 3.5 km from a CO [2]. It is fair to say that DSL copper networks do not allow sufficient data rates at required distances due to elevated expense caused by technical limitations.

As far as the cable service is concern, it was originally designed for analog video broadcast. The cable television companies responded to Internet service demand by integrating data services over their coaxial cable networks. Typically, these Hybrid Fiber

Coax (HFC) networks have fiber running between a video head-end or a hub to a curbside optical node, with the final drop to the subscriber being coaxial cable, repeaters, and tap couplers. The drawback of this architecture is that only a few radio frequency (RF) channels are dedicated for data, while the majority of bandwidth is tied up servicing legacy analog video. For instance, each shared optical node has less than 36 Mbps effective data throughput, which is typically divided among 2000 homes, resulting in frustrating slow speed during peak hours.

Therefore, it is apparent that neither DSL nor cable modems can keep up with such unprecedented bandwidth demand. Both technologies are built on top of existing communication infrastructure not optimized for data traffic. Network communities have come to realize that a new, data-centric broadband solution is necessary to alleviate access bottleneck.

2.1 Current Access Solutions

To face the challenge of access bottleneck, optical fibers are penetrating deeper into the first mile. 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. Few variations of FTTH are (FTTx): fiber-to the-building (FTTB), fiber-to-the-PC (FTTPC), and fiber to-the-curb (FTTC). Unlike previous architectures, where fiber is used as a feeder to shorten the lengths of copper and coaxial networks, these new deployments use optical fiber throughout the access network, reaching all the way to customer premises.

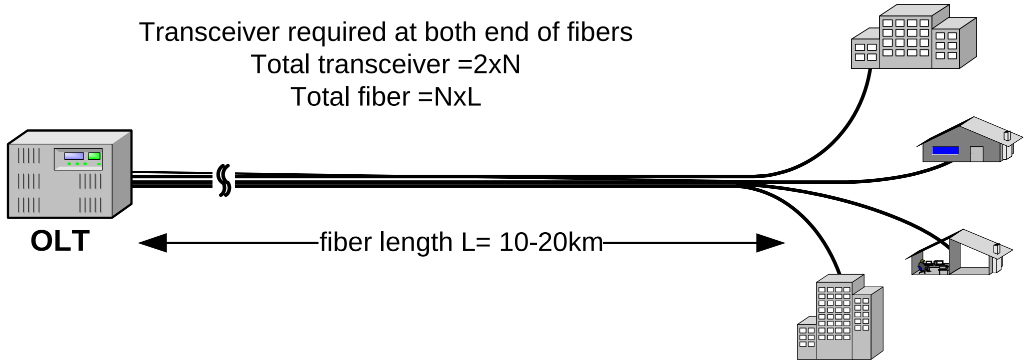
These emerging architectures are capable of supporting gigabit per second (Gbps) speeds, at costs comparable to those of current solutions.

Table 2.1: Comparison of FTTx architectures

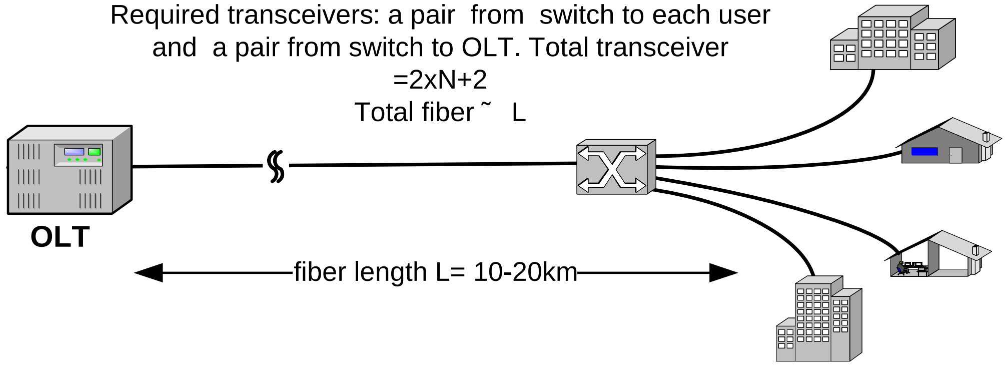
Architecture	Advantage	Disadvantages
Point-to-Point (PtP)- Fig. 2.2 (a)	dedicated fiber to each node gives unlimited bandwidth	significant outside fiber deployment (NxL), CO needs termination connectors for each nodes (N), requires increased number of transceivers (pair of transceivers for each line (N), total 2xN)
Curb-Switched Fig. 2.2 (b)	minimized fiber deployment	requires bandwidth sharing among the nodes, requires electrical power as well as back-up power at the curb switch, increased number of transceivers (pair of transceivers from switch to N users and a pair from switch to OLT, total 2xN+2) in system
Point-to-Multipoint (PtMP) Fig. 2.2 (c)	simple, minimized fiber deployment, least number of transceiver in system (a transceiver for each N user and 1 for OLT, total N+1), maintenance free passive splitter	requires bandwidth sharing among the nodes

Optical access network, namely FTTx, could be build upon (i) powerful and expensive Active Optical Network (AON) or (ii) simple and cost effective Passive Optical Network (PON). Most of today's optical networks are AON. The biggest hurdle towards the realization of the infinite bandwidth of fiber is the required conversions from electrical-to-optical (E/O) and optical-to-electrical (O/E). Telecommunication companies have built large high capacity AON. These high capacity networks have complex and

(a) PtP



(b) Switched



(c) PtMP - PON

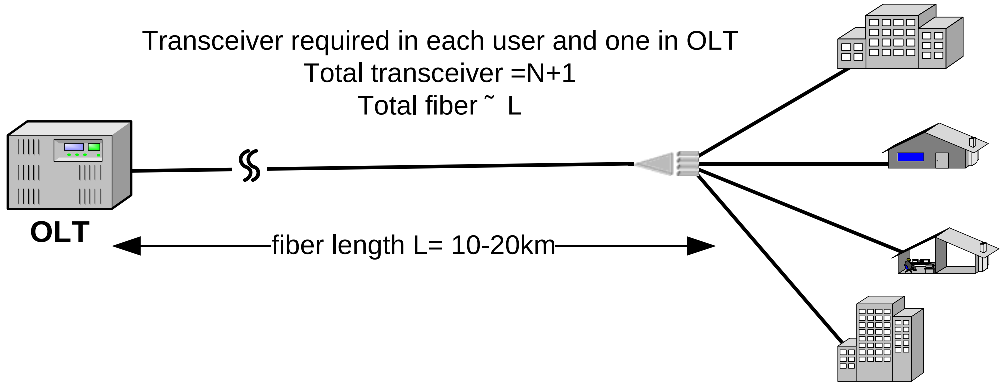


Fig. 2.2 FTTx deployment scenarios

expensive active devices such as Add-Drop Multiplexers (ADMs), Digital Cross Connects (DCS), Regenerators (repeaters), etc. These networks bring unlimited bandwidth to support large number of users over a vast area. The complexity and cost of AON is feasible in MAN, where numerous users share the expense. Compared to MANs, subscriber access networks serve a relatively small number of users within short distance, requiring much less bandwidth. Therefore, access network is very cost sensitive. Due to the mentioned technical complexity and higher cost of AON, current FTTH systems are not commercially viable; thus have prevented the deployment. FTTH design and implementation must reflect simplicity and cost effectiveness. Now the other option is the PON. To lower the cost and expedite the implementation of FTTH, PON-based solutions have drawn attention of research community and industry [3-5]. PONs are point-to-multipoint fiber optical networks with no active elements in the signal's path. Table 2.1 depicts a comparative analysis of the few FTTH design options (see Fig. 2.2). Note that PtMP architecture (Fig.2.2(c)) is a PON based architecture, which eliminated all active components from the distribution network. It brings reasonable amount of shared bandwidth to the each user's door step in a simple and cost effective manner.

2.2 Passive Optical Network (PON)

The idea of Passive Optical Network (PON) was first conceived nearly 20 years ago [6–8] as a response to the problem of limited bandwidth available in the local loop between a customer and the local exchange. At the time, the quality of optical fibers was improving and efficient transmitters and receivers appeared and it seemed possible to build an access network that would be based on optical technology. Due to the absence of any active elements in the light path such as amplifiers, signal regenerators, the architecture of the system was simple and offered bandwidth far beyond other access methods. However, the initial progress in the development of optical networks was slowed to a halt by economic and technological factors. The Internet was not as widespread then as it is now and customers were reluctant to pay higher prices for their broadband access. At the time it was foreseen that narrow band ISDN as well as DSL access would be sufficient for a while. Furthermore, the cost of replacing existing copper infrastructure with optical cables was not justified from an economic point of view. But as the bandwidth hungry internet became too cumbersome for existing access technology as well as the fiber optical technology became sophisticated and economically feasible, the PON technology drew new interest as a solution to the access bottleneck.

PONs are point-to-multipoint (PtMP) optical networks with no active elements in the signals' path from source to destination. The major network elements used in such networks are passive combiners, couplers, and splitters. A PtMP PON for the access network can be deployed in few configurations (see Fig. 2.3): tree-and branch, ring, or bus [9]. It will require 1×2 optical tap couplers and 1×N optical splitters. The tree branch architecture is widely used as the reference model [5, 9-11]. In this architecture, OLT resides in the local exchange (central office), connecting the optical access network to a backbone. The ONU is located at the end-user location. Generally, it consists of a single shared optical fiber/trunk connecting a service provider's central office (head end) to a passive optical star coupler/splitter (SC), which is located near residential customers (ONUs). The SC is intentionally positioned a substantial distance away from the central office (CO), but close enough to the ONUs in order to save fiber. Each ONU receives a dedicated short optical fiber but shares the long distribution trunk fiber. All transmissions in a PON are performed between Optical Line Terminal (OLT) and Optical Network Units (ONU). Therefore, in the downstream direction (from OLT to ONUs), a

PON is a point-to-multipoint network, and in the upstream direction it is a multipoint-to-point network.

PON based PtMP architecture seems to be the most suitable solution to access bottleneck. Some of the advantages of PON over existing access solutions are as follows:

(i) PON supports longer distance (~20km) between central office and customer premise. While with the DSL the maximum distance between the central office and the customer is only ~5.5 km.

(ii) PON minimizes fiber deployment in both the local exchange and the local loop (Fig. 2.2).

(iii) PON provides higher bandwidth due to deeper fiber penetration.

(iv) PON is a point-to-multipoint network; a PON supports downstream video broadcasting.

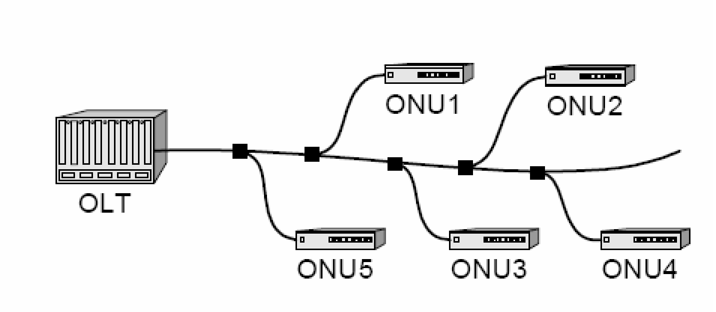
(v) PON eliminates the necessity of installing multiplexers and de-multiplexers in the splitting locations, thus relieving network operators from the gruesome task of maintaining them and providing power to them. Instead of active devices in these locations, a PON utilizes passive components that can be buried into the ground at the time of deployment.

(vi) PON allows easy upgrades to higher bit rates or additional wavelengths.

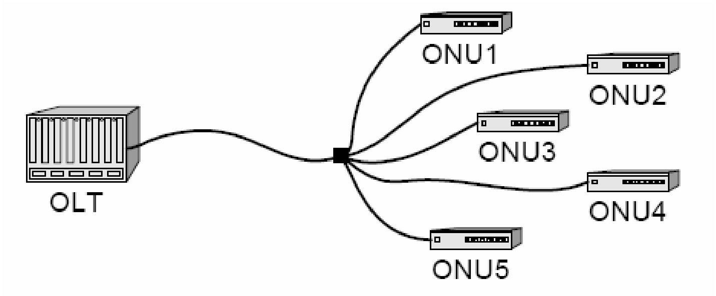
2.3 Available PON Technologies

There are few types of PONs, such as Broadband PON (BPON)[12], Ethernet PON (EPON)[5], Gigabit PON (GPON)[13], Asynchronous Transfer Mode (ATM)-based PON (APON)[14-16]. PON can utilize any of the two physical layer technologies, Wavelength Division Multiplexing (WDM) or Time Division Multiplexing (TDM). BPON uses WDM technology, where as EPON, GPON, and APON use TDM technology.

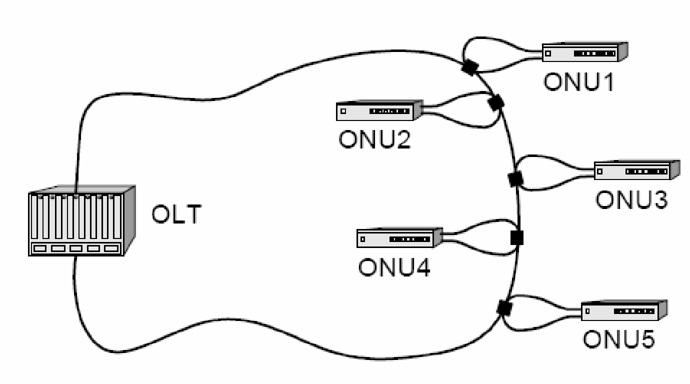
In the downstream direction (from OLT to ONUs), a PON is a point-to-multipoint network. The OLT typically has the entire downstream bandwidth available to it at all times. In the upstream direction (from ONUs to OLT), a PON is a multipoint-to-point network: multiple ONUs transmit towards one OLT. For all ONUs to be able to communicate with OLT, either each ONU must have a dedicated wavelength (WDM) or they all must share a single wavelength through a time sharing mechanism (TDM). While WDM is a very good solution, offering unlimited bandwidth and simplicity, but it remains cost-prohibitive for an access network. A WDM solution would require either a tunable receiver, or a receiver array at the OLT to receive multiple channels. An even more serious problem for network operators would be wavelength-specific ONU inventory: instead of having just one type of ONU, there would be multiple types of ONUs based on their laser wavelength. Using tunable lasers in ONUs may solve the inventory problem, but is too expensive at the current state of technology. For these reasons, a WDM PON network is not an attractive solution in today’s environment.



(a) Bus



(b) Tree



(c) Ring

Fig 2.3: PON deployment scenarios

In a TDM PON, all ONUs are identical (transmitter/ receiver) and OLT needs only one receiver. But it requires certain mechanism to time share the single wavelength among the ONUs. Any simple channel sharing mechanism will work and even allow statistical multiplexing for efficient use of bandwidth. TDM appears to be the preferred method today for optical channel sharing in an access network as it allows for a single upstream wavelength, and a single transceiver in the OLT, resulting in a cost-effective solution. Single fiber TDM based PONs are of few kinds. But, mainly they are classified into two networking architectures [according to layer 2/Medium Access Control (MAC) layer specifications], namely, ATM-based PON (APON) and Ethernet-based PON (EPON) [1-2].

2.3.1 APON vs. EPON

APONs were standardized and developed around 1995 through the work of the Full Services Access Network (FSAN) initiative [16]. 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. In APON, data is transmitted in fixed-length 53 byte cells (with 48 byte payload and 5 byte overhead). The Internet protocol calls for data to be segmented into variable-length packets of up to 65,535 bytes. For an ATM PON to carry IP traffic the packets must be broken into 48 byte segments with a 5-byte header attached to each one. This process is time consuming and complicated and adds additional cost to the OLT and ONUs. Moreover, 5 bytes of bandwidth are wasted for every 48-byte segment creating a burdensome overhead that is commonly referred to as the ATM cell tax. One of ATM's major shortcomings is the fact that a dropped or corrupted ATM cell will invalidate an entire IP datagram. However, the remaining cells carrying the portions of the same IP datagram will propagate further, thus consuming network resources unnecessarily. And finally, perhaps most importantly, ATM did not live up to its promise of becoming an inexpensive technology.

EPON is based on Ethernet. Ethernet currently accounts for more than 85 percent of all installed connections. Since it is simple and inexpensive, it has become a universally accepted standard, with over 320 million ports deployed worldwide [17]. The length structure of the Ethernet frame is variable (up to 1518B). Ethernet is tailor-made for transporting IP traffic and dramatically reduces the overhead incurred by ATM. Unlike ATM, Ethernet is an inexpensive technology, which is scaleable. With IP firmly entrenched as the dominant internetworking protocol, Ethernet is gaining new grounds in the access, MAN, and wide area network (WAN).

EPON is much efficient than the high cell tax APON. Furthermore, considering the fact that 95% of LANs use Ethernet, it becomes clear that ATM PON may not be the best choice to interconnect two Ethernet networks. As far as the economical feasibility is concern, ATM switches and network cards are significantly more expensive than Ethernet switches and network cards [17].

From the functional prospective, the key advantage of EPON over APON is that it eliminates complex and expensive ATM and Synchronous Optical Network (SONET) elements to simplify their networks dramatically. Traditional telecom networks use a complex, multilayered architecture, which overlays IP over ATM, SONET and WDM. This architecture requires a router network to carry IP traffic, ATM switches to create virtual circuits, add/drop multiplexers (ADM and digital cross-connects (DCS) to manage SONET rings, and point-to-point DWDM optical links. There are a number of limitations inherent to this architecture:

(i) It is difficult to provision because each network element (NE) in an ATM path must be provisioned for each service.

(ii) It is optimized for TDM voice, not data; so its fixed bandwidth channels have difficulty handling bursty data traffic.

(iii) It requires inefficient and expensive optical-to-electrical-to-optical (O-E-O) conversion at each network node.

(iv) It requires installation of all nodes up front (because each node is a regenerator).

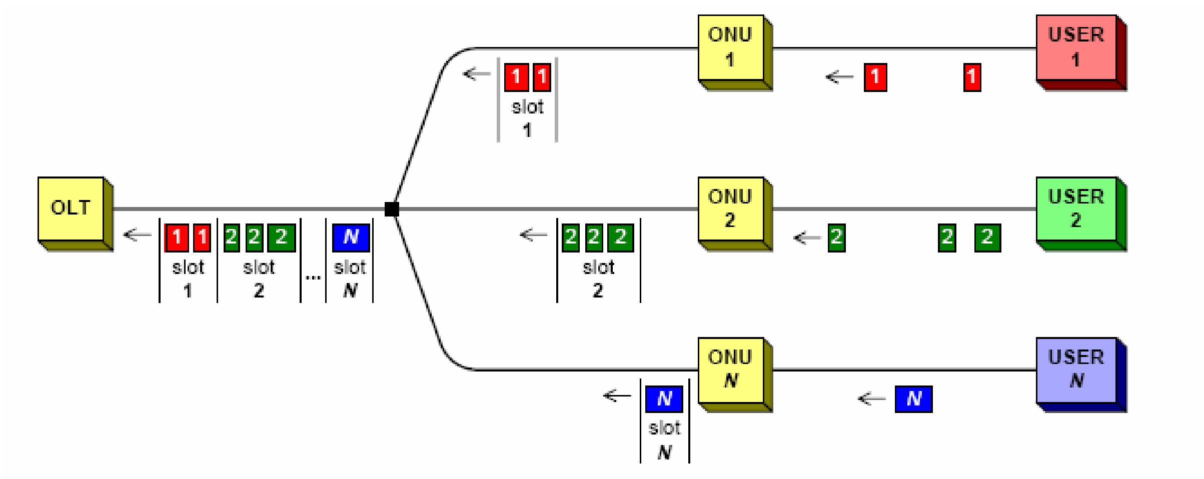
(v) It does not scale well because of its connection-oriented virtual circuits (VC).

Ethernet looks like the logical choice for an IP data-optimized access network. To substitute the ATM's costly and complex quality-of-service (QoS) techniques, the newly-adopted QoS mechanisms have made Ethernet networks capable of supporting voice, data, and video. These techniques include full-duplex transmission mode, prioritization (P802.1p), and virtual LAN (VLAN) tagging (P802.1Q). Ethernet Passive Optical Network (EPON), which represents the convergence of low-cost Ethernet equipment, firmly grounded technological advantages of Ethernet (suitable for access network) and low-cost fiber infrastructure, appears to be the best candidate for the next-generation access network.

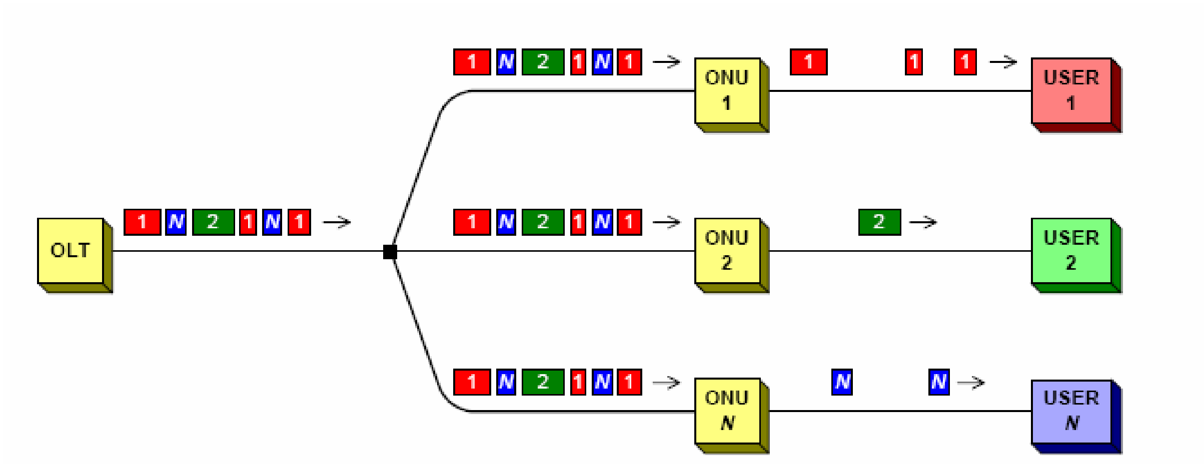
2.3.2 EPON Overview

EPON is a single fiber PtMP PON that uses TDM technology to carry data traffic encapsulated in Ethernet frames as defined in the IEEE 802.3 standard [18]. Generally, it consists of a single, shared optical fiber connecting a service provider's CO (head end) to a passive SC, which is located near residential customers. The SC is intentionally positioned a substantial distance away from the 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 ONUs using typical Ethernet frames. 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). Two wavelengths are used: typically 1310 nm (λ_{up}) for the upstream transmission and 1490 nm (λ_d) for the downstream transmission [3].

In the downstream direction (point-to-multipoint), an EPON operates as a broadcast and select network (Fig. 2.4b). Ethernet packets transmitted by the OLT pass through a 1×N passive splitter and reach each ONU. The value of N is typically between 4 and 64 (limited by available power budget). This behavior is similar to a shared-medium network. Because Ethernet is broadcasting by nature, in the downstream direction (from OLT to user), it fits perfectly with the EPON architecture; packets are broadcasted by the OLT and extracted by their destination ONU based on the MAC address.



(a) Upstream operation



(b) Downstream operation

Figure 2.4: EPON operation

In the upstream direction (multipoint-to-point), all ONUs share the wavelength (Fig. 2.4a). Thus, the ONUs need to employ some arbitration mechanism to avoid collisions. Note that in most cases the OLT has the overall network information due to centralized intelligence. Normally under TDMA scheme, the OLT arbitrates upstream transmissions by allocating an appropriate timeslot to each ONU. Only one ONU is allowed to transmit during a designated time slot. The start time and the duration of such a time slot is calculated via a *Dynamic Bandwidth Allocation* (DBA) module, residing at the OLT. To facilitate the implementation of bandwidth allocation schemes, the OLT and ONUs exchange control messages, namely, REPORT message and GATE message. These control messages are defined by the IEEE 802.3ah task force through the development of *Multi-Point Control Protocol* (MPCP) [3]. A REPORT message is sent by an ONU to OLT informing it about its bandwidth requirements. Upon receiving a REPORT, the OLT passes the message to a DBA module to perform the bandwidth allocation computation. The OLT then grants the ONU a transmission time slot by sending a GATE message indicating the start time and the duration of such a time slot. Each ONU transmits in its allocated time slot in sequence. Signals are multiplexed at the optical splitter (Nx1) towards OLT. 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 1-B code that is transmitted every 2 ms to synchronize the ONUs with the OLT [3]. 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.

2.4 Conclusion

In this chapter, we discussed that the growing internet demands over burdened the access network and existing solutions (DSL, ISDN and CM) failed to keep up. The present bottleneck could be alleviated thought the unlimited bandwidth of fiber optical technology. To face the challenge of access bottleneck, optical fibers are penetrating deeper into the first mile. FTTH is the ultimate level of access, allowing end users to access the backbone networks through the gigabit capacity of a fiber optic cable. Due to technical complexity and economical constrain of FTTH, it was not feasible as the optical solution for access. PON drew the attention of network community as the access solution, because it is simple, efficient and economical. Among the few types of PONs, EPON which represents the convergence of low-cost Ethernet equipment, firmly grounded technological advantages of Ethernet and low-cost fiber infrastructure, appears to be the best candidate for the next-generation access network.

Chapter 3

Proposed Ring-Based PON Architecture

3.1 Introduction

Ethernet-based *Passive Optical Network* (EPON) technology is emerging as a viable choice for the next-generation broadband access network [1-4]. A PON is a point-to-multipoint fiber optical network with no active elements in the signal's path. 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) using typical Ethernet frames. 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). Two wavelengths are used: typically 1310 nm (λ_{up}) for the upstream transmission and 1490 nm (λ_d) for the downstream transmission. In the downstream direction, an EPON operates as a broadcast and select network. In the upstream direction, multiple ONUs share the transmission channel. Thus, the ONUs need to employ some arbitration mechanism to avoid collisions. To date, PONs and almost all proposed EPON architectures reported in the literature are deployed in tree-and-branch configurations.

In general, the OLT arbitrates upstream transmissions by allocating an appropriate timeslot to each ONU. Only one ONU is allowed to transmit during a designated time slot. The start time and the duration of such a time slot is calculated via a *Dynamic Bandwidth Allocation* (DBA) module, residing at the OLT. To facilitate the implementation of bandwidth allocation schemes, the OLT and ONUs exchange control messages, namely, REPORT message and GATE message. These control messages are defined by the IEEE 802.3ah task force through the development of *Multi-Point Control Protocol* (MPCP) [5]. A REPORT message is sent by an ONU to OLT informing it about its bandwidth requirements. Upon receiving a REPORT, the OLT passes the message to a DBA module to perform the bandwidth allocation computation. The OLT then grants the ONU a transmission time slot by sending a GATE message indicating the start time and the duration of such a time slot.

Deploying point-to-point customer communication links to emulate a shared *Local Area Network* (LAN) within a single PON infrastructure is an important feature for providing a private networking capability, which has recently received some attention [6-8]. In general, standard upper layer shared LAN emulation techniques require the use of bridges/routers at the OLT to redirect data frames back to the ONUs [6]. These techniques are effective but reduce the available downstream transmission bandwidth and increase end-to-end delay of LAN traffic. Several physical layer LAN emulation techniques have been proposed to achieve intercommunication among ONUs within the same PON setup [7-8]. These include a physical layer solution that employs a fiber Bragg grating at the trunk feeder fiber to reflect a predefined wavelength channel for intercommunication between ONUs [7]. Another approach assumes that each ONU is connected to a *Star Coupler* (SC) via two distribution fibers, where one fiber is used to transport both downstream and upstream traffic, and the second fiber is used to deliver redirected frames back to the ONU for detection [8]. Most of these solutions suffer from high splitting loss as the redirected signal traverses through the SC once or twice, resulting in a lower power budget that limits the number of ONUs that can be attached to a single PON.

To provide a simple PON-based local access infrastructure that addresses these limitations and supports private networking capability, this work proposes a novel ring-based EPON architecture that supports a truly shared LAN capability as well as upstream access to the OLT (MAN/WAN traffic). Unlike a typical ring-based PON topology in which OLT and ONUs are interconnected via a long fiber ring, under the proposed architecture, ONUs are interconnected via a short distribution fiber ring in the local loop but share the standard trunk feeder fiber for long reach connectivity to the OLT. This minimizes fiber deployment in both the CO and the local loop. This architecture is well suited for an autonomous access environment such as a university campus or a private corporation where several buildings are closely dispersed within a 0.5-1 km diameter area. Note this distributed architecture evolved from the centralized ring-based architecture which we presented in [4].

The main characteristic of the proposed architecture is that it supports a fully distributed control plane among the ONUs for ONU-ONU communication as well as upstream access to the OLT. Specifically, it utilizes a fully distributed *Time Division Multiple Access* (TDMA) arbitration scheme in which the OLT is excluded from the arbitration process. In the proposed decentralized scheme, the ONUs exchange signaling and control information concerning their queue status and their transmission needs amongst themselves (REPORT messages). Then, the ONUs concurrently 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 including both LAN and MAN/WAN traffic without any collisions, eliminating the OLT's centralized task of processing requests and generating grants for bandwidth allocations.

In addition to the added flexibility and reliability of a distributed scheme, the proposed ring-based PON architecture introduces the following several advantages over a typical star-based centralized PON architecture: (1) supports a truly shared LAN capability within a single PON setup; (2) eliminates the typical utilization of an upstream burst-mode receiver and associated design challenges at the OLT; (3) increases the available downstream and upstream channel bandwidth; (4) supports more efficient upstream channel utilization; and (5) alleviates the typical limited power budget problem associated with physical layer LAN emulation techniques described above.

The rest of the chapter is organized as follows: Section 3.2 gives a brief overview of the ring-based centralized architecture, which leads to the decentralized scheme. Section 3.3 describes the proposed distributed architecture. Section 3.4 discusses the upstream and

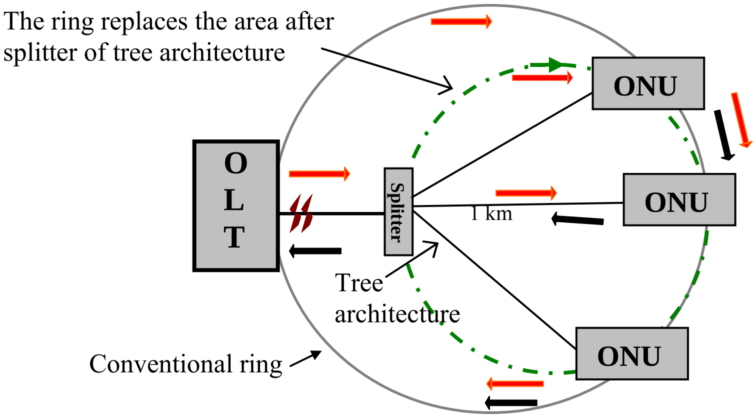
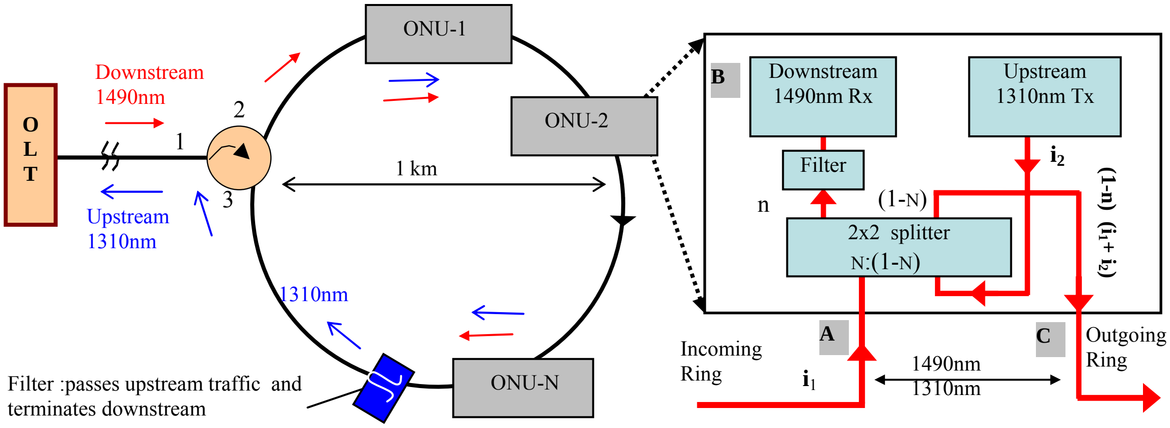


Figure 3.1: Centralized ring among other architectures

downstream power budget of the proposed architecture. Section 3.5 discusses distributed bandwidth allocation algorithm. Section 3.6 provides performance results for the proposed scheme and Section 3.7 offers some concluding remarks.

3.2 Ring-Based Centralized Architecture

The proposed distributed architecture of this chapter evolved from the ring-based centralized architecture which we presented in [4]. In this section, a brief overview of the ring-based centralized architecture is presented. Fig. 3.1 shows the centralized ring-based architecture in comparison to the typical PON architectures. Details of this architecture are shown is Fig. 3.2, where a passive 3 port circulator connects the trunk fiber of OLT to a number of ONUs on the small unidirectional ring. The small ring at the end of the trunk is of diameter ~1 km. There are N ONUs connected to this small ring. The coverage area of this ring is similar to the other architectures. The OLT is 10-20km away from the circulator. As in tree architecture, traffic from OLT to an ONU (1490 nm-λ_d) is called ‘downstream’ (point-to-multipoint) and the traffic from an ONU to the OLT (1310 nm-



(a) (b)

Figure 3.2: (a) Overview of centralized ring (b) ONU architecture

λ_{up}) is called ‘upstream’ (multipoint-to-point). Like the tree architecture, the downstream signal is broadcasted to all the ONUs. The long trunk of the OLT connects to the circulator port 1 at ring-trunk junction (Fig 3.2a). The downstream traffic from OLT enters circulator port1 and comes out of the circulator port 2 into the ring. Once the signal enters the ring, it circulates in the ring in a drop and go manner; a fraction of downstream signal is dropped at every ONU. The traffic from each ONU to OLT (upstream traffic) is transmitted by ONU in an assigned time slot by OLT. This upstream traffic travels in the ring towards the circulator to reach the OLT through the trunk fiber.

The incoming ring to an ONU (point A in Fig.3.2b) connects to ONU’s (n: 1-n) 2x2 passive splitter's input port. The 2x2 splitter (n is a small arbitrary percentage assumed here to be 10%) splits the incoming combined (upstream and downstream) signal at each node into a small (10%) “Drop signal component” and a major (90%) “Go signal component”. The small fraction of the circulating combined signal dropped at each node (Drop signal component) is passed through a filter that removes the upstream signal and passes only the downstream broadcast signal, which is then received by the 1490 nm downstream receiver (point B in Fig. 3.2b). As in tree architecture, if the downstream broadcast traffic matches that ONU's MAC address, then it is processed by that ONU; otherwise it is discarded. An ONU transmits its upstream traffic in a time slot assigned by the OLT. The upstream transmitter (1310 nm) of the ONU is connected to the other input port of the 2x2 splitter and is coupled with the incoming traffic. The combined traffic comes out of 90% output port of 2x2 splitter (Go signal component); it becomes the ONU's outgoing ring (point C in Fig. 3.2b). This outgoing ring now becomes the incoming ring for next ONU on the ring. Both signals travel in the ring in drop-and-go manner and 10% of the optical signal is dropped at every ONU until they face a filter at the end of the ring where the downstream signal is terminated from going back to OLT (Fig. 3.2a). Then only upstream traffic passes through the circulator port 3 to port 1 to join the trunk fiber towards OLT.

Table 3.1: Comparison of fiber usage (km)

D=t+s Max distance (ONU to OLT)	Tree t+N*s	Conventional Ring 2*pi*r	Novel Ring t+ 2*pi*r
10+1=11	26	33	13
15+1=16	31	50	19
20+1=21	36	65	24

Compare to the other architectures, the ring-based architecture requires less fiber deployment. Table 3.1 compares fiber usage of architectures (see Fig. 3.1). In Table 3.1, t = the trunk fiber length, s = distance of ONU from splitter in the tree architecture, D= the furthest distance between OLT and ONU as well as the diameter of the conventional

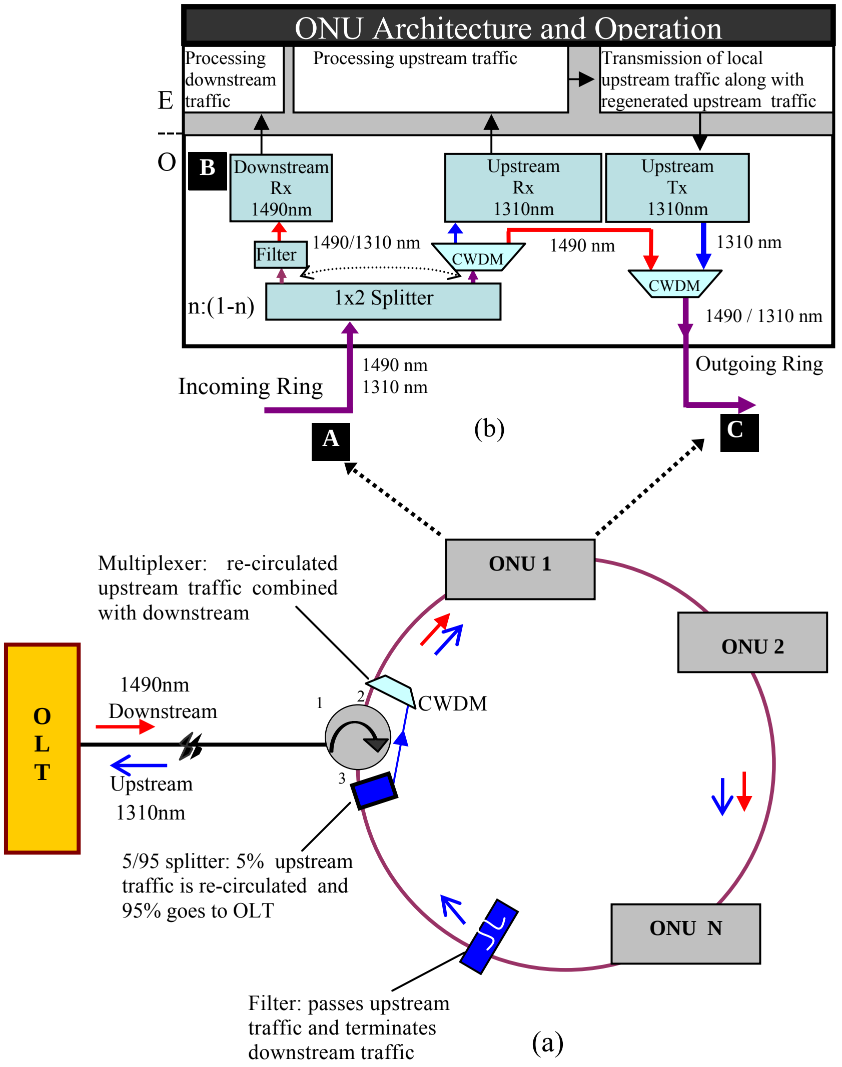


Figure 3.3: (a) Proposed ring-based architecture (b) ONU architecture

ring, r is the radius of a small ring and N is ONU count (16). Table 3.1 demonstrates that the centralized ring architecture requires the least amount of fiber among the architectures.

3.3 Proposed Distributed Architecture

Fig. 3.3a illustrates the proposed distributed ring-based PON architecture. An OLT is connected to N number of ONUs via a 20 km trunk feeder fiber, a passive 3 port optical circulator, and a short distribution fiber ring. To cover the same local access area as that covered by a conventional tree-based architecture [3-4], the small ring at the end of the trunk is assumed to have a 1 km diameter. The set of ONUs are joined by point-to-point links in a closed loop. The links are unidirectional: both downstream and upstream signals (combined signal) are transmitted in one direction only. Fig. 3.3b shows detailed ONU architecture. Each ONU attaches to the ring at a (n: 1-n) 1x2 passive star coupler (incoming signal at point A in Fig. 3.3b) and can transmit data onto the ring through the output port of a 2x1 *Coarse Wave Division Multiplexing* (CWDM) combiner (outgoing signal at point C in Fig. 3.3b). Note that in addition to the conventional transceiver maintained at each ONU (a λ_{up} upstream transmitter and a λ_d downstream receiver), this approach requires an extra receiver tuned at λ_{up}.

Downstream signal is coupled to the ring at port 2 of the optical circulator. After recombining with the re-circulated upstream signal via another 2x1 CWDM combiner (Fig. 3.3a) placed on the ring directly after the optical circulator, the combined signal then circulates around the ring (ONU 1 through ONU N) in a Drop-and-Go fashion. The downstream signal is then removed at the end of the ring using a filter (located directly after the last ONU) that passes only the 1310 nm upstream signal. The upstream signal emerging from the filter at the end of the ring is split into two components via a 1x2 passive splitter (Fig. 3.3a) placed on the ring directly after the filter. The first component is directed towards the OLT via circulator ports 1 and 3, while the second component is allowed to re-circulate around the ring after recombining with the downstream signal (originating from the OLT) via the 2x1 CWDM combiner Fig. 3.3a.

The (n: 1-n) 1x2 coupler (n is a small arbitrary percentage assumed here to be 10%) splits the incoming combined signal at each node into a small (10%) “Drop-signal-portion” and a large (90%) “Go-signal-portion”. The small portion of the circulating combined signal dropped at each node (Drop-signal) is passed through a filter that removes the upstream signal and passes only the downstream broadcast signal, which is then received and processed by the 1490 nm downstream receiver. The remaining portion of the combined signal emerging from the 90% coupler’s port (Go-signal) is first separated into its two constituent: downstream and upstream signals via a CWDM filter. The separated upstream signal (second component) is received and processed via the 1310 nm upstream optical receiver housed at the ONU, where it is then regenerated and retransmitted along with the ONU’s own local control and data traffic. Finally, the separated downstream signal is re-combined again with the retransmitted upstream signal (regenerated plus local) via the 2x1 CWDM of Fig. 3.3b to form the outgoing combined signal (incoming combined signal for next ONU) that circulates around the ring.

Since upstream transmission is based on a TDMA scheme, inter-ONU traffic (LAN data and control messages exchanged among ONUs) is transmitted along with upstream traffic destined to the OLT (MAN/WAN data) within the same pre-assigned time slot. The first component of the upstream signal destined to the OLT is received and processed by the 1310 nm upstream optical receiver (housed at the OLT), which accepts only MAN/WAN traffic, discards LAN traffic, and may discard or process (for reasons to be given below) the control messages. On the other hand, the second component of upstream signal is transmitted sequentially, bit by bit, around the ring from one node to the next where it is regenerated and retransmitted at each node.

Since the ring is a closed loop, upstream traffic will circulate indefinitely unless removed. The process of removing, regenerating and retransmitting the second component of the upstream signal at each node (ONU) is implemented as follows: first, the 1310 nm upstream optical receiver (housed at each ONU) terminates all upstream traffic, examines the destination MAC address of each detected Ethernet frame, and then performs one or more of the following functions: (1) all re-circulated upstream traffic addressed to the OLT is removed by the first ONU (ONU that is physically located on the ring directly after the 2x1 CWDM coupler of Fig. 3.3a); (2) all control messages (REPORTs) must be processed, regenerated, and then retransmitted by each node; (3) the source node removes its own transmitted inter-ONU control messages that complete one trip around the ring through re-circulation; (4) transient LAN traffic, terminated at an intermediate node, but destined to other nodes are regenerated and then retransmitted along with the node's own local upstream traffic within the designated proper time slot; (5) once the destination address of the LAN traffic matches the node's MAC address, it is copied and delivered to the end users and then discarded (not retransmitted to the next ONU). Note that only inter-ONU control traffic completes one trip around the ring (returns to the source node); the LAN traffic is removed at the destination ONU and the MAN/WAN traffic destined to the OLT is always removed before returning to the source node except for first ONU's traffic, which is removed by the source node (first ONU).

Note that any fiber or node (ONU) failure may bring down the whole network. Thus, protecting both node and fiber failures are essential for proper operation of the proposed network. For instance, to ensure failure recovery, there could be redundancy of the ONU transceiver and fiber. However, failure recovery is beyond the scope of this chapter and will be addressed in Chapter 6.

3.4 Downstream and Upstream Optical Power Budget

Table 3.2: Typical losses incurred at an ONU

Type of Loss	Point A to B	Point A to C
Connector	0.15	2x0.15
Splitter(10/90)	10.0	0.45
Filter	0.15	-
CWDM	-	0.3
TOTAL (dB)	10.3	1.05

Since the upstream signal is regenerated at every node, typical limited power budget problem (due to the splitting loss at each node) as well as receiver dynamic range problem (long/short optical network paths and different splitting factors) are totally eliminated. To examine the performance impact of the downstream power budget, we consider the best and worst-case scenarios by calculating the total *Optical Distribution Network* (ODN) loss (passive optical elements such as splitters, combiners, fibers, connectors, and splices forming an optical path) incurred by the downstream signal on its shortest and longest optical paths from the OLT to the first and last ONUs, respectively.

We assume that the trunk feeder fiber length is 20 km, the first ONU is 20 km away from the OLT, and the last ONU is 23.2 km away from the OLT (ring circumference is about 3.2 km). We consider losses of components based on recent industry standards: circulator insertion loss is 0.3dB and trunk SMF fiber loss at 1490 nm is 0.2 dB/km. There are two types of losses encountered by the downstream signal at each node. The first type is from point A to B in Fig. 3.3b (Drop-component) and the second type is from point A to C in Fig. 3.3b (Go-component). Table 3.2 quantifies both types of losses. Thus, total ODN loss incurred by downstream signal on its path to the first and the last ONU is:

$$L_{Total\_Loss}^{1st\_ONU} = L_{trunk}^{fiber} + L_{trunk}^{circulator} + L_{A-B}^{ONU} = 14.6\ dB \tag{3.1}$$

$$L_{Total\_Loss}^{last\_ONU} = L_{trunk}^{fiber} + L_{trunk}^{circulator} + (N-1)L_{A-C}^{ONU} + L_{Ring}^{fiber} + L_{A-B}^{ONU} \tag{3.2}$$

For example, if N = 16 ONUs, the total ODN loss using equation 3.2 above (worst case) is 30.9 dB, leading to a 16.3 dB of ODN differential attenuation. Assuming that the signal power transmitted by the OLT is 5 dBm, a receiver sensitivity of – 25.9 dBm will be required at the last ONU. Thus, a downstream receiver with a –25.9 dBm minimum sensitivity combined with a 16.3 dB dynamic range is required to support 16 ONUs. Note that downstream power budget demands on receiver parameters are still within practical and commercial reach.

Note also that the 16 nodes limit is not a shortcoming of an architecture that is specifically devised to support a private ring-based local access infrastructure within a 1 km diameter area. A typical large private organization would have at most 10-15 buildings within such geographically bounded area. To scale beyond 16 ONUs, a much higher transmitted power, highly sensitive receiver and lower components losses are required. Alternative complex solutions to support large number of ONUs by virtually eliminating the downstream power budget problem can be achieved by regeneration of downstream signal at ONUs; but that will require additional downstream transmitter at ONUs, slightly increasing the ONU complexity and cost.

3.5 Distributed DBA Scheme

3.5.1 Overview of Centralized DBA Schemes

Several centralized tree-based DBA schemes have recently been reported in the literature [9-13]. An OLT-based polling scheme, called *Interleaved Polling with Adaptive Cycle Time* (IPACT) based on *Grant* and *Request* messages, has been presented in [9]. Using IPACT, several DBA schemes were studied in [9]; namely fixed, limited, gated, constant credit, and linear credit. Amongst these algorithms, the limited was shown to exhibit the best performance. The limited DBA scheme is cycle-based, where a cycle (T_{CYC}) 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) respectively. Thus, the algorithm schedules between B_{MIN} and B_{MAX} (bytes) at a time, where B_i is determined 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 B_{MAX}. If R_i is the requested bandwidth of ONU_i, then the granted bandwidth ($B_{Granted}$) is equal to:

$$B_{Granted} = \begin{cases} R_i & \text{If } R_i \leq B_{MAX} \\ B_{MAX} & \text{If } R_i > B_{MAX} \end{cases} \qquad (3.3)$$

B_{MAX} is determined by the maximum cycle time T_{MAX} [9]:

$$B_{max} = \frac{1}{N}[R_{EPON}(T_{MAX} - (N*T_G)]$$

where N is the number of ONUs, T_G is the guard band slot, and R_{EPON} is EPON line rate.

All of the above referenced DBA schemes are OLT-based, that is the OLT has centralized intelligence. The performance of most of these centralized schemes, including the limited scheme, suffers from several limitations, including: (1) the bandwidth granted by the OLT, during cycle n, to ONU_i is only determined by the content of a single REPORT message transmitted in the previous cycle n-1 by ONU_i (i.e., the bandwidth computation module does not take into account the remaining requests of other ONUs). 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 granted to ONU_j, however, due to limitation # 1 cited above, the maximum bandwidth that may be granted to ONU_j is only B_{max}.

3.5.2 Proposed DBA Scheme under a Distributed Control Plane

The proposed scheme assumes a cycle-based upstream link, where the cycle size can be either fixed, or variable length confined within certain lower and upper bounds to accommodate the dynamic upstream traffic conditions. During a given cycle, each ONU transmits its control (REPORT) message (within its assigned time slot) around the ring from one node to the next, where it is finally removed by the source ONU after making one trip around the ring. It typically contains the desired size of the next timeslot based on the current ONU's buffer occupancy. Since these message frames are typical Ethernet frames, the ONU should also account for additional Ethernet overhead when requesting the next time slot; this includes 8-byte frame preamble and 12-byte *Inter Frame Gap* (IFG) associated with each frame. Note that since this overhead is typical to Ethernet frame, there is no additional overhead due to this architecture. Thus, this framing has no impact on our network beside what is standard to any EPON.

Since the REPORT messages are processed and retransmitted at each node, ONUs can directly communicate their status and exchange signaling and control message information with one another. Because both REPORT message and data frames (LAN and MAN/WAN data) are transmitted within the same time slot, a REPORT message can be transmitted either at the beginning or the end of a timeslot, depending on the bandwidth request approach implemented by the ONU. For reasons to be explained below, this work assumes that control messages (REPORTs) are always transmitted at the beginning of a timeslot, followed by LAN data, and finally followed by MAN/WAN data.

Each ONU maintains a database that contains the state of the queues of all the ONUs. This information is updated each cycle whenever the ONU receives new REPORT messages from all other ONUs. The DBA module housed at each ONU uses this information to calculate a new set of time slot assignments at each cycle. The ONUs sequentially and independently run instances of the same DBA algorithm outputting identical bandwidth allocation results. The execution of the algorithm at each ONU starts immediately once all REPORT messages have been collected. Thus, all ONUs must execute the DBA algorithm prior to the expiration of the current cycle so that bandwidth allocations scheduled for the next cycle are guaranteed to be ready by the end of current cycle. Note that normally a cycle starts after the DBA calculation for ONU timeslots is completed. Each cycle consists of N timeslots, where N is the number of ONUs in the network. Once all ONUs of a cycle completed their transmissions, then that cycle is over. An execution of the DBA algorithm produces a ***unique*** and identical set of ONU assignments. It is critical that the algorithm produces a unique outcome for any arbitrary set of inputs. Once the algorithm is executed, 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 allocations.

To carry out a general and fair comparison between the performances of a centralized DBA scheme and the proposed distributed scheme, this work uses the centralized limited service scheme reported in [9], along with the appropriate changes needed to accommodate the distributed architecture, as the basis for the decentralized DBA scheme presented here. As mentioned above, to globally optimize the bandwidth allocation process, the proposed DBA algorithm execution is performed only after each ONU receives and processes all other ONUs requests.

Based on bandwidth demands, ONUs can be classified into two groups, namely: lightly loaded ONUs that have bandwidth demands less than B_{MAX}; and heavily loaded ONUs that have bandwidth demands more than B_{MAX}. During each cycle, the DBA module must keep track of the unclaimed bandwidth from the set of lightly loaded ONUs. It then must redistribute this excess bandwidth to other heavily loaded ONUs based on their requested bandwidth, i.e. two ONUs requesting bandwidths B1 and B2 more than B_{MAX} will be assigned excess bandwidths proportional to B1 and B2.

During each cycle, the lightly loaded ONUs with $R_i < B_{MAX}$ will contribute a total remainder cycle bandwidth:

$$B_{Cycle\_Remainder} = \sum_{i}^{L} (B_{MAX} - R_i) \qquad \text{L: Number of lightly loaded ONUs}$$

The heavily loaded ONUs with $R_i > B_{MAX}$ will require a total over the limit cycle bandwidth:

$$B_{Cycle\_OverLimit} = \sum_{i}^{H} (R_i - B_{MAX}) \qquad \text{H: Number of heavily loaded ONUs}$$

The total remainder cycle bandwidth can be fairly distributed amongst the heavily loaded ONUs to expand their maximum transmission window as follows [10]:

$$\Delta B_i^{extra} = B_{Cycle\_Remainder} \left[\frac{R_i - B_{MAX}}{B_{Cycle\_OverLimit}} \right]$$

where ΔB_i is the extra bandwidth allocated to ONU$_i$. The granted bandwidth, B_{GH}, for a heavily loaded ONU$_i$ is given by:

$$B_{GH} = \Delta B_i^{extra} + B_{MAX} \tag{3.4}$$

If R_i is the requested bandwidth of ONU$_i$, $B_{Granted}$ is the bandwidth granted using the proposed limited service-based distributed DBA scheme (Eqs. 3.3 and 3.4), then $B_{Granted}$ can be expressed as:

$$B_{Granted} = \begin{cases} R_i & If\ R_i \leq B_{MAX} \\ R_i & If\ R_i > B_{MAX}\ \&\ B_{Cycle\_Remainder} \geq B_{Cycle\_OverLimit} \\ B_{GH} & If\ R_i > B_{MAX}\ \&\ B_{Cycle\_Remainder} < B_{Cycle\_OverLimit} \end{cases}$$

Note that the lightly loaded ONUs ($R_i < B_{MAX}$) can be scheduled instantaneously “on-the-fly” without waiting for DBA module to perform its end of cycle computations. Whereas, the heavily loaded ONUs ($R_i > B_{MAX}$) will have to wait until all REPORT messages have been received and the DBA algorithm has computed their bandwidth allocations. Thus, lightly loaded ONUs can be scheduled ahead of heavily loaded ones. The drawback of this approach is that it introduces some idle time between two consecutive cycles, n and n+1, provided that the ONUs follow the conventional approach [9-13] and transmit their REPORT messages after data frames. This idle time, during which the PON channel is not utilized, is equal to the time of one trip around the ring (to collect the REPORT) plus the time to perform the DBA computation.

To eliminate this idle time (as proposed above), for a given time slot, the REPORT message is always transmitted before data frames. Assuming that the duration of the time slot allocated only to the data transmission of the last ONU (ONU that is scheduled to transmit last) is always longer than idle time. This assumption is always valid since an ONU that is scheduled last must always be a heavily loaded ONU. In this case, all other ONUs would have an ample time to complete the execution of the DBA algorithm prior to the completion of the last ONU's data transmission (within the same cycle). This ensures that the process of bandwidth allocations to cycle n+1 is always executed before the expiration of cycle n.

Note that, in contrast to centralized architectures where the order of the ONUs transmission is fixed (sequential) in each cycle, the distributed architecture has the added flexibility of varying the order of the ONUs transmission according to ONUs traffic demands and priority. Thus, the order of ONUs transmission may be different in each cycle and need not be fixed. As mentioned above, this work assumes that lightly loaded ONUs are always scheduled ahead of heavily loaded ones. If there is more than one lightly loaded ONU at a given cycle n, we assume that their transmission order is based on previous cycle (n – 1) transmission order; that is, an ONU that reports its queue status first is scheduled first. We further assume that the order of heavily loaded ONUs transmission is based on the shortest request first (SRF); that is, ONU with the shortest bandwidth request is scheduled first (in ascending order) and ties are broken by the ONU ID.

Note that, if ONUs transmissions are scheduled in an ascending order, upstream data transmissions assigned to any consecutive time slots are always collision-free. On the contrary, if ONUs transmission are scheduled in a descending order (for instance, ONU5 is scheduled to transmit first and then ONU2), collision between two consecutive time slots is possible (re-circulated upstream data transmitted at the end of the time slot assigned to ONU5 might collide with data transmitted at the beginning of the next scheduled time slot assigned to ONU2). To avoid this problem, as suggested above, upstream MAN/WAN data are always transmitted at the end of a timeslot. Since MAN/WAN traffic destined to the OLT is always removed by the first ONU before returning to the source node, i.e., it is not allowed to re-circulate around the ring, collision-free transmission is ensured. Furthermore, since LAN traffic is terminated at the destination node and is not allowed to re-circulate either (except when traffic is addressed to an ONU located opposite to the direction of traffic flow), this further ensures a collision-free transmission.

It is important to emphasize that while the present work considers only distributed control plane among the ONUs to achieve direct intercommunication between them as well as upstream access to the OLT, a simple OLT-based centralized control plane can also be implemented along with the proposed distributed one, resulting in a hybrid control plane. In a system that supports a hybrid control plane, the OLT, rather than discarding upstream control messages, can process these messages to track and monitor both LAN and upstream MAN/WAN traffic levels. The OLT may issue a downstream control signal (GRANT message) to halt or prioritize certain traffic/services if an ONU violates its predetermined service level agreement.

3.5.3 Synchronization

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 [5]. The TDM controller at 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 synchronization between different ONUs is required for the appropriate operation of the distributed DBA algorithm.

3.6 Performance Evaluation

In this section, we assess the feasibility and compare the performance of the proposed architecture using the distributed DBA scheme with that of the centralized architecture using the limited DBA scheme. An event-driven packet-based simulation model was developed using C++. Two simulation programs with identical network parameters were developed, one for the centralized IPACT architecture and the other for the decentralized architecture. The performance metrics used here are average packet queuing delay and upstream channel utilization.

To compare the performance results of the proposed distributed scheme with that of the centralized scheme of [9], we use the same system parameters used therein; a system with 16 ONUs, access link data rate from users to an ONU of 100 Mb/s, and 1 Gb/s upstream

EPON line rate (from an ONU to the OLT). The distance between the OLT and the

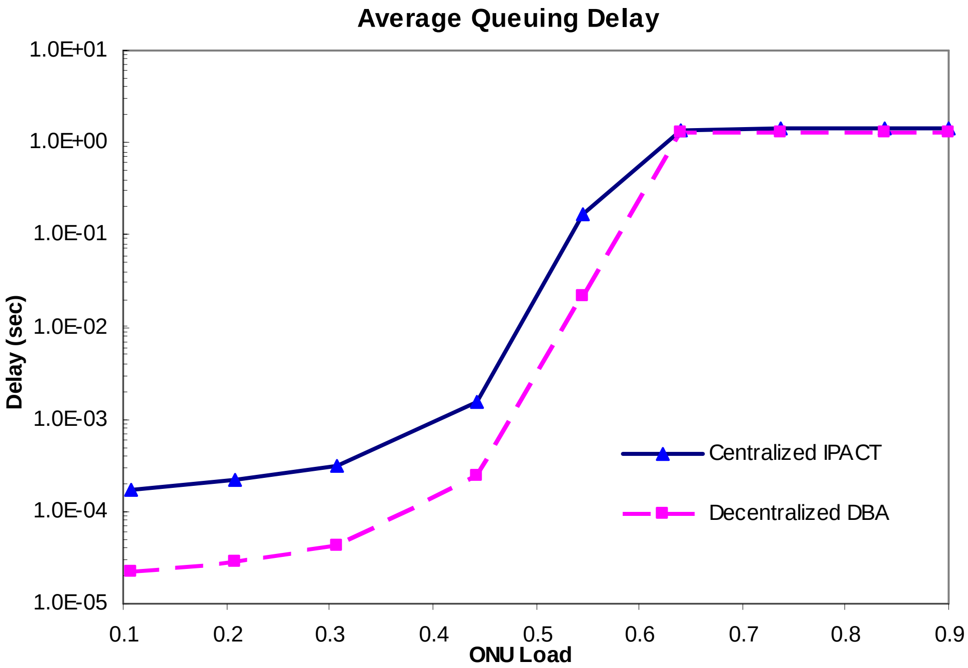


Figure 3.4: Centralized vs. Decentralized: average queuing delay

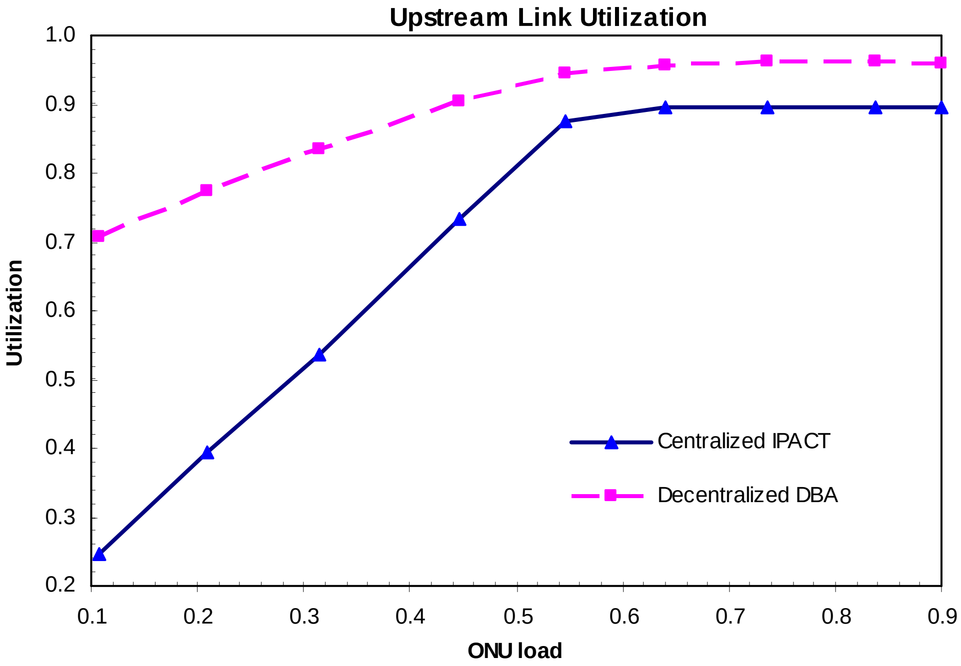


Figure 3.5: Centralized vs. Decentralized: upstream link utilization

ONUs varies from 20 km to 23 km (ring circumference ≈ 3 km). Maximum cycle time T_{MAX} is 2ms. For the Centralized architecture, the guard time, T_G, separating two consecutive transmission windows is set to 5 μs [9]. Whereas for the distributed architecture, we set $T_G = 0$ (since guard time slots are not needed). Buffer size in each ONU is 10 Mbytes.

The traffic model used here is the same as that reported in [9] 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 [14-15]. All arriving frames are then queued in a first-in-first-out buffer. Each point on the following plots corresponds to a sample of 50 million packets averaged over four different runs. In the following two subsections, the simulations will be repeated for two types of traffic. In section 3.6.1, all upstream traffic is assumed to be destined to the OLT (MAN/WAN traffic). In section 3.6.2, upstream traffic is assumed to be a mixture of MAN/WAN and shared LAN traffic.

3.6.1 Performance of Typical Upstream MAN/WAN Traffic

In this section, we assume that all upstream traffic is destined to the OLT (MAN/WAN traffic). Figures 3.4 and 3.5 compare the average packet queuing delay and channel utilization as a function of offered ONU load (OOL) for both the centralized and distributed DBA schemes. As can be seen from Figures 3.4 and 3.5, the distributed architecture demonstrates an improvement over the centralized approach in terms of average packet delay and channel utilization. At a very low load (0.1-0.2), average packet

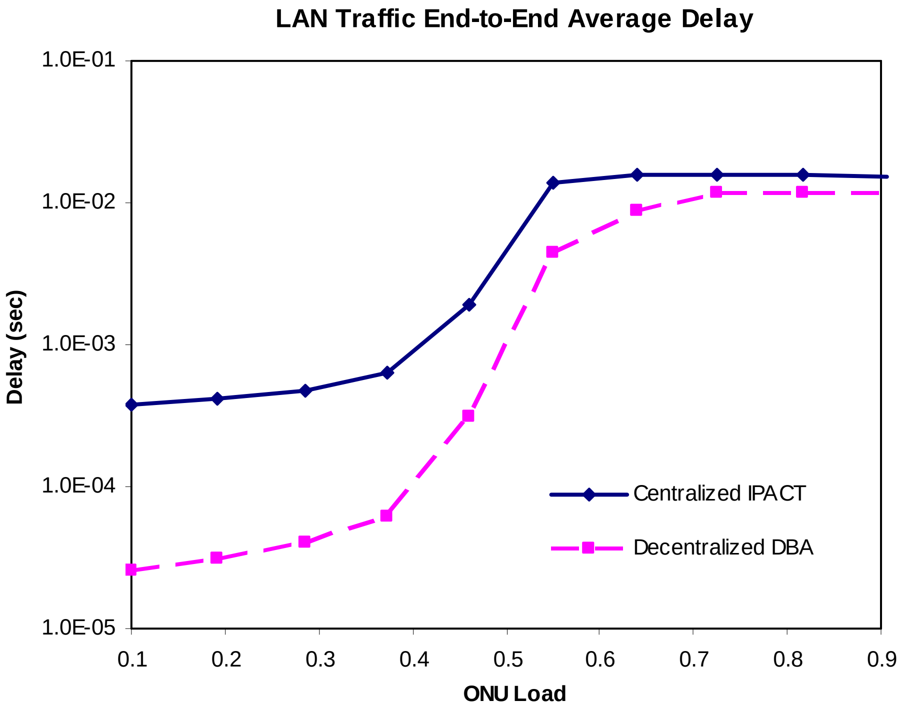


Figure 3.6: Centralized vs. Decentralized: LAN traffic end-to-end average delay

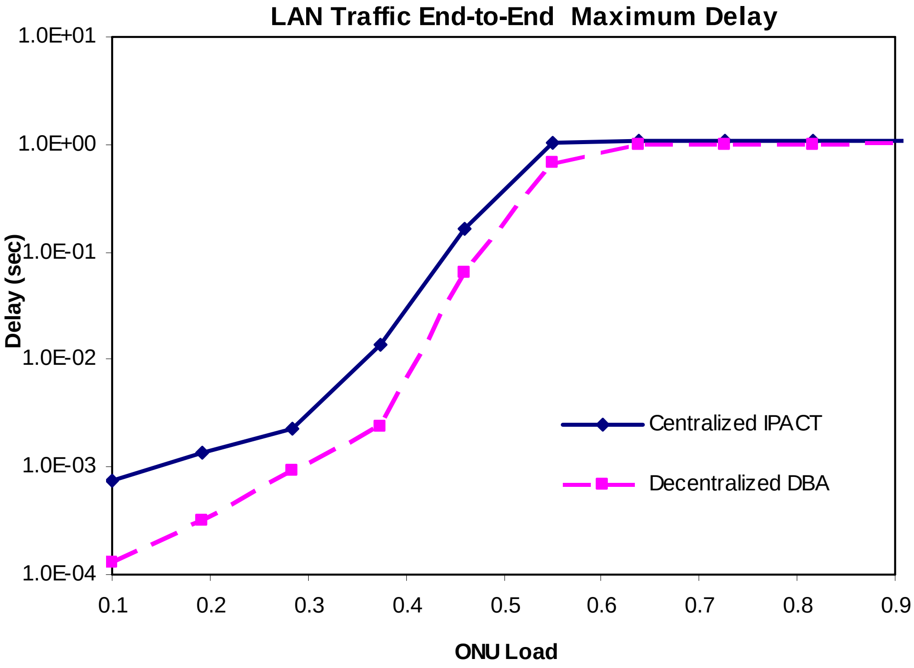


Figure 3.7: Centralized vs. Decentralized : LAN traffic end-to-end maximum delay

delay of the distributed DBA scheme is about 120-180 μs less than that of the centralized scheme (Fig. 3.4). This is mainly due to the savings of guard band slots (5μs/ONU *16 ONUs = 80 μs) as well as interchanging the order of ONUs transmissions, since queuing (TDM) delay is minimal for both architectures.

As the load increases beyond 0.5, the delay difference between the two schemes progressively changes from microseconds to milliseconds (queuing delay is now significant). This is mainly due to the following reason. As offered ONUs load increases, more and more ONUs may request bandwidth greater than B_{MAX} (heavily loaded ONUs) while some other ONUs may still request less than B_{MAX} (lightly loaded ONUs). In this case, contrary to centralized limited scheme, the proposed DBA scheme reallocates unclaimed cycle bandwidth to heavily loaded ONUs (see Eq. 3.4). This results in more efficient upstream channel utilization and consequently minimizes the average packet queuing delay. Finally, as the network saturates (OOL > 0.6), the performance of the two schemes are almost identical. This is because most ONUs now request more than the predetermined maximum allowed bandwidth (B_{MAX}), and thus most ONUs get the same allocation (B_{MAX}). This, in turn, eliminates the main advantage at higher loads, which is the reallocation of the unclaimed cycle bandwidth.

The results of Fig. 3.5 corroborate the results of Fig. 3.4, since lower average packet queuing delay indicates more efficient upstream channel utilization. Eliminating guard band slots increases the available upstream channel bandwidth and redistributing unclaimed bandwidth of lightly loaded ONUs to highly loaded ONUs leads to more efficient channel utilization.

3.6.2 Performance of Upstream LAN and MAN/WAN Traffic

In this section, we divide upstream traffic into 20% LAN traffic and 80% MAN/WAN traffic. Figures 3.6 and 3.7 compare the average and maximum end-to-end packet delay of LAN traffic, as a function of OOL, for both centralized IPACT and distributed DBA schemes. As can be seen from both figures, at all network loads, the average and maximum end-to-end packet delays of the distributed scheme are always less than those of the centralized one. As can be seen from Fig. 3.6, at low load, the centralized scheme has about 250-350 μs longer average end-to-end delays than that of the distributed one. This is mainly due to the *Round-Trip-Time* (RTT) delay from the ONUs to the OLT (over 210 μs for a 20 km trunk). In the case of the proposed ring-based architecture, the maximum propagation delay between the two most distant ONUs on the ring is about 15 μs. At higher load, as queuing delay becomes higher and dominant, the delays get higher and difference between the two schemes are now in milliseconds.

3.7 Conclusion

This chapter has proposed a novel ring-based local access PON architecture that has addressed some of the limitations of current tree-based PON architectures including supporting private networking capability. Specifically, we have proposed and devised a simple, novel ring-based EPON architecture that supports a truly shared LAN capability as well as upstream access to the OLT. The proposed architecture supports a fully distributed control plane among the ONUs for ONU-ONU communication as well as upstream access to the OLT. A distributed DBA scheme that supports the proposed decentralized architecture has been developed and its performance has been compared with that of a tree-based centralized scheme. The simulation results have indicated that the overall performance of the proposed distributed scheme, including average packet queuing delay and channel utilization, outperforms that of the centralized one, particularly for LAN traffic.

While the proposed architecture slightly increases the complexity and the cost of the ONU, however, its main feature of supporting a truly shared LAN capability among end users within a PON-based local access infrastructure might justify the extra cost. In addition, the proposed architecture offers several key advantages over typical star-based centralized PON architectures, including: (1) since the OLT always receives a fully regenerated upstream signal from the last ONU, upstream power levels received by the OLT are always guaranteed to be almost constant. This eliminates the typical near-far problem and, consequently, the need for a burst-mode receiver at the OLT; (2) regeneration and retransmission of upstream traffic at each node ensures: (a) collision-free upstream data transmission without resorting to the typical use of guard time bands. This saving of guard time’s overhead increases the available upstream transmission bandwidth and minimizes queuing delay; (b) signal level attains full power at every node. This eliminates the typical limited power budget problem associated with physical layer LAN emulation techniques described above.

Chapter 4

Delivering Triple Play Services

4.1 Introduction

EPON is expected to serve voice, video and data (triple play) over a single line with a given Quality of Service (QoS) requirements [1]. Each type of traffic has a different quality constraint and requires differentiated Class of Service (CoS). To support CoS, Ethernet networks must be able to classify traffic into classes of service and provide differentiated treatment to each class. Three classes of services are defined [2], namely the best effort (BE), the assured forwarding (AF), and expedited forwarding (EF). While EF services (delay sensitive service such as voice) require bounded end-to-end delay and jitter specifications, AF is intended for services that are not strictly delay sensitive like EF but which require bandwidth guarantees. Finally, BE applications (such as e-mail services) are neither delay sensitive nor do they require any jitter specifications.

Few centralized DBA schemes were recently introduced [3-7]. We previously mentioned the inherent drawbacks of centralized architecture in Chapter 3, such as lack of global optimization in upstream DBA, inefficiency in bandwidth utilization, etc. Based on these centralized schemes, upstream QoS support was introduced in EPON, where the intra-ONU scheduling of traffic classes takes place in ONU and upstream inter-ONU scheduling (DBA) takes place in OLT [8-10]. Since the two scheduling schemes are independent of each other, the final bandwidth allocated to a particular class of traffic for a given ONU may not be the optimum choice.

To address the above mentioned limitations of centralized scheme, in this chapter (Section 4.3) we introduce decentralized DBA schemes capable of supporting upstream QoS through differentiated CoS. In contrast to the centralized approach, the proposed QoS aware distributed DBA supports differentiated services through the integration of both scheduling mechanisms (intra-ONU and inter-ONU) at the ONU. This integration of both scheduling can only be supported by a decentralized architecture. The decentralized architecture is detailed in Chapter 3. Generally, in the distributed scheme, the ONUs exchange signaling and control information amongst themselves (within the ring). 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. With the support of this decentralized scheme, we develop QoS-based algorithm where intra-ONU (priority queuing) and inter-ONU bandwidth allocation takes place in ONUs.

The proposed decentralized scheme addresses some of the limitations of the centralized scheme cited earlier and can further provide several advantages as follows:

a) Since the decentralized bandwidth allocation computation is performed after receiving and processing all ONUs requests (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.

b) 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 the 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.

c) Since the decentralized 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.

d) Given that decentralized inter-ONU and intra-ONU 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.

e) Since the decentralized scheme’s available upstream bandwidth is higher than the centralized scheme, it results in minimized delay for the time sensitive traffic.

We demonstrate, in addition to the added flexibility and reliability, that the distributed approach has characteristics that make it far better suited than its centralized counter part for provisioning QoS necessary for properly handling voice, video, and data services over a single line.

4.2 Overview of Centralized QoS Scheme

An OLT-based polling scheme, called *Interleaved Polling with Adaptive Cycle Time* (IPACT) based on *Grant* and *Request* messages, has been presented in [4]. ONUs request OLT for upstream bandwidth; OLT being the upstream arbitrator, allocates upstream bandwidth to each ONU according to an algorithm. Using IPACT, several DBA schemes were studied in [4]; namely fixed, limited, gated, constant credit, and linear credit. Amongst these algorithms, the limited was shown to exhibit the best performance. The OLT based DBA (inter-ONU scheduling) was enhanced in [8] to support QoS through intra-ONU scheduling (priority queuing) at the ONUs (Fig. 4.1). Priority queuing with queue management facilitates class level traffic policing to allow traffic into ONU queues as well as class level transmission scheduling as per OLT's inter-ONU bandwidth allocation. Because the centralized limited IPACT scheme was shown to exhibit the best performance in [4], we will consider the QoS scheme detailed in [8] as a reference model for comparing the performance of our proposed distributed QoS scheme. An overview of QoS enabling mechanisms are detailed in the following paragraph.

4.2.1 Scheduling at OLT (inter-ONU)

The limited IPACT DBA scheme is cycle-based, where a cycle (T_{CYC}) 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) respectively. Thus, the algorithm schedules between B_{MIN} and B_{MAX}

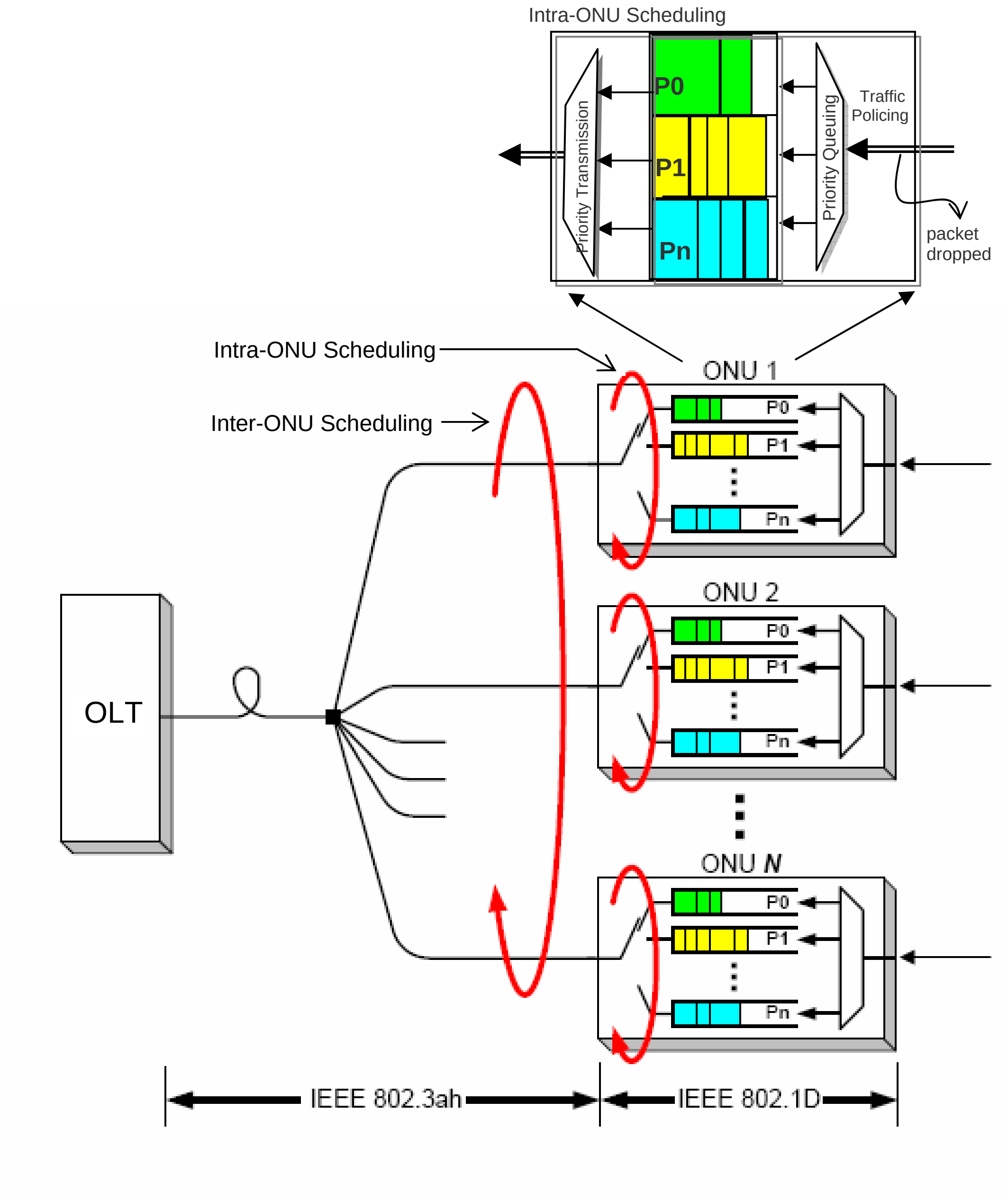


Figure 4.1: QoS mechanisms in centralized scheme

(bytes) at a time, where B^i is determined 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 B_{MAX}. If R^i is the requested bandwidth of ONU^i, then the granted bandwidth ($B^i_{Granted}$) is equal to:

$$B^i_{Granted} = \begin{cases} R^i & if \quad R^i \leq B_{MAX} \\ B_{MAX} & if \quad R^i > B_{MAX} \end{cases}$$

B_{MAX} is determined by the maximum cycle time T_{MAX} :

$$B_{MAX} = \frac{1}{N}[R_{EPON} \ \ (T_{MAX} - (N*T_G)]$$

where N is the number of ONUs, T_G is the guard band time between two consecutive ONU transmission slots, and R_{EPON} is EPON line rate. The bandwidth allocation information ($B^i_{Granted}$) is sent to ONUs by the OLT through a GATE message. $B^i_{Granted}$ is used by ONUs for intra-ONU scheduling to arrange class level transmission.

4.2.2 Scheduling at ONU (intra-ONU)

It is also known as Queue Management with Priority Queuing. In this case, queue management and priority queuing are used to divide an ONU's timeslot (allocated by the OLT) to the different classes of traffic supported by that ONU. In [8], authors focused on strict priority scheduling, because of its status as a default scheduling algorithm in IEEE 802.1D-compliant bridges and switches. Priority queuing is easy to implement. It provides low delay to high-priority traffic, but it has some performance shortcomings such as better-than-needed performance for high-priority queues and starvation of low-priority queues. Each ONU is equipped with *n* queues serving *n* priority classes (denoted $P_0, P_1, \ldots, P_n$), with P_0 being the highest priority and P_n being the lowest. When a packet is received at ONU, the ONU classifies its type and places it in the corresponding queue. The queues in each ONU share common memory space. If an arriving packet with priority *Pi* finds the buffer full in the ONU, it can preempt one or more lower-priority packets P_j $(j > i)$ from their queues, such that the P_i packet can itself be placed into the P_i queue. Between transmission slots, an ONU stores all the packets received in their respective queues. When the ONU slot starts, the ONU serves a higher-priority queue to exhaustion before serving a lower-priority queue. We assume there are three classes of traffic. The scheme can be summarized as follows:

Let, total reported queue size of an ONU, $R_t = (R_{p0} + R_{p1} + R_{p2})$, where R_{p0}, R_{p1}, R_{p2} are queue sizes of P_0, P_1, P_2 classes respectively. The class level bandwidth allocations (B_{p0}, B_{p1}, B_{p2}) within an ONU are as follows:

$$B_{p0} = R_{p0}, B_{p1} = R_{p1}, B_{p2} = R_{p2} \quad , \qquad \text{when} \quad R_t \leq B_{MAX}$$

$$\left.\begin{array}{l} B_{p0} = R_{p0} \\ B_{p1} = \begin{cases} R_{p1} & \text{if } R_{p1} \leq (B_{MAX} - B_{p0}) \\ (B_{MAX} - B_{p0}) & \text{if } R_{p1} > (B_{MAX} - B_{p0}) \end{cases} \end{array}\right\} \text{otherwise}$$

$$B_{P2} = \begin{cases} R_2 & if R_2 \le (B_{MAX} - B_{P0} - B_{P1}) \\ (B_{MAX} - B_{P0} - B_{p1}) & if R_2 > (B_{MAX} - B_{P0} - B_{P1}) \end{cases}$$

Note the above referenced DBA scheme is OLT-based and OLT has the centralized intelligence. The (inter-ONU scheduling) performance of most of the centralized schemes, including the limited IPACT scheme, suffers from several limitations, including: (1) the bandwidth granted by the OLT, during cycle n, to ONU^i is only determined by the content of a single REPORT message transmitted in the previous cycle n-1 by ONU^i (i.e., the bandwidth computation module does not take into account the remaining requests of other ONUs). 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 granted to ONU^j, however, due to limitation # 1 cited above, the maximum bandwidth that may be granted to ONU^j is only B_{MAX}; (3) since the centralized scheme requires typical guard band between two consecutive ONU transmissions, it reduces available upstream bandwidth. These lead to overall inefficient utilization of upstream bandwidth as well as inefficiency in intra-ONU scheduling; (4) furthermore, the intra-ONU scheduling takes place in ONU and upstream inter-ONU scheduling (DBA) takes place in OLT; since the two scheduling schemes are independent of each other, the final bandwidth allocated to a particular class of traffic for a given ONU may not be the optimum choice.

4.3 Proposed Decentralized QoS Scheme

The architecture and operation of the decentralized scheme is detailed in Chapter 3. For the purpose of clarity, we introduce a short overview of the general principles of decentralized operation. The proposed scheme assumes a cycle-based upstream link, where the cycle size can be either fixed, or variable length confined within certain lower and upper bounds to accommodate the dynamic upstream traffic conditions. During a given cycle, each ONU transmits its control (REPORT) message (within its assigned time slot) around the ring from one node to the next, where it is finally removed by the source ONU after making one trip around the ring. It typically contains the desired size of the next timeslot based on the ONU's current buffer occupancy. Since the REPORT messages are processed and retransmitted at each node, ONUs can directly communicate their status and exchange signaling and control message information with one another. A REPORT message can be transmitted at the beginning of a timeslot to save DBA calculation time [11]. Each ONU maintains a database that contains the state of the queues of all the ONUs. This information is updated each cycle whenever the ONU receives new REPORT messages from all other ONUs. The DBA module housed at each ONU uses this information to calculate a new set of time slot assignments at each cycle. The ONUs sequentially and independently run instances of the same DBA algorithm outputting identical bandwidth allocation results. The execution of the algorithm at each ONU starts immediately once all REPORT messages have been collected. Thus, all ONUs must execute the DBA algorithm prior to the expiration of the current cycle so that bandwidth allocations scheduled for the next cycle are guaranteed to be ready by the end of current cycle. Once all ONUs completed their transmissions, then that cycle is over.

An execution of the DBA algorithm produces a unique and identical set of ONU assignments. It is critical that the algorithm produces a unique outcome for any arbitrary set of inputs. Once the algorithm is executed, 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 allocations.

In Chapter 3, we saw that the inter-ONU scheduling of decentralized scheme is advantageous over the centralized one. Now we are to extend that indispensable advantage along with additional enhancements to support QoS. In addition to the task of arbitrating upstream bandwidth (inter-ONU), the ONUs are required to perform intra-ONU scheduling to support CoS/QoS. Intra-ONU scheduling plays an important role in supporting QoS in the EPON (through differentiated services). Through traffic policing, the incoming traffic is segregated and prioritized for entry into corresponding ONU queues. Then the class level transmission scheduling takes place. We bring into play six variations of the integrated decentralized scheduling to observe QoS performance

4.3.1 Integrated Scheduling at ONU

4.3.1.1 Priority Queuing: 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 (CBR)] and P2 being the lowest (non real-time). These classes are used for delivering voice (CBR), video stream [variable-bit-rate (VBR)], and best-effort data and they allow easy mapping of DiffServ's

(IETF framework for classifying network traffic into classes) expedited forwarding (EF), assured forwarding (AF), and best effort (BE) classes into 802.1D classes [2]. Packets are first segregated and classified (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 a common memory space. A high priority packet is always allowed in the ONU's buffer even by displacing existing low priority traffic from the buffer. At the time of ONU reporting, ONU^i reports the queue size of each class as follows: $R^i_{p0}, R^i_{p1}, R^i_{p2}$ corresponding to P_0, P_1, P_2 classes of ONU^i respectively. Total demand of the ONU is $R^i_t = (R^i_{p0} + R^i_{p1} + R^i_{p2})$.

Note the ONU reporting can be accomplished in two ways (a) an ONU can send its report within its time slot as in [11], saving the DBA time but report will be untimely (b) ONUs can send report at the end of each cycle, costing DBA time but benefiting from up-to-date report (see details in Table 4.1 and in Section 4.6).

4.3.1.2 Transmission Scheduling: Based on bandwidth demands, ONUs can be classified into two groups, namely: lightly loaded ONUs that have bandwidth demands less than B_{MAX}; and heavily loaded ONUs that have bandwidth demands more than B_{MAX}. Note each ONU is allowed up to B_{MAX} without any arbitration scheme.

During each cycle, the DBA module must now keep track of the unclaimed bandwidth from the set of lightly loaded ONUs. It then must redistribute (in addition to B_{MAX}) this excess bandwidth to other heavily loaded ONUs based on certain scheme.

During each cycle, the lightly loaded ONUs with $R_t^i < \mathrm{B_{MAX}}$ will contribute a total cycle bandwidth:

$$\mathrm{B}_{Cycle\_Remainder} = \sum_{\mathrm{i}}^{\mathrm{L}} (\mathrm{B_{MAX}} - R_t^i)$$, L: is the number of lightly loaded ONUs

The heavily loaded ONUs with $R_t^i > \mathrm{B_{MAX}}$ will require a total over the limit cycle bandwidth:

$$B_{Cycle\_OverLimit} = \sum_{\mathrm{i}}^{\mathrm{H}} (R_t^i - \mathrm{B_{MAX}})$$, H: is the number of heavily loaded ONUs

An ONU can transmit as much as reported queue size when any of (a) or (b) is true:

a) $R_t \leq B_{MAX}$, note ONU can transmit without waiting for DBA calculation, as per reporting sequence [11].

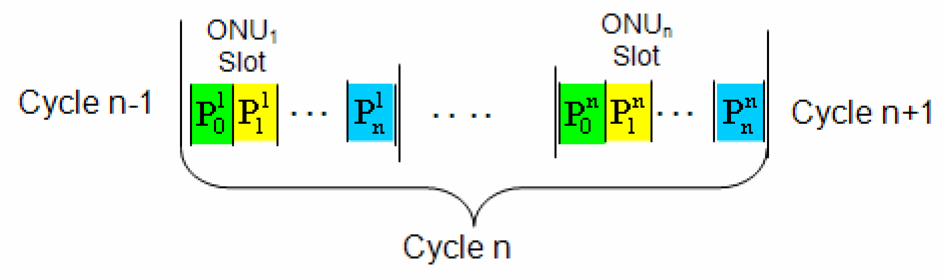
b) $R_t > B_{MAX}$ & $B_{Cycle\_\mathrm{Re}mainder} \geq B_{Cycle\_OverLimit}$

It implies that an ONU will be allowed bandwidth (B_{p0}, B_{p1}, B_{p2}) to transmit all traffic from each class as reported, as shown below $B_{p0} = R_{p0}, B_{p1} = R_{p1}, B_{p2} = R_{p2}$.

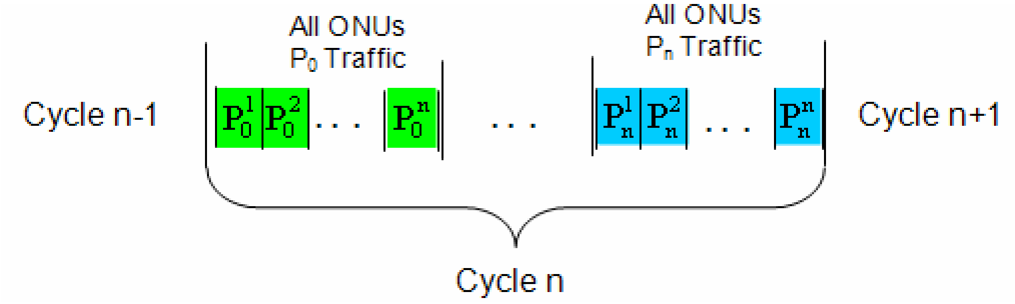
On the contrary when none of the above holds, then it requires the ONUs to invoke detail algorithm to distribute cycle bandwidth among the ONU classes considering fairness and QoS restrictions. Six variations of these algorithms are introduced in this section. We will call the decentralized DBAs (DDBA) as DDBA1 through DDBA6.

Note the various classes of traffic within an ONU can be scheduled in any of the two sequences (Fig. 4.2):

1. ONU_n transmits entire timeslot (all of its three classes-Fig. 4.2a) first, then ONU_{n+1} does the same until all ONUs are done with transmission in that cycle. Note, with in an ONU, the high priority class is always transmitted first.
2. ONU_n transmits its high priority P_0 traffic, followed by the P_0 traffic of ONU_{n+1} and this process continues until all the ONUs complete P_0 traffic transmission in that cycle (Fig. 4.2b). Then ONU_n starts to transmit its P_1 traffic followed by ONU_{n+1} P_1 traffic, until all ONUs complete P_1 traffic transmission in that cycle.



(a)



(b)

Figure 4.2: Transmission scheduling (a) each ONU transmits the entire time slot (all classes) (b) high priority traffic of each ONU is transmitted first

Table 4.1: Overview of Distributed DBAs

DBA	**Reporting**	**DBA Idle Time**	**Bandwidth Allocation**	**Transmission Sequence**
Centralized IPACT	Within ONU Slot	No	Strict priority: when the higher priority is satisfied with bandwidth, then the lower priorities are considered in sequence of their priority.	ONUs transmit their entire timeslot at once and higher priority is transmitted first within the slot.
DDBA1	Within ONU slot	No	P0 is allocated as reported, then P1 is allocated (up to B_{max}). Any additional P1 need is allocated from cycle remainder. Then P2 is allocated from any available bandwidth (P2 in disadvantage).	Transmission option 1 (Fig.4.2a): like centralized.
DDBA2	Within ONU slot	No	Similar to DBA1 but transmission sequence based on classes (best for highest priority).	Transmission option 2 (Fig.4.2b): All ONUs highest priority is transmitted first, then the next priority of all ONUs.
DDBA3	Within ONU slot	No	P0 is allocated as needed. Then the rest of the bandwidth from Bmax is equally divided between P1 and P2. Then cycle remainder is distributed among P1 classes of all ONUs by Max-Min Fair Distribution. Then any cycle remainder is given to all ONUs' P2 class by proportional distribution. P2 gets a fairer share than DBA 1& 2, but P1 looses its advantage.	Transmission option 2.
DDBA4	End of cycle (fresh) report	Yes	Similar to DBA2, but reporting is at the end of cycle. Time sensitive class benefits from fresher report, but costs DBA time.	Transmission option 2.
DDBA5	End of cycle (fresh) report	Yes	Similar to DBA3, but reporting is at the end of cycle. Time sensitive class benefits by fresher report, but costs DBA time.	Transmission option 2.
DDBA6	End of cycle (fresh) report	Yes	P0 is allocated bandwidth as needed. Then no priority is given in between P1 and P2 in distribution of bandwidth from Bmax and cycle remainder. P1 and P2 shares based on Max-Min Fair. It is the fairest in terms of P1 and P2 class, but P1 looses advantage.	Transmission option 2.

3. This process continues for P_2 traffic as well. The cycle ends when all classes of traffic of all ONUs are transmitted.

For a general overview of all the schemes, refer to Table 4.1. All the DDBAs are detailed next.

4.3.1.2.1 DDBA1: Note that reporting is scheduled within ONU time slot. P_0 class demand is low ($B_{max} >> P_0$) and it results in granted bandwidth equal to the report, $B_{P0} = R_{p0}$. Rest of the bandwidth of B_{max} and cycle remainder is allocated to P1 first and any left over goes to P2. First, P_1 demands of all ONUs that exceed (B_{max}-B_{p0}) is derived as: $B_{P1\_Cycle\_OverLimit} = \sum_{i \in H1} Y^i$, where $Y^i = (R^i_{p0} + R^i_{p1}) - B_{MAX}$ and H1 is the number of heavily loaded ONUs with $(R^i_{p0} + R^i_{p1}) > B_{MAX}$. Then P1 class is allocated bandwidth as follows:

$$B^i_{p1} = \begin{cases} R^i_{p1} & if\, R^i_{p1} \leq (B_{\max} - B^i_{p0}) \\ R^i_{p1} & if \quad B_{P1\_Cycle\_OverLimit} \leq B_{Cycle\_Remainder} \\ (B_{\max} - B^i_{p0}) + B^i_{p1\_extra} & if\, R^i_{p1} > (B_{\max} - B^i_{p0}) \,\&\, B_{P1\_Cycle\_OverLimit} > B_{Cycle\_Remainder} \end{cases},$$

where $B^i_{p1\_extra}$ = Max-Min Fair ($B_{Cycle\_Remainder}$, Y^i), where $\forall i \in H1$

Max-Min Fair [12–16] is a resource distribution scheme. The principle of this scheme is as follows: intuitively, a fair share allocates a queue with a "small" demand that it wants, and evenly distributes unused resources to the "high-demand" queues. While sharing C bandwidth among n queues, where $\sum_{i=1}^{n} Q^i > C$, the order of calculation is as follows:

1) resources are allocated ***in order of increasing demand*** ($Q^1 \le Q^2 \le ... \le Q^n$)

2) no queue gets a resource share larger than its demand ($B^i_{Share} = \min(Q^i, C/n)$)

3) no other allocations satisfying (2) has a higher minimum allocation

4) condition (3) ***recursively holds*** as we remove the minimal user and reduce the total resource accordingly

5) queues with unsatisfied demands get ***an equal share of the resources***

($B^i_{Share} = B^{i+1}_{Share}$ ***)***.

Note that, $C = \sum_{i=1}^{n} B^i_{Share}$. The pseudo code of max-min fair share is shown in Fig. 4.3.

Then the P_2 demands of all ONUs which exceeds (B_{max}-R_{p0}-R_{p1}) is derived:

$B_{P2\_Cycle\_OverLimit} = \sum_{i \in H2} X^i$ where H2: # of heavily loaded ONUs and X is defined as

$$X^i = \begin{cases} R^i_{p2} & if (B_{\max} - R^i_{po} - R^i_{p1}) \le 0 \\ R^i_{p2} - (B_{\max} - R^i_{po} - R^i_{p1}) & if (B_{\max} - R^i_{po} - R^i_{p1}) > 0 \end{cases}$$

Then the present cycle remainder is derived as follows:

$$\bar{B}_{Cycle\_Remainder} = B_{Cycle\_Remainder} - \sum_{i \in H1} Y^i$$

Max-Min Fair Share ()

```
Let:  1. There are n queues ($Q^1$ , $Q^2$ ... $Q^n$)
      2. The queues sizes are in ascending order ($Q^1 \le Q^2 \le \ldots \le Q^n$)
      3. C is total available bandwidth
      4. $B^i_{Share}$ is Max-Min Fair share bandwidth for $Q^i$

    k=n;

    for (i=1; i≤n;  i++)
     {

    if ($Q^i \le C/k$)     {
                           $B^i_{Share} = Q^i$ ;
                           }

            else
                           {
                           $B^i_{Share}$ = C/k ;
                           }

            C=C- $B^i_{Share}$ ;
            k=k-1;
     }
```

Figure 4.3: Pseudo code for Max-Min Fair Share

This leads to the P2 class allocations as follows:

$$B^i_{p2} = \begin{cases} 0\,, & if\,(B_{\max} - R^i_{po} - R^i_{p1}) \le 0 \,\&\, \bar{B}_{Cycle\_Remainder} \le 0 \\ \min(R^i_{p2}, (B_{\max} - R^i_{po} - R^i_{p1})), & if\,(B_{\max} - R^i_{po} - R^i_{p1}) > 0 \,\&\, \bar{B}_{Cycle\_Remainder} \le 0 \\ (B_{\max} - R^i_{po} - R^i_{p1} + B^i_{p2\_extra}), & if\,(B_{\max} - R^i_{po} - R^i_{p1}) > 0 \ \&\ R^i_{p2} > (B_{\max} - R^i_{po} - R^i_{p1}) \,\&\, \bar{B}_{Cycle\_Remainder} > 0 \\ \min(R^i_{p2}, B^i_{p2\_extra}), & if\,(B_{\max} - R^i_{po} - R^i_{p1}) \le 0 \,\&\ R^i_{p2} > 0 \,\&\, \bar{B}_{Cycle\_Remainder} > 0 \end{cases}$$

where $B^i_{p2\_extra} = \bar{B}_{Cycle\_Remainder} \left[\dfrac{X}{B_{P2\_Cycle\_OverLimit}} \right]$

Note that all transmissions scheduled as per transmission option1.

***4.3.1.2.2 DDBA2*:** It is the same as DDBA1, except the transmission is not scheduled based on ONU, but based on classes (transmission option 2).

***4.3.1.2.3 DDBA3*:** Note that reporting is scheduled within ONU time slot. Since the P0 demand is small, $B^i_{p0} = R^i_{p0}$ and then the rest of Bmax is divided between P1 and P2. Then the cycle remainder is distributed among P1 of all the ONUs through Max-Min Fair allocation; then the cycle remainder is distributed among P2 of all the ONUs through proportional distribution. Process is as follows:

$$B^i_{p1} = \begin{cases} R^i_{p1} & if\ R^i_{p1} \le \dfrac{(B_{\max} - R^i_{p0})}{2} \\ \dfrac{(B_{\max} - R^i_{p0})}{2} + B^i_{p1\_extra} & if\ R^i_{p1} > \dfrac{(B_{\max} - R^i_{p0})}{2} \end{cases}$$

$B^i_{p1\_extra}$ = Max-Min Fair ($B_{Cycle\_Remainder}, Y^i$), $\forall i \in H1$ where H1: # of heavily loaded ONUs

and $Y^i = R^i_{p1} - \dfrac{(B_{\max} - R^i_{p0})}{2}$, Y>0. Now the P2 allocation is

$$B_{p2}^{i}=\begin{cases} R_{p2}^{i} & if\ R_{p2}^{i}\le\frac{(B_{\max}-R_{p0}^{i})}{2} \\ (B_{\max}-R_{p0}^{i})/2+B_{p2\_extra}^{i} & if\ R_{p2}^{i}>\frac{(B_{\max}-R_{p0}^{i})}{2} \end{cases}$$

where $B_{p2\_extra}^{i}=\overline{\mathrm{B}}_{Cycle\_Remainder}\left[\frac{X^{i}}{B_{P2\_Cycle\_OverLimit}}\right]$ and accumulative over demand of P2 is

$B_{P2\_Cycle\_OverLimit}=\sum_{i\in H2}X^{i}$, H2: # of heavily loaded ONUs, where

$X^{i}=R_{p2}^{i}-\frac{(B_{\max}-R_{p0}^{i})}{2}$, X>0 and

$$\overline{\mathrm{B}}_{Cycle\_Remainder}=B_{Cycle\_Remainder}-\sum_{i\in H1}Y^{i}$$

Note that all transmissions scheduled as per transmission option 2.

4.3.1.2.4 DDBA4: It is the same as DDBA2, but the difference is in ONU reporting scheme. Unlike reporting at the beginning of slot of each ONU, reporting of all ONUs takes place at the end of cycle. It costs DBA idle time, but facilitates up-to-date reporting.

4.3.1.2.5 DDBA5: It is the same as DDBA3, but the difference is in ONU reporting scheme. Unlike reporting at the beginning of slot of each ONU, reporting of all ONUs takes place at the end of cycle as in DDBA4.

4.3.1.2.6 DDBA6: Reporting of all ONUs take place at the end of cycle. Needy ONUs get their proportional share of the cycle bandwidth remainder. Then P_0,P_1,P_2 classes share that bandwidth by Max-Min Fair. Note this is the only DDBA where P1 class has no privilege over P2.

$\mathrm{B}_{Cycle\_Remainder} = \sum_{\mathrm{i}}^{\mathrm{L}} (\mathrm{B}_{\mathrm{MAX}} - R_t^i)$, L: is the number of lightly loaded ONUs and where

$R_t^i = R_{p0}^i + R_{p1}^i + R_{p2}^i$.

The heavily loaded ONUs with R_t^i > $\mathrm{B}_{\mathrm{MAX}}$ will require a total over the limit cycle bandwidth:

$B_{Cycle\_OverLimit} = \sum_{\mathrm{i}}^{\mathrm{H}} (R_t^i - \mathrm{B}_{\mathrm{MAX}})$, H: is the number of heavily loaded ONUs.

Each needy ONU's proportional share of the cycle remainder is:

$B_{extra}^i = \mathrm{B}_{\mathrm{Cycle\_R}emainder} \left[\frac{R_t^i - B_{MAX}}{B_{Cycle\_OverLimit}} \right]$ and $B_{Granted}^i = B_{extra}^i + \mathrm{B}_{\mathrm{MAX}}$.

$B_{Granted}^i$ is shared among the queues of traffic classes (R_k^i) in ONU$^{\mathrm{i}}$ using Max-Min Fair scheme as follows: B_k^i = Max-Min Fair ($B_{Granted}^i$, $\mathrm{R}_k^{\mathrm{i}}$), where $\forall k \in P$ and P=*{p0,p1,p2}*.

Note that, in contrast to centralized IPACT where the order of ONUs transmission is fixed (i.e., sequential) in each cycle, the distributed schemes has the added flexibility of varying the order of ONUs transmission according to ONUs traffic demands and priority.

4.4 Performance Evaluation

In this section, we compare the simulation performance of the proposed QoS aware decentralized schemes with that of the centralized one. An event-driven packet-based simulation model was developed using C++. Two simulation programs with identical network parameters were developed, one for the QoS aware centralized IPACT scheme and the other for the decentralized QoS scheme. The performance metrics used here are average packet queuing delay, average queue size and packet loss ratio.

To compare the performance results of the proposed distributed scheme with that of the centralized scheme, we used identical network parameters: a system with 16 ONUs, access link data rate from users to an ONU of 100 Mbps, and a 1 Gbps upstream link data rate (from an ONU to the OLT). The distance between the OLT and the ONUs is ~21 km for the centralized tree architecture and 20km to 23km (ring circumference 3km) for decentralized architecture. Maximum cycle time is 2 ms. The guard time for centralized scheme, separating two consecutive ONU transmissions, is 5 µs. There is no guard time for the decentralized architecture. Buffer size in each ONU is 10 MB.

The traffic model used here is the same as that reported in [11] 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 [17-18]. All arriving frames are then queued as per queuing restrictions. Each point on the following plots corresponds to a sample of 50 million packets averaged over three different runs. Note we generated uneven loads from the ONUs, where half of the ONUs are heavily loaded and other half are lightly loaded.

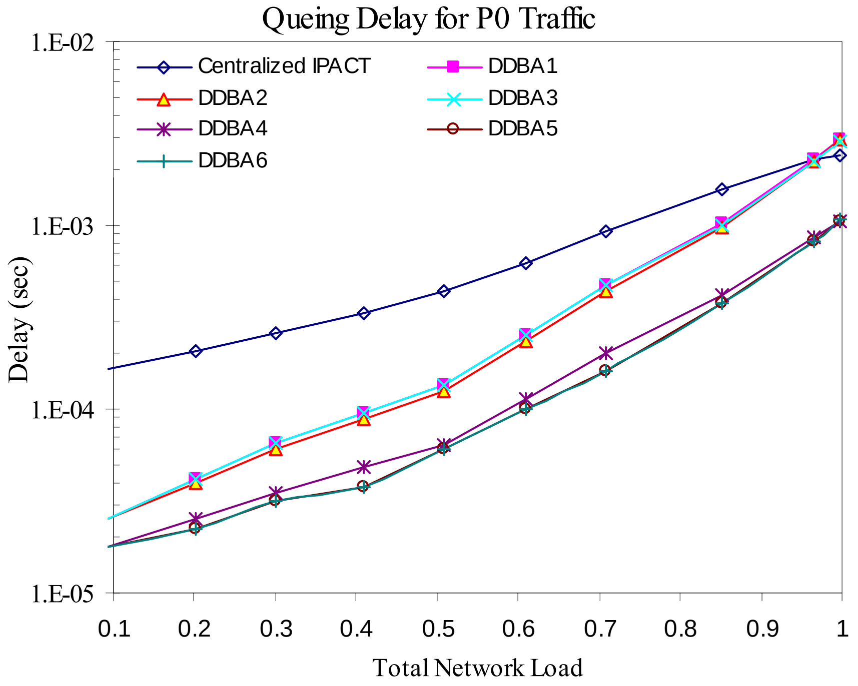


Figure 4.4: Queuing delays for P0 traffic vs. TNL

We consider three priority classes P0, P1, and P2. Here P0 is the highest priority and P2 is the lowest. These classes are used for delivering voice, video stream, and data. Each ONU maintains three separate priority queues that share the same buffering space: i) Class P0 is used to emulate a circuit over packet connection. P0 traffic has CBR. In our model, we chose to emulate a Tl connection. The Tl data arriving from the user is packetized at the ONU by placing 24B of data in a packet. Including Ethernet and UDP/IP headers, it results in a 70 bytes frame generation every 125us. Hence, the P0 data consumed 4.48 Mbps of bandwidth [9]. This is the highest priority traffic. ii) Class P1 consisted of VBR video streams that exhibit properties of self-similarity. Packet sizes in P1 streams is standard Ethernet frame ranged from 64 to 1518 B. iii) Class P2 is same as P1, but not time sensitive. This class has the lowest priority. As we varied the ONU offered load, P0 was always kept constant [4.48Mbps/100 Mbps=0.0448 of an ONU offered load (OOL)]. The remaining load was split equally between P1 and P2. For n ONUs, the total network load (TNL) = $0.1\sum_{i}^{n} \mathrm{OOL}_i$.

Figure 4.4 shows the queuing delays of the highest priority class (P0). Note the P0 traffic demand is very low, therefore under any scheme, this class always receives bandwidth equal to its report. Generally all variations of DDBAs outperform centralized scheme, because:

> (i) DDBA has more available bandwidth (no inter-ONU guard time).
>
> (ii) Their DBA decisions are globally optimized due overall network demand analysis.
>
> (iii) Redistribution of cycle remainder.
>
> (iv) Since the ONUs decide their inter-ONU and intra-ONU bandwidth, their class level allocations are more efficient.

The exception to that is at network saturation (TNL ~1), when DDBA1-3 cause slightly more delay than the centralized one. It is because, the DDBA1-3 queue reports are sent at the start of the ONU slot, not as timely as the end of cycle reporting like DDBA4-6. At network saturation the cycle length gets longer and the reports do not reflect the present queue status of the ONUs . Therefore, P0 allocation is not timely enough to outperform centralized scheme. Note the centralized scheme used strict priority, where an ONU first serves the P0 class regardless of the reported queue size to the OLT. It shows better performance for P0, but does not establish fairness among the queues. On the other hand, we are using Fair Queuing scheme [9]; once DBA allocates class level bandwidth, newly arrived P0 traffic are not considered at transmission time. They have to wait until next

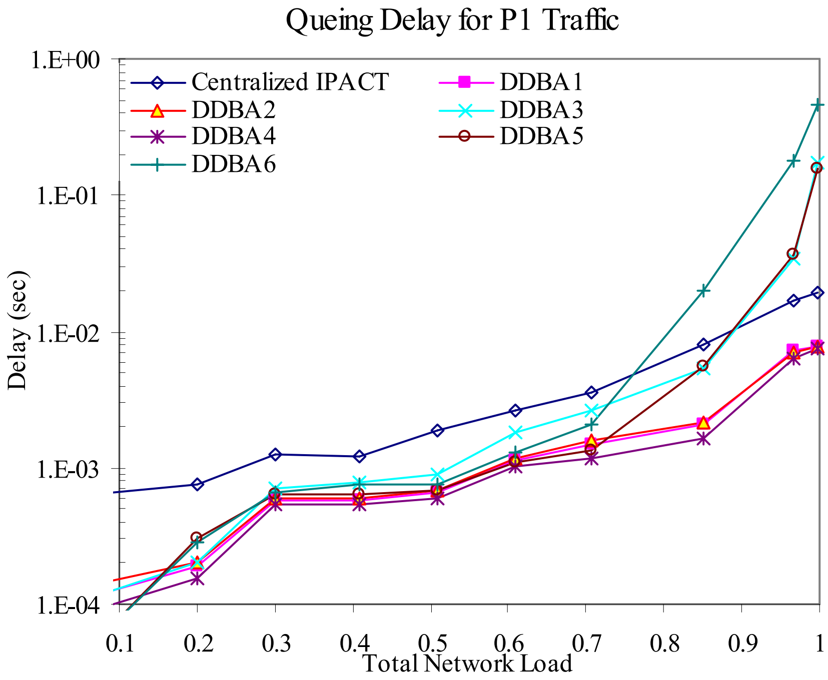


Figure 4.5: Queuing delays for P1 traffic vs. TNL

cycle, causing additional delay until next cycle. It slightly lags the P0 performance at network saturation, but establishes fairness among queues. To keep the fairness as well as enhance the P0 performance, we introduced DDBA4, 5, and 6, where ONUs exchange reports at the end of cycle. Thus the reports are timely and allows efficient transmission of traffic, resulting in lower delay of P0 traffic than DDBA1, 2, 3 and IPACT. Note it will cost some idle upstream time, but the overall performance improves.

Figure 4.5 shows the P1 traffic delays for centralized scheme and various decentralized schemes. Generally, all DDBAs outperforms centralized scheme due to the reasons stated previously. The exception to that is DDBA3, 5 and 6; at TNL~0.75, DDBA6 has higher delay than the centralized scheme. DDBA3, 5 also has similar effect at TNL~0.9. At low TNL, all traffic classes are allocated bandwidth equal to their report and the bandwidth sharing preferences have little impact. But at higher TNL, when the traffic demands exceeds available bandwidth, the sharing preferences among traffic classes impact the delays of each traffic class. As we see in higher TNL, since DDBA6 allocates comparatively lesser bandwidth to P1 (than other DBAs), it results in higher delay of P1 traffic. DDBA3 and 5 allocate more bandwidth to P1 compare to DDBA6 but less than DDBA1, 2 and 4. Consequently, in higher TNL, DDBA3 and 5 perform better than DDBA6, but worse than DDBA1, 2 and 4. Note, normally P1 traffic gets only 50% of (B_{max}-P0) in DDBA3, and 5, contrary to other DBAs (DBA1, 2, 4) where P1 gets the most bandwidth out of B_{MAX}.

The P1 performance in DDBA6 is worst than all DDBAs; but at low TNL, the DDBA6 performance improves, because of the up-to-date reporting at the end of cycle. It also applies to DDBA5. Despite the DDBA3 and 5 used same bandwidth sharing scheme, the DDBA5 outperforms DDBA3 until TNL~0.85 due to the up-to-date reporting at the end of cycle. After that they both perform the same (and delay exceeds the centralized scheme). Because, at higher load, regardless of freshness of report, the queues are mostly full and exceed B_{MAX}; the timely report cannot help P1 traffic any further. Among the better performing s (1, 2, 4), all of them gives preference to P1 traffic without any consideration to P2. DDBA4 has the best performance among them, because it uses fresh reporting. Note that this advantage will cause P2 higher delays. For the opposite reason, DDA6 will cause P2 to have lower delays.

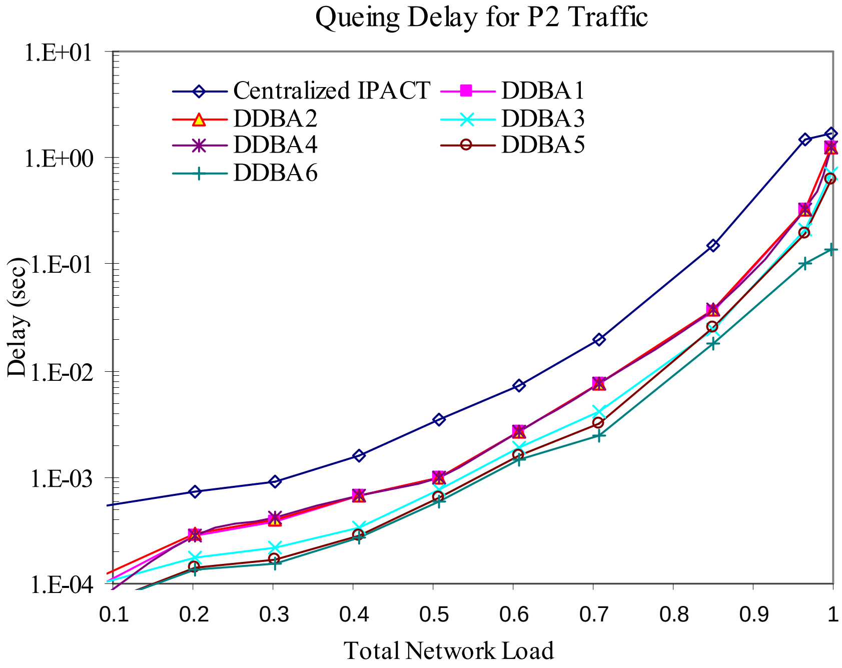


Figure 4.6: Queuing delays for P2 traffic vs. TNL

Figure 4.6 shows the P2 traffic queuing delays for all DBAs. Generally, any scheme which gave preference to P1 traffic over P2 has to pay the price with higher delay of P2 traffic. All DDBAs outperform centralized scheme, because it allocates bandwidth to P1 first, then any left over goes to P2. In addition to that, the other factors are untimely reporting, DBA is not globally optimized, no reuse of cycle remainder and independent intra-ONU and inter-ONU scheduling.

Unlike DDBA3, 5 and 6, the DDBA1, 2, and 4 gave less preference to P2 traffic resulting in a comparatively higher delay of P2 traffic. Among DDBA3, 5, 6, DDBA6 perform slightly better. Because, DDBA3 and DDBA5 equally divide bandwidth to P1 and P2, but gives cycle leftover only to P1; but DDBA6, without giving any preference to P1, always shares all bandwidth among P0, P1 and P2 using Max-Min Fair Distribution scheme. That gives P2 in DDBA6 a slight advantage. DDBA5 performs slightly better than

DDBA3, because of end of cycle fresher report. Note that we are using priority queuing in a fixed buffer size. At higher load, it causes P2 class traffic to be dropped (Fig. 4.9) resulting in lower delays than expected. For an infinite buffer case the delay would be higher.

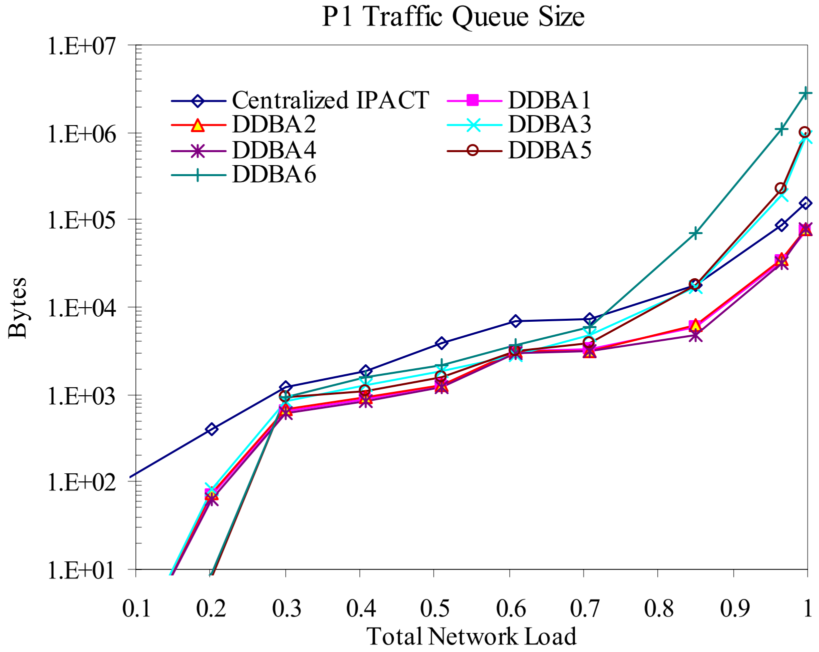


Figure 4.7: Queue size of P1 traffic vs. TNL

Figure 4.7 shows the P1 traffic queue size. The queue size is a direct reflection of how long the packets stay in the buffer. In another word, longer the queuing delay, larger is the queue size. Figure 4.5 clearly relates to Figure 4.7. The centralized scheme along with DDBA1, 2 and 4 allocate bandwidth to P1 first. After the P1 allocation, P2 may get bandwidth when available. It clearly gives an advantage to P1 and causes smaller P1 queue size. DDBA1, 2 and 4 have smaller queue size than centralized scheme due to inherent advantages of decentralized scheme, specially the redistribution of cycle remainder in the presence of heavily loaded and lightly loaded ONUs.

DDBA3, 5 and 6 do not give total preference to P1 over P2 as in rest of the DBAs. Therefore, they are going have larger queue size than rest of them. But the bandwidth sharing scheme may not have much impact at low load, because of ample available bandwidth. Also due to decentralized advantages, DDBA 3, 5, and 6 have smaller queues size than centralized scheme at low TNL. As the TNL grows, the decentralized advantages fade out at the face of massive P1 demand. DDBA3 and 5 allocate 50% of (B_{MAX} –P0) to P1 and other 50% to P2 and any cycle remainder is also allocated to P1. Due to that, under high demand, DDBA3 and 5 queue sizes out grow DDBA1, 2, 4 and the centralized scheme. DDBA6 is even worse for P1 but fairer than all; it gives no preference between P1 and P2, causing the largest P1 queue size among all DBAs (in high load).

Figure 4.8 shows the queue size of P2 traffic. Figure 4.8 relates to Figure 4.6, due to the proportional relation of queuing delay and queue size. The centralized scheme has the largest queue size because of no preference to P2 traffic. P2 traffic waits for cycles before it is transmitted (especially at higher load). Note, at higher load, it causes P2 class traffic to be dropped and resulting in smaller than expected buffer size. For an infinite buffer case the queue size would be higher. DDBA1, 2 and 4 gives no preference to P2 as the centralized scheme, but due to inherent advantages of the decentralized scheme, they have lesser queue size than centralized scheme. Since DDBA3 and 5 give more preference to P2 than DDBA1, 2 and 4, their queue sizes are smaller than DDBA1, 2 and 4. Between DDBA3 and 5, the DDBA5 performs slightly better due to fresher reports.

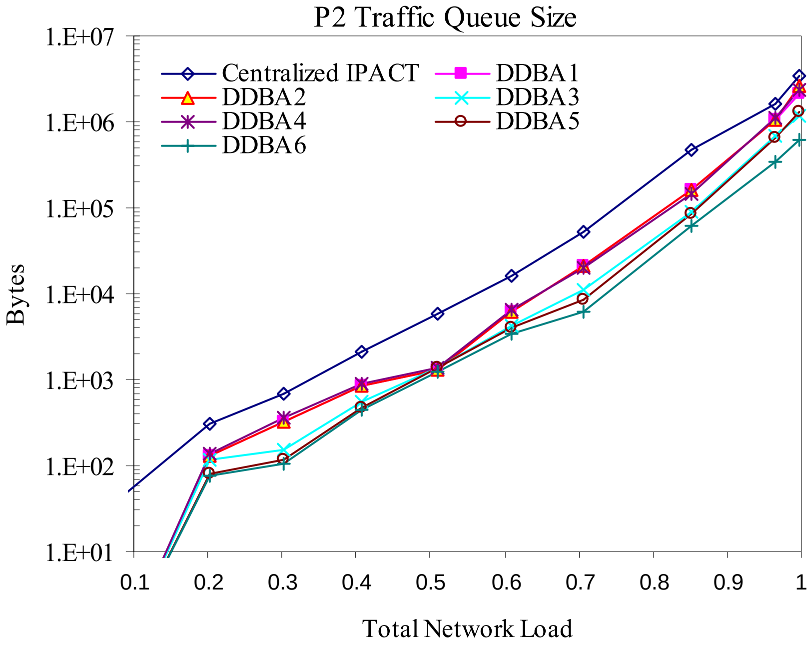


Figure 4.8: Queue size of P2 traffic vs. TNL

DDBA6 has the smallest queue size, because it is fairer to P2 than any other DBAs. Note that the accumulative average queue size of P1 and P2 do not reach the maximum size due to the two types of ONU loads (lightly loaded and heavily loaded ONUs). At TNL~1, the queues of the heavily loaded ONUs are saturated and start dropping traffic; on the other hand, the queues of the lightly loaded ONUs are not full, bringing the average queue size down. For the evenly loaded ONUs, at TNL~1, all ONUs fills up almost evenly and starts to drop traffic around the same time, increasing the average queue size to maximum.

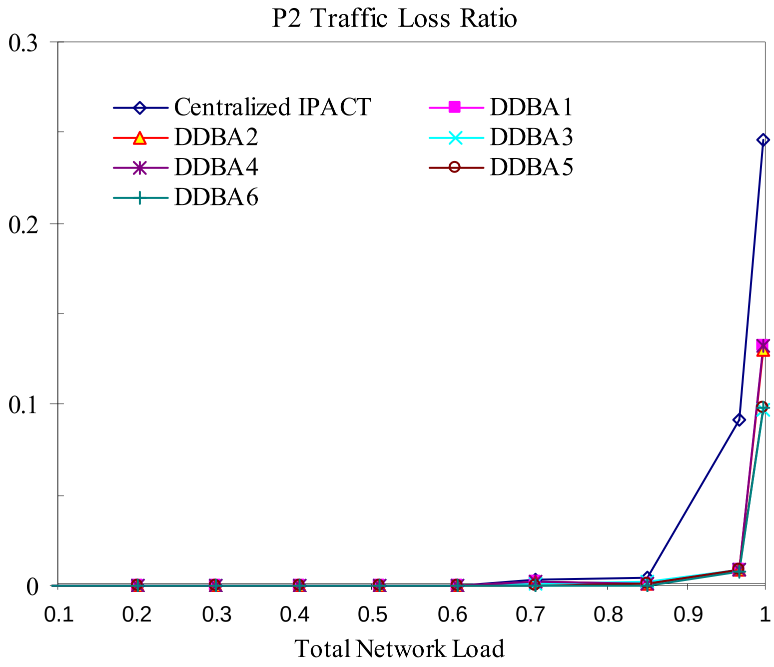


Figure 4.9: Packet loss ratio of P2 traffic vs. TNL

Figure 4.9 shows packet loss ratio of P2 traffic for all DBAs. Note due to priority queuing, the lower priority traffic is dropped to make space for higher priority traffic. Centralized P2 queue size is the largest among all (see Fig. 4.8). At higher load, the arrival of higher priority traffic (P0 and P1) displaces the P2 traffic. It results in highest P2 packet loss ratio. The decentralized schemes (P2) perform better, because they have shorter P2 queuing delay resulting in smaller P2 queue sizes than centralized scheme. Among the DDBAs, the DDBA 3, 5 and 6 perform slightly better, due to their fairer bandwidth allocation towards P2 traffic.

4.5 Conclusion

The centralized DBA scheme generally suffers from number of drawbacks, such as lack of global optimization in upstream DBA, lesser upstream bandwidth, inefficiency in bandwidth utilization etc. In terms of QoS support, the intra-ONU scheduling of traffic classes takes place in ONU and upstream inter-ONU scheduling (DBA) takes place in OLT; since the two scheduling schemes are independent of each other, the final bandwidth allocated to a particular class of traffic for a given ONU may not be the optimum choice.

In this chapter we proposed a decentralized scheme to support QoS through differentiated CoS. Fully distributed time division multiple access schemes were developed based on the ring-based decentralized architecture. It allowed collision-free upstream data transmission without resorting to the typical use of guard time. Furthermore, in contrast to the centralized approach, the proposed QoS-based distributed DBA supported differentiated services through the integration of both scheduling mechanisms (intra-ONU and inter-ONU) at the ONUs. This integrated scheduling feature that can only be supported by a decentralized architecture, provides better QoS guarantees for properly handling voice, video, and data services over a single line. The higher available upstream bandwidth, redistribution of unused cycle bandwidth, and integrated scheduling resulted in minimized queuing delay, buffer size and packet loss ratio. Finally, in addition to the added flexibility and reliability, the overall performance of the proposed decentralized EPON architecture and the associated QoS schemes are shown to be more efficient than their centralized counterparts. Among the variations of distributed DBAs, some have a specific advantage over the others. Rather than converging to one solution, we demonstrated that according to system needs one solution could be chosen over the others to handle specific circumstances.

Chapter 5

Downstream Bandwidth Allocation under Decentralized Architecture

5.1 Introduction

All transmissions in an Ethernet Passive Optical Network (EPON) are performed between an Optical Line Terminal (OLT) and an Optical Network Unit (ONU). Traffic from an ONU to an OLT is called ‘upstream’ (multipoint-to-point) traffic, and traffic from an OLT to an ONU is called ‘downstream’ (point-to-multipoint) traffic [1-5]. Typically OLT administers network control plane, known as centralized architecture [1-10]. In recent literature, upstream Dynamic Bandwidth Allocation (DBA) schemes drew a lot of attention [6-11]; the downstream DBA issues are yet to be explored. In the downstream direction, an EPON operates as a broadcast and select network. OLT houses N number of queues, each corresponding to an ONU. Incoming traffic is first sorted according to destination ONU addresses and then accepted into the corresponding queues. OLT divides available downstream bandwidth among the queues and allocates transmission timeslots. The OLT then broadcasts frames from a queue (corresponding to a destination ONU) in its allocated timeslot. An ONU accepts the broadcasted frame matching its Medium Access Control (MAC) address to frame destination address. In the typical centralized architecture, the OLT is required to perform an additional function of arbitrating among the ONUs for sharing upstream bandwidth. To facilitate the implementation of upstream bandwidth allocation schemes, the OLT and ONUs exchange control messages, namely, REPORT message and GATE message. These control messages are defined by the IEEE 802.3ah task force through the development of Multi-Point Control Protocol (MPCP) [5]. A REPORT message is sent by an ONU to OLT informing it about its bandwidth requirements. Upon receiving a REPORT, the OLT passes the message to OLT's DBA module to perform the bandwidth allocation computation. The OLT then grants the ONU a transmission time slot by sending a GATE message indicating the start time and the duration of such a time slot. The frequent generation of GATE messages overburdens the OLT, thereby consuming and reducing downstream bandwidth, hindering the downstream performance.

Deploying point-to-point customer communication links to emulate a shared *Local Area Network* (LAN) within a single PON infrastructure is an important feature for providing a private networking capability [12-15]. In general, within the centralized control plane, standard upper layer shared LAN emulation techniques require the use of bridges/routers at the OLT to redirect data frames back to the ONUs [12]. These techniques are effective but reduce the available downstream transmission bandwidth and increase end-to-end delay of LAN traffic [11]. Similar to the GATE message, the reduction of available downstream bandwidth to support LAN capability also hinders the downstream performance.

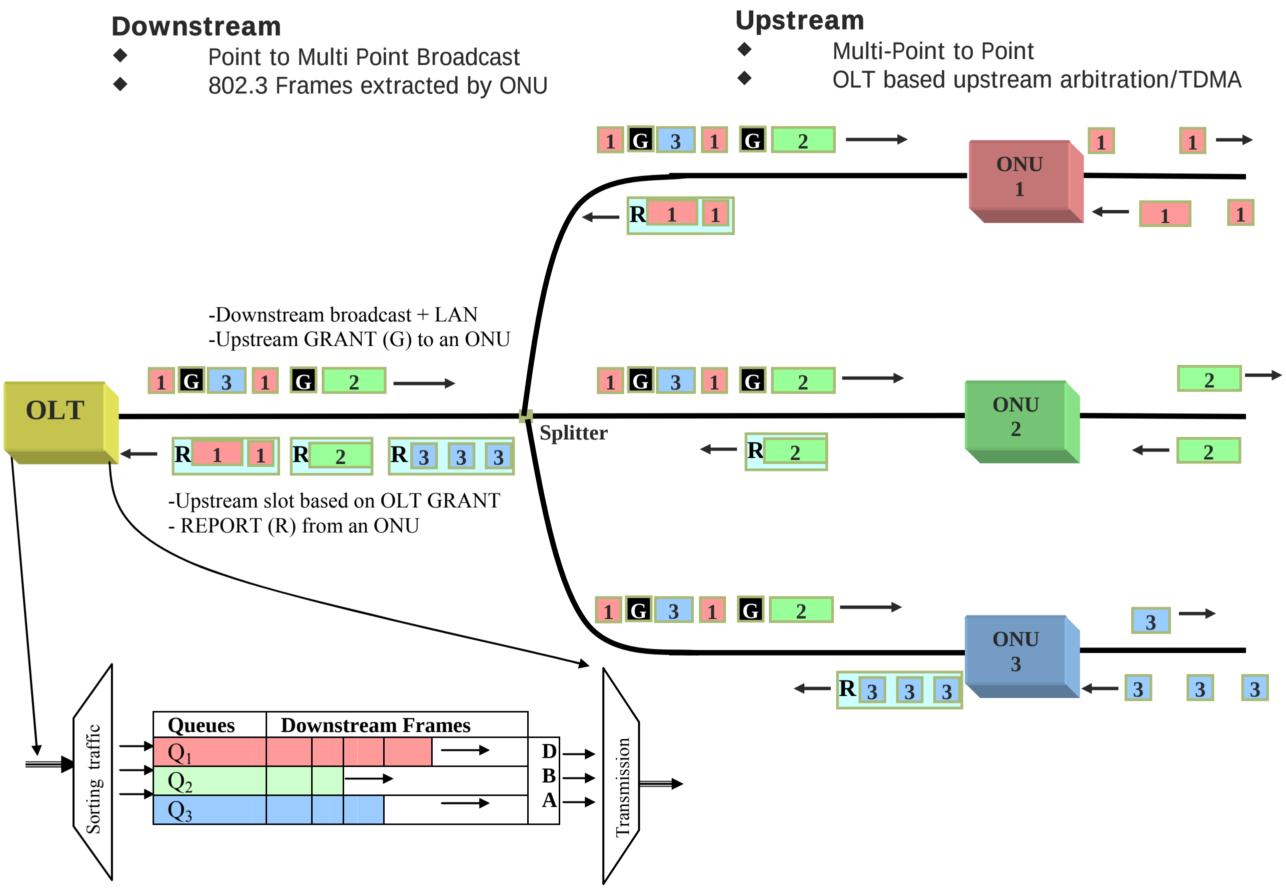


Figure 5.1: Upstream/Downstream operation in a centralized tree based EPON

The concern of this chapter is to explore the solution to the above mentioned two impediments (GATE and LAN consumption of downstream bandwidth) to downstream performance. We propose a downstream DBA scheme over a decentralized upstream control plane, where OLT is excluded from the function of upstream bandwidth allocation and LAN traffic re-routing. The proposed decentralized scheme eliminates downstream bandwidth wastage and increases downstream performance significantly. The main characteristic of the proposed scheme is that it supports a fully distributed upstream control plane among the ONUs for ONU-ONU communication (LAN connectivity) as well as upstream access to the OLT. Specifically, it utilizes a fully distributed *Time Division Multiple Access* (TDMA) arbitration scheme in which the OLT is excluded from the arbitration process. In the proposed decentralized scheme, the ONUs exchange signaling and control information concerning their queue status and their transmission needs amongst themselves (REPORT messages). Then, the ONUs concurrently 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 including both LAN and OLT traffic without any collisions, eliminating the OLT's centralized task of processing requests and generating grants for bandwidth allocations. Therefore, unlike centralized scheme, the entire downstream bandwidth is utilized for downstream transmission purpose. In the following sections, we developed few variations of decentralized downstream DBAs and conducted detailed simulations in order to study the performance and validate the effectiveness of the proposed DBAs.

5.2 Downstream Operation within Centralized Scheme

OLT has N number of queues, each corresponding to an ONU, refer to Fig. 5.1 depicting centralized tree based EPON operation, for more information. Incoming traffic is sorted according to destination ONU address and entered into corresponding queues. Periodically, the OLT checks the downstream queue occupancy and makes DBA decisions for each queue destined for an ONU, then the OLT transmits downstream traffic from the queues (the order of transmission from the queues is not significant at this point). Note that in centralized approach, such as Interleaved Polling with Adaptive Cycle Time (IPACT), OLT is responsible for scheduling both upstream and downstream transmission slots. Within a given downstream cycle, OLT receives ONUs request (REPORT) for upstream bandwidth. While OLT transmits downstream traffic, it also calculates upstream bandwidth for ONUs and broadcasts GRANTs to the ONUs. These upstream GRANTs consume downstream bandwidth, contributing to inefficient use of downstream bandwidth.

Similarly, LAN emulation within a single centralized PON infrastructure also causes wastage of downstream bandwidth due to standard upper layer shared LAN emulation techniques requiring the use of bridges/routers at the OLT to redirect data frames back to the ONUs using downstream bandwidth [12]. These two types of downstream bandwidth wastages are discussed in greater detail as follows:

5.2.1 GRANT wastage

The wastage of downstream bandwidth due to cyclic upstream GRANT depends on ONU counts and upstream cycle length. Upstream cycle length is variable within a certain minimum and maximum bound to accommodate variable traffic load. At higher upstream load the cycle time will be up to the maximum length. The maximum upstream cycle length is generally set to 2ms [6-11]. At a lower upstream load, cycle lengths are shorter. Note that the minimum cycle length cannot be shorter than the round-trip-time (RTT) between an ONU and OLT [6]. Because cycle length is the time between two consecutive GRANTS (sequential) to a specific ONU, it requires at least one RTT between OLT and ONU. The RTT between OLT and ONU depends on the distance between OLT and ONU (i.e. for ~20km trunk the RTT is ~ 200μs; similarly for 10km the RTT is ~ 100 μs). Minimum upstream cycle length is defined as follows:

$$T_{upstream\_cycle\_\min} = MAX\begin{cases} T_{rtt} \\ \mathrm{N}\left(\mathrm{T}_{\mathrm{gaurd}} + (\mathrm{R}/\mathrm{R}_{\mathrm{EPON}})\right) \end{cases}$$

R_{EPON} = ONU to OLT Transmission line rates (1Gbps)

R = ONU reports to OLT (84B)

N = number of ONUs (32)

T_{guard} = guard time between two ONUs upstream transmissions=1 μs

T_{rtt} = round trip time between ONU and OLT

While the upstream cycle length gets shorter, the number of upstream cycles per second increases significantly. Thus, at a low network load the total number of grants/second issued by the OLT increases. The number of upstream grants per cycle is equal to the number of ONUs in the system. Therefore, increased ONU count also results in more wastage of downstream bandwidth. The maximum wastage of downstream bandwidth due to GATE message is as follows:

$$B^{GATE}_{downstream\_waste} = \frac{1}{T_{upstream\_cycle\_\min}}\left(\mathrm{N} * \mathrm{G}\right), \text{ where}$$

G= size of GRANT (84B*8=672bits)

$B^{GATE}_{downstream\_waste}$ = wastage of downstream bandwidth

$T_{upstream\_cycle\_\min}$ = minimum upstream cycle time (upstream network load<0.1)

N = number of ONUs

To illustrate this point, Table 5.1 shows the worst case wastage of downstream bandwidth under various conditions.

Table 5.1: Centralized scheme's bandwidth wastage

ONU count	**16**		**32**	
Trunk (km)	10	20	10	20
$T_{upstream\_cycle\_\min}$ (μs)	100	200	100	200
Maximum $B^{GATE}_{downstream\_waste}$ (Mbps)	~108	~54	215	~108

In low upstream load (TNL<0.1 from 32 ONUs with 10km trunk), the downstream bandwidth loss could be as much as 200Mbps (20% of total bandwidth). As the upstream load grows the cycle length reaches the maximum (~2ms, regardless of OLT distance from ONUs), causing the downstream bandwidth loss to be minimum. For 32 ONUs, the minimum loss is as follows:

$$B^{GATE}_{downstream\_waste} = \frac{1}{T_{upstream\_cycle\_\max}} 32 * 672b \sim 11Mbps$$

Under the given conditions, the GATE wastage could be 11 Mbps to 215 Mbps.

5.2.2 LAN wastage

The other type of downstream bandwidth wastage is caused by upstream LAN support. Due to LAN emulation, ONU LAN traffic is sent to OLT and OLT broadcasts then back to ONUs [12]. The downstream bandwidth wastage due to LAN traffic is equivalent to the amount of upstream LAN traffic. We define LAN wastage as follows:

$$B^{LAN}_{downstream\_waste} = \mathrm{P} * \mathrm{L}_{\mathrm{upstream}}$$

where,

P = % of upstream network load (LAN)

$\mathrm{L}_{\mathrm{upstream}}$ = total upstream network load

Assuming 20% of the total upstream load is inter-ONU LAN traffic [11], then the wasted downstream bandwidth would be as high as 200 Mbps in maximum network load (TNL~1). Note that both GATE and LAN wastages are dependent on upstream network load due to centralization at OLT. At low upstream TNL the downstream loss is dominated by GATE messages. On the other hand, at high TNL the downstream loss is dominated by LAN traffic broadcasted back to ONUs. As the GATE message loss decreases the LAN traffic loss increases. The accumulative loss of the LAN and GATE traffic is always more than 200Mbps, regardless of TNL. Reduction in available downstream bandwidth is a major drawback for the centralized system that supports LAN emulation.

OLT has to distribute the reduced downstream bandwidth among the queues. OLT's downstream DBA module executes the distribution/allocation scheme. Since the centralized limited IPACT DBA scheme was shown to exhibit good performance in upstream direction [6], we will consider this simple distribution scheme as a reference model for centralized downstream bandwidth distribution. Periodically OLT checks the queues occupancy to distribute the available downstream bandwidth. Since these queues hold typical Ethernet frames, the OLT also accounts for additional Ethernet overhead for each frame; this includes 8-byte frame preamble and 12-byte *Inter Frame Gap* (IFG) associated with each frame. Cyclic bandwidth distribution by OLT is as follows:

$$B_i^{Granted\_} = \begin{cases} Q_i & if Q_i \leq B_{\max} \\ B_{\max} & if Q_i > B_{\max} \end{cases}$$

where,

$$B_{max} = \frac{1}{N} [(R_{EPON} - B^{LAN}_{downstream\_waste} - B^{GATE}_{downstream\_waste}) * T_{dowsntream\_cycle\_max}]$$

R_{EPON} = OLT to ONU (downstream) transmission line rate

N = number of queues in OLT

$T_{downstream\_cycle\_max}$ = maximum downstream cycle time

$B^{LAN}_{downstream\_waste}$ = wastage of downstream bandwidth due to LAN traffic

$B^{GATE}_{downstream\_waste}$ = wastage of downstream bandwidth due to GATE message

Note that wastage by GATE and LAN varies according to upstream load, but the accumulative average always above certain amount.

5.3 Proposed Downstream Operation

In Chapter 3 we detailed the decentralized architecture with LAN capability. The decentralized scheme is based on ONU-ONU connectivity. Inter-ONU connectivity is used to exchange REPORT among the ONUs. The DBA module housed at each ONU uses this REPORT to calculate a new set of timeslot assignments at each cycle. The ONUs sequentially and independently run instances of the same DBA algorithm outputting identical bandwidth allocation results. The OLT is excluded from this operation. Furthermore, the inter-ONU connection allows the LAN traffic exchange among the ONUs, relieving OLT from the burden of receiving and re-broadcasting LAN traffic back to the ONUs. Since OLT is excluded from the functions of upstream bandwidth allocation and LAN support, the total downstream bandwidth is available for downstream function. Thus, proposed distributed scheme eliminates the centralized scheme's downstream impediments.

OLT houses N number of queues, each corresponding to an ONU. Incoming traffic is sorted according to destination ONU addresses and accepted into corresponding queues. Periodically, the OLT checks the downstream queue occupancy and makes DBA decisions for each queue. In other words, OLT simply divides available downstream bandwidth among the queues to allocate downstream transmission timeslot. The three variants of decentralized downstream bandwidth distribution schemes will be further discussed: Fixed Maximum Limit (FML), Cycle Remainder Redistribute (CRR), and Max-Min Fair (MMF).

5.3.1 Fixed Maximum Limit (FML):

It is a simple distribution scheme which uses the same distribution approach as used in limited IPACT [6]. Since the centralized limited IPACT scheme was shown to exhibit the good performance in upstream direction, its bandwidth distribution scheme as a reference model for comparing the performance of decentralized scheme with the centralized scheme will be considered. In every cycle, the available downstream bandwidth is divided among the queues in following manner:

$$B_i^{Granted} = \begin{cases} Q_i & if Q_i \leq B_{\max} \\ B_{\max} & if Q_i > B_{\max} \end{cases} \qquad (5.1)$$

$$\text{where, } B_{\max} = \frac{1}{N}[R_{EPON} * T_{downstream\_cycle\_max}]$$

R_{EPON} = OLT to ONU (downstream) transmission line rate

N = number of queues in OLT

$T_{downstream\_cycle\_max}$ = maximum cycle time

In other words, queue is granted the requested number of bytes, but no more than a given predetermined $B_{\max}$. Queue size over B_{max} will be granted bandwidth in next cycle.

5.3.2 Cycle Remainder Redistribute (CRR)

Initially CRR works like FML (eq. 5.1), but eradicates its shortcoming. FML scheme suffers from a limitation due to the bursty nature of Ethernet traffic. In an instant, some queues might have less traffic to transmit while others may require more bandwidth than B_{max}. For instance, assume that Q_i size is small ($Q_i < B_{max}$), while Q_j size is large ($Q_j > B_{max}$). Although there is an excess amount of bandwidth ($B_{max} - Q_i$) that can be granted to Q_j, however, due to limitation of the distribution method, the maximum bandwidth that may be granted to Q_j is only B_{max}. Note that to support CRR DBA, unlike FML, OLT has to consider overall queue occupancy of the system (i.e. bandwidth allocation cannot be done one to one basis). Therefore, we propose the redistribution of unused cycle remainder from the low occupancy queue to the high occupancy queue to enhance efficiency without neglecting fairness. Based on OLT queue sizes, they can be classified into two groups, namely: low occupancy queues that whose size is smaller than B_{max}; and high occupancy queues that have size more than B_{max}. During each cycle, the DBA module must keep track of the unused cycle bandwidth from the set of low occupancy queues. It then must redistribute this excess bandwidth to other high occupancy queues according to the ratio of their sizes, i.e. two queues sizes Q1 and Q2 more than B_{max} will be assigned excess bandwidths proportional to Q1 and Q2.

During each cycle, the low occupancy queues with $Q_i < B_{max}$ will contribute a total remainder cycle bandwidth:

$$B_{Cycle\_Remainder} = \sum_{i}^{L} (B_{max} - Q_i) \qquad \text{L: Number of low occupancy queues}$$

The high occupancy queues with $Q_i > B_{max}$ will require a total over the limit cycle bandwidth:

$$B_{Cycle\_OverLimit} = \sum_{i}^{H} (Q_i - B_{max}) \qquad \text{H: Number of high occupancy queues}$$

The total remainder cycle bandwidth can be fairly distributed amongst the high occupancy queues to expand their maximum transmission window as follows [7]:

$$\Delta B_i^{extra} = B_{Cycle\_Remainder} \left[\frac{Q_i - B_{max}}{B_{Cycle\_OverLimit}} \right]$$

where ΔB_i is the extra bandwidth allocated to Q_i. The granted bandwidth, B_{GH}, for a high occupancy queues Q_i is given by:

$$B_i^{GH} = \Delta B_i^{extra} + B_{max} \qquad (5.2)$$

If Q_i is the queue size, $B_{Granted}$ is the bandwidth granted using the proposed limited service-based distributed DBA scheme (Eqs. 5.1 and 5.2), then $B_{Granted}$ can be expressed as:

$$B_i^{Granted} = \begin{cases} Q_i & if\, Q_i \le B_{\max} \\ Q_i & if\, Q_i > B_{\max} \;\&\, B_{Cycle\_\mathrm{Re}mainder} \ge B_{Cycle\_OverLimit} \\ B_i^{GH} & if\, Q_i > B_{\max} \;\&\, B_{Cycle\_\mathrm{Re}mainder} < B_{Cycle\_OverLimit} \end{cases}$$

5.3.3 Max-Min Fair (MMF)

We introduce another bandwidth distribution scheme named Max-Min Fair share [16–20]. The scheme is as follows: intuitively, a fair share allocates a queue with a "small" demand that it wants, and evenly distributes unused resources to the "high-demand" queues. While sharing C bandwidth among n queues, bandwidth among n queues, where $\sum_{i=1}^{n} Q_i > C$, the order of calculation is as follows:

1) resources are allocated **in order of increasing demand** ($Q_1 \le Q_2 \le \ldots \le Q_n$)

2) no queue gets a resource share larger than its demand ($B_i^{Granted} = \min(Q_i, C/n)$)

3) no other allocations satisfying (2) has a higher minimum allocation

4) condition (3) **recursively holds** as we remove the minimal user and reduce the total resource accordingly

5) queues with unsatisfied demands get **an equal share of the resources** $B_i^{Granted} = B_{i+1}^{Granted}$.

Note that, $C = \sum_{i=1}^{n} B_i^{Granted}$.

The pseudo code of max-min fair share is shown in Fig. 5.2.

Max-Min Fair Share ()

```
Let :  1. OLT has n queues ($Q_1, Q_2, ... Q_n$)
       2. The queues sizes are in ascending order ( $Q_1 \le Q_2 \le ... \le Q_n$ )
       3. C is total available bandwidth
       4. $B_i^{Granted}$ is max-min fair share bandwidth for $Q_i$

    k=n
    for (i=1; i≤n; i++)
    {
        if ($Q_i \le C/k$) {
            $B_i^{Granted} = Q_i$ ;
        }
        else {
            $B_i^{Granted} = C/k$ ;
        }

        C=C- $B_i^{Granted}$ ;
        k=k-1;
    }
```

Figure 5.2: Pseudo code for Max-Min Fair

Note since OLT houses all queues and allocates bandwidth, the order of transmission is not important in our discussion.

5.4 Simulation Results

In this section, we compare the simulation performance of the proposed distributed downstream scheme with that of centralized downstream scheme. An event-driven packet-based simulation model was developed using C++. Two simulation programs with identical network parameters were developed, one for the centralized architecture and the other for the decentralized architecture. Initially we employed Fixed Maximum Limit

(FML) distribution scheme for both the centralized and decentralized schemes to demonstrate the architectural advantage of the proposed scheme. Then we repeat the decentralized downstream simulation with enhance DBAs (CRR, FMM) for a comparative analysis among the decentralized downstream DBAs. The performance metrics used here are average packet queuing delay, utilization, packet loss ratio and queue size.

To compare the performance results of the proposed distributed scheme with that of the centralized scheme, we use the following system parameters: an OLT housing 32 queues corresponding to 32 ONUs, downstream link rate is 1 Gb/s, incoming data rate to OLT queues is 50Mbps, each queue size is 5MB, the trunk length ~ 10km, and maximum downstream cycle time T_{MAX} is 2ms. Parameters specifically associated with the centralized architecture are: upstream link rate is 1Gb/s, inter-ONU upstream transmission guard time (T_{guard}) is 1 μs, downstream GRANT message size is 84B and the LAN traffic is assumes to be 20% of total upstream traffic [11].

The traffic model used here is the same as that reported in [6] where each queue 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 [21-22]. All arriving frames to the OLT are then queued in a first-in-first-out buffer. Each point on the following plots corresponds to a sample of 50 million packets averaged over four different runs. In the following two subsections, the simulations will be repeated for different distribution schemes. In Section 5.4.1 the centralized scheme is compared to decentralized scheme using Fixed Maximum Limit distribution scheme and in Section 5.4.2 few variants of decentralized DBAs (CRR, FMM) are compared among themselves.

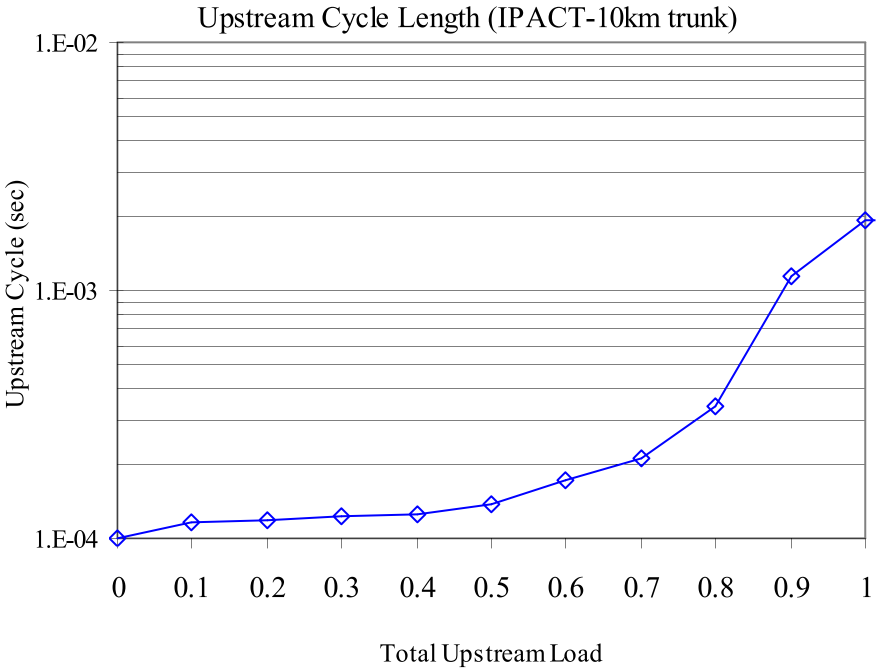


Figure 5.3: Upstream cycle time of centralized scheme

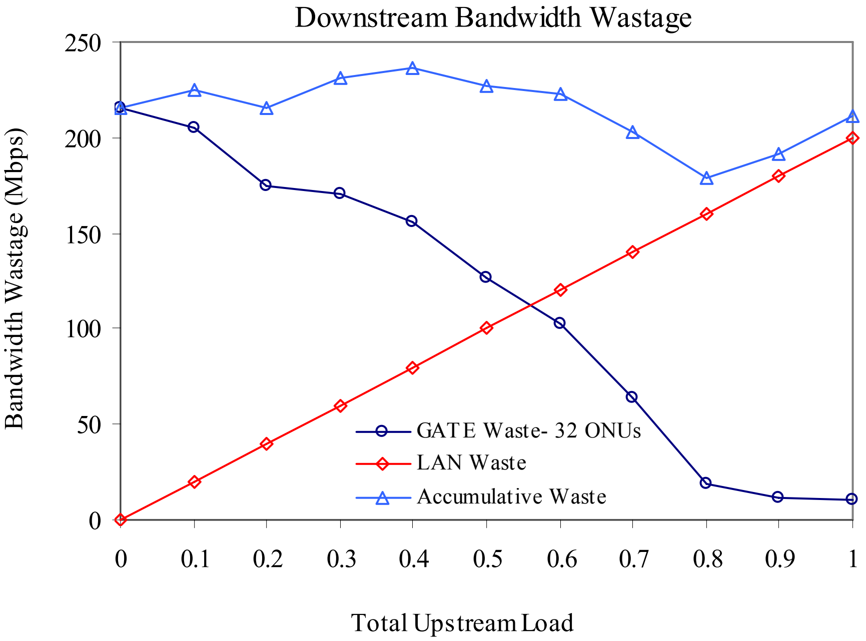


Figure 5.4: Centralized downstream bandwidth wastage

5.4.1 Performance of Centralized vs. Decentralized Scheme

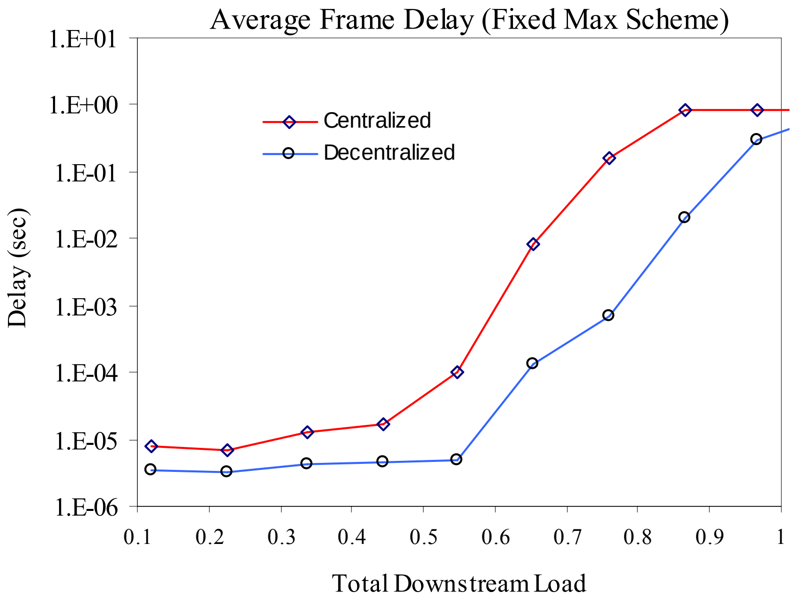


Figure 5.5: Centralized vs. Decentralized: frame queuing delay

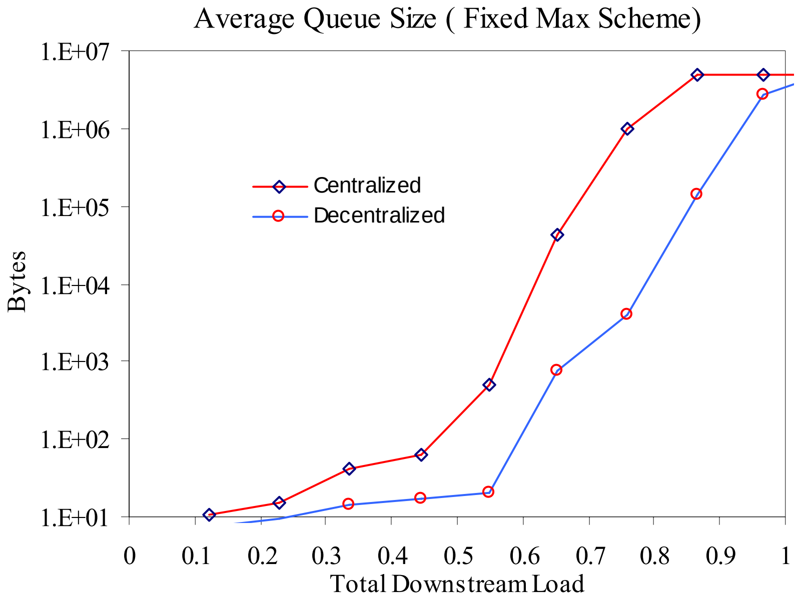


Figure 5.6: Centralized vs. Decentralized average queue size

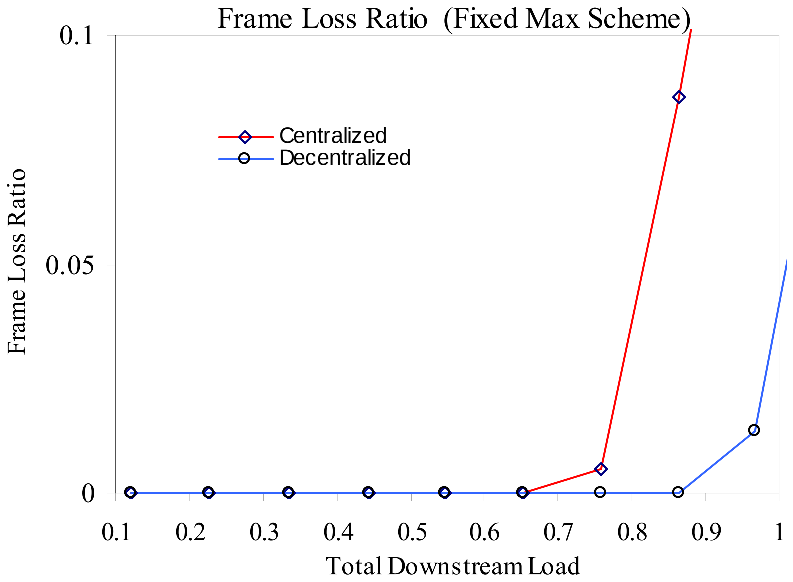


Figure 5.7: Centralized vs. Decentralized frame loss ratio

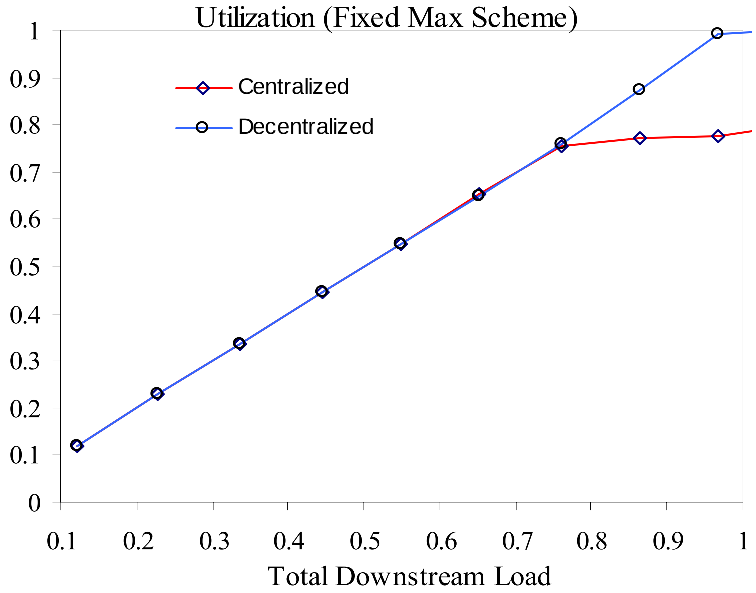


Figure 5.8: Centralized vs. Decentralized utilization

In centralized scheme the entire downstream bandwidth is not available for use and that directs us to determine the available downstream bandwidth in centralized scheme. Total reduction in available bandwidth is constituted by the GATE and LAN traffic. Since the

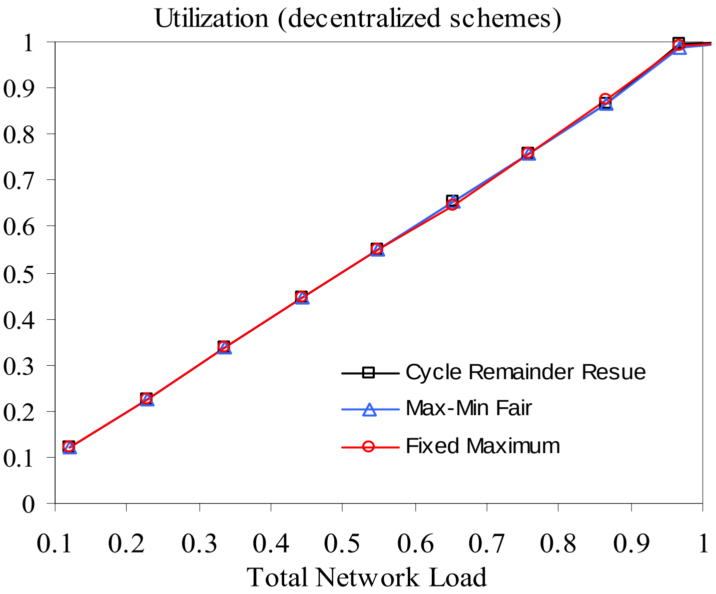


Figure 5.9: Decentralized schemes: utilization

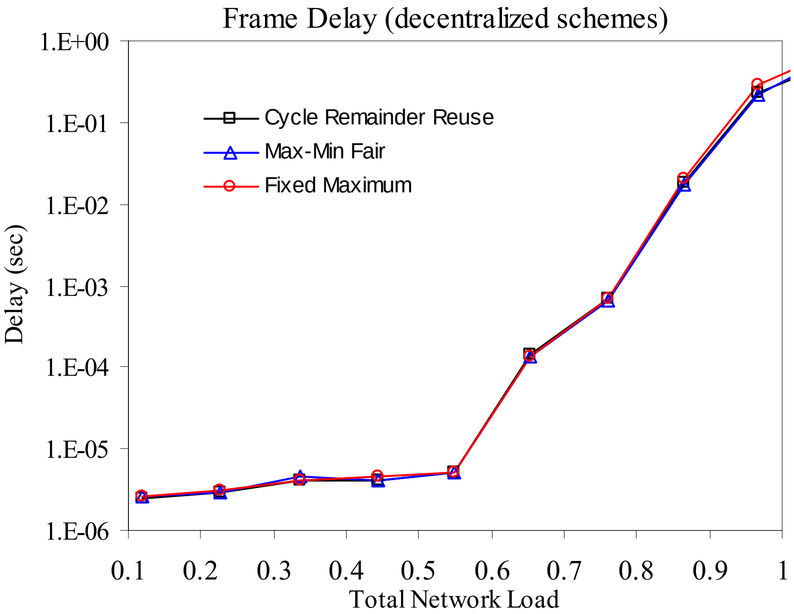


Figure 5.10: Decentralized schemes: frame queuing delay

GATE message wastage is a function of upstream cycle length, we show the upstream cycle time as a function of TNL in Fig. 5.3. It shows that at the lowest TNL the cycle length is the shortest and at the highest TNL the cycle length is the longest. Note that shorter cycle length translates to more cycles per second and longer cycle length translates to less cycles per second. Each cycle generates GATE messages equivalent to the number of ONUs. Thus, low TNL requires frequent GATE messages, resulting in maximum GATE loss and on the contrary, at high TNL number of GATE messages decrease, causing lesser GATE loss. Therefore, GATE loss is inversely proportional to the cycle length. Conversely, the LAN loss is proportional to the cycle length/TNL. The LAN traffic load increases as the TNL increases causing more downstream bandwidth wastage. Fig. 5.4 depicts the centralized scheme's downstream bandwidth wastage due to GATE message, LAN traffic, as well as the accumulative wastage due to both losses. As we see in the figure, regardless of upstream network load, the minimum downstream bandwidth wastage is over 210Mbps for the centralized architecture with LAN emulation.

Fig 5.5 shows the average frame queuing delay for centralized and decentralized scheme. The centralized delay is always longer than the decentralized one. At low downstream TNL, the effective downstream bandwidth in centralized scheme is at least 20% less than the decentralized scheme. Note that it is still adequate to serve the queues at low load. But the OLT's additional GATE and LAN function in centralized scheme causes the downstream cycle to get longer than decentralized scheme. The consequence is comparatively longer buffer stay; resulting in longer queuing delay in centralized network in spite of adequacy of bandwidth. On the other hand, at higher TNL, the centralized scheme suffers from lack of available bandwidth; therefore it cannot serve as much traffic in a given cycle as in decentralized scheme. Packets have to wait in queue for next cycle, causing more queuing delay. This longer queue occupancy results in comparatively larger queue size for centralized scheme then that of decentralized scheme (see Fig. 5.6). At higher load the queue sizes get larger and at TNL~1, queues are saturated causing packet drop. Fig 5.7 depicts this fact; since centralized scheme has comparatively larger queue size, it causes more packet drop than decentralized scheme.

Note that the utilization is similar in both cases until the TNL~.75 (Fig. 5.8). Due to low TNL, most of the cycle is unused for both cases. The centralized architecture uses that time for GATE and LAN messages, where as decentralized scheme just waits around. The average usage of bandwidth is similar, until the demand gets higher and centralize scheme cannot serve as much as the decentralized scheme due to over 20% reduction of available downstream bandwidth. This reduction of bandwidth confines the centralized utilization to below 80% at high TNL.

5.4.2 Performance of various Decentralized Schemes

In this section, we show few variants of decentralized schemes and compare their performance. All the three decentralized DBAs use the same available bandwidth, but difference is in distribution scheme. Their performance is same as far as the utilization is concern (Fig. 5.9). Because in all three cases, same amount of traffic is served in a given time. As far as the delay is concerned, the FML has comparatively little more delay at higher load. It is because, FML has a maximum limit of allowable bandwidth for a queue in each cycle. When there is unused cycle bandwidth (due to low demand ONUs), which is not given to high demand ONUs in that cycle, it causes extra delay for those queues that have to wait for another cycle to transmit. Due to this additional delay, at high load,

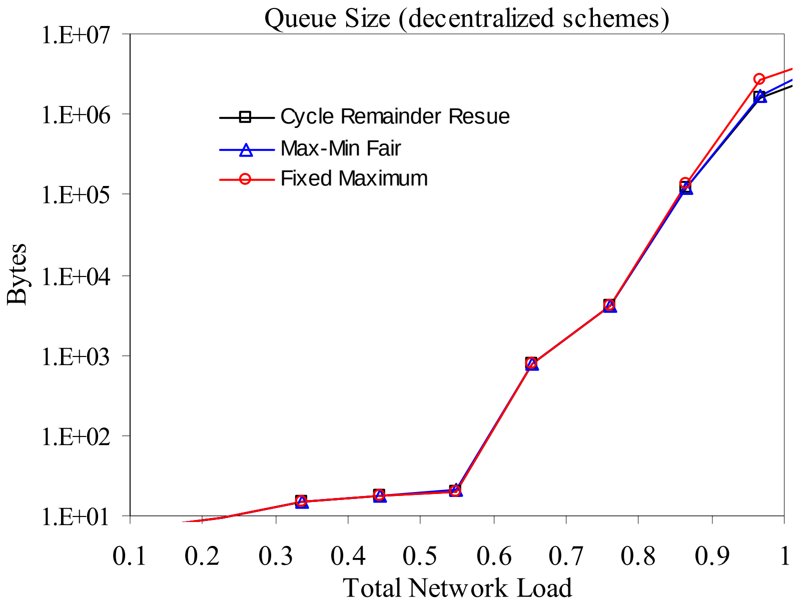


Figure 5.11: Decentralized schemes: queue size

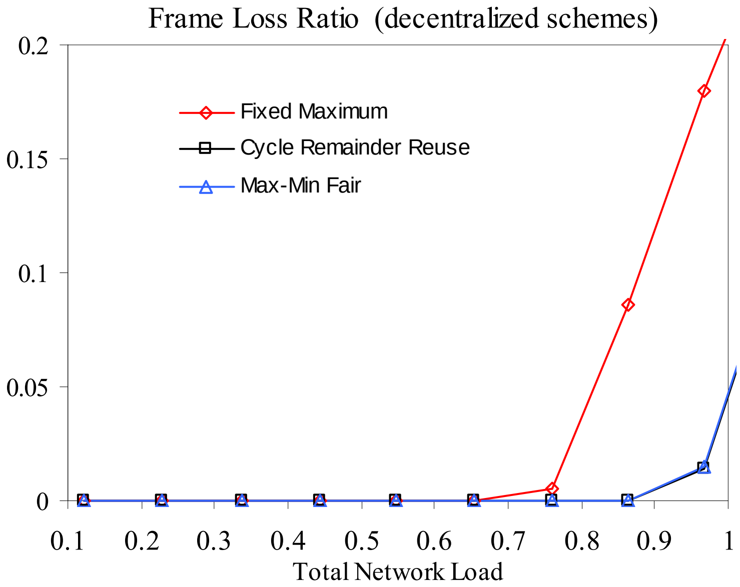


Figure 5.12: Decentralized schemes: frame loss ratio

the queue size and frame loss for FML DBA is little higher than other two schemes (Fig.5.10-5.12), where queues can transmit more than their maximum limit through redistribution of unused cycle bandwidth. In another word, due to global optimization in distribution of bandwidth, the CRR and FMM have slightly better performance.

5.5 Conclusion

In term of downstream bandwidth allocation, we pointed out the disadvantages of centralized architecture with LAN support. The available downstream bandwidth in the centralized scheme is reduced due to LAN support and upstream GRANT generation. The reduction of downstream bandwidth results in poor downstream performance. The proposed decentralized scheme supports distributed upstream DBA as well as LAN, where OLT is excluded from these functions; it results in burden free OLT operation in downstream direction with 100% available bandwidth. This advantage of the decentralized scheme demonstrated better performance in downstream queuing delay, utilization and packet loss ratio.

Chapter 6

Distributed Failure Detection and Recovery

6.1 Introduction

Since EPON brings numerous advantages to hundreds of subscribers, it is imperative that EPON architecture is efficiently protected from any failure. PON architectures are traditionally deployed as tree topologies, which provides simplicity in the deployment of access networks. However, tree-based topologies has several inherent drawbacks; one of them is lack of a simple and cost-effective protection and/or restoration capabilities against failures in the distribution network. The mainstream PON DBA and protection schemes have been centralized – relying on a component at the distant OLT to arbitrate upstream transmission and to detect and recover distribution and trunk fiber breaks [1-9]. The centralized processes of upstream bandwidth allocation as well as detecting and restoring distribution fiber breaks at the distant OLT are lengthy and complex processes and require many changes at each ONU [10].

To improve the reliability of PONs, optical layer protection schemes including Automatic Protection Switching (APS) are used to handle fiber failures [5-9]. In APS schemes, the optical path of the transported signals is switched to a predetermined path upon detection of either a fiber or an equipment failure [11-12]. ITU-T G.983.1 recommended four possible protection schemes, which duplicate fibers and equipment at the ONUs and OLT [8]. These schemes are all centralized and can significantly alter the cost-effectiveness of the PONs since they require many redundant components as well as many spare fibers connections to each ONU [9-10]. Recently, the authors in [10] demonstrated an APS scheme in conjunction with two different LAN-emulation techniques, where the detection and subsequent recovery of a distribution fiber break is carried out at the affected ONU. In addition to the required fiber Bragg grating and/or secondary distribution fibers for the redirection of the upstream signals back to the ONUs, adjacent ONUs are interconnect using additional fiber via the optical switches. This further increases both the cost and complexity of the system. Finally, those proposed schemes can only protect against distribution fiber breaks. Note that a distribution fiber break in the tree architecture affects only an ONU; contrary to that, in a ring based architecture (as in Chapter 3), any such failure will cripple the entire network. Therefore, the significance of protection in a ring-based architecture carries even more emphasis.

This chapter proposes a 2-fiber self healing ring-based local access PON architecture that addresses some of the limitations of current tree-based PON architectures including supporting private networking capability as well as providing a simple and cost-effective fully distributed resilience capabilities against any and all kinds of networking failures. Specifically, this work proposes and devises a simple self-healing EPON architecture that supports a truly shared LAN capability among end users [10, 13-16]. The proposed decentralized APS scheme is capable of protecting against both node (ONU) and distribution/trunk fiber failures. The scheme enables the restoration of all network traffic including upstream, downstream, and LAN data. In addition, the APS scheme can also protect against any combination of concurrent double failures including trunk/distribution fiber breaks and node failures. Note we previously presented ring-based centralized protection architecture in [17], which leads to the evolution of the proposed distributed protection architecture.

Unlike a typical ring-based PON topology in which the OLT and ONUs are distributed around a metro fiber ring, the proposed architecture interconnects ONUs via a short distribution fiber ring in the local loop but allow them to share a standard trunk fiber for long reach connectivity to the OLT [16]. This minimizes fiber deployment in both the CO and the local loop. This architecture is well suited for an autonomous access environment such as a university campus or a private corporation where several buildings are closely dispersed within a 1-2 km diameter area.

The main characteristic of the proposed architecture is that it supports a fully distributed control plane among the ONUs for ONU-ONU communication. The control plane supports fully distributed fault detection and recovery mechanisms as well as a decentralized DBA scheme in which the OLT is excluded from both the arbitration and detection/recovery processes. In the proposed scheme, ONUs exchange signaling and control information (REPORT messages) including the state of nodes and distribution/ trunk fibers as well as their queue status and transmission needs. Each ONU has an APS module that monitors the state of its adjacent distribution fiber paths and its own state via the presence or absence of upstream signal. Thus, detection of most networking fault scenarios and subsequent protection switching are initiated and managed independently by the affected ONU in a distributed manner.

The simulation results indicate that the performance of the proposed distributed recovery scheme, including traffic loss and restoration speed, outperforms that of a typical centralized scheme. Furthermore, the recovery time associated with any and all different distribution network/trunk failures is still within the delay-bound limit required for delivering guaranteed triple play services. In addition to the added flexibility and reliability of a distributed scheme, the proposed architecture eliminates the OLT's centralized task of failure detection and subsequent recovery scenarios as well as the task of processing requests and generating grants (GATE messages) for bandwidth allocations. This reduces the additional processing complexities and delays at the OLT.

The rest of the chapter is organized as follows: Section 6.2 gives brief overview of the initial ring-based centralized protection architecture, which lead to the evolution of the proposed distributed protection architecture. Section 6.3 describes the proposed distributed self-healing architecture. Section 6.4 presents the distributed fault detection and recovery mechanisms. Section 6.5 presents concurrent double failures recovery process. Section 6.6 discusses the scalability of the proposed architecture. Section 6.7 provides analysis of the recovery time. Section 6.8 presents simulation results and Section 6.9 offers some concluding remarks.

6.2 Ring-Based Centralized Protection Architecture

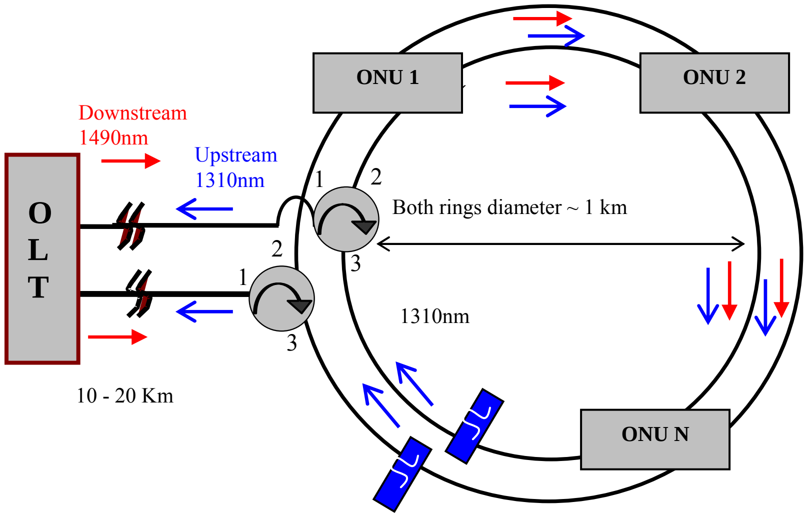


Figure 6.1: Centralized protection architecture

In this section, we briefly describe the overview of the original centralized ring-based protection architecture, which leads to the evolution of the proposed distributed protection architecture. This protection architecture is applicable to the ring-based centralized architecture briefly discussed in Section 3.2 of Chapter 3. We employ the 1+1 protection in the fiber network. Advantage of 1+1 over 1:1 is that the traffic is

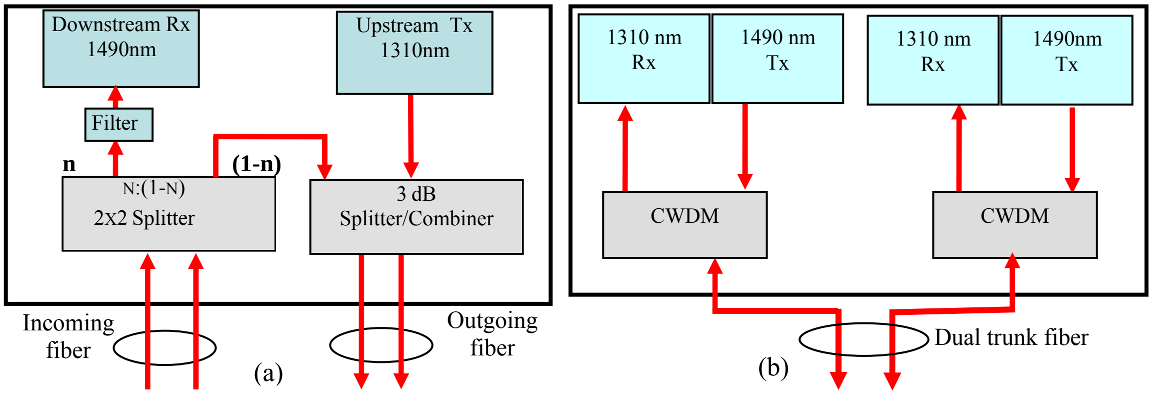


Figure 6.2: Protection architectures of (a) ONU (b) OLT

always running on both fibers concurrently. Thus, no switching is required in a situation of a fiber cut and there is no delay or packet loss associated with this type of failure detection. Furthermore, it is widely deployed and simple to implement. Along with additional transmitters and receivers in OLT, it will require twice the fiber required for the single ring operational architecture. Thus, this ring-based architecture requires redundancy of the following elements: trunk fiber, ring fiber, OLT PON interface and circulator (Fig. 6.1 and 6.2). Table 6.1 shows the comparative fiber requirement of ring-based centralized architecture with others. The ring architecture requires less fiber than other architectures.

Table 6.1: Comparative fiber usage in protected arhitectures

D=t+s (maximum distance between OLT and ONU)	**Tree t+N*s**	**Conventional Ring 2*pi*r**	**Novel Ring T+ 2*pi*r**
10+1=11	2*26=52	33*2 =66	13*2=26
15+1=16	2*31=62	50*2=100	19*2=38
20+1=21	2*36=72	65*2=130	24*2=48

ONUs are connected to one incoming and outgoing ring fiber in a single ring operation. But for 1+1 protection, ONUs are connected to two incoming and outgoing fibers, which requires internal modification of ONUs; rest of the architecture are almost similar to single ring operation (Section 3.2). The two incoming ring fibers are connected to ONU’s 2x2 coupler/splitter and the 10% of that combined (3dB gain) signal (“Drop component”) is dropped for ONU downstream receiver. The 90% of that signal (“Go component”) is coupled with ONU upstream transmitter output through a 2x2 passive splitter/coupler. Then the combined signal splits (3dB loss) and exits through the two outgoing rings (Fig. 6.2a) towards the next ONU. Signals travel through all ONUs in a drop-and-go manner up to the end of the rings, where the filters terminate downstream traffic (1490 nm) from going back to OLT. Only the upstream (1310 nm) traffic passes through the circulators to two independent trunk fibers towards OLT receivers. In single ring operation, OLT has one transmitter (1490 nm) and one receiver (1310 nm). Alike ONU, the OLT is also modified for protection scheme. Since OLT is the center of intelligence, there is redundant transmitter and receiver in OLT; traffic is independently transmitted from them into the two trunk fibers (Fig. 6.2b). Traffic of two separate trunk fibers travel towards the circulators attached to the end of trunk to enter the two independent rings connecting all the ONUs (Fig. 6.1).

ONUs periodically send REPORT messages to OLT stating their queue information. In return, OLT sends GATE messages to ONUs granting transmission time slot. This flow of messages allows detection of any problem in the line. In a trunk failure, OLT receives no upstream signal on that trunk. In a ring failure, OLT receives 3dB less power from all the ONUs up to the failure. Finally, in case of an ONU upstream transmitter failure, OLT receives no report from that ONU. Based upon these indications OLT detects failure without any loss of traffic. For details of failure detection scheme refer to [17].

6.3 Proposed Self-Healing Architecture

Figure 6.3 illustrates the proposed self-healing distributed ring-based EPON architecture. The solid lines represent the normal state architecture while the dotted lines represent the redundant protection components. The protected architecture is identical to that of the normal working architecture except for the following additional components (dotted lines): i) a redundant short distribution fiber ring and a trunk fiber; ii) a redundant transceiver pair located at the OLT; and iii), an APS module attached to each ONU.

6.3.1 Overview of Normal State Operation

As can be seen from Fig. 6.3 (solid lines only), an OLT is connected to N number of ONUs via a 20 km trunk feeder fiber, a passive 3 port optical circulator, and a short distribution fiber ring [16]. To cover the same local access area as that covered by a conventional tree-based architecture, the small ring at the end of the trunk is assumed to have a 1-2 km diameter. The set of ONUs are joined by point-to-point links in a closed loop. The links are unidirectional: both downstream and upstream signals (combined signal) are transmitted in one direction only. The top of Fig. 6.3 shows detailed ONU architecture. Each ONU attaches to the ring at a (n: 1-n) 1x2 passive star coupler (incoming signal at point A in Fig. 6.3) and can transmit data onto the ring through the output port of a 2x1 CWDM combiner (outgoing signal at point E in Fig. 6.3). Note that in addition to the conventional transceiver maintained at each ONU (a λ_{up} upstream transmitter and a λ_d downstream receiver), this approach requires an extra receiver tuned at λ_{up}.

Downstream signal is coupled to the ring at port 2 of the optical circulator. After recombining with the re-circulated upstream signal via another 2x1 CWDM combiner (Fig. 6.3) placed on the ring directly after the optical circulator, the combined signal then circulates around the ring (ONU 1 through ONU N) in a Drop-and-Go fashion. The downstream signal is then removed at the end of the ring using a filter (located directly after the last ONU) that passes only the 1310 nm upstream signal. The upstream signal emerging from the filter at the end of the ring is split into two components via a 1x2 passive splitter (Fig. 6.3) placed on the ring directly after the filter. The first component is

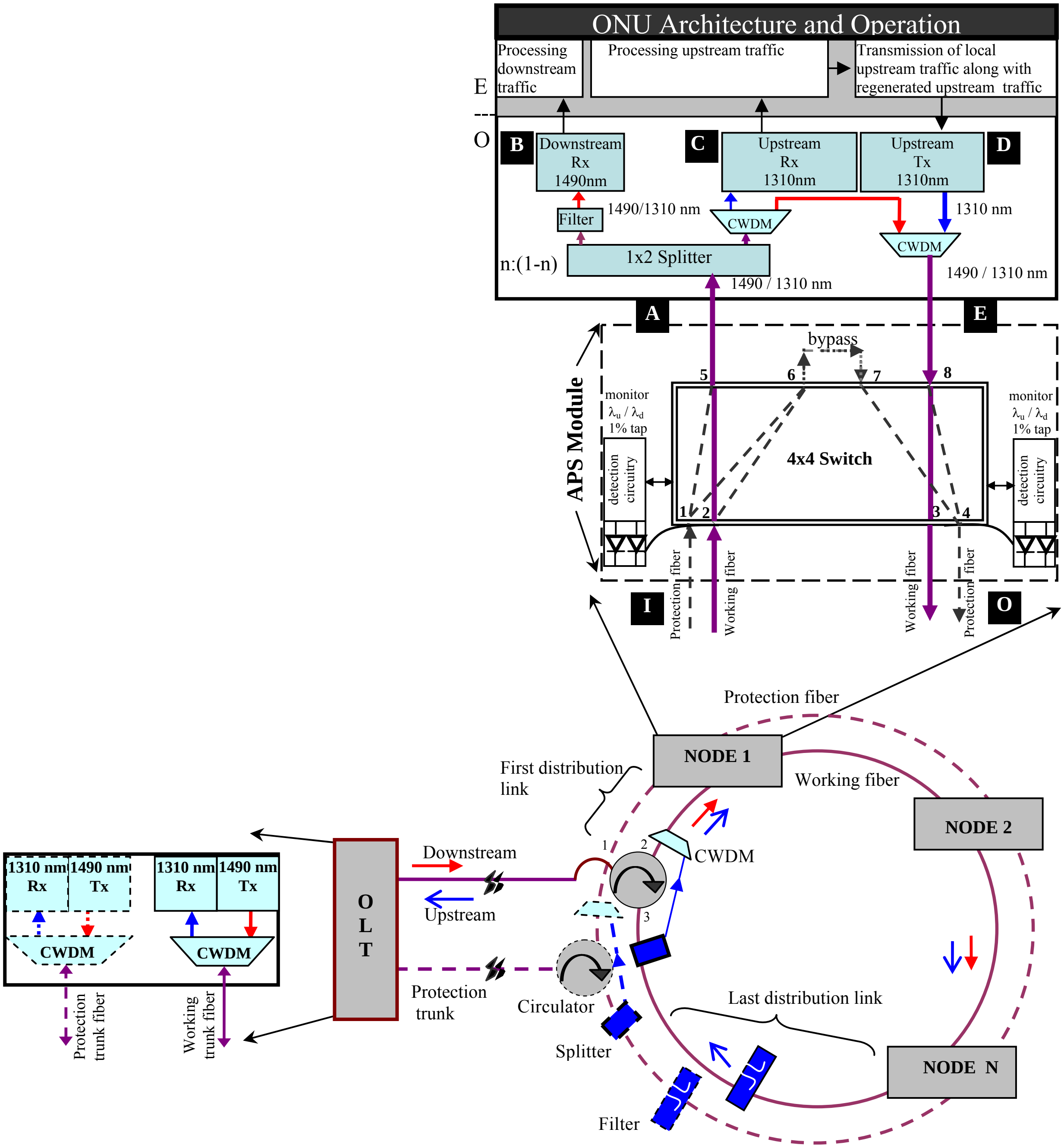


Figure 6.3: Protected network

directed towards the OLT via circulator ports 3 and 1, while the second component is allowed to re-circulate around the ring after recombining with the downstream signal (originating from the OLT) via the 2x1 CWDM combiner.

The (n: 1-n) 1x2 coupler (n is a small arbitrary percentage assumed here to be 10%) splits the incoming combined signal at each node into a small (10%) “Drop-signal-portion” and a large (90%) “Go-signal-portion”. The small portion of the circulating combined signal dropped at each node (Drop-signal) is passed through a filter that removes the upstream signal and passes only the downstream broadcast signal, which is then received and processed by the 1490 nm downstream receiver. The remaining portion of the combined signal emerging from the 90% coupler’s port (Go-signal) is first separated into its two constituent: downstream and upstream signals via a CWDM filter. The separated upstream signal (second component) is received and processed via the 1310 nm upstream optical receiver housed at the ONU, where it is then regenerated and retransmitted along with the ONU’s own local control and data traffic. Finally, the separated downstream signal is re-combined again with the retransmitted upstream signal (regenerated plus local) via the 2x1 CWDM to form the outgoing combined signal (incoming signal for next ONU) that circulates around the ring.

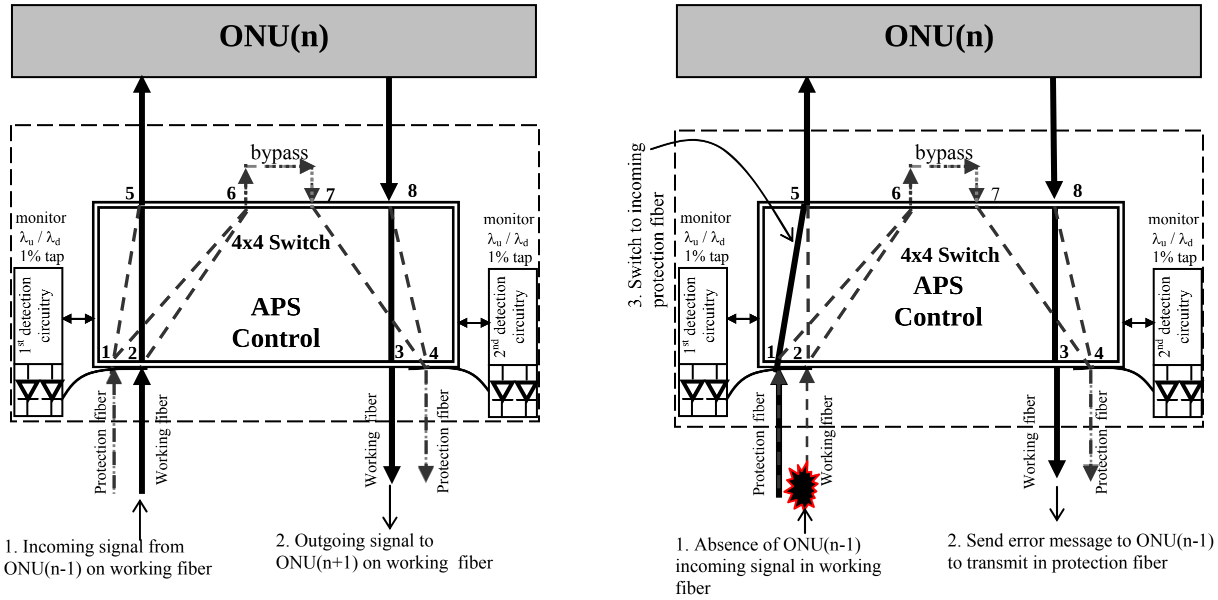
Since upstream transmission is based on a TDMA scheme, inter-ONU traffic (LAN data and control messages exchanged among ONUs) is transmitted along with upstream traffic destined to the OLT (MAN/WAN data) within the same pre-assigned time slot. The first component of the upstream signal destined to the OLT is received and processed by the

1310 nm upstream optical receiver (housed at the OLT), which accepts only MAN/WAN traffic, discards LAN traffic, and process the control messages. On the other hand, the second component of upstream signal is transmitted sequentially, bit by bit, around the ring from one node to the next where it is regenerated and retransmitted at each node.

Since the ring is a closed loop, upstream traffic will circulate indefinitely unless removed. The process of removing, regenerating and retransmitting the second component of the upstream signal at each node (ONU) is implemented as follows: first, the 1310 nm upstream optical receiver (housed at each ONU) terminates all upstream traffic, examines the destination MAC address of each detected Ethernet frame, and then performs one or more of the following functions: (1) the source node removes its own transmitted frames that complete one trip around the ring through re-circulation; (2) once the destination address of the LAN traffic matches the node's MAC address, it is copied and delivered to the end users; (3) all upstream traffic (including LAN and control frames) is processed, regenerated, and then retransmitted by each node (except those that matches functions 1 and 2 above).

6.3.2 Protected State Operation

The APS module attached to each ONU is the basic building block of the proposed self-healing mechanisms that performs both fault detection and automatic switching processes. The APS module connects to both incoming and outgoing working and protection fibers. Each APS module houses a commercially available low loss 4x4



1. Incoming signal from ONU(n-1) on working fiber

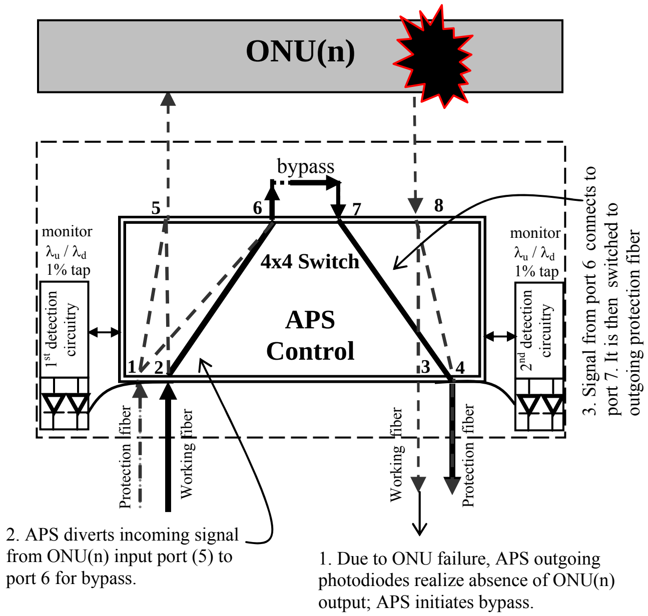
2. Outgoing signal to ONU(n+1) on working fiber

1. Absence of ONU(n-1) incoming signal in working fiber

2. Send error message to ONU(n-1) to transmit in protection fiber

(a) Normal Operation

(b) Incoming fiber failure



3. Signal from port 6 connects to port 7. It is then switched to outgoing protection fiber

2. APS diverts incoming signal from ONU(n) input port (5) to port 6 for bypass.

1. Due to ONU failure, APS outgoing photodiodes realize absence of ONU(n) output; APS initiates bypass.

(c) ONU failure (bypass operation)

Figure 6.4: ONU(n) APS switching scenarios

bidirectional Optical Switch (OS) that is capable of switching from any port to any port [18]. It also includes two detection circuits, where each circuit comprises a 1x2 CWDM filter, two p-i-n detectors, and a control circuit to configure the OS. The first detection circuit is used to detect both incoming upstream and downstream signals via taping a small portion (about 1%) of the incoming combined signal. Likewise, the second detection circuit is used to detect both outgoing upstream and downstream signals via taping a small portion (about 1%) of the outgoing combined signal. As can be seen from Fig. 6.3 and 6.4a, under normal operation, the combined signal (both downstream and upstream signals) traverses the incoming and outgoing working fibers via ports 2-5 and 8-3, respectively.

In general, we classify failure scenarios into three different classes, a trunk link failure, a general distribution link failure and a general node (ONU) failure. A general distribution link is defined here as a fiber segment that connects two adjacent ONUs. All links connecting adjacent ONUs are general distribution links except the following two special links, which require different detection and recovery mechanisms: 1) the distribution fiber segment that connects the first ONU (ONU_1) and the circulator; this link will be referred here as the first link; 2) the distribution fiber segment that connects the last ONU (ONU_N) and the circulator; this link will be referred here as the last link. All nodes distributed around the ring are general nodes except the last ONU, which requires only different recovery mechanisms. As will be shown below, all links and nodes that are at the trunk-ring junction (trunk, first and last links, first and last ONUs) have special significance.

Figure 6.4 details various switching scenarios. In the case of a general link failure, the first control circuit of the affected node (the ONU that detects and manages the failure) configures the OS such that the incoming signal is switched from the working fiber to the protection fiber via ports 1-5 of the affected ONU's OS (Fig. 6.4b). In the case of a general node failure, the incoming signal totally bypasses the failed node and traverses to the next node (Fig. 6.4c). This is achieved by switching the incoming signal directly to the outgoing protection fiber of the failed node via ports 2-6-7-4.

6.3.3 Synchronization & Resynchronization

Under normal operation, all ONUs are synchronized to a common reference clock extracted from OLT's 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 [2-3]. The TDM controller at 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.

In the even of a failure, normal ONUs transmissions stop and the sync between ONUs and OLT is lost. This causes granted timeslots of that cycle to be nullified. Once the failure recovery is complete, a new cycle's grants must be recalculated via the DBA module located at each ONU. This would require establishment of resynchronization

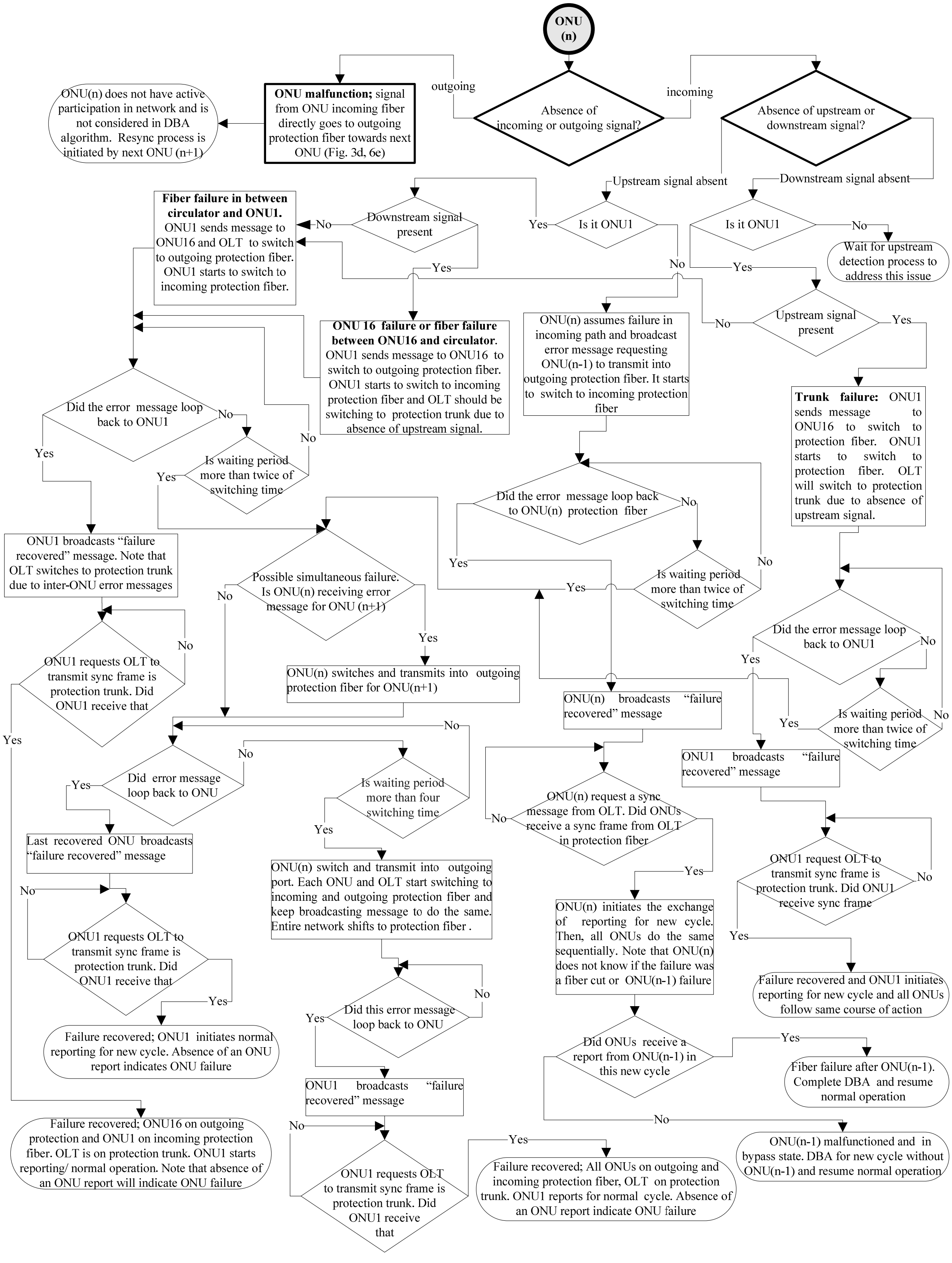


Figure 6.5: Overall failure detection and recovery algorithm of ONU

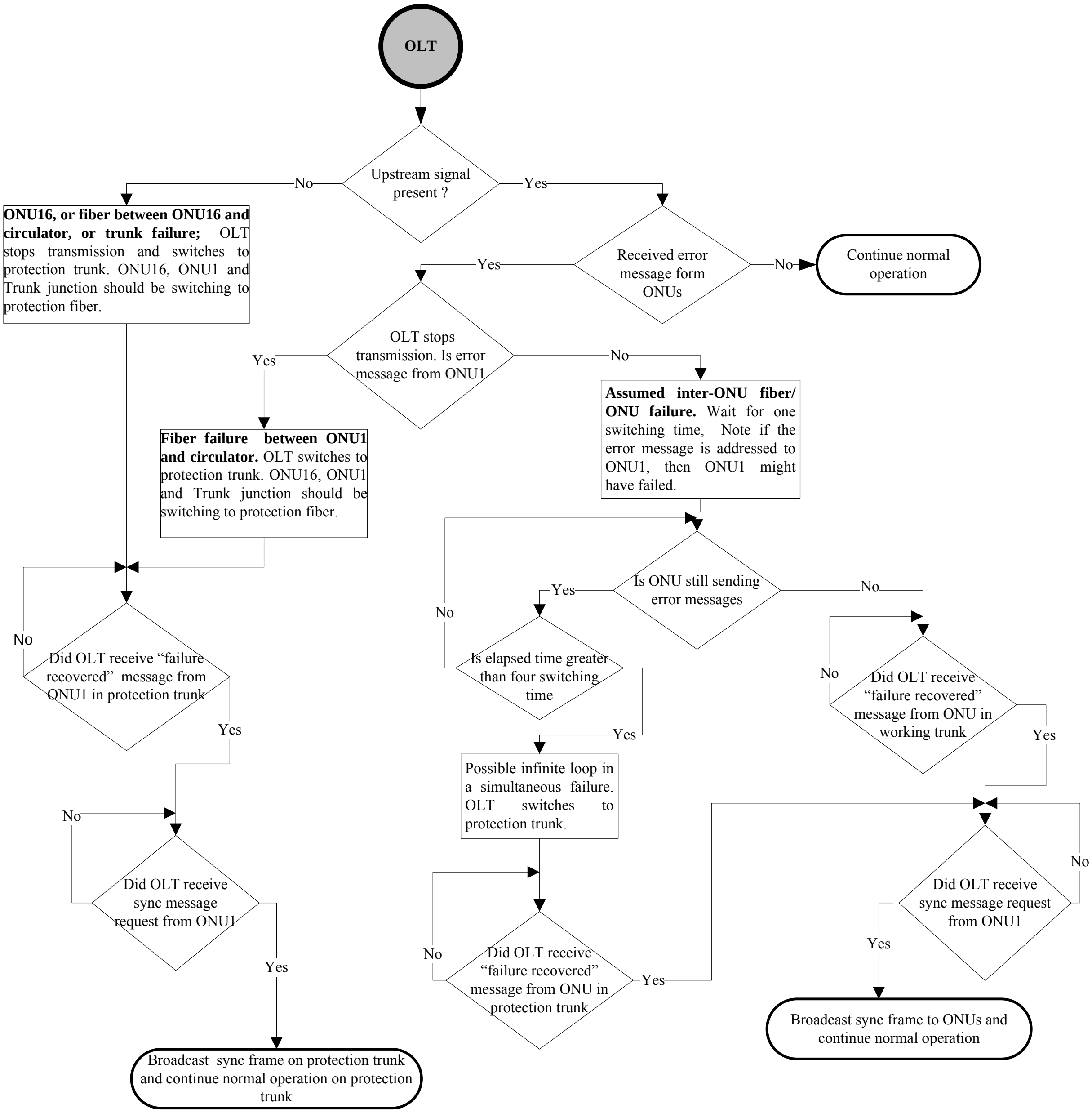


Figure 6.6: Overall failure detection and recovery algorithm of OLT

between ONUs and OLT again. As will be shown below, this is accomplished as follows: once affected node (ONU_n) confirms recovery of failure, it floods the network with a second REPORT message requesting resynchronization frames from OLT. Once ONU_n receives resync frames from OLT, it initiates a new cycle by transmitting its normal REPORT to all other ONUs. Then, all ONUs sequentially send their REPORTs; once all reports are exchanged for DBA calculation of the new cycle, new grants are calculated and normal operation resumes.

6.4 Fully Distributed Fault Detection and Recovery Mechanisms

6.4.1 Decentralized Fault Detection Mechanisms

The proposed distributed scheme utilizes a time division multiple access (TDMA) arbitration scheme in which the OLT is excluded from the arbitration process. It assumes a cycle-based upstream link, where the cycle size can be either fixed or variable length confined within certain lower and upper bounds to accommodate the dynamic upstream traffic conditions. During each cycle, the order of ONUs transmission is assumed to be fixed (sequential in an ascending order, i.e., ONU_1, ONU_2,..ONU_N). Each ONU transmits its control (REPORT) message (along with both LAN and upstream data) within its pre-assigned cycle's time slot around the ring from one node to the next, where it is finally removed by the source ONU after making one trip around the ring. Since the REPORT messages are processed and retransmitted at each node, ONUs can directly communicate their status and exchange signaling and control message information with one another.

Under normal operation, the REPORT typically contains the desired size of next timeslot based on the current ONU's buffer occupancy. ONUs exchange REPORT messages concerning their queue status and their transmission needs amongst themselves. Then, ONUs sequentially and independently run instances of the same DBA algorithm outputting identical bandwidth allocation results [16]. Once the algorithm is run, the ONUs sequentially and orderly transmit their data including both LAN and MAN/WAN traffic without any collisions, eliminating the OLT's centralized task of processing requests and generating grants for bandwidth allocations.

Table 6.2: ONU1 failure detection table

Special Fiber Failure	Upstream signal	Downstream Signal
Trunk	√	X
After the last ONU	X	√
Before the first ONU	X	X

In the event of a failure scenario, the REPORT typically contains a failure indication alarm message that includes specific instructions to both the OLT and a remote node that will be involved in the recovery process. Figures 6.5 and 6.6 depict the over all failure detection and recovery algorithm of ONU and OLT respectively. Since upstream signal is always present on the ring and the trunk (cyclic control message is always transmitted independent of the presence or absence of LAN/WAN data), general failure detection scenarios (general distribution link and node failures) will primarily be based on detecting absence/presence of upstream signal only. Thus, all ONUs are continuously monitoring the status of upstream signal on both incoming and outgoing fibers. If the first control circuit of a given ONU_n detects the absence of upstream signal on its incoming working fiber, a general distribution link failure is assumed. This is the link that interconnects ONU_{n-1} and ONU_n.

On the other hand, each of the three special links (trunk link, first and last distribution links) that are the junction of OLT, ONU_1, and ONU_N requires its own different failure detection mechanism. As shown in table 6.2, all three failure scenario are detected and managed by the first ONU and each requires monitoring both downstream and upstream signals. Thus, the first ONU is the only node that is required to monitor both upstream and downstream signals.

If the first control circuit of a given ONU_n detects the presence of upstream signal on its incoming working fiber while the second control circuit detects absence of same signal (after being processed, regenerated, and retransmitted by ONU_n) on its outgoing working fiber, a node (ONU_n) failure is assumed. While ONU_n detects its own failure (via APS module attached to it), however, as will be shown, managing the failure is delegated to the next node on the ring (ONU_{n+1}). Thus, failure of the last ONU is also managed by the first ONU. This means that, overall; the first ONU manages all four special failure scenarios including the three special links plus last node failures. The detection and recovery process for each and every node on the ring is identical, except for the last node, which requires only different recovery mechanisms.

6.4.2 Recovery Process

The recovery process is implemented via the participation of three cooperating network nodes including the affected node (ONU_n), OLT, and either ONU_{n-1} (in the case of a link failure) or ONU_{n+1} (in the case of a node failure).

6.4.2.1 General Link Recovery

The successful completion of the recovery process of a given general link failure scenario involves the following steps:

1) To avoid false failure detection, once the affected node (for instance ONU_n) detects a given fiber break, it must wait for a predetermined timeout.

2) ONU_n then sequentially or simultaneously performs the following three functions: i) stops upstream traffic (LAN and MAN) transmission; ii) switches to incoming protection fiber (Fig. 6.4b); and iii) floods the network with a failure indication alarm message that includes specific instructions to both ONU_{n-1} (to switch its transmission from outgoing working fiber to outgoing protection fiber) and OLT (to stop downstream transmission).

3) ONU_n keeps flooding the network with the failure message expecting its failure frame to loop back to it via ONU_{n-1}'s outgoing protection fiber.

4) Each and every ONU on the ring that receives the failure message stops all upstream transmission; likewise, OLT stops downstream transmission upon receiving the failure message.

5) Once ONU_n receives back its failure frame (assuming ONU_{n-1} has already switched to outgoing protection fiber), it starts flooding OLT with a second REPORT message requesting downstream resynchronization frames.

6) Once OLT receives resynchronization request from ONU_n, it resumes downstream transmission.

7) Once ONU_n receives resynchronization frames from OLT, it initiates a new cycle (recovery process is now complete) by transmitting its normal REPORT control message to all other ONUs. Then, all ONUs sequentially send their REPORTs; once all reports are

exchanged for DBA calculation of the new cycle, new grants are calculated and normal operation resumes.

6.4.2.2 General Node Recovery

The recovery process of a general node failure involves the following steps:

1) Once the APS module attached to a given node (for instance ONU_n) detects its failure and waits for a timeout, it then reconfigures the OS to the bypass mode by switching the incoming signal directly to the outgoing protection fiber via ports 2-6-7-4 (Fig. 6.4c). Due to its failure, ONU_n can’t broadcast a failure indication message to its adjacent node (ONU_{n+1}) or to any other node.

2) While ONU_n’s APS module is initiating the switching process, ONU_{n+1} will detect the absence of upstream signal on its incoming working fiber and erroneously assumes a distribution fiber break between itself and ONU_n.

3) ONU_{n+1} then starts the process of a general link recovery (7 steps listed above).

4) ONU_{n+1} performs the following three functions: i) stops upstream traffic transmission; ii) switches to incoming protection fiber; and iii) floods the network with a failure indication alarm message that includes specific instructions to both ONU_n (to switch its transmission from outgoing working fiber to outgoing protection fiber) and OLT (to stop downstream transmission).

5) ONU_{n+1} keeps flooding the network with the failure message waiting to receive back its failure frame via ONU_n’s outgoing protection fiber.

6) ONU_n can’t receive or process ONU_{n+1}’s request message. However, ONU_n’s control circuit is already reconfiguring the OS to the bypass mode (switching to the outgoing

protection fiber). In other words, ONU_n's APS module is indirectly implementing the message.

7) Once ONU_{n+1} switches to incoming protection fiber and ONU_n switches to outgoing protection fiber (i. e., ONU_n is bypassed), ONU_{n+1} receives back its failure message and proceeds with steps 5 and 6 of the general link recovery listed above.

8) Once ONU_{n+1} receives resynchronization frames from OLT, it initiates a new cycle by sending its normal REPORT control message to all other ONUs. Then, all ONUs sequentially send their REPORTs. In this case, however, due to ONU_n failure, ONU_{n+1} will not timely receive its REPORT. ONU_{n+1} then realizes that absence of upstream signal is in fact due to ONU_n failure (and is not due to a fiber break) and that it is now in bypass state.

9) ONU_{n+1} starts DBA calculation for the new cycle without ONU_n's REPORT; new grants are calculated and normal operation resumes.

Two comments are in order here. First, note that when ONU_n fails, it is ONU_{n+1} that undertakes the management of the recovery process assuming erroneous failure scenario (fiber break). Second, Note that OLT's role in the detection and recovery process of both general link and node failures has been so far limited to receiving and processing both the failure indication and synchronization messages.

6.4.2.3 Special Links and Node Recovery

The recovery mechanism of each of the four special failure scenarios including the three special links (trunk, first, and last links) as well as last node failure is almost identical and, in each case, requires the participation of all three nodes at the junction (OLT, ONU_1, and ONU_N), where all the involved parties must switch to protection fiber. Note that it is the first ONU that manages all of these four special failure scenarios. All steps associated with the recovery of a general link/node failure described above are also applicable to these four special link/node failure scenarios except that the OLT's role must now be extended to include switching its transmission to the trunk protection fiber. There are two options to implement OLT switching process. The first option is to extend ONU_1's failure indication alarm message to include an additional request to the OLT to switch its transmission to the protection fiber. The second option is that OLT itself must independently implement the switching process.

However, due to loss of connectivity, ONU_1 can't communicate at all with OLT in three cases (trunk, last link and last node failures) out of these four special failure scenarios. This immediately excludes the first option and leaves the second option as the only viable choice. In other words, in each of these three cases, OLT must independently both stops downstream transmission and switches its transmission to the protection fiber. Note that each and every failure scenario of these three cases cuts off the flow of upstream signal to OLT. Thus, in all of these three cases the OLT can independently detect the absence of upstream signal and then automatically stops downstream transmission and switches its transmission to the protection trunk fiber.

The fourth remaining special failure scenario, "first link failure", is the only special case where ONU_1 can communicate with OLT (connectivity is not lost) and, hence, flow of upstream signal to the OLT is not interrupted. In this case, the second option is no longer applicable and OLT must rely on ONU_1's additional request to switch its transmission to the protection trunk fiber.

***6.4.2.3.1 First Link Recovery*:** 1) the control circuit of ONU_1's APS module detects the absence of both downstream and upstream signals and, thus, identifies the failure as a first link failure (see table 6.2 above). 2) ONU_1 then performs the following three functions concurrently: i) stops upstream transmission; ii) switches to incoming protection fiber; and iii) floods the network with an extended failure indication alarm message that includes specific instructions to both ONU_N (to switch its transmission from outgoing working fiber to outgoing protection fiber) and OLT (to both stop downstream transmission and switch its transmission to the protection trunk fiber). 3) Once all involved parties (OLT, ONU_1 and ONU_N) switch to protection fibers, upstream and downstream signal path with OLT is restored.

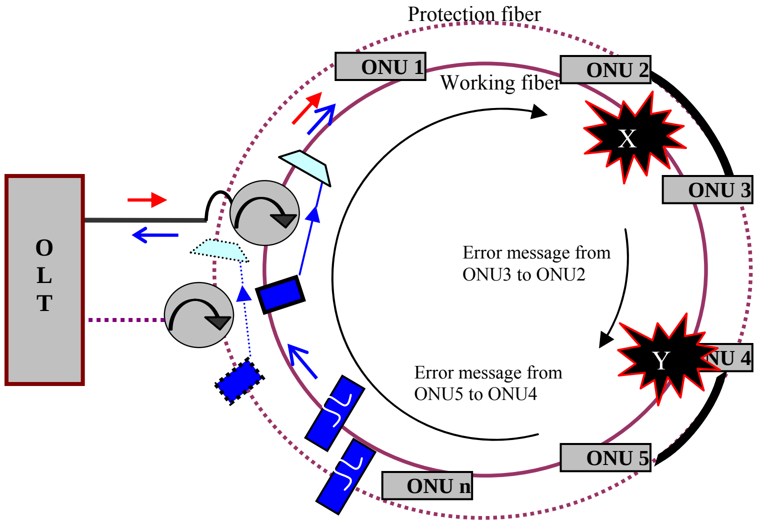
The recovery process of each of the other two special link failures (trunk and last link) is identical to that of the first link except that in step 2-iii above; ONU_1 now floods the network with rather a short failure indication alarm message that includes specific instructions only to ONU_N. The additional command directed to the OLT in the case of the first link failure is now absent and is independently implemented by OLT via detecting absence of upstream signal.

6.4.2.3.2 Last ONU (ONU_N) Recovery: The recovery process of this node is identical to that of a special link failure. This is because, as in the case of a general node failure, when ONU_N fails, ONU_1 undertakes the management of the recovery process assuming erroneous failure scenario (last link failure). Specifically, when ONU_N fails, ONU_1 detects the absence of upstream signal and the presence of downstream signal and, thus, erroneously assumes that the failure is a last link failure (see table 6.2 above). ONU_1 then floods the network with a short failure indication alarm message that includes specific instruction to ONU_N (to switch its transmission from outgoing working fiber to outgoing protection fiber). OLT detects the absence of upstream signal and automatically stops downstream transmission and switches its transmission to the protection fiber.

Although ONU_N can't receive and/or process this message due to its failure, however, as in the case of a general node failure, ONU_N's control circuit is already reconfiguring the OS to the bypass mode (switching to the outgoing protection fiber). In other words, ONU_N's APS module is indirectly implementing the message.

6.5 Concurrent Double Failures Recovery Process

The proposed single failure fault detection and APS recovery mechanisms can also combined to recover from any combination of concurrent double failures including trunk, distribution fiber, and node failures. Depending on the combination of double failures, the recovery time is mainly dominated by switching time and varies from just one switching time (two adjacent ONU failures, trunk and distribution fiber failures) to a maximum of five switching time (two disjoint distribution fiber breaks). Number of illustrative examples shown as follows (Fig. 6.7 to Fig. 6.14).



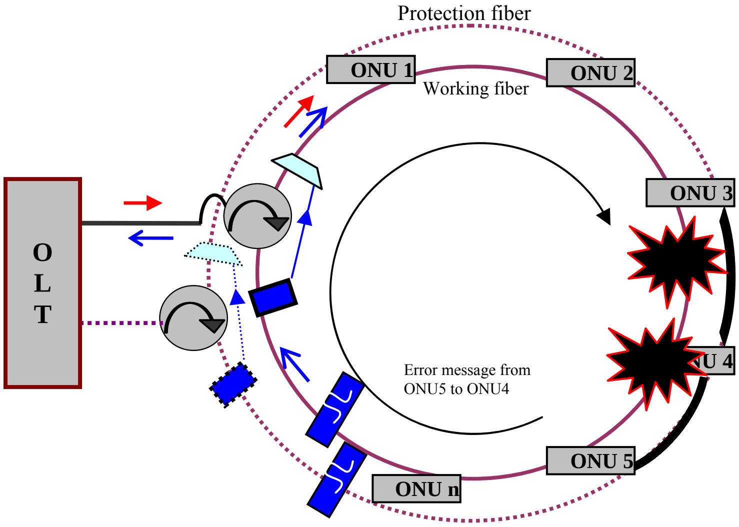
Step1:
-(Y) ONU5 detects absence of upstream signal at incoming fiber and generates error message to ONU4 to switch to outgoing protection fiber and its APS start switching to incoming protection fiber.
-(Y) ONU4 fails to generate outgoing signal. Then APS realizes ONU4 failure and starts to switch to outgoing protection fiber (bypass).
-(X) ONU3 detects absence of upstream signal at incoming fiber and generates error message to ONU2 to switch to outgoing protection fiber and ONU3 APS start switching to incoming protection fiber.

Step2 (after T_{switch}):
-ONU4 is switched to outgoing protection fiber and ONU5 is switched to incoming protection fiber. Thus ONU4 and ONU5 are linked and ONU3's error message reaches ONU2. ONU2 starts switching to outgoing protection fiber.

Step3 (after two T_{switch}):
-ONU2 switched to outgoing protection fiber.
-Optical path is reestablished between ONU2, ONU3 and ONU4, ONU5. Failure recover after two switching time.

Figure 6.7: Distribution fiber break and a disjoint ONU failure

Step1:
-Absence of signal at ONU4 incoming fiber requires ONU4 to generate error message and APS start switching to incoming protection fiber. ONU4 fails to generate error message and then APS realizes ONU4 failure and starts to switch to outgoing protection fiber (bypass).
-Absence of signal at ONU5 incoming fiber causes ONU5 to constantly send error message to ONU 4 to switch to outgoing protection fiber. But message cannot go beyond ONU3 due to simultaneous failure. ONU5 starts to switch to incoming protection fiber.



Step2 (after T_{switch}):
-ONU4 is switched to incoming and outgoing protection fiber and ONU5 is switched to incoming protection fiber . Thus ONU4 and ONU5 are linked. ONU3 keeps receiving error message for ONU4.

Step3 (after two T_{switch} time):
-After waiting two switching time (T_{switch}), ONU3 responds to the error message addressed to ONU4 (assuming failure on outgoing path in a simultaneous failure). ONU3 starts to switch to outgoing protection fiber.

Step4 (after three T_{switch} time):
After the 3 switching time (T_{switch}), the path between ONU3, ONU4 and ONU5 is reestablished in protection fiber; and the error message loops back to ONU5. Failure recovered after 3 switching time.

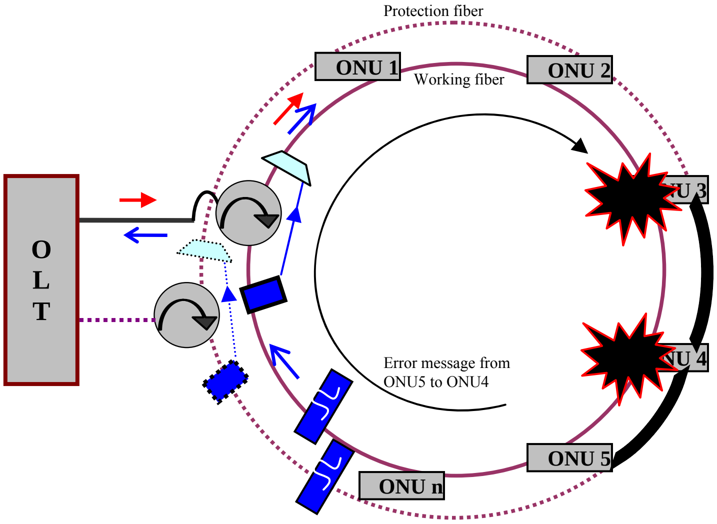
Figure 6.8: Fiber failure and adjacent ONU failure

Step 1

-APS realizes ONU3 failure due to absence of outgoing signal at APS outgoing photodiode and starts to switch to outgoing protection fiber (bypass).

-Absence of signal at ONU4 incoming fiber requires ONU4 to generate error message and APS start switching to incoming protection fiber. ONU4 fails to generate error message and then APS realizes ONU4 failure and starts to switch to outgoing protection fiber (bypass).

-Absence of signal at ONU5 incoming fiber causes ONU5 to constantly send error message to ONU 4 to switch to outgoing protection fiber. But message cannot go beyond ONU3 due to simultaneous failure. ONU5 starts to switch to incoming protection fiber to receive signal from ONU4 (assuming ONU4 will receive its message).

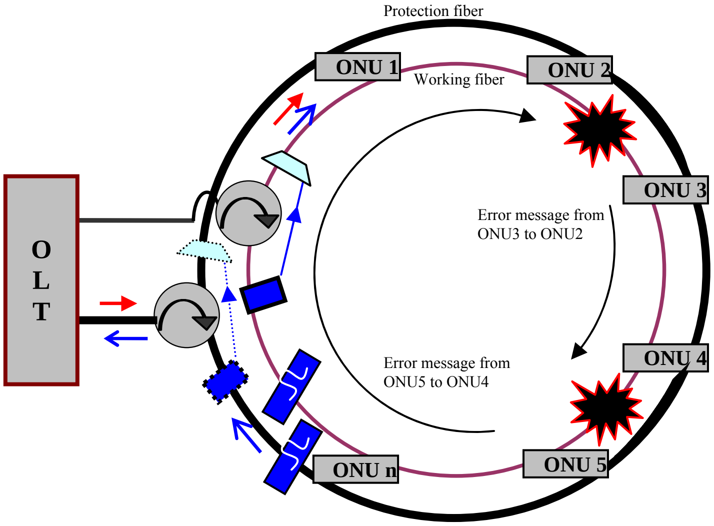


Step 2 (after one switching time)

-ONU 3 and ONU4 are in bypass state (outgoing protection fiber). ONU4 and ONU5 both are on incoming protection fiber.

-Optical path from ONU3 to ONU5 is established. Now the ONU5 error message loops back to source. Failure recovered after one switching time.

Figure 6.9: Two adjacent ONU failures



Step 1:
- Absence of signal at ONU5 and ONU3 incoming fiber compel them to generate consistent error message. APS start switching to their incoming protection fiber. The error message of ONU5 to ONU4 and ONU3 to ONU2 to switch to outgoing protection fiber do not reach their destinations due to simultaneous failure on the path.

Step 2 (after one switching time):
-ONU3 and ONU5 are switched to their incoming protection fiber. They are waiting for their error message to loop back.

Step 3(after four switching time):
-ONU2 and ONU4 never know that they need to transmit into outgoing protection fiber, because they never receive the messages which were intended for them. This infinite loop goes on, until the maximum timeout period for simultaneous failure recovery expires ($4*T_{switch}$). All ONUs keep receiving error message and ONU3 and ONU5 do not receive their loop back message, then all ONUs and OLT realize that it is infinite loop. Now, ALL ONUs switch to incoming and outgoing protection fiber, and OLT switches to protection trunk fiber. It costs one switching time (all at once).
-Now all failed path are re-established. Failure recovered after five switching time.

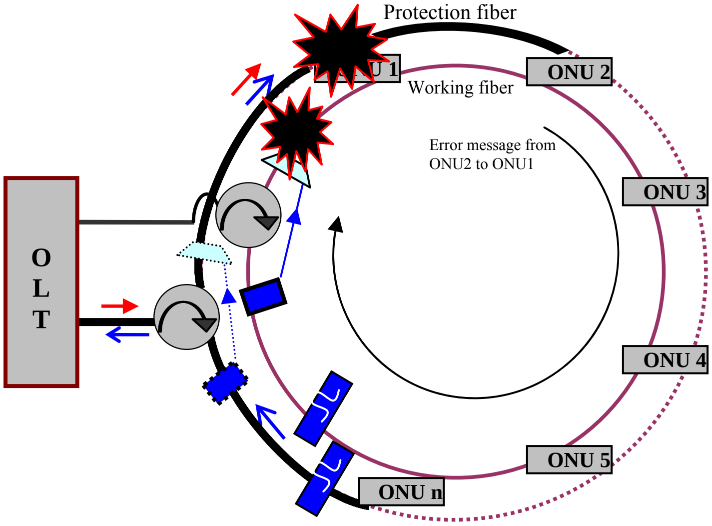
Figure 6.10: Two disjoint fiber failure

Following double failures are of special cases involving the junction of the first ONU, the last ONU (16th) and the trunk.

Step1:

-Due to the absence of both upstream and downstream signal at ONU1 incoming fiber, APS starts switching to incoming protection fiber. Then, ONU1 tries to generate error message but fails. Since ONU1 failed to generate output, APS assumes ONU failure and starts to switch to outgoing protection fiber (bypass).

-Absence of signal at ONU2 incoming fiber causes ONU2 to constantly send error message to ONU 1 to switch to outgoing protection fiber. ONU2 starts to switch to incoming protection fiber.



Step2 (after T_{switch}):

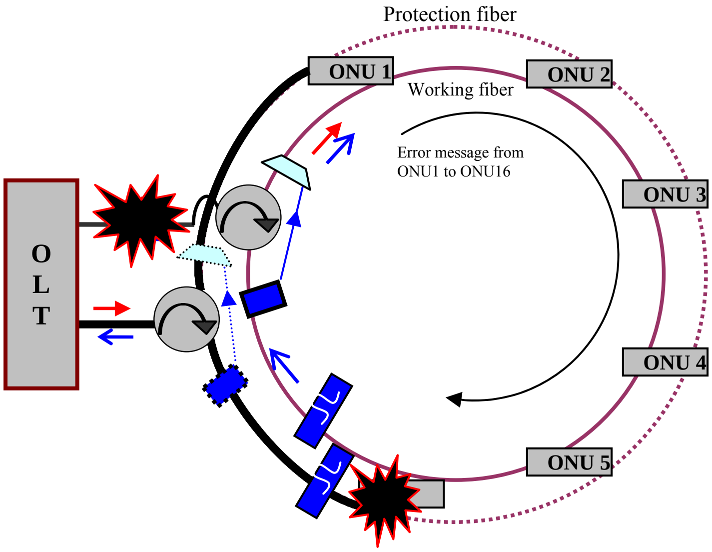
-ONU1 is switched to incoming and outgoing protection fiber and ONU2 is switched to incoming protection fiber. Thus ONU1 and ONU2 are linked. ONU1 cannot generate error message to notify ONU16 to switch to outgoing protection fiber.

-ONU16 keeps receiving error message from ONU2 addressed to ONU1.

Step3 (after two T_{switch} time):

-After waiting two switching time, ONU16 responds to the error message addressed to ONU1 by ONU2 (assuming failure on outgoing path in a simultaneous failure). ONU16 starts to switch to outgoing protection fiber.

Figure 6.11: ONU1 and incoming fiber failure



Step1:

-Absence of both upstream and downstream signal at ONU1 incoming fiber requires ONU1 to generate error message asking ONU16 to switch to protection fiber. ONU1 APS start switching to incoming protection fiber.

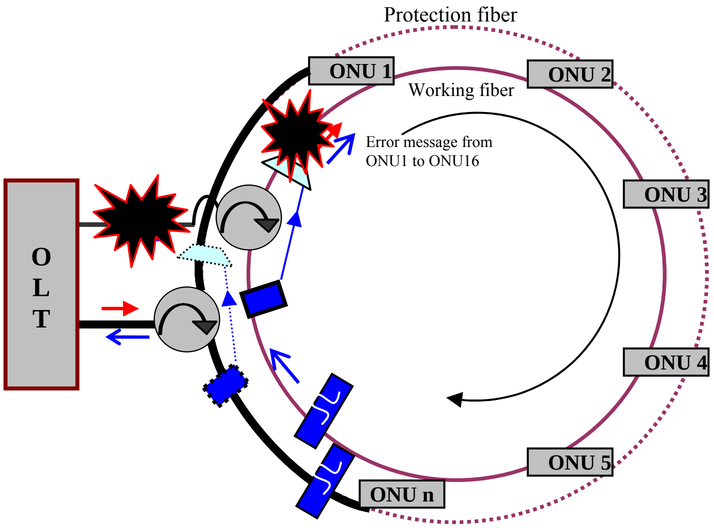
-Absence of upstream signal at OLT causes OLT to start to switch protection trunk fiber.

-ONU16 fails to generate outgoing signal and switches to outgoing protection fiber for bypass.

Step2 (after T$_{switch}$):

-ONU1 is switched to incoming protection fiber and ONU16 is switched to outgoing protection fiber. OLT is switched to protection trunk fiber. Thus OLT, ONU1 and ONU16 are linked. Failure recovered after one switching time.

Figure 6.12: ONU16 and trunk failure



Step1:

-Absence of both upstream and downstream signal at ONU1 incoming fiber requires ONU1 to generate error message asking ONU16 to switch to protection fiber. ONU1 APS start switching to incoming protection fiber.

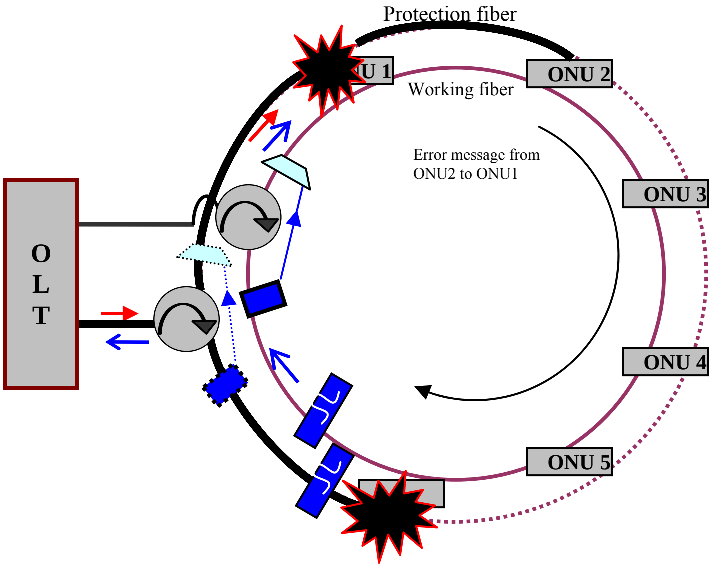
-Absence of upstream signal at OLT causes OLT to start to switch protection trunk fiber.

Step2 (after T$_{switch}$):

-ONU1 is switched to incoming protection fiber and ONU16 is switched to outgoing protection fiber due to ONU1 error message. OLT is switched to protection trunk fiber. Thus OLT, ONU1 and ONU16 are linked. Failure recovered after one switching time.

Figure 6.13: ONU1 incoming fiber and trunk failure

Step1:
-ONU16 APS goes to bypass (outgoing protection fiber).
-Due to ONU16 failure there is absence of upstream signal at ONU1 incoming fiber and that requires ONU1 to generate error message and switch to incoming protection fiber. Since ONU1 fails to generate error message, APS also starts bypass process (switch to outgoing protection fiber).
-Due to ONU1 failure, ONU2 fails to receive signal at incoming working fiber and generates error message for ONU1 and starts to switch to incoming protection fiber.
-Absence of upstream signal at OLT causes OLT to start to switch protection trunk fiber.



Step2 (after T_{switch}):
-ONU1 is switched to incoming and outgoing protection fiber and ONU16 is switched to outgoing protection fiber. ONU2 is on incoming protection fiber. ONU16 to ONU2 are linked and ONU2 error message loops back to itself. OLT is switched to protection trunk fiber. Failure recovered after one switching time.

Figure 6.14: ONU1 and ONU16 failure

6.6 Power Budget & Scalability of the Proposed Architecture

The scalability of the proposed architecture is mainly limited by the concatenated 10:90 1x2 splitter losses encountered by downstream signal at each node. Compared to the normal state, the signals encounter the additional OS and tap losses at each node.

Since upstream signal is regenerated at every node, typical limited upstream power budget problems (due to splitting loss at each node) as well as receiver dynamic range problem (long/short optical network paths and different splitting factors) are totally eliminated. To examine the performance impact of downstream power budget, we consider the worst-case scenario by calculating the total *Optical Distribution Network* (ODN) loss (passive optical elements such as splitters, combiners, fibers, connectors, switches and splices forming an optical path) incurred by the downstream signal on its longest optical path from the OLT the last ONU.

We assume a 20 Km trunk feeder fiber, first ONU is 20 km away from OLT, and last ONU is 23.2 km away from OLT (ring circumference is about 3.2 km). We consider low loss components based on recent industry standards: low loss switch with 0.3dB/channel insertion loss [18], circulator insertion loss of 0.3dB, and trunk SMF fiber loss at 1490 nm is 0.2 dB/km. There are two types of losses encountered by downstream signal at each node. The first type is along the path I-A-B in Fig. 6.3 (Drop-component) and the second type is along the path I-A-E-O in Fig. 6.3 (Go-component). Table 6.3 quantifies both types of losses.

Table 6.3: Typical losses incurred at an ONU (Fig. 6.3)

Type of Loss	Path I-A-B (normal)	Path I-A-E-O (normal)	Path I -O (by pass)
Connector (I/O)	0.1	2x0.1	2x0.1
Splitter-10/90 (A)	10.0	0.45	-
Filter (B)	0.15	-	-
CWDM (C/E)	-	2x0.3	-
Switch (I-A)/(E-O)	0.3	2x0.3	2x0.3
Tap (I/O)	0.05	2x0.05	2x0.05
TOTAL (dB)	10.6	1.95	0.9

Thus, total ODN loss incurred by downstream signal on its path to the last ONU is:

$$L_{Total\_Loss}^{last\_ONU} = L_{trunk}^{fiber} + L_{trunk}^{circulator} + L_{recirculator}^{CWDM} + (N-1)L_{I-O}^{ONU} + L_{Ring}^{fiber} + L_{I-B}^{ONU} \qquad (.61)$$

For example, if N = 8 ONUs, total ODN loss using equation 6.1 above (worst case) is ~ 29.45 dB. Assuming that OLT transmitted power into the fiber is 4.6 dBm and allowing a 3 dB power margin, a downstream receiver sensitivity of – 27.85 dBm will be required at the last ONU (worst case). Note that the downstream power budget is within a practical and commercial reach to support 8 - 10 ONUs. On the other hand, if N = 16 ONUs, the total ODN loss using equation 6.1 above (worst case) is ~ 45 dB.

To accommodate such a high loss budget, a semiconductor optical amplifier (SOA) can be used at the OLT to amplify the downstream signal [19-20]. A gain clamped SOA has been used in a 1.25 Gb/s EPON uplink experiment at 1.3 with a 48 dB loss budget [19]. In another experiment, using a commercially available non-gain clamped SOA in a 2.5 Gb/s downstream link, a total loss budget of 45 dB as well as a gain of 18 dB is achieved without utilizing forward error-correction (FEC) [20].

Thus, in the case of N = 16 ONUs, if we assume the same parameters of the SOA used in [20] along with same OLT transmitter power into the fiber of 4.6 dBm, and allowing a power margin of 3 dB; a downstream receiver sensitivity of – 25.40 dBm will be required at the last ONU. Note that downstream power budget is still within a practical and commercial reach to support 16-20 ONUs. Note also that the 8-20 nodes limit is not a shortcoming of an architecture that is specifically devised to support a private ring-based local access infrastructure within a 1-2 km diameter area. A typical large private organization would have at most 10-15 buildings within such geographically bounded area. To scale beyond 20 ONUs, a higher transmitted power and/or the use of FEC schemes are also required.

6.7 Recovery Time Analysis

Recovery time is defined here as the time from when failure occurs to when service is fully restored and a new cycle resumes. The total recovery time is the sum of the following delay components:

1) Fault detection time (T_{fail_det}), which is assumed here ~ 0 for all fault detection scenarios, which are performed independently via an ONU without OLT participation. This is true for any and all network failure detection scenarios except for the three special failure cases (trunk, last link, and last node failures) where both ONU_1 and OLT jointly participate in the detection process. In each of these three cases, the longer of the two

detection times between ONU and OLT is used (time it takes for OLT to detect absence of upstream signal, time it takes for ONU_1 to detect the fault).

Noting that it takes for ONU_1 ~ 0 time to detect either of the last link or last node failure; but it takes for OLT ~ 100 μs to detect absence of upstream signal in either case, detection time for last link or last node failure is ~ 100 μs. On the other hand, since it takes for ONU_1 an average ~ 50 μs to detect a trunk failure (assuming that the cut occurs at the middle of the trunk) and for OLT an average of about 50 μs to detect absence of upstream signal, average detection time for a trunk failure is ~ 50 μs.

2) Timeout ($T_{timeout}$) to avoid false fault detection, which is assigned here an arbitrary value of 5 μs.

3) REPORT/GATE transmission time ($T_{msg\_transm} =$ 64*8 (bits) /1Gbits/s) = 0.5 μs). The recovery process requires two REPORT messages (failure indication message and sync request) and one GATE message (sync frame).

4) Time it takes REPORT/GATE messages ($T_{msg\_prop\_ring}$ and $T_{msg\_prop\_trunk}$) to travel from the source node/OLT to OLT/source node. This is the propagation delay between the source node (worst case scenario when the source node is ONU_1) and OLT, which has a maximum value of about 15 + 100 = 115 μs (corresponding to 3.14 km ring circumference + 20 km trunk feeder). Since there are two REPORTs from the source node to OLT and one GATE from OLT to source node, worst case total propagation

delay due to this item, $T_{tot\_mess\_prop} \approx 3 \times 115$ μs = 345 μs. This is true for any and all network failure scenarios except for the three special failure cases listed in item 1 above. In each of these special cases, first REPORT (failure indication message) flooded by ONU_1 travels a shorter distance (new destination is now the last ONU). Thus, First REPORT's travel time has now been reduced from a 115 μs to ~ 15 μs (one trip around the ring).

Thus, in the case of a trunk failure, $T_{tot\_mess\_prop}$ is reduced to about 245 μs. However, in the case of a last link or last node failure, $T_{tot\_mess\_prop}$ is further reduced to about 2 x 115 μs = 230 μs (15 μs short travel time is not accounted for). This is because the total sum of 1) time it takes for ONU_1 to detect either of last link/node failure, 2) first REPORT transmission time, and 3) time it takes the first REPORT to arrive at the last ONU (5 + 0.5 + 15 = 20.5 μs) is still much shorter than the time it takes for OLT to detect absence of signal (105 μs).

5) REPORT/GATE message processing times ($T_{msg\_process}$), which is assumed here ~ 0.
6) One switching time (in the case of a single failure) can be as long as 12 ms; it varies according to the switch (T_{switch}) [18].

Thus, the general failure recovery time, $T_{Gene\_Fail\_Reco}$, for any and all network failure scenarios except for the three special failure cases (trunk, last link, and last node failures) is given by:

$$\mathbf{T}_{\mathbf{Gene\_Fail\_Reco}} = T_{fail\_det} + T_{timeout} + 3 * T_{msg\_transm} + 3\,(T_{msg\_prop\_ring} + T_{msg\_prop\_trunk}) + 3 * T_{msg\_process} + T_{switch}$$

$$= 0.0 + 5.0 ++3*0.5 + 3\,(15 +100) + 0.0 + T_{switch}$$

$$= 351.5\ \mu s + T_{switch}$$

On the other hand, in the case of either last link or last node failure, the failure recovery time, $T_{last\_link\_node}$, is given by:

$$\mathbf{T}_{\mathbf{last\_link\_node}} = T_{fail\_det} + T_{timeout} + 3 * T_{msg\_transm} + 2\,(T_{msg\_prop\_ring} + T_{msg\_prop\_trunk}) + 3 * T_{msg\_process} + T_{switch}$$

$$= 100.0 + 5.0 + 3*0.5 + 2\,(15 + 100\,) + 0.0 + T_{switch}$$

$$= 336.5\ \mu s + T_{switch}$$

Finally, in the case of a trunk failure, T_{trunk}, is given by:

$$\mathbf{T}_{\mathbf{trunk}} = T_{fail\_det} + T_{timeout} + 3 * T_{msg\_transm} + 2\,(T_{msg\_prop\_ring} + T_{msg\_prop\_trunk}) + T_{msg\_prop\_ring} + 3 * T_{msg\_process} + T_{switch}$$

$$= 50.0 + 5.0 + 3*0.5 + 2\,(15 + 100\,) + 15 + 0.0 + T_{switch}$$

$$= 301.5\ \mu s + T_{switch}$$

In general, the switching time is much longer than all other delay components combined and, therefore, the total recovery time is mainly dominated by switching time.

6.8 Simulation Results

In this section, we assess the feasibility and compare the performance of the proposed distributed detection and recovery schemes with that of a typical centralized recovery scheme based on ITU-T G.983.1 recommendations [8-9]. An event-driven packet-based simulation model was developed using C++. The traffic model used here is the same as that reported in [2, 16] where each ONU is modeled as an on/off source, with Pareto distribution to capture the self-similar nature of Ethernet traffic [21-22]. The performance metric used here is traffic loss for both upstream and downstream transmissions.

The following are the system parameters assumed in the simulation [2, 16]: 1) an EPON system with 16 ONUs; 2) access link data rate from users to an ONU is 100 Mb/s; 3) EPON downstream line rate is assumed to be same as that of upstream = 1 Gb/s; 4) For the distributed architecture, the distance between OLT and ONUs varies from 20 km to 23 km (ring circumference ≈ 3 km), where as for the centralized scheme the distance between the OLT and all ONUs is ~ 21km; 5) buffer size in each ONU is 10 Mbytes, 6) OLT houses 16 queues, each corresponds to a given ONU with 10 Mbytes size; 7) access link data rate from backbone to each queue at the OLT is 100 Mb/s; 8) maximum cycle time is 2 ms for both upstream and downstream transmission; 9) for the centralized scheme, the guard time, T_G, separating two consecutive upstream transmission windows (timeslot) is set to 5 μs (maximum upstream timeslot per ONU is 120 μs), whereas for the distributed architecture, T_G is set to zero (upstream traffic is terminated/regenerated at each node); 10) IEEE 802.3ah Multi-Point Control Protocol (MPCP) REPORT/GATE message is 64 bytes; 11) for the centralized scheme, the limited service DBA scheme

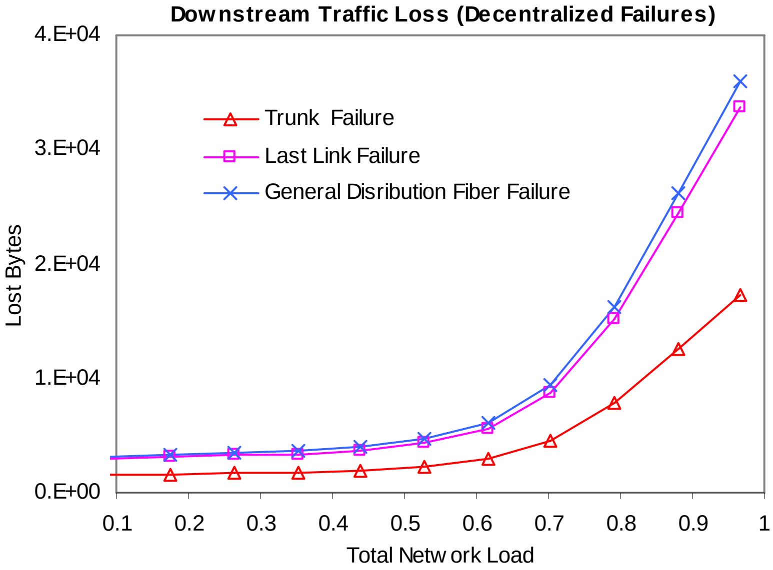


Figure 6.15: Downstream traffic loss vs. network load

reported in [2] is used here for both upstream and downstream traffic, whereas for the distributed scheme, the DBA scheme reported in [16] is used.

At each ONU, all arriving frames are queued in a first-in-first-out (FIFO) buffer. Similarly, all arriving frames at OLT are queued in corresponding ONU's FIFO buffer. For simplicity, we assume that all ONUs have equal average traffic load in both directions. All upstream traffic is addressed to OLT except inter-ONU control traffic. We randomly generate various types of failures and measure the corresponding upstream and downstream traffic loss. Each point in the simulation results corresponds to a sample of 40 million packets averaged over three different runs.

Fig. 6.15 shows downstream traffic loss versus network load for three different scenarios of network failures including trunk failure, general distribution link/node failure, and last link failure. Note that downstream traffic loss depends on three factors: 1) Offered

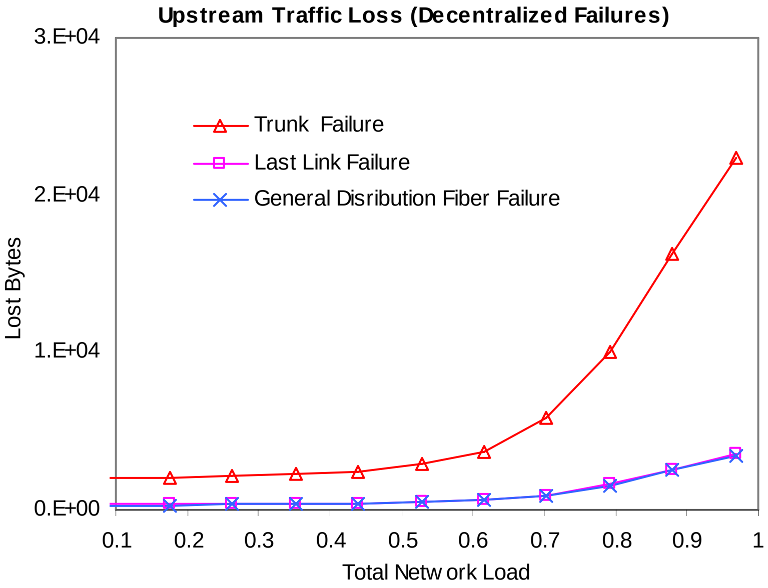


Figure 6.16: Upstream traffic loss vs. network load

network load; the higher the load the higher the traffic loss. 2) Existing traffic on the trunk and distribution ring at the instant of failure. 3) The duration from when failure occurs to when OLT detects absence of upstream signal (only in the case of the 3 special failures, trunk, last node, and last link) or receives a failure indication message from the affected ONU instructing it to stop further downstream transmission (in all other failure cases). This duration is the sum of only the first four delay components listed above.

As can be seen from Fig. 6.15, a trunk failure causes the least amount of traffic loss while a general link/node failure causes the most. Assuming that trunk cut occurs at the middle of the trunk, downstream traffic loss includes existing traffic already on the fiber at the time of failure (~50 μs) as well as traffic transmitted by OLT during the duration from the instant of failure occurrence up to the time when it is able to detect the failure and stops further transmission (detection time + timeout ~55 μs). Thus, total downstream traffic loss in the case of a trunk failure is equivalent to ~105 μs of transmitted traffic by OLT.

On the other hand, in the case of a general link (link connecting ONU_1 and ONU_2) or general node (ONU_2) failure, total downstream traffic loss is equivalent to ~ 221 μs of transmitted traffic by OLT. In this case, traffic loss includes existing traffic on the trunk fiber at the time of failure (~100 μs) as well as traffic transmitted by OLT during the duration from the instant of failure occurrence until it receives alarm indication message (from ONU_2 or ONU_3) instructing it to stop further transmission (5 μs + 115 μs + 0.5 μs ~ 121 μs).

Fig. 6.16 shows upstream traffic loss versus network load for the same three different failure scenarios of Fig. 6.15. Note that upstream traffic loss also depends on the same three factors listed above except now that it is the affected ONU that must detect the failure and notifies all other ONUs to stop upstream transmissions. As can be seen from Fig. 6.16, in contrast to downstream traffic loss, a trunk failure causes the most amount of upstream traffic loss while a general link/node failure causes the least. This is because the average time it takes for ONU_1 to detect a trunk failure (50 μs) is much longer than that taken by any other ONU to detect any and all other failure scenarios (~ 0).

Assuming that trunk cut occurs at the middle of the trunk, upstream traffic loss includes existing traffic on the fibers (trunk + distribution ring) at the time of failure (~ 65 μs), traffic lost during ONU_1's detection time (55 μs), traffic lost during ONU_1's REPORT transmission time (0.5 μs), plus traffic lost during REPORT'S propagation time around the ring to notify all other ONUs to stop upstream transmission (15 μs). Thus, total

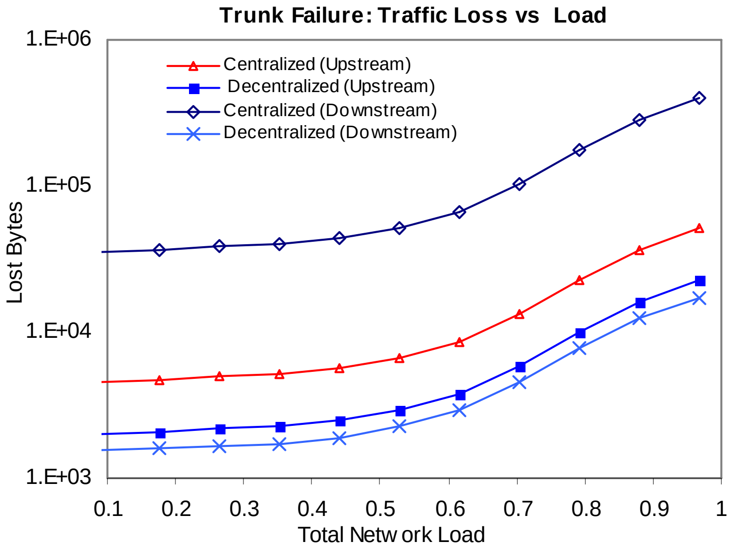


Figure 6.17: Traffic loss in trunk failure (centralized vs. decentralized)

upstream traffic lost in the case of a trunk failure is equivalent to ~ 136 µs of transmitted traffic by ONUs.

On the other hand, in the case of a general link/node (ONU_2) failure, total upstream traffic loss is equivalent only to ~ 22 µs of transmitted traffic by ONUs. In this case, upstream traffic lost includes existing traffic on the ring at the time of failure (~ 1 µs), traffic lost during $ONU_{2,3}$'s detection time (5 µs), traffic lost during $ONU_{2,3}$'s REPORT transmission time (0.5 µs), plus traffic lost during REPORT's propagation time around the ring to notify all other ONUs to stop upstream transmission (15 µs).

Fig. 6.17 compares the performance (downstream and upstream average traffic loss) of tree-based centralized architecture versus that of the ring-based distributed architecture, in the case of a trunk failure. As can be seen from Fig. 6.17, the distributed scheme exhibits less traffic loss compared to that of the centralized scheme for both upstream and downstream transmissions. Under the centralized scheme, downstream traffic loss is mainly dominated by the time it takes for OLT to detect the failure. OLT’s detection time comprises three components [9]: 1) time it takes for OLT to send the GATE message to all ONUs and receives REPORT message from all ONUs (entire MPCP cycle of 2.0 ms); 2) propagation delay between OLT and the ONU; and 3) GATE message transmission time. This is equivalent to ~ 2.4 ms (OLT’s detection time) of lost traffic transmitted by OLT [9]. Compare this value to the much smaller value of 105 µs obtained above for a trunk failure under the distributed scheme (Figs. 6.15 and 6.17).

On the other hand, upstream traffic loss under the centralized scheme includes existing traffic on the fibers at the time of failure (~ 70 µs), plus traffic lost during two successive ONU time slots transmission (~ 240 µs; maximum time slot per ONU ~ 120 µs excluding guard time). Thus, total upstream traffic lost in the case of a trunk failure is equivalent to ~ 310 µs of transmitted traffic by ONUs. Compare this value to the smaller value of 136 µs obtained above for upstream traffic loss in the case of a trunk failure under the distributed scheme (Figs. 6.16 and 6.17).

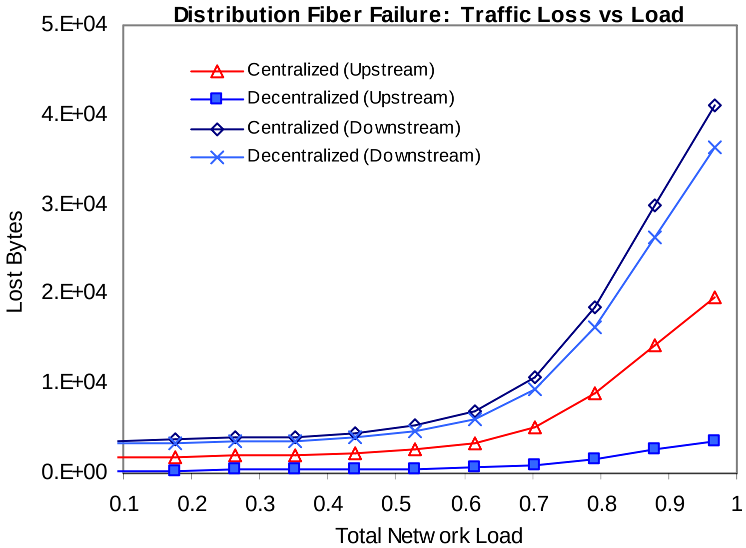


Figure 6.18: Traffic loss in distribution fiber failure (centralized vs. decentralized)

Fig. 6.18 is similar to Fig. 6.17 except that the results shown are now for a general distribution link failure. The results of Fig. 6.18 further corroborate those of Fig. 6.17 and indicate that the distributed scheme generally exhibits less traffic loss compared to that of the centralized scheme under any and all different failure scenario. Note that under the centralized scheme, distribution link failure will only affect a single node (the ONU connected to the failed link). Only downstream traffic directed to that particular ONU is lost.

For an OLT's detection time of 2.4 ms [9], there could be two successive timeslots of downstream transmissions to the affected ONU (one at the beginning cycle n and next one at the beginning of cycle n+1, both falls within 2.4ms). Since the maximum downstream transmission window (timeslot) per ONU is 125 μs (no guard time), maximum downstream traffic lost in the case of a general link failure is equivalent to ~ 250 μs of transmitted traffic by OLT. This value is still slightly higher than the 221 μs obtained above for a general link failure under the distributed scheme (Figs. 6.15 and 6.18).

Similarly, under the centralized scheme, only upstream traffic transmitted by the affected ONU is lost. Independent of OLT's failure detection time, upstream traffic lost in the case of a general distribution link failure can't exceed a maximum of a single timeslot of the affected ONU (120 μs). This is because, within the 2 ms cycle period in which the failure occurs, the affected ONU can only receive one grant (GATE message) from OLT before the failure and can’t receive any more after the failure. Compare this worst case scenario value to that obtained above (22 μs) for upstream traffic loss in the case of a general link/node failure under the distributed scheme (Figs. 6.16 and 6.18).

6.9 Conclusion

This chapter has proposed a 2-fiber self healing ring-based local access PON architecture that has addressed some of the limitations of current tree-based PON architectures including supporting private networking capability as well as providing a simple and cost-effective fully distributed resilience capabilities against any and all kinds of networking failures. The main characteristic of the proposed architecture is that it supports a fully distributed control plane among the ONUs for ONU-ONU communication. The control plane is shown capable of supporting fully distributed fault detection and recovery mechanisms.

The proposed scheme is shown capable of restoring all network traffic including upstream, downstream, and LAN data. In addition, the APS can also protect against any combination of concurrent double failures including trunk/distribution fiber breaks and node failures. The simulation results have indicated that the performance of the proposed distributed recovery scheme, including traffic loss and restoration speed, outperforms that of a typical centralized scheme. Furthermore, the recovery time associated with any and all different distribution network/trunk failures is still within the delay-bound limit required for delivering guaranteed triple play services.

The proposed architecture offers several key advantages over typical star-based centralized PON architectures, including: 1) supports a truly shared LAN capability among end users within a single PON setup; 2) in addition to the added flexibility and reliability of a distributed scheme, the proposed architecture eliminates the OLT's centralized task of failure detection and subsequent recovery scenarios as well as the task of processing requests and generating grants for bandwidth allocations. This reduces the additional processing complexities and delays at the OLT; 3) eliminates typical near-far problem as well as utilization of an upstream burst-mode receiver and associated design challenges at the OLT (upstream signal is terminated and fully regenerated at each node); 4) alleviates the typical limited upstream power budget problem associated with physical layer LAN emulation techniques described above; 5) increases available upstream channel bandwidth (typical guard time bands overhead is eliminated).

Chapter 7

Conclusions and Future Work

The growing demand for broadband access to the Internet has led to new interest in access schemes based on optical fiber. Amongst them, the PON architecture is considered the most promising from technological and economic point of view. Among the few types of PON, EPON which represents the convergence of low-cost Ethernet equipment and low-cost fiber infrastructure appeared to be the best candidate for the next-generation access network. To take full advantage of the EPON in an access-network environment, a whole host of existing issues, specially related to centralized architecture, must be addressed. In this work an innovative survivable ring-based distributed EPON architecture was presented that addressed some of the limitations of current tree-based centralized EPON architectures including supporting private networking capability.

7.1 Conclusions

In Chapter 3, we proposed the ring-based distributed EPON architecture with LAN capability. The details of distributed architecture along with its operational principles are described. A distributed DBA scheme was proposed, where ONUs exchanged control information among them to perform DBA (OLT was relieved from that burden). The distributed DBA globally optimized decision making process. The LAN emulation problems and increased end-to-end delay of LAN traffic were also addressed. In addition to the added flexibility and reliability of a distributed scheme, this architecture demonstrated the following several advantages over a typical star-based centralized PON architecture: (i) eliminated the typical utilization of an upstream burst-mode receiver and associated design challenges at the OLT; (ii) increased the available downstream and upstream channel bandwidth; (iii) improved LAN and WAN traffic performance significantly; (iv) supported more efficient upstream channel utilization.

In Chapter 4, we extended the globally optimized distributed DBA of the architecture to support QoS through differentiated class of service. Contrary to centralized approach, the distributed DBA integrated the inter-ONU and intra-ONU scheduling mechanisms in ONU; it improved overall network performance. We examined various approaches of bandwidth division and prioritization.

In Chapter 5, we analyzed the downstream bandwidth reduction problem of centralized architecture due to the LAN emulation and GRANT messages. Solution to this problem was proposed, where it demonstrated notable improvement in downstream performance.

In Chapter 6, we discussed the importance of survivability of ring-based PON architecture. A distributed failure detection scheme was proposed for the ring-based distributed EPON architecture. An APS protection scheme was introduced, whereby in the event of a failure in an ONU, trunk fiber, or distribution fiber, its detection and subsequent protection switching was efficiently initiated by the affected ONU. The proposed scheme demonstrated notable improvement in failure detection time and recovery as well as traffic loss over the centralized scheme.

7.2 Future Work

In this dissertation, we proposed the ring-based solution to some of the obstacles of centralized tree based architecture. The theoretical analysis and simulation results was presented to demonstrate the advantages of the proposed approach. Although results of simulation experiments were shown, the ultimate test would require implementing the proposed algorithms in real network with realistic traffic. Therefore, physical experiments (test bed) to observe real life implications of the proposed approach would be an avenue of future investigation.

Another area of future work would be network security. The way the proposed architecture works is that one ONU traffic (LAN/WAN) passes through the other ONUs. It raises security concern. Therefore, one of the areas to explore in future would be security related issues on this architecture.

Abbreviations

AF Assured Forwarding

ATM Asynchronous Transfer Mode

BPON Broadband Passive Optical Network

BE Best Effort

CRC Cyclic Redundancy Check

DA Destination Address

DBA Dynamic Bandwidth Allocation

DDBA Distributed DBA

DiffServ Differentiated Services

DSL Digital Subscriber Loop

EF Expedited Forwarding

EFMA Ethernet in the First Mile Alliance

EPON Ethernet Passive Optical Network

FCS Frame Check Sequence

FSAN Full Service Network Access

FTTH Fiber-to-the-Home

GPON Gigabit Passive Optical Network

IFG Inter-frame Gap

IEEE Institute of Electric and Electronic Engineers

IETF Internet Engineering Task Force

IPACT Interleaved Polling with Adaptive Cycle Time

ISDN Integrated Services Digital Network

ITU International Telecommunications Union

LAN Local Area Network

LLC Logical Link Control

MAC Media Access Control

MPCP Multi-Point Control Protocol

OAM Operations and Maintenance

OLT Optical Line Terminal

ONU Optical Network Unit

OOL Offered ONU Load

OSI Open Systems Interconnection

PON Passive Optical Network

QoS Quality of Service

SA Source Address

SBA Static Bandwidth Allocation

SLA Service Level Agreement

RTT Round Trip Time

TDMA Time Division Multiplexing Access

TNL Total Network Load

WDM Wavelength Division Multiplexing

Publications

- **ASM Delowar Hossain** et al, “Supporting Private Networking Capability in EPON”, *accepted in ICC’06 conference*, Istanbul, Turkey, June 2006
- **ASM Delowar Hossain** et al, “A Novel Ring-Based Local Access PON Architecture for Supporting Private Networking Capability,” *Journal of Optical. Networking*, vol. 5, no.1, January 2006.
- **ASM Delowar Hossain et al, “**A Novel Ring-Based EPON Architecture”,” in *Proceedings of BroadNets’05 Workshop(COMNETS),* Boston, Massachusetts, Oct. 3-7 2005.
- **ASM Delowar Hossain** et al, “A Distributed Control Plane Architecture for EPON: Simulation Study and Feasibility Experiment*,*” in *Proceedings of IIT’05 Conference*, Dubai, UAE, Sept. 26-28 2005.
- Erkan, **A. D. Hossain**, M. F. Arend, R. Dorsinville, M. A. Ali, " On the merits of implementing a novel decentralized ethernet-based PON architecture for next-generation broadband access networks", Proc. SPIE Vol. 6012, p. 31-38, Oct 2005
- **ASM Delowar Hossain** et al, “ Protection for a Ring-Based EPON Architecture”, in *Proceedings of BroadNets’05 Workshop(COMNETS),* Boston, Massachusetts, Oct. 3-7 2005.
- **ASM Delowar Hossain** et al, “Efficient Dynamic Bandwidth Allocation for EPON ,” in *Proceedings of IIT’05 Conference*, Dubai, UAE, Sept. 26-28 2005.

Bibliography

Chapter 1

[1] Full Services Access Networks, http://www.fsanet.net/.

[2] B. Lung, “PON Architecture ‘Future proofs’ FTTH,” Lightwave, vol. 16, no. 10, Sept. 1999, pp. 104-7.

[3] Chang-Hee Lee, W. V. Sorin, and Byoung Yoon Kim, “Fiber to the Home Using a PON Infrastructure” IEEE Journal of Ligthwave Technology, Vol. 24, No. 12, December 2006.

[4] Donald. E.A. Clarke and Tetsuya Kanada, “Broadband: The Last Mile,” IEEE Communications Magazine, vol. 31, no. 3, pp. 94–100, Mar. 1993.

[5] G. Pesavento and M. Kelsey, “PONs for the broadband local loop,” Lightwave, PennWell, vol. 16, no. 10, pp. 68 – 74, September 1999.

[6] IEEE 802.3 Ethernet in the First Mile Study Group, http://www.ieee802.org/3/efm/public/index.html.

[7] Alloptic, “Ethernet Passive Optical Networks,” The International engineering Consortium, http://www.iec.org .

[8] 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.

[9] G. Kramer et al., “Ethernet PON: design and analysis of an optical access network,” Photon. Network Commun. J., vol. 3, no. 3, July 2001.

[10] G. Kramer, B. Mukherjee, and G. Pesavento, “IPACT: A dynamic Protocol for an Ethernet PON (EPON),” IEEE Comm. Mag., pp. 74-80, Feb. 2002.

Chapter 2

[1] K. G. Coffman and A. M. Odlyzko, “Internet growth: Is there a "Moore's Law" for data traffic?” Handbook of Massive Data Sets, J. Abello, P. M. Pardalos, and M. G. C. Resende, eds., Kluwer, 2001.

[2] “Access Network Systems: North America – Optical Access. DLC and PON Technology and Market Report,” Report RHK-RPT-0548, RHK Telecommunication Industry Analysis, San Francisco, June 2001.

[3] G. Pesavento and M. Kelsey, “PONs for the broadband local loop,” Lightwave, PennWell, vol. 16, no. 10, pp. 68 – 74, September 1999.

[4] B. Lung, “PON architecture ‘futureproofs’ FTTH,” Lightwave, PennWell, vol. 16, no. 10, pp. 104 –107, September 1999.

[5] IEEE 802.3 Ethernet in the First Mile Study Group [Online]. Available: http://www.ieee802.org/3/efm/public/index.html.

[6] J.R. Stern, J.W. Ballance, D.W. Faulkner, S. Hornung, and D.B. Payne, “Passive Optical Local Networks for Telephony Applications and Beyond,” Electronics Letters, vol. 23, no. 24, pp. 1255–1257, Nov. 1987.

[7] Donald. E.A. Clarke and Tetsuya Kanada, “Broadband: The Last Mile,” IEEE Communications Magazine, vol. 31, no. 3, pp. 94–100, Mar. 1993.

[8] Yih-Kang Maurice Lin, Dan R. Spears, and Mih Yin, “Fiber-Based Local Access Network Architectures,” IEEE Communications Magazine, vol. 27, no. 10, pp.64–73, Oct. 1989.

[9] Ivan Andonovic and Deepak Uttamchandani, Eds., Principles Of Modern Optical Systems, chapter 11, Artech House, 1989.

[10] G. Kramer and G. Pesavento, “Ethernet passive optical network (EPON): building a next generation optical access network,” IEEE Commun. Mag., pp. 66–73, Feb. 2002.

[11] C. Assi et al., “Dynamic bandwidth allocation for quality of service over Ethernet PONs,” IEEE J. Select. Areas Commun., Dec. 03.

[12] ITU-T, "G.983.1 - Broadband Passive Optical Networks (BPON): General characteristics," June 1999.

[13] ITU-T, "G.984.1 - Gigabit-capable Passive Optical Networks (GPON): General characteristics," Mar. 2003.

[14] David Faulkner, Rajendrakumar Mistry, Tom Rowbotham, Kenji Okada, Wsewolod Warzanskyj, Albert Zylbersztejn, and Yves Picault, "The Full Services Access Networks Initiative," IEEE Communications Magazine, vol. 35, no. 4, pp. 58–68, Apr. 1997.

[15] YoichiMaeda, Kenji Okada, and David Faulkner, "FSAN OAN-WG and future issues for broadband optical access networks," IEEE Communications Magazine, vol. 39, no. 12, pp. 126–132, Dec. 2001.

[16] ITU-T Recommendation G.983.1, Broadband optical access systems based on Passive Optical Networks (PON), in Series G: Transmission Systems and Media, Digital Systems and Networks, Telecommunication Standardization Sector of ITU, October 1998.

[17] S. Clavenna, "Metro Optical Ethernet," Lightreading (www.lightreading.com), November 2000.

[18] IEEE Standard for Information technology - Telecommunications and information exchange between systems - Local and metropolitan area networks – Specific Requirements. Part 3: Carrier sense multiple access with collision detection (CSMA/CD) access method and physical layer specification, ANSI/IEEE Std 802.3-2002, Available at http://standards.ieee.org/getieee802/download/802.3-2002.pdf.

Chapter 3

[1] B. G. Pesavento and M. Kelsey, “PONs for the broadband local loop,” Lightwave, PennWell, vol. 16, no. 10, pp. 68 – 74, September 1999.

[2] Alloptic, “Ethernet Passive Optical Networks,” The International Engineering Consortium, http://www.iec.org.

[3] G. Kramer et al., “Ethernet PON: design and analysis of an optical access network,” Photon. Network Commun. J., vol. 3, no. 3, July 2001.

[4] ASM Delowar Hossain et al, “A Novel Ring-Based EPON Architecture”,” in Proceedings of BroadNets’05 Workshop(COMNETS), Boston, Massachusetts, Oct. 3-7 2005.

[5] IEEE 802.3ah Task Force [Online].Available: http://www.ieee802.org/3/efm/public/index.html.

[6] D. sala, et. Al., IEEE EFM, May 2002 materials, www. ieee802.org/3/efm.

[7] C.-J. Chae et al., “A PON system suitable for internetworking optical network units using a fiber Bragg grating on the feeder fiber,” IEEE Photon. Technol. Lett., vol. 11, pp. 1686–1688, Dec. 1999.

[8] E. Wong et al., “CSMA/CD-based EPON with optical internetworking capability among users,” IEEE PTL, Vol 16, No 9, pp. 2195-97, Sep. 2004.

[9] G. Kramer et al., “Interleaved Polling with Adaptive Cycle Time (IPACT): A Dynamic Bandwidth Distribution Scheme in an Optical Access Network, Photonic Network Comm, vol. 4, no. 1 pp. 89-107, Jan. 2002.

[10] C. Assi et al., “Dynamic bandwidth allocation for quality of service over Ethernet PONs,” IEEE J. Select. Areas Commun., Dec. 2003.

[11] D. Nikolova, B. Van Houdt, and C. Blondia, “Dynamic bandwidth allocation algorithm in EPON: a simulation study,” in Proc. Opticomm’03, Dallas, TX, Oct. 2003.

[12] A. Shami et. al. , "Jitter Performance in Ethernet Passive Optical Networks," IEEE/OSA Journal of Lightwave Technology, Vol. 23, No. 4, pp. 1745-1753, April 2005.

[13] A. Shami et al. "QoS Control Schemes for Two-Stage Ethernet Passive Optical Access Networks," IEEE Journal on Selected Areas in Communications (IEEE JSAC), Vol. 28, No.3, pp. 1467-1478, August 2005.

[14] V. Paxson and S. Floyd, "Wide area traffic: the failure of Poisson modeling," IEEE/ACM Trans. Networking, vol. 3, pp. 226–244, June 1995.

[15] 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., 1996, pp. 339–366.

Chapter 4

[1] K. Rege et al., “QoS management in trunk-and-branch switched Ethernet networks,” IEEE Commun. Mag., vol. 40, pp. 30–36, Dec. 2002.

[2] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, An Architecture for Differentiated Services, IETF, RFC 2475, Dec. 1998.

[3] G. Kramer et al., “Ethernet PON: design and analysis of an optical access network,” Photon. Network Commun. J., vol. 3, no. 3, July 2001.

[4] G. Kramer, B. Mukherjee, and G. Pesavento, “IPACT: a dynamic protocol for an Ethernet PON (EPON),” IEEE Commun. Mag., pp. 74–80, Feb. 2002.

[5] G. Cramer, B.Mukherjee, and A. Maislos, “Ethernet passive optical network (EPON): a missing link in an end-to-end optical internet,” in Multiprotocol Over WDM: Building the Next Generation Internet, S. Dixit, Ed. New York: Wiley, Mar. 2003.

[6] D. Nikolova, B. Van Houdt, and C. Blondia, “Dynamic bandwidth allocation algorithm in EPON: a simulation study,” in Proc. Opticomm’03, Dallas, TX, Oct. 2003.

[7] 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.

[8] G. Kramer et al., “Interleaved Polling with Adaptive Cycle Time (IPACT): A Dynamic Bandwidth Distribution Scheme in an Optical Access Network, Photonic Network Com vol. 4, no. 1 pp. 89-107, Jan. 02.

[9] G. Cramer et al., “On supporting differentiated classes of service in EPON-based access network,” J. Opt. Networks, 2002.

[10] Assi et al., “Dynamic bandwidth allocation for quality of service over Ethernet PONs,” IEEE J. Select. Areas Commun., Dec. 03.

[11] A. Hossain, R. Dorsinville, M. Ali, A. Shami, and C. Assi, "Ring-based local access PON architecture for supporting private networking capability," J. Opt. Network. 5, 26-39, 06.

[12] J.M. Jaffe "Bottleneck Flow Control.", IEEE Transactions on Communications, 29(7), 1981.

[13] Hou, H. Tzeng and S. Panwar, "A generalized max-min rate allocation policy and its distributed implementation using the ABR flow control mechanism," IEEE INFOCOM '98, pp. 1366-1375, Apr. 1998.

[14] A. Malla, M. El-Kadi, S. Olariu and P. Todorova, "A fair resource allocation protocol for multimedia wireless networks," IEEE Trans. Parallel and Distributed Systems, vol. 14, no. 1, pp. 63-71, Jan. 2003.

[15] Y. Zhou and H. Sethu, "On achieving fairness in the joint allocation of processing and bandwidth resources: principles and algorithms," IEEE/ACM Trans. Networking, vol. 13, no. 5, pp. 1054 - 1067, Oct. 2005.

[16] M. Hosaagrahara and H. Sethu, "Max-min fairness in input queued switches," ACM SIGCOMM Student Poster Session, August 2005, Philadelphia, PA, USA.

[17] V. Paxson and S. Floyd, "Wide area traffic: the failure of Poisson modeling," IEEE/ACM Trans. Networking, vol. 3, pp. 226–244, June 1995.

[18] 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., 1996, pp. 339–366.

Chapter 5

[1] B. Lung, “PON architecture ‘future proofs’ FTTH,” Lightwave, PennWell, vol. 16, no. 10, pp. 104 –107, September 1999.

[2] Alloptic, “Ethernet Passive Optical Networks,” The International Engineering Consortium, http://www.iec.org.

[3] G. Kramer and G. Pesavento, “Ethernet passive optical network (EPON): building a next generation optical access network,” IEEE Commun. Mag., pp. 66–73, Feb. 2002.

[4] G. Kramer et al., “Ethernet PON: design and analysis of an optical access network,” Photon. Network Commun. J., vol. 3, no. 3, July 01.

[5] IEEE 802.3ah EFM Study Group, http://www.ieee802.org/3/efm/public/index.html.

[6] G. Kramer et al., “Interleaved Polling with Adaptive Cycle Time (IPACT): A Dynamic Bandwidth Distribution Scheme in an Optical Access Network, Photonic Network Comm, vol. 4, no. 1 pp. 89-107, Jan. 2002.

[7] C. Assi et al., “Dynamic bandwidth allocation for quality of service over Ethernet PONs,” IEEE J. Select. Areas Commun., Dec. 2003.

[8] D. Nikolova, B. Van Houdt, and C. Blondia, “Dynamic bandwidth allocation algorithm in EPON: a simulation study,” in Proc. Opticomm’03, Dallas, TX, Oct. 2003.

[9] A. Shami et. al. , "Jitter Performance in Ethernet Passive Optical Networks," IEEE/OSA Journal of Lightwave Technology, Vol. 23, No. 4, pp. 1745-1753, April 2005.

[10] A. Shami et al. "QoS Control Schemes for Two-Stage Ethernet Passive Optical Access Networks,” IEEE Journal on Selected Areas in Communications (IEEE JSAC), Vol. 28, No.3, pp. 1467-1478, August 2005.

[11] A. Hossain, R. Dorsinville, M. Ali, A. Shami, and C. Assi, "Ring-based local access PON architecture for supporting private networking capability," J. Opt. Network. 5, 26-39, 06.

[12] D. sala, et. Al., IEEE EFM, May 2002 materials, www. ieee802.org/3/efm.

[13] C.-J. Chae et al., “A PON system suitable for internetworking optical network units using a fiber Bragg grating on the feeder fiber,” IEEE Photon. Technol. Lett., vol. 11, pp. 1686–1688, Dec. 1999.

[14] E. Wong et al., “CSMA/CD-based EPON with optical internetworking capability among users,” IEEE PTL, Vol 16, No 9, pp. 2195-97, Sep. 2004.

[15] N. Nadarajah, M. Attygalle, E. Wong, and A. Nirmalathas, “Novel schemes for local area network emulation in passive optical networks with RF subcarrier multiplexed customer traffic,” J. Lightvw. Technol., vol. 23, no. 10, pp. 2974–2983, Oct. 2005.

[16] J.M. Jaffe “Bottleneck Flow Control.”, IEEE Transactions on Communications, 29(7), 1981.

[17] Y. Hou, H. Tzeng and S. Panwar, “A generalized max-min rate allocation policy and its distributed implementation using the ABR flow control mechanism,” IEEE INFOCOM ’98, pp. 1366-1375, Apr. 1998.

[18] A. Malla, M. El-Kadi, S. Olariu and P. Todorova, “A fair resource allocation protocol for multimedia wireless networks,” IEEE Trans. Parallel and Distributed Systems, vol. 14, no. 1, pp. 63-71, Jan. 2003.

[19] Y. Zhou and H. Sethu, “On achieving fairness in the joint allocation of processing and bandwidth resources: principles and algorithms,” IEEE/ACM Trans. Networking, vol. 13, no. 5, pp. 1054 - 1067, Oct. 2005.

[20] M. Hosaagrahara and H. Sethu, “Max-min fairness in inputqueued switches,” ACM SIGCOMM Student Poster Session, August 2005, Philadelphia, PA, USA.

[21] V. Paxson and S. Floyd, “Wide area traffic: the failure of Poisson modeling,” IEEE/ACM Trans. Networking, vol. 3, pp. 226–244, June 1995.

[22] 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., 1996, pp. 339–366.

Chapter 6

[1] D. Sala and A. Gummalla, “PON functional requirements: services and performance,” in IEEE 802.3ah Meeting in Portland OR, July 2001.

[2] G. Kramer, B. Mukherjee, and G. Pesavento, “IPACT: A dynamic Protocol for an Ethernet PON (EPON),” IEEE Comm. Mag., pp. 74-80, Feb. 2002.

[3] G. Kramer et al., “Ethernet PON: design and analysis of an optical access network,” Photon. Network Commun. J., vol. 3, no. 3, July 2001.

[4] C. Assi et al., “Dynamic bandwidth allocation for quality of service over Ethernet PONs,” IEEE J. Select. Areas Commun., Dec. 2003.

[5] M. K. Abdullah, W. T. P’ng, P. W. Lau, and E. R. Tee, “FTTH access network protection using a switch,” in Proc. 9th Asia-Pacific Conf. Commun., 2003, pp. 1219–1222.

[6] W.-P. Lin, M.-S. Kao, and S. Chi, “The modified star-ring architecture for high-capacity subcarrier multiplexed passive optical networks,” J. Lightw. Technol., vol. 19, no. 1, pp. 32–39, Jan. 2001.

[7] W.-P. Lin, M.-S. Kao, and S. Chi, “A DWDM/SCM self-healing architecture for broad-band subscriber networks,” J. Lightw. Technol., vol. 21, no. 2, pp. 319–328, Feb. 2003.

[8] ITU-T recommendation G.983.1, "Broadband optical access systems based on passive optical networks (PON): 1998.

[9] Yu-mi Kim et al., “Cost Effective Protection Architecture to Provide Diverse Protection Demands in Ethernet Passive Optical Network”, Proceedings of ICCT 2003.

[10] N. Nadarajah, M. Attygalle, E. Wong, and A. Nirmalathas, “Novel schemes for local area network emulation in passive optical networks with RF subcarrier multiplexed customer traffic,” J. Lightw. Technol., vol. 23, no. 10, pp. 2974–2983, Oct. 2005.

[11] O. Gerstel and R. Ramaswami, “Optical layer survivability: A post-bubble perspective,” IEEE Commun. Mag., vol. 41, no. 9, pp. 51–53, Sep. 2003.

[12] T.-H. Wu, “Emerging technologies for fiber network survivability,” IEEE Commun. Mag., vol. 33, no. 2, pp. 62–74, Feb. 1995.

[13] C. J. Chae et al., “A PON system suitable for internetworking optical network units using a fiber Bragg grating on the feeder fiber,” IEEE Photon. Technol. Lett., vol. 11, pp. 1686–1688, Dec. 1999.

[14] E. Wong et al., “CSMA/CD-based EPON with optical internetworking capability among users,” IEEE PTL, Vol 16, No 9, pp. 2195-97, Sep. 2004.

[15] N. Nadarajah, M. Attygalle, A. Nirmalathas, and E. Wong, “A novel local area network emulation technique on passive optical networks,” IEEE Photon. Technol. Lett., vol. 17, no. 5, pp. 1121–1123, May 2005.

[16] ASM Delowar Hossain, A. Shami, C. Assi, and M. A. Ali “Supporting Private Networking Capability in EPON”, ICC’2006, Istanbul, Turkey, June 2006.

[17] ASM Delowar Hossain et al, “ Protection for a Ring-Based EPON Architecture”, in Proceedings of BroadNets’05 Workshop(COMNETS), Boston, Massachusetts, Oct. 3-7 2005.

[18] http://www.dbmoptics.com/products/gpoti/switches.php.

[19] N. Suzuki et. Al, ECOC 2005, paper TU1.3.3.

[20] D. Nesset, D. B. Payne, R. Davey, and T. Gilfedder, “Demonstration of Enhanced Reach and Split of a GPON System Using Semiconductor Optical Amplifiers”, in Proceedings of ECOC 2006, France, September 2006.

[21] V. Paxson and S. Floyd, “Wide area traffic: the failure of Poisson modeling,” IEEE/ACM Trans. Networking, vol. 3, pp. 226–244, June 1995.

[22] 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., 1996, pp. 339–366.