uOttawa

L'Université canadienne
Canada's university

FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTOCTORALES

uOttawa

L'Université canadienne
Canada's university

FACULTY OF GRADUATE AND POSDOCTORAL STUDIES

Ali Nouroozifar
AUTEUR DE LA THÈSE / AUTHOR OF THESIS

M.A.Sc. (Electrical Engineering)
GRADE / DEGREE

School of Information Technology and Engineering
FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

Priority Queueing Scheduling Techniques for QoS Management in Ethernet Passive Optical Networks
TITRE DE LA THÈSE / TITLE OF THESIS

H. Mouftah
DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

EXAMINATEURS (EXAMINATRICES) DE LA THÈSE / THESIS EXAMINERS

C.H. Lung

J. Zhao

Gary W. Slater
LE DOYEN DE LA FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTDOCTORALES /
DEAN OF THE FACULTY OF GRADUATE AND POSTDOCORAL STUDIES

Priority Queueing Scheduling Techniques for QoS Management in Ethernet Passive Optical Networks

By

Ali Nouroozifar (B.Sc.E.E.)

A thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
In partial fulfillment of the requirements for the

Master of Applied Science (MASc) degree
in
Electrical Engineering

Ottawa-Carleton Institute for Electrical and Computer Engineering (OCIECE)
School of Information Technology and Engineering (SITE)
University of Ottawa
Ottawa, Ontario, Canada
August, 2005

Your file Votre référence
ISBN: 0-494-11365-0
Our file Notre référence
ISBN: 0-494-11365-0

In the name of God, who gave us knowledge.

Abstract

With the vast availability of Ethernet in Local Area Networks (LAN) deployed globally and connecting LANs to Metropolitan Area Networks (MAN) and Wide Area Networks (WAN) via the current transport technologies (i.e., Synchronous Optical Network (SONET) in Americas and Synchronous Digital Hierarchy (SDH) in the rest of the world), there exists a need to resolve bandwidth gap between access networks and transport networks. Ethernet Passive Optical Network is one of the potential solutions offering to solve the bottleneck issue of access networks and adaptability of carrying Ethernet frames from subscribers to service providers. This study reviews Ethernet Passive Optical Networks (EPON) and then proposes a Priority Queueing (PQ) Algorithm in order to offer differentiated service and Quality of Service (QoS) in EPON.

The scope of this study aims to encompass in detail some of the most prominent features involved in the priority scheduling in Ethernet passive optical networks (EPONs). The main items presented in this study include a review of the current developments in the access networks, a detailed review of the passive optical networks (PONs), and analysis of scheduling in EPON. In addition, the simulation results are presented to verify and validate the proposed scheme. Special attention is also given to the delay and bandwidth requirements of the service disciplines.

"No problem can stand the assault of sustained thinking." –VOLTAIRE

Acknowledgement

The author wishes to express his deep gratitude and respect to Prof. H.T.Mouftah of the School of Information Technology and Engineering (SITE), University of Ottawa, for his guidance and assistance in the preparation of this thesis. His deep understanding of the subject coupled with dedicated hard work will be a source of inspiration to the author in his future endeavors.

My first teacher in optics that happened to be my father, I am in debted to him for all his teaching and technical discussions. He profoundly shaped a foundation that made my journey more comfortable. My mother and my family also deserve more credit than simple words can express for their love, encouragement and support during the difficult periods that led to the production of this thesis. I am thankful to my son for all the hours I stole from him that I should have spent with him.

Grateful acknowledgement is also extended to Prof. H. Naser of Lakehead University, for his much appreciated assistance and technical comments.

Table of Contents

List of Figures

List of Tables

Acronyms

ADSL: Adaptive Digital Subscriber Loop
AF: Assured Forwarding
ANSI: American National Standards Institute
AON: Active Optical Network
APON: ATM Passive Optical Network
ATM: Asynchronous Transfer Mode
BE: Best Effort
BPON: Broadband Passive Optical Network
bps: Bits per Second
CBR: Constant Bit Rate
CO: Central Office
CoS: Class of Service
DBA: Dynamic Bandwidth Allocation/Assignment
DiffServ: Differentiated services
DSL: Digital Subscriber Line
EF: Expedited Forwarding
EFM: Ethernet in the First Mile
EPON: Ethernet Passive Optical Network
FTP: File Transfer Protocol
FTTX: Fiber To The X (business, curb, home, MTU/MDU, neighborhood, office, premises, and user)
Gbps, Gb/s: Giga Bits per Second
GE/Gig E: Gigabit Ethernet
GFP: Generic Framing Procedure
GPON: Gigabit Passive Optical Network
GS: Guaranteed Service
IEEE: International Electrical & Electronic Engineer
IntServ: Integrated Services
IFG: Inter Frame Gap
IP: Internet Protocol
IPG: Inter packet Gap
ISDN: Integrated Services Digital Network
ITU: International Telecommunication Union
Kbps, Kb/s: Kilo Bits per Second
LAN: Local Area Network
MAC: Medium Access Control
MAN: Metropolitan Area Network
Mbps, Mb/s: Mega Bits per Second
MP2P: Multi Point-to-Point
MPCP: Multi-Point Control Protocol
Nm: Nanometer (10^{-9} Meters)
OAN: Optical Access Network

ODN:	Optical Distribution Network
OLT:	Optical Line Terminating unit/Terminal
ONT:	Optical Network Terminating unit/Terminal
ONU:	Optical Network Unit
P2MP:	Point-to-Multi Point
P2P:	Point-to-Point
PON:	Passive Optical Network
POTS:	Plain Old Telephone System/Service
PQ:	Priority Queueing
QoS:	Quality of Service
RED:	Random Early Detection
RR:	Round Robin
SBA:	Static Bandwidth Allocation
SDH:	Synchronous Digital Hierarchy
SLA:	Service Level Agreement
SLS:	Service Level Specification
SONET:	Synchronous Optical Network
SP:	Service Provider
TCP:	Transmission Control Protocol
TDM:	Time Division Multiplexing
TDMA:	Time Division Multiple Access
UDP:	User Datagram Protocol
VDSL:	Very High Bit Rate (Speed) Digital Subscriber Line
VLAN:	Virtual Local Area Network
VoD:	Video on Demand
VoIP:	Voice over Internet Protocol
WAN:	Wide Area Network
WDM:	Wavelength Division Multiplexing
WFQ:	Weighted Fair Queueing
WRR:	Weighted Round Robin
xDSL:	See DSL (ADSL, and VDSL)

Chapter 1 Introduction

1.1 Background

Telecommunication networks may be divided into three main categories of access networks, metropolitan area networks (MANs), and core networks. Access networks link end users to central offices (COs). The main objective of an access network is to concentrate generated upstream traffic, to reduce the size of the access switches and interfaces. Currently users' traffic is transported to metropolitan area networks (MANs) in the CO via synchronous optical network/synchronous digital hierarchy (SONET/SDH) systems. Core networks are responsible of the long haul transport of all kinds of traffic.

Access networks originally were twisted-pair loops running to subscribers. They were designed to support a voice frequency with a 4 kHz bandwidth. Dial-up modems utilize this low frequency band for data transfer speeds up to 56 kbps. To increase the data rate further, digital subscriber line (DSL) technologies were developed. The first DSL technology was basic Rate Integrated services Digital Network (ISDN), transporting two 64 kbps payload channels and one 16 kbps signalling channel through telephony network. Subsequent higher speed DSLs were aimed at business users with speeds in the T1/E1 range and geographical coverage of a few kilometers without any need to a repeater.

However, for high speed residential access the signal must be diverted from the telephony network that having class 5 voice frequency switches with data rates support of up to 56 kbps to a high speed data network. Asymmetric Digital Subscriber Line (ADSL) carries traffic from 0.5 to 6 Mbps in the downstream and with lower data rates in the upstream, bypasses the telephony switch at the central office (CO). Similarly, very high speed DSL (VDSL) carries asymmetric traffic with up to 50 Mbps in the downstream.

Next generation access networks must comply with the requirements dictated from new applications (e.g., Voice over Internet Protocol (VoIP), video on demand (VoD), multimedia broadcasting, real-time audio/video communications, etc.). Such requirements are lower delay, higher bandwidth and low jitter. The bandwidth in the current access networks cannot provide a reliable and wide enough bandwidth to all subscribers. Access networks are the last segment to be upgraded to fiber-based technologies in order to provide decent future-proof services and applications.

Broadband services have been the center point of many research activities in industry and academia due to this overwhelming pressure dictated by the subscribers to the service providers and carriers. Current broadband networks are not capable of supporting the tremendous forthcoming needs for newer access networks providing the ultimate broadband services to the subscribers with sufficiently high speeds. Due to copper-based and wireless-based constraints, they are not suitable candidate for this purpose. The only decent solution filling up the shortcomings of the previous mentioned techniques is fiber-based technologies. These technologies are commonly referred as Optical Access Networks (OANs). This family of networks consists of active solutions and passive solutions. By active solutions we mean that there will be a need for providing power to the equipment in the transmission path between subscribers and central office (CO) and main signal is actively switched. In passive solutions on the other hand, main signal is passively split and there is no requirement for providing power to the equipment in the transmission path. Active optical networks (AONs) are not considered here and the focus of this thesis is on the passive optical networks (PONs). The range of technologies developed based on these networks consists mostly of Asynchronous Transfer Mode (ATM) based, and Ethernet-based systems. Due to the fact that more than 90% of the LANs worldwide are Ethernet-based [CHI04], Ethernet passive optical networks (EPONs) have been one of the center points of broadband networks research. Current EPON systems can operate at up to 1 Gbps over distances of up to 20 km that is 40 times greater bandwidth delivery than ADSL2 can achieve at 1 km. Therefore it is an unchallenged fact that fiber, as a communication medium, offers almost infinite bandwidth over far greater distances relative to all other transport mediums.

This thesis addresses passive optical fiber systems for the access network. "Access," network is the so-called last mile (or recently re-named as the first mile) network that connects subscribers to the service provider's backbone networks.

1.2 Motivation and Objective

The unprecedented and unpredictable path of Internet expansion and new wave of its applications is opening new directions in the way we live and we do business [NOU04]. Bundling the services tailored for individual users are placing increasing demands from both business and private customers for flexible, transparent, high speed, customized high bandwidth services. Huge technological advances have been witnessed in the long haul in the last decade; however, access networks due to their wide penetration represent one of the most crucial problems in providing broadband services to subscribers. Increasing competition is one of the main drivers for the network evolution in the last mile segment of subscribers' battle ground for more cost effective bandwidth and services. The growing demand in high-bandwidth multimedia applications and the abundance of fiber optic have created a fertile environment for optical access network to seed and grow. However the survivals of proposed optical access techniques rely on their speed, performance, reliability, flexibility, and cost.

Telecom carriers are motivated to increase revenues by transporting various types of client traffic across their networks and optimizing their bandwidth utilization [FIS04]. Current access networks cannot cope with the increasing demand from both businesses and residential subscribers. There is an inevitable need to replace and/or upgrade current access networks with a newer one, being able to handle current and future traffic for the foreseeable future. Although having the most suitable characteristics for broadband communications, fiber has the lowest physical medium penetration in the access networks.

In this thesis, we examine thoroughly the technical and architectural aspects of the various alternative solutions for the optical access networks. We focus on the type of optical access networks that can offer bidirectional interactive services to residential users and businesses in the subscriber loop. In order to provide Quality of Service (QoS) for subscribers, we introduced a scheduling algorithm to differentiate users and applications.

1.3 Thesis Contributions

This thesis studies Ethernet passive optical networks as one of the candidate for the next generation access networks, being able to cope with the high speed and bandwidth demand. As various users are requiring different services, it is important to support quality of service (QoS) in access segment. In order to do so, due to the fact that in upstream transmission (i.e. from user to CO), medium is shared amongst users, there must be a mechanism to schedule users' transmission in a collision-free stream. Thus, the study of upstream scheduling algorithms in EPON is the main focus of this study. We concentrate on a kind of upstream scheduling algorithm supporting quality of service (QoS). We analytically model our algorithm and calculate the experienced delay by users for different types of services. In order to differentiate customers, we propose using priority scheduling to have preferential treatment for different services, due to the fact that every service has different requirements. Discrete event simulation then verifies the model and validates the obtained results. Summary of this contribution can be found in references [NOU04], [NOU05a], [NOU05b] and [NOU05c].

1.4 Thesis Overview

The thesis covers systems based on Ethernet passive optical networks (EPONs) that take advantage of a high-bandwidth, low maintenance, unpowered outside plant. The distinctions between them and the relative advantages/disadvantages of the various fiber-to-the-premises networks: asynchronous transfer mode PON (APON), broadband PON (BPON), Ethernet PON (EPON) and Gigabit PON (GPON), is the subject of the second chapter. The third chapter presents a study of quality of service in EPON in depth and in breadth, and various ways to provision QoS are investigated. In Chapter 4, performance analysis of the proposed scheme is provided and validation of results based on a discrete-event simulation program is shown. Finally conclusion of the study and closing thoughts are presented in Chapter 5.

Chapter 2 Ethernet Passive Optical Networks - State-of-the-Art

2.1 Introduction

There is a substantial need from both the users and the networks, dictating novel access infrastructure installations. The next generation access networks must be capable of carrying higher volumes of traffic to a poll of preferential customers with their individual set of requirements. The only suitable medium coping with the increased traffic is fiber optics that is the least deployed transmission medium in the access networks. Fiber-based solutions are now viewed as the cost-effective and future-proof conduit to provide high-bandwidth “triple play” service (voice, video, and data) to residential and business customers.

2.2 Passive optical network (PON)

One of the best solutions to bypass the capacity problems of access networks is by using a Passive Optical Network (PON). It is called a passive network because there is no active element in the signal path from Central Office (CO) to the subscriber. In the downstream direction (toward user) it is like a broadcast signal, i.e., point-to-multipoint (P2MP). Every Optical Network Unit (ONU) receives the transmitted signal by Optical Line Terminal (OLT)-but it picks up the signal destined for it by checking the destination address in the Medium Access Control (MAC). In the upstream direction (toward CO) it is a Multi Point-to-Point (MP2P) communication. With some arbitration mechanism ONUs transmit to OLT in a collision free manner.

Deploying PONs in access network will have several advantages:

- Longer distances between central office (CO) and customer premises (CP) can be covered, around 20 km.

- Fiber deployment is minimized in local exchange office and local loop.

- Higher bandwidth is available due to more fiber optic penetration.

- Video broadcast in downstream as either Internet Protocol (IP) video or analog video.

- No need for active elements in the optical signal path between central office and customer premises, minimizing the maintenance and operation costs.

- End-to-End transparency, providing higher bit rates upgrades [YON03].

Typically a PON consists of OLT, ONU, and passive optical splitters/combiners. The optical line terminal (OLT) is the host of the optical access network, consisting of optical access interfaces to service provider network [GUM04]. ONUs are located at or near the user site and transforming the broadband optical signal to the appropriate signal(s). Splitters/Combiners are connecting ONUs to OLT without any need for power provisioning.

A typical PON configuration is depicted in Figure 2.1.

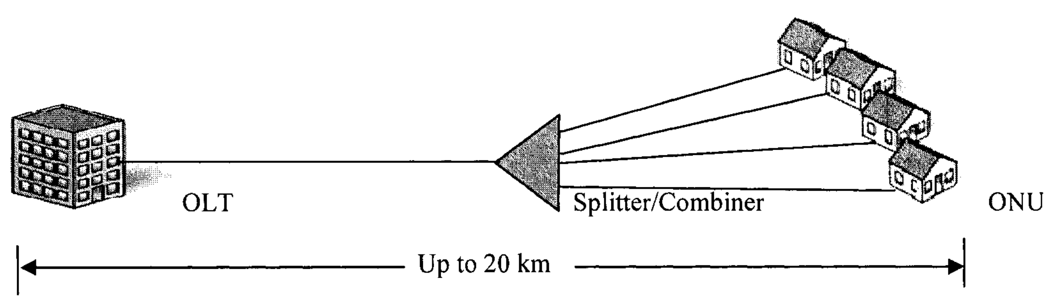

Figure 2.1 A typical PON Configuration.

PON can be implemented in various topologies, (e.g., tree, bus, and ring). Various PON topologies are depicted in Figure 2.2. The most common topology between OLT and ONU is tree topology, branching is achieved via splitters/combiners. In a bus topology all ONUs are connected to a common fiber from OLT. In a ring configuration ONUs are joined to a fiber running from OLT in a ring like connection. Redundancy can be added to any part of the network, like trunk of the tree in Figure 2.2 (d).

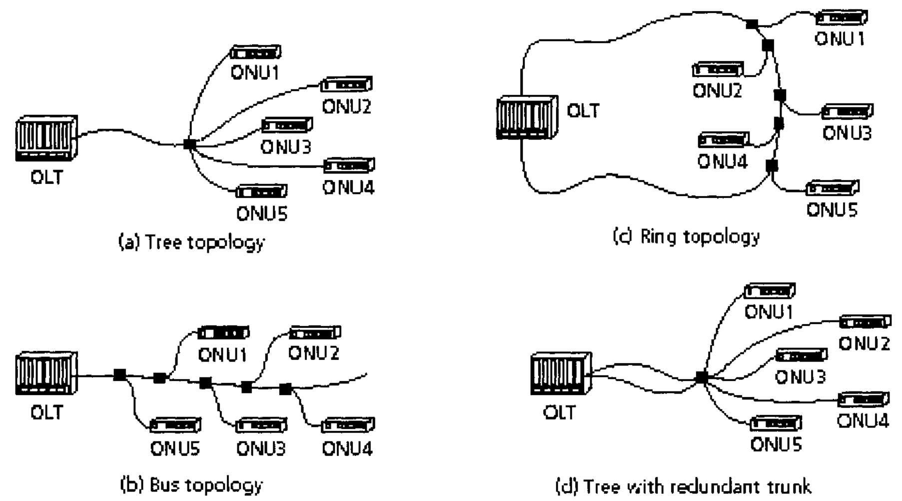

Figure 2.2 Various topologies of EPON [KRA02a].

In the following we briefly review current PON technologies.

APON was developed based on the ATM signaling layer. It supports downstream data rates of 155 Mbps and 622 Mbps, with ATM protocol overheads subtracted, generally support up to 448 Mbps of usable bandwidth in the downstream direction and in upstream 155 Mbps. Downstream traffic band is 1480 to 1580 nm and upstream band is 1260 to 1360nm, laying one single fiber between OLT and ONU [KAM02]. With having two fiber cables connecting OLT to each ONU, the downstream/upstream band is 1260 nm to 1360 nm. By Wavelength Division Multiplexing (WDM) overlaying around 1550 nm (i.e., 1539 nm to 1565 nm), an additional band is provided to enhance downstream traffic delivery capability.

BPON is a broadband version of PON supporting data rates of up to 622 Mbps in the upstream and 1.2 Gbps in the downstream direction. Being an APON variant, ATM signalling is used as in APON. In upstream every frame consists of 53 cells, each containing 53 byte data and 3 byte overhead. Downstream frames are filled with 56 cells each with 53 bytes.

GPON is a standard that supports both legacy traffic in its native format and IP traffic. GPON supports two methods of encapsulation: the ATM and GPON encapsulation method (GEM). The ATM method is an evolution of existing APON/BPON standards. With GEM, all traffic is mapped across the GPON network using a variant of SONET/SDH generic framing procedure (GFP). GPON supports asymmetric traffic rates of up to 2.5 Gbps and using generic framing procedure (GFP) that is SONET compatible enables end-to-end (E2E) Time Division Multiplexing (TDM) hierarchy. GPON supports both Ethernet signaling required for quality of service (QoS), and also supports the traditional mapping of voice trunks via SONET/SDH signaling via GEM.

Contention in upstream is resolved by using a Time Division Multiple Access (TDMA) approach via "*Grant*" mechanism in response to ONU's "*Request*" for sending upstream traffic. ONUs must be synchronized in order to avoid contention; a ranging process is used, so OLT will have an estimate of every ONU distance from it. After ranging, OLT put equalization delay at each ONU, and by that all ONU are virtually placed at same distance from OLT that is the distance of the furthest ONU [KIM03].

Due to point-to-multipoint (P2MP) nature of downstream traffic, every ONU is receiving even the traffic that it was not destined to that ONU, leading to security flaws. To alleviate the issue, scrambling is used by encrypting each payload with a periodically changing 24-bit key. Originally in standards ITU-T G983.1 through G983.3, there is no

QoS provisioning. However with new G983.4, with the use of dynamic bandwidth assignment (DBA), allowing the allocation of different bandwidth to different customers based on their agreed service level agreements (SLAs). The new set of standards, ITU-T G 983.5, with redundancy features, is addressing the reliability and survivability issues.

2.3 Ethernet PON (EPON)

With the emergence of Ethernet as the convergence layer in access networks, dominating 95 % of the LANs, speed increase from 10 Mbps to currently 10 Gbps, a new breed of PON, namely Ethernet PON (EPON) has evolved. Based on IEEE802.3ah standard it supports symmetric data rates of up to 1.25 Gbps [PES03]. It transports data encapsulated in variable-length Ethernet frames to and from users. Addressing the speed and bandwidth gap between WAN and LAN in a cost effective and efficient manner relying on the widely deployed Ethernet protocol as a promising solution.

In downstream, packets are transmitted to ONUs, but ONUs will filter the packets intended to them if the MAC address is in match with their address. All other received non-system packets are discarded. In upstream, TDM is used, by assigning different timeslots, ONUs are able to transmit in a collision-free fashion.

Maturity of Ethernet and its wide deployment, cost, speed scalability, simplicity, and efficiency, are all resulting EPON to be a real contender among PON technologies battling for their access market stake. End-to-end QoS provisioning in Ethernet-based networks remain an open problem due to the lack of scalability of Virtual Local Area Networks (VLANs) [GAG03]. Scalability is one of the major advantages of Ethernet compared to ATM and Frame Relay. ATM scales up to 622 Mbps, and Frame Relay run from 1.5 Mbps to 45 Mbps, but Ethernet offers rates from 1 Mbps to 10 Gbps in increments of 1 Mbps. In EPON, IP is the sole service layer. This makes it possible to deploy a uniform service model throughout the network. This also homogenizes packet delivery as traffic originates and gets transported as IP over Ethernet. Therefore packet life cycle all is in Ethernet.

In the downstream, packets are broadcasted by the OLT in a Point-to-Multi Point (P2MP) fashion. After passing through splitter stage(s), ONUs receives packets. Packets are selectively extracted based on their MAC address. A typical downstream packet scheduling is depicted in Figure 2.3.

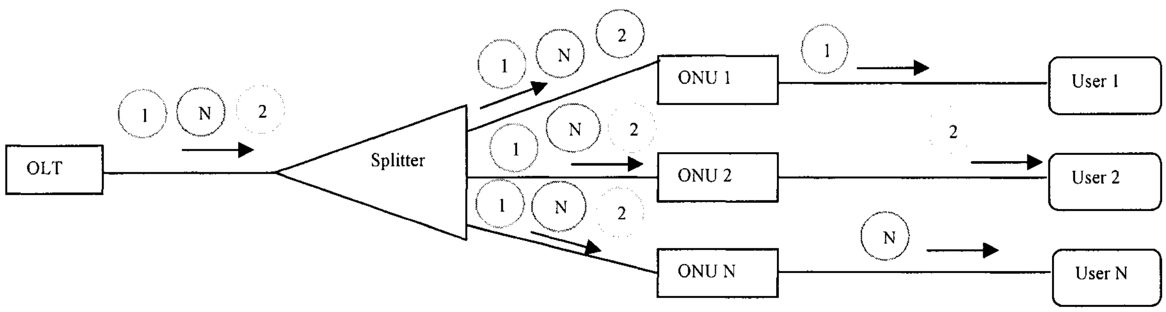

Figure 2.3 Downstream scheduling.

In the upstream direction (toward CO) it is a Multi Point-to-Point (MP2P) communication. Because of the directional coupling of combiner, ONUs' transmitted packets will reach only the OLT, and not any other ONUs. EPON's upstream characteristic in this sense is similar to that of Point-to-Point (P2P) architecture. However, in upstream ONUs may transmit simultaneously causing packet collision. To alleviate this issue some arbitration mechanism in the upstream is required. A typical upstream scheduling is depicted in Figure 2.4.

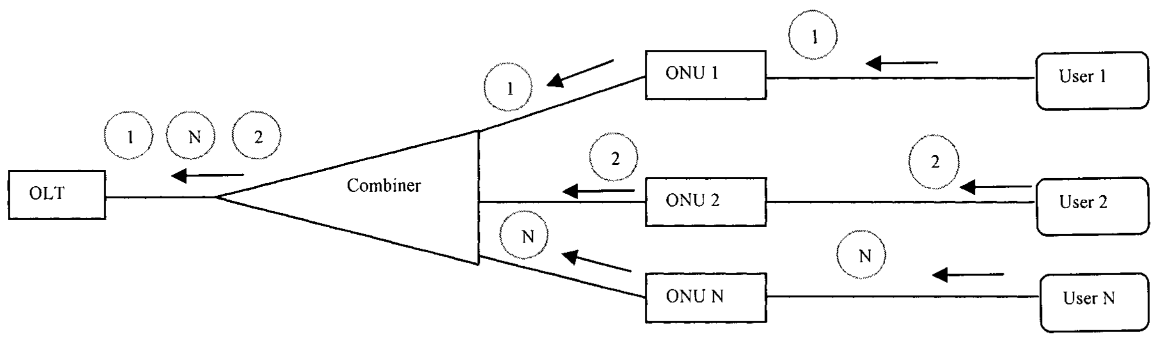

Figure 2.4 Upstream scheduling.

Various passive optical network (PON) technologies reviewed here and their main characteristics are compared in Table 2.1 [NOU04].

Table 2.1 Various PON characteristic parameters.

PON Technology	APON	BPON	EPON	GPON	
Protocol	ATM	ATM	Ethernet +FEC	ATM	ATM and GEM
Standard	ITU-T G.983.1 (including Amendment 1)	ITU-T G.983.3	IEEE802.3ah	ITU-T G.983.1 (Amendment 2)	ITU-T G.984
Data Packet Cell Size	56 bytes (53 ATM cell-load and 3 overhead)	53 bytes (48 payload and 5 overhead)	1,518 bytes	Variable size, from 53 bytes up to 1,518	
Maximum Bandwidth	155Mbps, 622Mbps downstream; 155Mbps upstream	1.2Gbps downstream; 622 Mbps upstream	Up to symmetric 1.25Gbps	Downstream configurable from 1.2Gbps to 2.5Gbps; upstream configurable in 155 Mbps, 622 Mbps, 1.25Gbps, or 2.5Gbps	
Downstream Wavelength	1480-1580 nm	1480 nm to 1500 nm	1550 nm	1480 nm to 1500 nm	
Upstream Wavelength	1260-1360 nm	1260 nm to 1360 nm	1310 nm	1260 nm to 1360 nm	
Traffic Modes	ATM	ATM	Ethernet	ATM, Ethernet, TDM	
Voice	TDM	TDM	VoIP or TDM	Native TDM	
Video	Not supported	1550 nm overlay	1550 nm overlay	Either as RF or over IP	
ODN Classes Supported	B, and C	A, B, and C	A and B	A, B, and C	
Split Ratio	Up to 32	Up to 32	16/up to 32	Up to 64	
Fiber Type **Avg. Attenuation**	ITU-T G.652 (single or Dual fiber) 1310 nm 0.36 dB/km 1550 nm 0.23 db/km	ITU-T G.652 (single fiber) 1310 nm 0.36 dB/km 1550 nm 0.23 db/km	1000BASE-PX10: single fiber 1000BASE-PX20: dual fiber (no fiber type specified)	ITU-T G.652 (single or Dual fiber) 1310 nm 0.36 dB/km 1550 nm 0.23 db/km	
Reachable Distance	20 km	20 km	1000BASE-PX10: 10 km 1000BASE-PX20: 20 km	20 km	10 km FP 20 km diff. fiber types
Optical Budget	25 dB	25 dB	PX10: 21 dB PX20: 26 dB	25 dB	
Data Rates	Symmetric: 155.52/622.08 Mbps Asymmetric: DS: 622.08Mbps US: 155.52 Mbps	Symmetric: 155.52/622.08 Mbps Asymmetric: DS: 622.08Mbps US: 155.52 Mbps	Symmetric: 1.25Gbps	Symmetric: 155.52/622.08 Mbps Asymmetric: DS: 622.08/1244.16 Mbps US: 155.52/622.08 Mbps	Symmetric: 1244.16 /2488.32 Mbps Asymmetric: DS: 1244.16/2488.32 Mbps US: 155.52/622.08/ 1244.16 Mbps
GigE Interfaces	No	No	Yes	Yes	

Class A: 5-20 dB between ONU and OLT
Class B: 10-25 dB between ONU and OLT
Class C: 15-30 dB between ONU and OLT

GEM: GPON Encapsulated Method
FEC: Forward Error Correction

Power Budget Calculation in PON

Optical power budget is calculated as the following;

$$P_{TX} = P_{RX} + L_L + M_{Sys}$$

where P_{TX} is the transmit power, P_{RX} is the receiver sensitivity, L_L is the link loss and M_{Sys} is the system margin, typically is around 6-8 dB.
Link loss is calculated as following;

$$L_L = \alpha_f * L + \alpha_{con} + \alpha_{splice} + \alpha_{dev}$$

where α_f is the fiber loss, L is the maximum fiber run length, α_{con} is the connector loss, α_{splice} is the fiber splice loss, and α_{dev} is the device loss.

2.4 EPON's principle of Operation

The IEEE802.3 defines two basic configurations for an Ethernet network. One being deployment over a shared medium, and using Carrier Sense Multiple Access with Collision Detection (CSMA/CD) protocol and the other one being connecting stations through a switch, using a full duplex point-to-point links [DIX03]. Properties of EPON are in a way that it is a combination of both of these configurations.

In the downstream path, medium is shared among all ONUs. Ethernet frames are broadcasted by OLT, passing through a 1: N passive splitter, to ONUs. Received packets are extracted if their media access control (MAC) address is a match with ONU MAC address.

Due to the directional properties of passive optical combiner (s) in the upstream path, transmission from any ONU will reach only to the OLT and it will not be received by any other ONUs. Hence in the upstream the medium is considered to be a multi point-to-point (MP2P). However without any arbitration mechanism to control upstream transmission, packets might collide and bandwidth may not be fairly shared among ONUs.

Again due to the directional properties of passive optical combiner, ONUs cannot detect any collision at OLT that makes the implementation of a contention-based media access control very difficult. OLT can detect the collision and sending a jamming signal for this purpose to ONUs, but because of the experienced propagation delays, efficiency is degraded. Contention-based schemes are providing statistical averages that timely delivery to nodes might not be guaranteed, which causes unsatisfactory results in residential access. However, this issue is less important in enterprise networks due to shorter and over-provisioned links carrying mostly data traffic.

Non-contention and TDM-based schemes are proposed to introduce determinism in frame delivery. OLT assigns timeslots to ONUs after they are synchronized to a common time reference. If the buffered frames fit into assigned time slot, they will be transmitted with the full link speed. If the buffer is empty, 10-bit idle characters are transmitted. Two main timeslot assignments are static TDMA with fixed timeslots and dynamic timeslot allocation based on the instantaneous queue occupancies of ONUs. Other timeslot schemes can be based on quality of service metrics, service level agreements contracts, service prioritizations, etc.

Timeslot assignments can be done in centralized and decentralized way. In centralized approach, OLT is responsible for allocating timeslots. However in decentralized approach, ONUs are allocating timeslots based on the arrived traffic. Due to the need for informing other ONUs of the intention for transmission and the EPON limitations that there is no direct communication between ONUs, this scheme is posing some constraint on the EPON. Centralized-schemes are realizable in EPON for allocating channel access to ONUs.

With centralized approach, OLT is aware of network status and can make decisions based on the ONU states. Hence, ONUs have much simpler functionalities and they will be cheaper that leads to less costly ONUs for subscribers. However, in centralized approach, for OLT in order to make a well informed time allocation decision requires of a priori knowledge of ONUs' buffer occupancy. The bursty nature of arrived traffic to ONUs rules out buffer occupancy prediction mechanisms. Polling schemes are used to inform OLT of ONUs buffer status with the use of *Grant* and *Request* messages in such a way to overcome this issue. ONUs will inform OLT of any status changes by sending a *Request* message. In response to *Request* messages received from ONUs, OLT will process these messages and assigns transmission timeslots with the start time, and duration of the granted window. This information is delivered using *Grant* messages.

2.5 Multi-Point Control Protocol (MPCP)

Multi-point control protocol was defined by IEEE 802.3ah task force. Its operation is based on the two *Gate* and *Report* messages. *Gate* is sent in downstream and it is granting ONUs upstream transmission timeslot. *Report* is an upstream message to inform the OLT of any ONU conditions, e.g., queue occupancy, arrived traffic class of service, arrival rate, etc.

Two modes of operation have been defined for MPCP: auto-discovery (initialization) and normal operation. The former is used to detect ONUs joining the network and discovering round-trip delay and MAC address of the specific ONU, and other future features are to be defined later. Normal mode is used to grant transmission timeslots to all initialized ONUs.

ONUs can report their local conditions simultaneously; auto discovery is a contention-based procedure to solve this issue. MPCP is described below:

1. OLT allocates an unassigned initialization timeslot with a certain time length depending on the roundtrip time and transmission window.

2. OLT informs ONU of its granted time slot start time and duration by sending a *Gate* message and timestamp this message.

3. Only new joining ONUs will respond to the initialization message. After receiving this message, ONU will adjust its local time to the arriving *Gate* message timestamp.

4. When ONUs local time reaches its transmission start time, the ONU will transmit its *Report* message containing its address and the time that the *Report* was sent.

5. OLT by receiving the *Report* message from newly joined ONUs will record their MAC addresses and their round trip times.

After receiving *Gate* message, newly joined ONUs transmit their *Report* messages and collision might occur. In that case, the ONUs that have collided *Report* messages will not get any timeslot. The collided ONUs will arbitrarily after some random number (e.g. exponential back-off) of initialization cycles will attempt to send another *Report* message.

Normal operation of MPCP is described below. The main purpose for MPCP is to deliver the operational messages from OLT to any ONU and vice versa, and it does not consider any bandwidth allocation schemes.

1. MPCP receives a request to transmit a *Gate* message to a particular ONU from MAC. It will consist of the start time and the duration of the transmission.

2. OLT and ONUs will adjust their clocks. MPCP will timestamp the *Gate* message.

3. ONUs after verifying the MAC address of the received *Gate* message will set their local registers with the transmission start time and duration that they received. ONU will check the arrival time of the *Gate* message with the timestamp of the *Gate* message. If the difference is more than a pre-defined threshold, the ONU will assume it has lost its synchronization and will switch itself to un-initialized mode. It will not be able to transmit and has to wait for the next initialization message.

4. If the arrival time of the *Gate* is close to the timestamp enclosed in the *Gate*, ONU will update its local clock to the timestamp. By reaching the transmission start time, ONU will start transmission for the allocated duration. Frames are not fragmented. Frames that will not fit into the timeslot will be deferred for next timeslot.

Report messages are sent in the assigned transmission windows along with the data frames. *Report* messages are sent either automatically or on-demand. A *Report* message is generated in the MAC control client layer and is time stamped in the MAC control [DIX03]. *Report* will contain ONU's local condition such as queue length. ONUs should take into consideration for following overhead, namely 64-bit frame preamble and 96-bit inter frame gap (IFG) associated with every frame.

After the *Report* arrival in OLT, it is passed to the MAC control client layer responsible for bandwidth allocation decision. OLT will recalculate the round trip time to that ONU. Due to changes in refractive index caused by temperature change, there might be some deviation from new calculated RTT to previously measured RTT. If this deviation is greater than some expected value, OLT will consider that ONU to be out of synchronization and it will not grant any further transmission window until that ONU is initialized again. For the sake of simplicity, other details are not explained here and more details are under discussion within the IEEE 802.3ah task force.

2.6 Literature Survey

It is very important to be able to provide quality of service in access networks in order to comply with different subscribers' service requirements. Due to the nature of passive optical networks, in the downstream we are having streamed broadcast from central office to subscribers, and they will be able to capture only the information that was intended to be received by them. However, in the upstream, due to the multipoint-to-point nature of the network, it is crucial to provide service prioritization among subscribers. In the following we will review researches that had been done and undergoing in this area.

Kramer et al. in [KRA02b, and KRA02c] presented one of the early papers in this area, based on a protocol called interleaved polling with adaptive cycle time (IPACT). It is an OLT-based polling scheme, in a manner that next ONU is polled before the transmission from the previous one has arrived. IPACT provides statistical multiplexing for ONUs resulting in efficient channel utilization.

In this scheme OLT has knowledge of round trip time (RTT) to/from any ONUs and knows how many bytes are waiting in each ONU's buffer. OLT will allow Optical Network Terminals (ONTs) to send with *Grant* message in response to their *Request* messages. OLT will broadcast data to all ONUs, ONUs will only capture the data destined to them based on the MAC addresses contained within data. Guard times provided in order to compensate for RTT fluctuations and variations of processing times.

In this scheme, ONU synchronization is not required. Also contrary to TDMA schemes, ONUs are not required to perform the ranging process that makes them to appear equidistant from OLT. OLT will perform all scheduling and bandwidth allocation functions with regard to provided information by every single ONU.

For solving the granted window size they propose several schemes (i.e. fixed service, limited service, constant credit, linear credit, and elastic service). In fixed service, regardless of requested window size, maximum window is always granted. In limited service, requested window size is granted if it is not greater than the maximum window size. In the constant credit, a constant credit is added to the requested window size. In the linear credit, size of the added credit is proportional to the requested window. And finally, in elastic service, the maximum window is granted in a manner that the accumulated size of last N Grants does not exceed N*maximum window size. IPACT minimizes unused bandwidth by using time polling messages that are interleaved on downstream traffic. There are two issues with IPACT, it does not consider the fact that different ONUs have different bandwidth requirements and IPACT's dynamic bandwidth allocation introduces burstiness in the upstream traffic. IPACT can provide on-demand bandwidth based on user queue length, but it has some difficulty to provide QoS guarantee to different users and it does not support different QoS levels for a single ONU. IPACT can provide a small delay under the light load, with an increase in the load, the average delay performance deteriorates quickly. Due to the fact that IPACT concentrates on bandwidth utilization, it also cannot provide different QoS guarantee to sources especially delay and delay variation bound.

Ma et al. in [MA03] provided a bandwidth guarantee polling (BGP) MAC protocol that allowing the upstream bandwidth to be shared based on the service level agreements (SLAs). This scheme provides premium services to business customers and best effort (BE) to residential customers preferring low-cost services. Their algorithm is mitigating two issues in EPON. First, difficulties in implementing conventional contention-based carrier sense multiple access protocols and secondly, issue of providing controllable and guaranteed bandwidth allocation to each ONU. A reservation-based time division multiple access scheduling scheme utilizing a common wavelength that makes every ONU using same kind of transceivers resulting in more cost effective equipments. ONUs are divided into two groups, one group has ONU members requiring guaranteed bandwidth and the other group contains members without the guaranteed bandwidth requirements.

By combining polling with TDM, two advantages of overcoming TDM synchronization problems and TDM inflexibility and poor efficiency are achieved. This polling scheme is a roll-call polling (i.e. OLT polls ONUs consequently in an adaptive manner to allow ONUs to send their data, and keeps an entry table of polled ONUs). Table entries include one bandwidth unit for either bandwidth guaranteed ONUs or non-bandwidth guaranteed ONUs. Table entries include ONUs' number and propagation delay to OLT. Bandwidth guaranteed ONUs will be polled more than once in a polling cycle and they will consume more bandwidth. Their proposed BGP algorithm has two parts, scheduling and even distribution algorithm (EDA). Scheduling is responsible for polling ONUs. On the other hand EDA, evenly distributes multiple entries of the same bandwidth guaranteed ONUs among all other table entries. Their simulation results indicate that the bandwidth guaranteed ONUs will face less average delay. And also non-bandwidth guaranteed ONUs will suffer more delay than the bandwidth guaranteed ONUs with more than one entry, and it is very similar to IPACT under light loads. Their results prove the more entries a bandwidth guaranteed ONU has, it results less average delay, higher throughput, shorter queue length and lower loss rate. The bandwidth guaranteed ONUs with multiple entries have a better performance than an ONU under IPACT and on contrary to IPACT scheme, it can provide differentiated services.

Nikolova et al. in [NIK03a], presented a novel dynamic bandwidth allocation (DBA) scheme using multipoint control protocol (MPCP) with threshold reporting and also capable for inter and intra-ONU priority scheduling. Due to the fact that frames are not fragmented in EPON, concept of threshold reporting is introduced in order to achieve higher bandwidth efficiency. In front of each frame there exists a preamble of 8 bytes and between two frames there is at least a 12 byte inter packet gap (IPG) [IEE02]. Time elapses between two executions of scheduling algorithm are defined as a cycle confined by lower and upper bounds ($T_{\min}$ and $T_{\max}$). Algorithm will allocate B_i between $B_{\min}$ and $B_{\max}$, where B_i is found by multiplying T_i by line rate. During each cycle, each ONU is granted exactly one transmission window, at least sufficiently large enough for sending one *Report* message. They defined two priority scheduling schemes, full priority scheduling (FPS) and Interval priority scheduling (IPS). The former is the normal priority scheduling scheme and in the latter scheme, the ONU remembers the total number of bytes (per queue) that it reported in last *Report* message. One threshold is assigned either statically or dynamically to each queue of any ONU. Three varieties of this algorithm were proposed, IPSA, R-FPSA and R-IPSA which, they differ in their Constant Bit Rate

(CBR) traffic allocation as well as in their intra-ONU scheduling. With IPSA, bandwidth losses were reduced, but with the expense of higher average queueing delay and jitter for highest priority in high loads. With R-IPSA (combining IPSA and CBR scheduling), high load delay increments were reduced. Combining FPS and CBR scheduling at ONU results R-FPSA, hence improving average queueing delay with having lower bandwidth efficiency.

Zhang et al. [ZHA03] suggested a dual deterministic effective bandwidth (DEB) generalized processor sharing (GPS) scheduler to provide delay constraint and lossless QoS guarantee and avoiding QoS degradation due to statistical multiplexing. Scheduling is performed in two locations, master scheduler at OLT and slave scheduler at ONU. In the former, bandwidth weights are assigned according to QoS mapping and in the latter, by optimizing delay variation, bandwidth utilization is improved. A positive real number, namely weight is assigned to each traffic stream and bandwidth is allocated dynamically based on their corresponding weights and traffic load. In addition to delay bounds, a bound on variation of delays of one source and (or) different sources is guaranteed. Delay variation bounds are providing source synchronization and service fairness is achieved and results in more efficient bandwidth and buffer utilization. Different types of traffic sources are mapped into QoS-aware and best-effort sources. With weight assignment, each QoS-aware traffic source receives a minimum service rate equal to its deterministic effective bandwidth, ensuring a lossless deterministic QoS and delay-constrained service. In order to control the inter-source delay variation constraint, a slave scheduler is added to each ONU. Slave scheduler will record all arrival and service time stamp of every packet of the incoming traffic. It also calculates each source delay and inter-source delay variations. Backlog clearing time is calculated by collecting the queue length information of each QoS source. Any source completing its backlog, the according bandwidth will be released and re-allocated among the still backlogged sources. On contrary to most of the DBA schemes that only OLT schedules upstream transmission, the proposed scheme has two schedulers, one at OLT and the other is located at each ONU. OLT scheduler receives requests from different QoS classes collected from different ONU and multiplexed to calculate bandwidth allocation among QoS classes. ONU scheduler, further select the proper bandwidth or grants allocated to it from OLT based on its own queueing status and QoS contract. This study did not investigate the performance issues corresponding to queue length effect on different services at ONUs.

Zheng et al. in [ZHE05] proposed an adaptive polling protocol called the earliest packet first (EPF) algorithms that schedules the transmission order of different ONUs based on the arrival time of the first packet waiting in the queue of each ONU. Due to the bursty nature of traffic, with fixed timeslot allocation for each ONU, high bandwidth utilization cannot be achieved. A multipoint control protocol (MPCP) has been widely adopted to implement dynamic bandwidth allocation, using two messages of *Gate* and *Report*. A *Gate* message is used by the OLT to allocate a transmission window to an ONU. A *Report* message is used by an ONU to report its local conditions to the OLT. A polling protocol must schedule not only the transmission start time and duration of each ONU transmission but also the transmission order of different ONUs. Instead of round robin (RR), which is simple to implement, but it is not taking into account the instantaneous traffic conditions at each ONU, they propose the earliest packet first (EPF) algorithm. In order to support this scheme, *Report* message from any ONU must carry additional information on the arrival time of the first packet waiting in the queue of that ONU and the polling table at the OLT must maintain this information for each ONU. The polling table maintains not only the ONUs' bandwidth demand and round trip time to each ONU but also the arrival time of the first packet waiting in the ONUs' queue. The OLT schedules the transmission of different ONUs in an ascending order of the arrival time in the polling table. In order to increase bandwidth utilization, an interleaved polling mechanism is used that allows the OLT to poll the next ONU before the transmission from the previous ONU arrives. It was shown that under medium traffic load the adaptive polling scheme can reduce the average packet delay and maximum packet delay compared to a polling protocol using round robin scheme. However, these improvements come with the price of having a more comprehensive polling table.

Lee et al. [LEE04] suggested a two-step scheduling algorithm to support dual bandwidth allocation policies. Their algorithm provides a simple scheduling process to combine static bandwidth allocation (SBA) and dynamic bandwidth allocation (DBA). The main idea of the proposed algorithm is to separate the process of G*rant* start-time calculation from the process of G*rant* generation. In the first step, G*rant* messages are generated by each module according to its own bandwidth allocation algorithm and multiplexed by the scheduler. In the second step, multiplexed G*rant* messages are converged with downstream data traffic. Four priorities in descending order of their precedence is SBA, minimum bandwidth queue, DBA, and auto-discovery *GATE* queue. With the simulation, they showed the proposed algorithm exhibits similar queueing delay performance characteristics and service characteristics as SBA-only or DBA-only schemes.

In [HSU03], Hsueh et al. introduced a hybrid slot size/rate (HSSR) media access control protocol to enforce quality of service and fairness in EPONs. For simplicity they consider two traffic classes, high priority and best effort. Each frame is divided into two parts. One steady part for high priority traffic and the other is the dynamic part for best effort traffic. Their proposed MAC protocol minimizes packet delays and delay variations for the higher priority traffic, while increasing throughput efficiency for the best effort traffic by reclassification of incoming traffic to take advantage of the available bandwidth in the guaranteed time slots. By dynamically allocating the excess bandwidth according to users' service rates, fairness is achieved. High priority traffic is always assigned in the fixed location of the frame to minimize delay variations, and best effort traffic can be reclassified and assigned in high priority time slots for efficient use of available bandwidth.

Luo et al. [LUO05] presented a dynamic bandwidth allocation with multiple services (DBAM) algorithm to support various types of traffic. Rather than providing multiple services among ONUs and among end users separately, DBAM integrate both of them into the *REPORT/GRANT* mechanism with class-based bandwidth allocation. DBAM takes frames arriving during waiting time into consideration, and hence reduces frame delay and queue length.

2.7 Summary

This chapter presented the basic concept behind passive optical network and gave a more detail discussion of EPON systems. EPON's principle of operation and the standardized protocol governing it was also presented. A comprehensive literature survey of the proposed algorithms to support Quality of Service (QoS) also provided different approaches to provision desired quality for applications. In the next chapter, Quality of Service in EPON and queueing analysis is provided.

Chapter 3 Priority Queueing Scheduling for QoS Management in EPONs

3.1 Introduction

In real-time communications [ADA97], the quality of communication depends on the time at which messages are successfully delivered to recipients. Quality of Service (QoS) refers to the capability of a network to provide better service to selected network traffic over various technologies, including Frame Relay, Asynchronous Transfer Mode (ATM), Ethernet and 802.1 networks, SONET, and IP-routed networks that may use any or all of these underlying technologies. The primary goal of QoS is to provide priority including dedicated bandwidth, controlled jitter and latency (required by some real-time and interactive traffic), and improved loss characteristics. Also of equal importance is making sure that providing priority for one or more flows does not make other flows fail. QoS technologies provide the elemental building blocks that will be used for future business applications in campus, WAN and service provider networks. Network Quality of Service (QoS) requirements are typically formulated in terms of performance metrics such as deadline, delay jitter, and loss rate.

The deadline is the specific maximum delivery delay bound, which is an application-layer, end-to-end timing constraints. Another important performance metric is delay jitter, which is the maximum variation in delay experienced by packets in a single connection. Some real-time applications such as non-interactive television and audio broadcasting require bounds on jitter but not on delay. The delay jitter is usually controlled, at cost of additional delay, by using buffers. For continuous media playback applications, the ideal case would be that the network introduces only constant delay, or zero delay-jitter. Having a bounded delay jitter service from network makes it possible for the destination to calculate the amount of buffer space needed to eliminate the jitter. The smaller the jitter bound, the less amount of buffer space is needed. Since it is more important to provide end-to-end delay and delay-jitter bounds than average low delay for guaranteed service (GS) class, packets arriving too earlier may not even be desirable in such an environment. In fact, the earlier a packet arrives before its delay bound or playback point, the longer it needs to occupy the buffer. This is one of the most important differences between the performance requirements of the guaranteed-performance service and the best effort service provided by the traditional computer networks. Performance bounds are more important for the guaranteed service while average performance indices are more important for the best effort service. The third important parameter is the loss probability. Loss rate will specify tolerable loss for certain traffic undergoing transmission in the medium. Packet loss can occur due to buffer overflow or delay bound violation. A statistical service allows a nonzero loss probability while a deterministic service guarantees zero loss. With a deterministic service, all packets will meet their performance requirements even in the worst case scenario. With a statistical service, stochastic or probabilistic bounds are provided instead of worst case bounds. Statistical service allows the network to overbook resources beyond the worst case requirements, thus may increase the overall network utilization by exploiting statistical multiplexing gain.

3.2 Quality of Service (QoS)

New challenges are dictating new user requirements from Service Providers (SP) in today's competitive telecom market worldwide. As a result of choosing broadband services by users and enterprises, an opportunity for offering a combination of feature-enhanced voice and bandwidth-intensive multimedia services is materialized by service providers.

3.2.1 Customer QoS preferences

Different customers can be classified based on their QoS preferences. Following groups are distinguished:

- Best effort traffic users: the amount of bandwidth and quality of service are not a main concern for the customers' applications: e.g. e-mail, file transfer protocol (ftp), and web traffic.

- Users with hard preferences: customers won't be satisfied if the available QoS is under certain threshold. For example we can consider an audio application that under certain bit rates the quality of transported audio will not be acceptable.

- QoS oriented users: this category of customers due to their applications requirements they are demanding higher QoS levels, e.g., real-time video/audio conferencing.

3.3 Service Differentiation

The basic goal of Differentiated Services (DiffServ) architecture [BLA98] is to fulfill the performance requirements of users. Users requesting a certain performance level (e.g. required bandwidth, delay, and delay variation) and hence network provides it as long as the user traffic has certain characteristics. The performance level provided and the characteristics of the traffic to be injected in the network are defined in SLA.

Differentiated Services define a model for implementing scalable differentiation of QoS in the Internet. The main approaches of service differentiation are to (a) classify traffic at the boundaries of network, and (b) regulate (condition) this traffic at the boundaries. The classification operation entails the assignment of the traffic to Behavioral Aggregate (BAs). These behavioral aggregates are a collection of packets with common characteristics, as far as how they are identified and treated by the network. The network classifies the packets based on the content of the packet headers. The idea is to have a small number of classifications to simplify the allocation of resources for the traffic classes. As can be seen in Figure 3.1 each one of the queue will provide different treatment for user packets based on the requested Service Level Agreement (SLA).

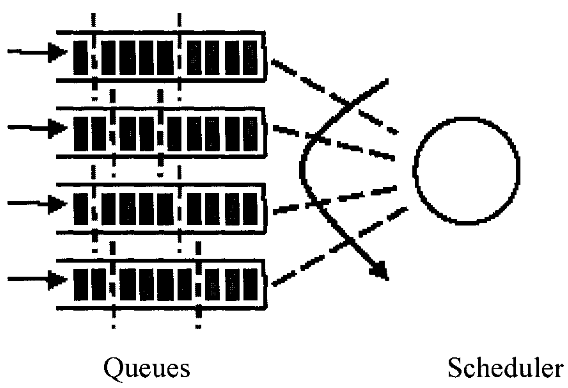

Figure 3.1.: Four different classes of service.

3.4 Queueing Analysis

Queueing algorithms can be thought of as allocating three nearly independent quantities: bandwidth (which packets get transmitted), promptness (when do those packets get transmitted), and buffer space (which packets are discarded by the gateway). The most common queueing algorithms are First-Come-First-Serve (FCFS) or First-In-First-Out (FIFO) which essentially relegates all congestion control to the sources, since the order of arrival completely determines the bandwidth, promptness, and buffer space allocations. That means FCFS inextricably intervenes these three allocation issues. There may indeed be flow control algorithms that, when universally implemented throughout a network with FCFS gateways can overcome these limitations and provide reasonably fair and efficient congestion control. With today's drivers and decentralized computing environment, it is unrealistic to expect universal implementation of any given flow control algorithm. This is not merely a question of standards, but also one of compliance. Even if a universal standard was adopted, malfunctioning hardware and software could violate the standard, and there is always the possibility that individuals would alter the algorithms on their own machine to improve their performance at the expense of others. Consequently, congestion control algorithm should function well even in the presence of ill-behaved sources.

Several algorithms, such as Drop Tail, Random Exponential Marking, Fair Queueing, Virtual Queue, etc., proposed recently try to provide an efficient solution to this problem. The design of a traffic scheduling algorithm involves an inevitable tradeoff among its delay, complexity of implementation, and fairness. Among the three, the delay and implementation complexity are clearly the most important criteria for the selection of an algorithm for use in a real system. While the fairness properties of the algorithm affect only the short-term distribution of service offered to the sessions sharing the link, a larger delay bound implies increased burstiness of the session at the output of the scheduler, thus increasing the amount of buffering needed in the switches to avoid packet losses. In addition to minimizing the end-to-end delay in a network of servers, the delay behavior of an ideal algorithm must include:

- Insensitivity to traffic patterns of other sessions (isolation);
- Delay bounds that are independent of the number of sessions sharing the outgoing link;

- Ability to control the delay bound of a session without depending on the internal parameters of the scheduler.

3.4.1 Priority Queueing

Consider the M/G/1 system with the difference that arriving customers are divided into n different priority classes. Class 1 has the highest priority amongst all; class 2 has the second highest and so on. The arrival rate and the first two moments of service time of each class k are denoted by λ_k, $\overline{X_k} = \frac{1}{\mu_k}$, and $\overline{X_k^2}$, respectively. The arrival processes of all classes are assumed independent, Poisson, and independent of the service times.

Non-preemptive Priority: We consider the Non-preemptive Priority rule whereby a customer undergoing service is allowed to complete service without interruption even if a customer of higher priority arrives in the meantime. A separate queue is maintained for each priority class. When the server becomes free, the first customer of the highest nonempty priority queue enters service. This priority rule is one of the most appropriate for modeling packet transmission systems. Here an equation for average delay for each priority class will be developed. This is similar to the P-K formula and admits a similar derivation. Denote

N_q^k = Average number in queue for priority k ;
W_q = Average queueing time for priority k ;
$\rho_k = \frac{\lambda_k}{\mu_k}$ = System utilization for priority k ;
R = Mean residual service time.

We assume that the overall system utilization is less than 1, that is

$$\rho_1 + \rho_2 + \ldots + \rho_n < 1 \tag{3.1}$$

When this assumption is not satisfied, there will be some priority class k such that the average delay of customers of priority k and lower will be infinite while the average delay of customers of priority higher than k will be finite.

As in the derivation of the P-K formula [KLE75a], we have for the highest priority,

$$W_1 = R + \frac{1}{\mu_1} N_Q^1 \tag{3.2}$$

The derivation is similar for all priority classes k > 1. The general formula for the queueing waiting time for priority k is:

$$W_k = \frac{R}{(1-\rho_1-\ldots-\rho_{k-1})(1-\rho_1-\ldots-\rho_k)} \tag{3.3}$$

Proof of these results can be obtained from any standard textbook on queueing theory. The average delay per customer of class k is,

$$T_k = \frac{1}{\mu_k} + W_k \tag{3.4}$$

The mean residual service time R can be derived as for the P-K formula. We have

$$R = \frac{1}{2} \sum_{i=1}^{n} \lambda_i \overline{X_i^2} \tag{3.5}$$

The average waiting time in queue and the average delay per customer for each class is obtained by combining formula obtained for W_k with the formula obtained for R:

$$W_k = \frac{\sum_{i=1}^{n} \lambda_i \overline{X_i^2}}{2(1-\rho_1-\ldots-\rho_{k-1})(1-p_1-\ldots-\rho_k)} \tag{3.6}$$

The analysis given above does not extend easily to the case of multiple servers, primarily because there is no simple formula for the mean residual time R. If, however, the service times of all priority classes are identically and exponentially distributed, there is a convenient characterization of R.

Note that it is possible to affect the average delay per customer by choosing the priority classes appropriately. It is generally true that average delay tends to be reduced when customers with short service times are given higher priority. (For an example from common experience, consider the supermarket practice of having special checkout counters for customers with few items. A similar situation can be seen in copying machine waiting lines, where people often give priority to others who need to make just a few copies.)

3.5 Quality of Service in EPON

A key issue for quality of service (QoS) provisioning in EPON is adding some enhancements to multipoint control protocol (MPCP) in order to provide QoS [NIK03b]. By adding these extensions, EPON will support users' quality of service and service differentiation is achieved. Based on the service and applications requirements variability, every user's SLA is distinct and distinguished from others. Services could be differentiated based on their delay, jitter, bandwidth, tolerable loss rate, and etc. in the following we review bandwidth allocation schemes in EPON.

3.5.1 Bandwidth Allocation Schemes

In order to distribute bandwidth amongst all users in upstream, an arbitration mechanism should be in place to distribute bandwidth. There are two main ways of bandwidth allocation: i) static bandwidth allocation (SBA) and ii) dynamic bandwidth allocation (DBA). In the former, each ONU is receiving a fixed timeslot with a fixed window size. In the latter, bandwidth is distributed in a dynamic manner based on the

ONU requirements. The OLT dynamically allocates bandwidth based on the queue status of the ONUs that is informed through a *REPORT* message. In the next sections these two schemes are described.

3.5.1.1 Static Bandwidth Allocation (SBA)

In this scheme bandwidth is allocated based on a static distribution mechanism and every ONU receives a fixed timeslot. The OLT generates a *GATE* message without a *REPORT* from the ONUs, allocating a fixed amount of bandwidth in a static manner. SBA works exactly like time division multiple access (TDMA) in that the time slot of each ONU is fixed ahead of time without traffic arrival rate considerations of ONUs. Without the overhead of the queue status *Report* and transmission G*rant*, SBA is the simplest method to distribute bandwidth among the ONUs and also it is easy to implement it. It suits the time-sensitive or constant-rate transmission services such as E1, T1 and plain old telephone service (POTS) quite well [LEE04]. Due to the fact that ONU will occupy the upstream bandwidth during its assigned time slot even if there is no frame to transmit, hence resulting in the increased delay for all other ONUs buffered packets. Due to lack of statistical multiplexing between the ONUs and the fact that upstream bandwidth might be either idle or lightly loaded, upstream bandwidth is not utilized efficiently.

3.5.1.2 Dynamic Bandwidth Allocation (DBA)

In this section we review the DBA algorithms proposed in the literature. As we stated earlier in upstream in order to avoid contention, TDMA is used. However, due to the lack of statistical multiplexing [KRA02a] and burstiness of traffic, some timeslots might not be filled and that is leading to underutilized bandwidth. Dynamic schemes on the other hand adaptively adjusting the time slot size, allowing the excess traffic to be carried in unused timeslots. We will briefly review various DBA approaches.

Interleaved Polling with Adaptive Cycle Time (IPACT)

An upstream bandwidth allocation using maximum window size was reported by [KRA02c]. They used an interleaving scheme in a way that multiple polling requests are overlapped in time to avoid accumulation of switchover times. Ranging is not required in this scheme as is required in traditional TDMA. By adapting the granted transmission window to ONU's respective load, bandwidth is dynamically distributed among ONUs.

Latency Classification Scheme

In order to provide service satisfaction, a DBA scheme was proposed based on the delay profiles [YOS02]. Two service classes of low delay and normal delay were defined. Low delay service class is favored delay sensitive applications like audio, real-time video, etc. On the other hand normal delay service class is suitable for delay insensitive applications like data transport, email, etc. Bandwidth allocation for low delay services has priority over normal delay services. Low delay services will be provided by small granted bandwidth and large *Grant* frequency, hence reduced latency but efficiency is degraded. Normal delay services will be getting high granted bandwidth but with small *Grant* frequency, hence delay is increased and efficiency is improved.

Generalized Processor sharing (GPS) –based DBA

A deterministic bandwidth estimation in order to comply with delay constraint and providing lossless QoS guarantee is presented by Zhang et. al. [ZHA03]. Obtaining a bound on delay and backlog is the main objective of a GPS scheduler. They proposed using two schedulers, one in OLT and the other one in ONU. Master scheduler, residing in OLT, is responsible for mapping the QoS specifications into deterministic effective bandwidth, and assigns the weight correspondingly. However, slave scheduler, residing in ONU is in charge of improving bandwidth utilization in order to optimize the delay variation.

3.5.2 Dynamic Bandwidth Allocation (DBA) Formulation

In this section we formulate our problem of dynamically allocating the bandwidth based on customer-tailored QoS metrics [NOU05a].

- $I_i(t)$ denotes the incoming traffic of i th flow, constrained by the service level specifications (SLSs) agreed and contracted between customer and service provider.

- P_i denotes the priority class of i th flow.

- $T_i(t)$ is the transported traffic of the i th flow.

- $L_i(t)$ is the i th flow queue length.

- R_i is the allocated rate of i th flow.

- D_i is the delay tolerance of the i th flow.

- $d_i(t)$ is the experienced delay of the i th flow.

- B is the available bandwidth.

N traffic streams with various QoS levels are considered, representing various delay and bandwidth requirements for different customers. Systems dynamics are described below,

$$L_i(t) = I_i(t) - T_i(t) \text{ , i= 1,2, ..., N} \tag{3.7}$$

denotes that queue length is the difference of the transported traffic from the incoming traffic for every flow.

$$\frac{T_i(t)}{T_j(t)} \geq \frac{P_i}{P_j} \text{, i, j = 1, 2,..., N} \tag{3.8}$$

indicate that the rates of transported traffic of various flows are related to their corresponding priorities.

$$d_i(t) \le D_i \text{, i=1,2,...,N} \tag{3.9}$$

defines that the experienced delay must be less or equal than the delay tolerance of a stream.

$$\sum_{i=1}^{N} R_i \le B \text{, i=1,2,...,N} \tag{3.10}$$

states that allocated rate cannot exceed the available bandwidth.

3.5.2.1 Policing

We consider three different QoS levels as priorities, and the customers select among these priorities based on their service requirements. After QoS assignments, the packets have to be scheduled. In the next section we will review the queueing discipline.

3.5.2.2 M/G/1 priority system

We propose a single server M/G/1 non-preemptive priority system. The higher priority packet(s) will be served immediately after finishing packet currently in service. The arrival of all traffic classes to any ONU is assumed independent, Poisson, and independent of the service times. QoS levels are considered to be of M levels mapped to M priorities. In the present work, delay D and customer throughput λ are considered as QoS criteria. In an M/G/1 non-preemptive priority system, the average waiting time W_k for a priority k can be estimated as:

$$W_k = \frac{R}{(1-\rho_{k-1})(1-\rho_k)} \tag{3.11}$$

where $\rho_0 = 0$, $\rho_k = \sum_{i=1}^{k} \frac{\lambda_i}{\mu}$, ($\rho_k$ is the utilization for priority k), $R = \frac{1}{2}\sum \lambda_i \overline{X_i}^2$, ($X_i$ is the service time for the i th customer), μ is the link capacity, λ_i is the i th customer's arrival rate. The average service time for a customer is $\frac{1}{\mu}$ (see [KLE75a, b, and TAK93]).

In order to satisfy every individual customer QoS level, we have,

$$D_k \geq W_k = \frac{R}{(1-\sum_{i=1}^{k-1}\frac{\lambda_i}{\mu})(1-\sum_{i=1}^{k}\frac{\lambda_i}{\mu})}, \forall k \in [1,...,M] \tag{3.12}$$

and the total delay for every traffic class will be,

$$T_k = W_k + D_{Lk} + D_P + S_k \tag{3.13}$$

Where D_{Lk}, is the round trip delay associated with each ONU for priority k, D_P is the processing delay (we assume this delay to be fixed and similar for every ONU), and S_k is the service time for class k traffic ($S_k = \frac{1}{\mu_k}$).

Let us assume that every customer based on different applications has a different set of link characteristics requirements. As an example for every customer's profile, two parameters of delay D_i and priority pare considered. We assume three different levels of priorities. Each arrived traffic to be transported based on the corresponding class of service must be verified to meet that CoS requirements. A typical matrix of N customers and three priorities will look like the following:

$$\begin{Bmatrix} (T_{11}, p_{11}), (T_{12}, p_{12}), (T_{13}, p_{13}) \\ (T_{21}, p_{21}), (T_{22}, p_{22}), (T_{23}, p_{23}) \\ (......,.......), (.....,......), (.....,......) \\ (T_{N1}, p_{N1}), (T_{N2}, p_{N2}), (T_{N3}, p_{N3}) \end{Bmatrix} \tag{3.14}$$

Based on different instances of arrival traffic, we want to maximize throughput with regard to two constraints of delay and priority. QoS metrics will be examined for the waiting traffic to be transported, if the constraints are satisfied traffic will be sent. Service allocation matrix based on priorities and delays of every individual traffic flow can be demonstrated by:

$$\begin{Bmatrix} s_{11}, s_{12}, s_{13} \\ s_{21}, s_{22}, s_{23} \\ \ldots\ldots,\ldots\ldots,\ldots\ldots \\ s_{N1}, s_{N2}, s_{N3} \end{Bmatrix} \tag{3.15}$$

Due to the fact that at every instance, one kind of traffic will arrive to the system, so we will have only one value of "1" in every row. After priority assignment to traffic flows, bandwidth and delay throughput are optimized.

3.6 Summary

This chapter presented Quality of Service (QoS) in EPON. Queueing analysis and priority queueing analysis were also provided. Formulas for finding delays for higher priority and lower priorities based on higher priorities were derived in this chapter. General formula for finding delays for priorities was obtained and presented in a closed form. In the next chapter discrete event simulation is used and results provided to verify our analysis.

Chapter 4 Performance Analysis and Results

4.1 Introduction

The basic objective of the mathematical modeling of computer communication system is the prediction and analysis of the system performance [HAY04]. Data traffic considered to have three properties of being bursty, asymmetric, and delicate. In many applications, there is short activity periods followed by long idle periods, hence burstiness of data is implied. Data burstiness have two implications, the long term average usage by a single source is small and making resource sharing as an economical method. The other issue is that, sources transmit at a relatively high instantaneous rate, reducing the resource dedication time to any data source. Data asymmetry implies that, in many of the applications the volume and speeds are different in outgoing and incoming traffic. Delicacy refers to data susceptibility to errors; hence there is a need for error counter measurement methods.

Network flowing traffic can be classified into several types. Two of the dominant traffic types are Ethernet packets/frames and ATM cells. All messages are broken down to either packets or cells; depending on the network segment that message is traversing it. The Ethernet packet length varies from 60 bytes to 1500 bytes. The ATM cells have a fixed length of 53 bytes. Aggregation of billions of these packets and/or cells is what is called network traffic. Fitting the appropriate traffic model to capture the stochastic nature of the telecommunication networks is an extremely important task before studying and evaluating their performance. Sniffers can be used to collect network data and analyze packets that were dumped during the time sniffer was used. With sniffing source and destination, its length, its type, etc. For the purpose of fitting a traffic model to a telecommunication network only arrival time and length are sufficient. In the next section QoS in EPON is reviewed.

4.2 Simulation and Results

In order to obtain a realistic and accurate performance analysis, it is important to simulate the system behavior with injecting appropriate traffic patterns to the system. A discrete event simulator was developed under C++. Windows 2000 was the operating system used. Simulations were run on an IBM PC, having an Intel Pentium 4, 3.20 GHz processor, 760 MB of ram and a 40GB hard drive capacity.

ONUs are distributed randomly away from OLT. Distances between OLT and ONUs are varied between 10 km to 20 km. The trunk capacity was set to 1000 Mbps and ONU's access rate was 100 Mbps. Each ONU was assigned a common buffer space of 10 MB for all three classes of service.

Three classes of service are considered and their corresponding traffic is generated based on some assumption taken from real data traffic. Each ONU have three queues for three classes of service. CoS 0 is used to emulate the DiffServ's Expedited Forwarding (EF) and is suitable for delay sensitive and low-loss applications. CoS 1 is used to emulate the DiffServ's Assured Forwarding (AF) suiting low-loss traffic. CoS 2 is representing Best Effort (BE) traffic. Class of service 0 considered to be CBR with an average rate of 4.5 Mbps and packet size of 70 bytes, emulating T1 services. The other two classes of service were considered as ON-OFF sources with ON-OFF intervals according to a Pareto distribution, to model self-similar traffic, with an average rate of approximately 47.5 Mbps. Three packet sizes of 64, 594, and 1518 bytes and their corresponding probabilities considered to be 0.62, 0.10, and 0.28. Each point on the plots were collected after several runs of the simulation program and corresponds to a sample of more than 100 million packets. Generated packets for every class of service is randomly generated and randomly distributed to different ONUs. To provide QoS, priority treatment of arrived traffic has been considered. Higher priority traffic will receive lower delay than other lower priority traffic.

4.3 Simulation Results

ONUs were distributed in a random manner by the simulation program, the following figure shows ONUs distances from the OLT.

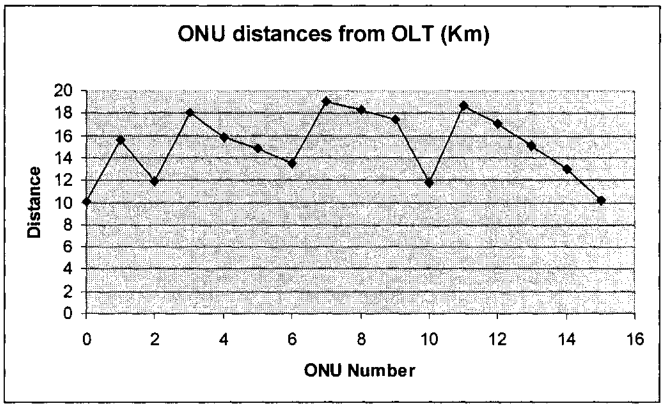

Figure 4.1 ONUs distances from OLT.

ONU's distances are ordered in ascending order from OLT and it is presented in Figure 4.2.

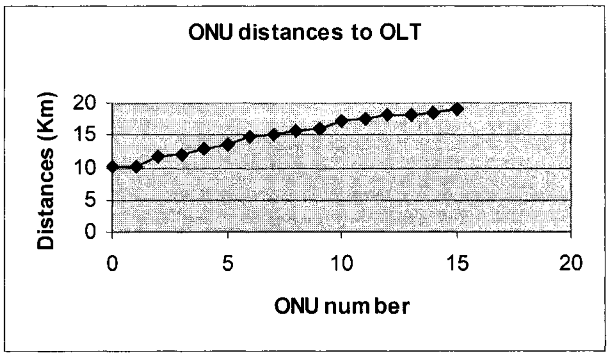

Figure 4.2 ONUs distances from OLT. (Ascending order)

In the following figure, average delays for ONUs are presented based on their initial random distribution prior to sorting ONUs in an ascending order. CoS 2 is showing most delay fluctuations compared to other class of services. This is due to two reasons, one being ONUs geographical distributions and the other is CoS 2 traffic has to wait for completion of other higher priorities traffic of other classes of service.

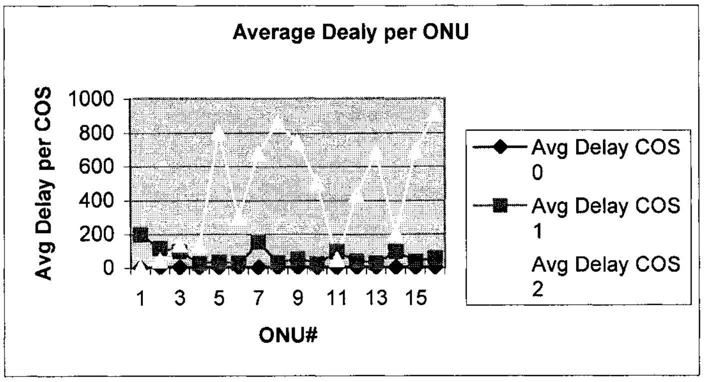

Figure 4.3 Average delay per ONU for various CoS.

In the following figure average delay for various class of service (CoS) is shown [NOU05c], and it can be seen from the figure that higher priorities as it is expected experiencing minimum delay and higher priorities are delayed after service of higher priority traffic are completed.

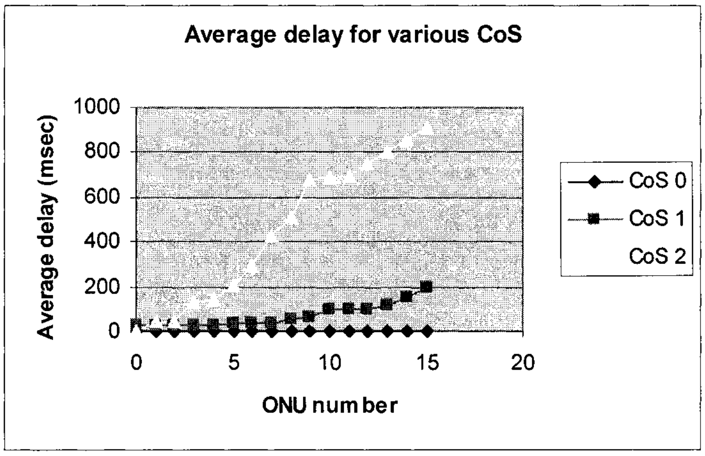

Figure 4.4 Average delay per ONU. (ordered)

In Figure 4.5, average delay for three classes of service is represented. Class of service 0, due to its higher priority comparing to other classes of service is experiencing much lower delays. Other classes of service depending on their traffic priorities they will experience delays substantially more than CoS 0 delays.

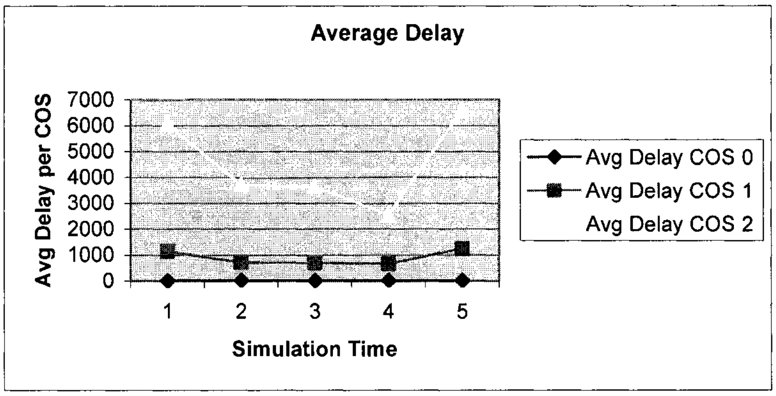

Figure 4.5 Average delay per ONU.

Average delay for three classes of service is presented in Figure 4.6. ONUs are ordered based on their distances from OLT. Class of service 0 is showing the lowest delay among all other classes of service and class of service 1 is having higher delay than

CoS 0 but lower delays than CoS 2. Class of service 2 is experiencing higher delays than other higher priority traffic and it is showing a sporadic behavior.

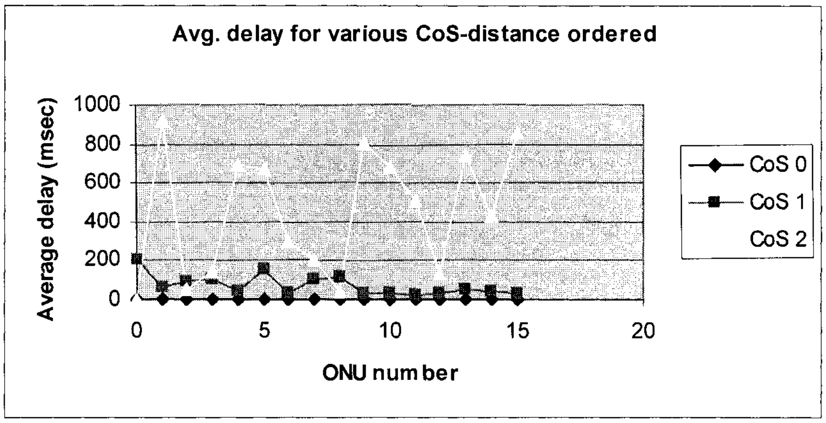

Figure 4.6 Average delay for various COS (ONUs are ordered)

Packets are generated in a random fashion and are arbitrarily distributed to different ONUs. Number of received packets for each ONUs for different CoS is presented in Figure 4.7.

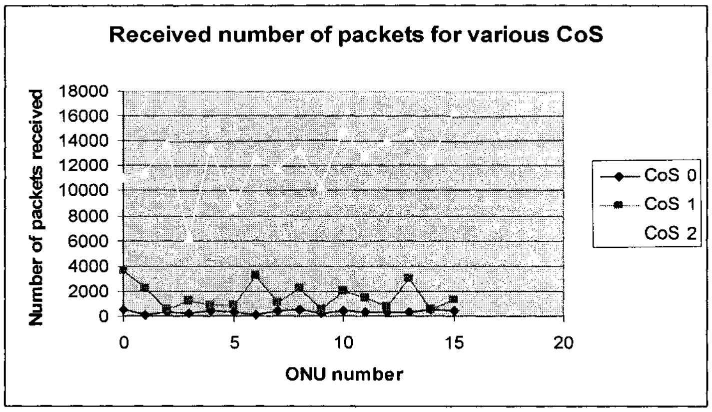

Figure 4.7 shows number of packets received in the ONUs for different CoS.

Figure 4.8 presents maximum number of packets in queues for different class of service for each ONU.

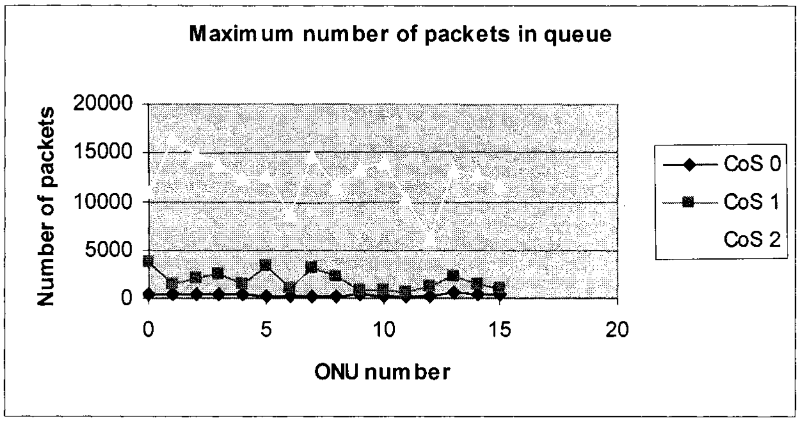

Figure 4.8 Maximum number of packets in queue for classes of service per ONU.

Maximum number of queued packets and delays for various classes of service for each ONU is presented in Figure 4.9. Higher priority traffic is having smaller share of the randomly generated packets and their delays are substantially lower. Lower priority traffic is receiving a larger number of packets and hence their delays are greater.

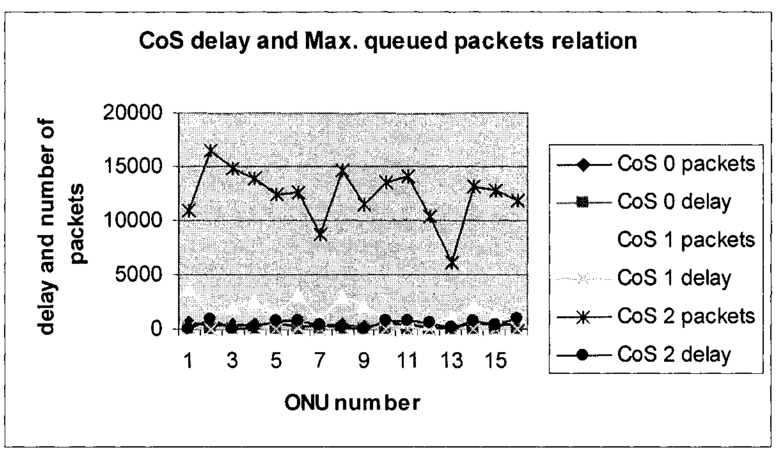

Figure 4.9 Classes of sevice delays and maximum queued packets for each ONU.

Maximum queued packets for each class of service for every ONU is presented in Figure 4.10.

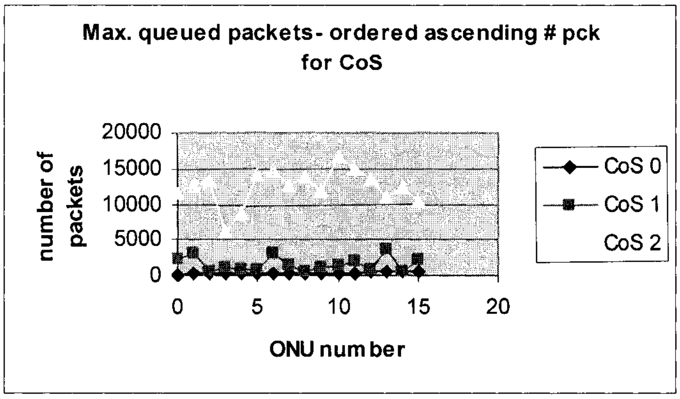

Figure 4.10 Maximum number of queued packets for each ONU.

From Figure 4.11, average throughput of three classes of service is presented [NOU05b]. Throughput of each class of service was graphed based on their corresponding transported traffic to the total input traffic. Higher priority traffic is having a lower throughput than other traffic streams due to the fact that percentage of higher priority traffic to other priorities is much lower, hence their link utilization is much lower than others. Class of service 1 is showing the highest throughput due to their highest traffic intensity.

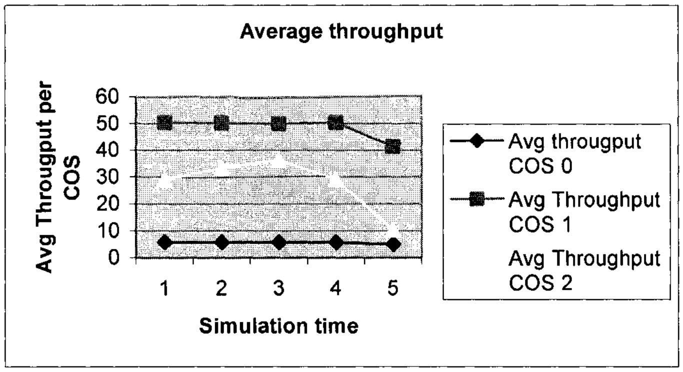

Figure 4.11 Average throughput for various CoSs [NOU05b].

Average delay versus number of packets for CoS 0, CoS 1, and CoS 2 is shown respectively in Figures 4.12, 4.13, and 4.14.

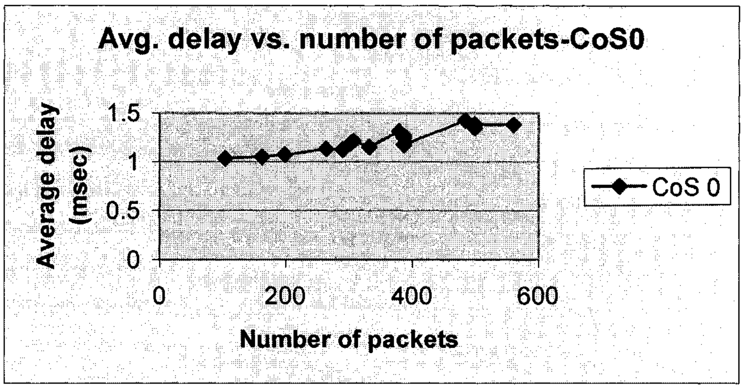

Figure 4.12 CoS 0: Average delay versus number of packets.

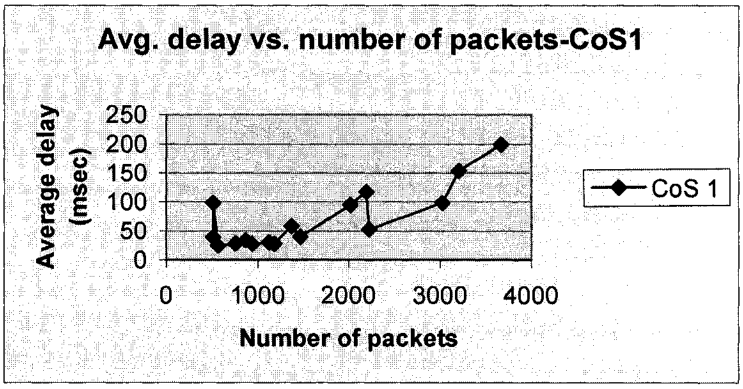

Figure 4.13 CoS 1: Average delay versus number of packets.

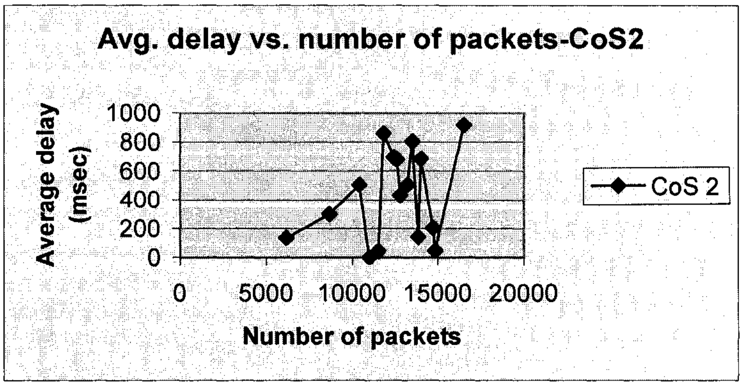

Figure 4.14 CoS 2: Average delay versus number of packets.

As can be seen from Figures 4.13 and 4.14, experienced delay by lower priorities depends on the higher priority traffic and cannot have a smooth relation between the arrived traffic and the delay.

4.4 Summary

This chapter presented an algorithm to provide preferential treatment between users and applications. Results from discrete event simulator verified our analysis and algorithm. A delay-based treatment depending on the user's agreed service level agreement (SLA) is realized. Next chapter is devoted to conclusion and future research.

Chapter 5 Conclusions

5.1 Concluding remarks

A variety of access technologies are under development due to growing customer demand for greater bandwidth, integrated services and rapid service deployment. Passive optical networks comprising of passive devices between subscribers and service providers' networks are one of the well-suited candidate for offering bandwidth-rich services for emerging bandwidth-hungry multimedia applications. PON technologies are eliminating the costly maintenance that is the case with any access networking technology. Among the PON technologies, the thesis concentrated on EPON. EPON is the first public network that was not developed from telephony or cable communications, but from enterprise data networks [KRA05]. As it was iterated in order to provide quality of service (QoS), an arbitration mechanism to provide bandwidth based on the customers service level agreements (SLAs) is required. There are two general way for this purpose, static and dynamic bandwidth allocation algorithms. With dealing with ever changing and bursty traffic, static bandwidth allocation schemes cannot provide QoS guarantee for users. In this regard, a novel dynamic bandwidth allocation (DBA) algorithm was proposed to support QoS for EPON subscribers. Proposed scheme provides customers with their required service requirements (i.e., bandwidth, and delay) and based on delay criteria customers can be differentiated and hence a preferential treatment is provided. Three classes of service were considered with different delay requirements. Applications with different delay requirement were classified and directed to the corresponding queue to receive the agreed upon delivery. Highest priority traffic were ahead of other traffic to be delivered and other lower priority traffic without the presence of a higher priority traffic in the queue were delivered to the OLT. Depending on the packet classes, packets experienced delays based on their priorities. Higher priority traffic received the least delays among all other traffic types. Lowest traffic experienced the most delay comparing to the two higher priority traffic classes. With this scheme, customers could receive preferential treatment based on their service requirements and service agreements, meeting their needs. Results presented here shows that different applications can be classified and receiving corresponding service based on the assigned priority. Highest achieved throughput belongs to class of service 1 due to their higher incoming traffic. Lowest throughput among all three classes was for class of service 0 representing their lower traffic among all. Class of service 2 achieved an average throughput between the other two classes of service.

5.2 Future Directions

With the adaptation of EPON in the access networks, a whole new stream of applications will follow. This will require new investment in the metro and long-haul networks, due to the tremendous increase in traffic. At the same time, EPON networks might need to be enhanced to support more bandwidth and higher data rates. Various system enhancements can be considered in a more comprehensive manner as the potential future research directions.

One possible extension is wavelength-based and will provide new wavelengths for both upstream and downstream traffic. Adding various wavelengths in the EPON network is an interesting topic to be studied. By doing so, fewer ONUs will share the available bandwidth, resulting in higher bandwidth efficiency. Currently major issue for supporting different wavelength is the cost of tunable lasers in the transceivers.

Another interesting research topic to be carried out is the support of higher data rates, with the finalization of the 10 Gbps Ethernet standards by the IEEE. In this scenario, only a portion of ONUs may support the higher data rates and further changes will take place with the time as the need arises from subscribers. In this aspect the network should be able to support various data rates, adding to the system complexity and cost. This transitional period from lower rate to higher data rates in the network and its effect on customers should be considered ahead of time in more details.

Another possible research scenario is the provision of multi fiber deployment from CO to the splitter. Some branches might be reconfigured, by attaching to the new fiber. This new fiber might be installed initially for future proofing to alleviate the issue of new fiber redeployment. Depending on the expansion scenario and geography of the area and service providers' implications, different research intensive studies should be carried out with the complete technical analysis for system under study.

Other issues to be consider is fair bandwidth allocation for users and service providers sharing broadband networks, and scheduling algorithms are important issues to be considered. However, scheduling algorithms utilized by different systems could either be interoperable or be standardized. Currently different vendors are providing their proprietary equipments and they are not interoperable. Interoperability of EPON networks is a major concern to be studied in order to provide a transparent service and enhancing system throughput. Providing fair and prioritized bandwidth to users in a way that some users will not be exhausted is an interesting research topic to be further investigated.

Security is another major concern, due to the fact that downstream traffic is forwarded to all subsribers. Users must be protected against malicious intruders. Providing security by some cryptographic or ciphering key will be one future research path to be implemented in a cost effective manner. Adding complexity to ONUs will have a cost burden on customers and their willingness to subscribe to the next generation access networks.

Survivability of the service is also important. As content becomes more rich and extensive, any service interruption will cause more data loss that is translated into more monetary loss. Recovery of the data and service, by some restoration mechanisms i.e., IP mechanisms or other methods, is another research direction to be investigated in more details.

It can be finally mentioned that implementing dynamic bandwidth allocation in a network is a very research intensive task. Imposing real customers' behavioral constraint on the working system in real time is a motivating task to be considered for future research.

References:

[ADA97]A. Adas, "Traffic models in broadband networks", IEEE Communication Magazine, Vol. 35, no. 7, July 1997, pp. 82–89.

[BLA98] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, W. Weiss, "An Architecture for Differentiated Services", RFC 2475, 1998.

[CHI04] G. Chiruvolu, A. Ge, D. Elie-Dit-Cosaque, M. Ali, and J. Rouyer, "Issues and Approaches on Extending Ethernet Beyond LANs", IEEE Comm. Mag., March 2004, pp 80-86.

[DIX03] S. Dixit, IP Over WDM: Building the Next Generation Optical Internet, Wiley Interscience 2003.

[FIS04] T. Fishchaber, K. Marks, H. Fallside, "Next-Generation Data Transport over Metro Area Networks", Xcell journal, No. 50, Fall (2004).

[GAG03] M. Gagnaire, Broadband Local Loops for High-Speed Internet Access, Artech House, 2003.

[GUM04] A. Gumaste, and T. Antony, First Mile Access Networks and Enabling Technologies, Cisco Press, 2004.

[HAY04] J.F. Hayes T.V.J. Ganesh Babu, Modeling and analysis of telecommunications networks, Wiley-Interscience, 2004.

[HSU03] Y.L. Hsueh, F.T. An, K.S. Kim, and L.G. Kazovsky, "A New Media Access Control Protocol with Quality of Service and Fairness Guarantee in Ethernet-Based Passive Optical Networks", 2nd Symposium on Photonics, Networking, and Computing, Cary, North Carolina, pp. 1392-1395, September, 2003.

[IEE02] Carrier sense multiple access with collision detection (CSMA/CD) access methods and physical layer specifications, ANSI/IEEE Std. 802.3, Part 3, 2002.

[ITU98] ITU-T, Recommendation G.983: Broadband Optical Access Systems Based on Passive Optical Networks, 1998.

[KAM02] I. Kaminow and T. Li, Optical Fiber Telecommunications, IVA-IVB, Academic Press, 2002.

[KLE75a] L. Kleinrock, Queuing Systems, Vol. 1, Wiley, 1975.

[KLE75b] L. Kleinrock, Queuing Systems, Vol. 2, Wiley,1975.

[KRA02a] G. Kramer, and G. Pesavento, "Ethernet Passive Optical Network (EPON): Building a Next-Generation Optical Access Network", IEEE Comm. Mag., Feb. 2002, pp. 66-73.

[KRA02b] G. Kramer, B. Mukherjee, and G. Pesavento, "Interleaved Polling with Adaptive Cycle Time (IPACT): A Dynamic Bandwidth Distribution Scheme in an Optical Access Network", Photonic Network Communications, vol. 4, no.1 pp. 89-107, January 2002.

[KRA02c] G. Kramer, B. Mukherjee, and G. Pesavento, "IPACT: A Dynamic Protocol for an Ethernet PON (EPON)", IEEE Comm. Mag. Feb. 2002 pp74-80.

[KRA05] G. Kramer, Ethernet Passive Optical Networks, McGraw-Hill, 2005.

[LEE04] H.-S. Lee, T.-W. Yoo, J.-H. Moon, H.-H.Lee, "A Two-Step Scheduling Algorithm to Support Dual bandwidth Allocation Policies in an Ethernet passive Optical Network", ETRI Journal, Vol. 26, No. 2, Apr 2004, pp. 185-188.

[LUO05] Y. Luo, and N. Ansari, "Bandwidth Allocation for Multiservice Access on EPONs", IEEE Optical Communications, Feb. 2005, pp. S16-S21.

[MA03] M. Ma, Y. Zhu, and T.H. Cheng, "A Bandwidth Guaranteed Polling MAC Protocol for Ethernet Passive Optical Networks", IEEE INFOCOM 03, San Francisco, California.

[NIK03a] D. Nikolova, B. Van Houdt, and C. Blondia, "Dynamic Bandwidth allocation Algorithms in EPON: A simulation Study", Proceedings of SPIE, Vol.5285, OptiComm October 2003, pp. 369-380.

[NIK03b] D. Nikolova, B. Van Houdt, and C. Blondia, "QoS issues in EPON", Proceedings of Community Nets and FTTH/P/x Workshop, Dallas, USA, October 2003.

[NOU04] A. Nouroozifar, H. Naser, H.T. Mouftah, "Optical access networks: limitations and prospects", Photonic North 2004, Ottawa, Canada, SPIE, Proceedings of the SPIE, Volume 5579, pp. 583-591.

[NOU05a] A. Nouroozifar, H.T. Mouftah, T. Saad, "A Novel Upstream Scheduling Algorithm for Quality of Service Provisioning in Ethernet Passive Optical Networks", NFOEC/OFC 2005, Anaheim, USA.

[NOU05b] A. Nouroozifar, H.T. Mouftah, H. Naser, "A Novel Differentiated Scheduling Algorithm for EPONs", 7th International Conference on Transparent Optical Networks (ICTON 2005), July 3 - 7, 2005, Barcelona, Spain, pp. 237-240.

[NOU05c] A. Nouroozifar, H. T. Mouftah, H. Naser, "A Novel Scheduling Algorithm for EPONs", International Symposium on Telecommunications (IST 2005) (IEEE), September 10-12, 2005, Shiraz, Iran.

[PES03] G. Pesavento, "Ethernet Passive Optical Network (EPON) Architecture for Broadband Access", Optical Networks Magazine, Jan/Feb 2003, pp. 1-7.

[TAK93] H. Takagi, Queueing Analysis: A Foundation of Performance Evaluation, North Holland/Elsevier, 1993, vol. 1&2&3.

[YON03] Z. Yongqing, "Multiaccess Scheme for Ethernet Passive Optical Network (EPON)", Nanyang Technological University, Jan. 2003.

[YOS02] O. Yoshihara, N. Oota, and N. Miki, "Dynamic Bandwidth Allocation Algorithm for GE-PON", Proceedings of the International Conference on Optical Internet 2002.

[ZHA03] L.Zhang, E-S. An, C-H. Youn, H-G. Yeo, and S. Yang, "Dual DEB-GPS Scheduler for Delay-Constrant Applications in Ethernet Passive Optical Networks", IEICE Trans. Commun., VOL. E86-B, NO.5, May 2003, pp. 1575-1584.

[ZHE05] J. Zheng, and H.T. Mouftah, "An Adaptive MAC Polling protocol for Ethernet Passive Optical Networks (EPONs)", Proc. IEEE ICC'05, May 2005.

Appendix A: Queueing Models

Packet Queueing Algorithms

In trying to provide an efficient solution to congestion collapse problem several Packet Queueing Algorithms have been proposed. In this section we deal with the descriptions of some of these algorithms.

- Drop Tail Algorithm: Drop Tail (DT) is the simplest and the most commonly used algorithm in the current Internet gateways, which drops packets from tail of the full queue buffer. Its main advantages are simplicity, suitability to heterogeneity and its decentralized nature. However, this approach has some serious disadvantages, such as lack of fairness, no protection against misbehaving or non responsive flows (i.e. flows that do not reduce their sending rate after receiving the congestion signals from gateway routers) and no relative Quality of Service (QoS).

- DEC-bit Algorithm: DEC-bit was developed for use on a connectionless network with a connection-oriented transport protocol. What takes place is each router monitors the load it is experiencing and explicitly notifies the end node when congestion is about to occur.

This notification is implemented by setting a bit (DEC-bit) in the header of the packet that flows through the router. The router sets this bit if the average queue length is greater than or equal to 1 at the time the packet arrives at the router. The last step involves the source adjusting its sending rate in order to avoid congestion. The sender decreases the congestion window by 0.875 times if 50 percent or more of the last windows worth of packets have the DEC-bit sent, otherwise the sender increments the congestion window by one packet.

- Random Early Detection Algorithm: Random Early Detection (RED) Algorithm is a mechanism that detects congestion and provides feedback to end hosts by dropping packets. The motivation behind RED is to keep the queue size small, reduce burstiness and solve the problem of global synchronization. RED drops packets before the actual physical queue is full. It operates based on the average queue length that is calculated using an exponential weighted average of the instantaneous queue length. RED drops packets with certain probability based on the average length of the queue. The drop probability increases from 0 to a maximum drop probability as the average queue size increases from a minimum threshold to the maximum threshold. If the queue size goes above the maximum threshold, all packets are dropped. The RED algorithm can be broken up into two separate parts. The first part is an algorithm to compute the average queue size. This determines the level of burstiness that will be allowed in the gateway. The second part is an algorithm for calculating the packet marking probability. This determines how frequently the gateway marks packets, given the current level of congestion.

- Fair Queueing Algorithm : Fair Queueing (FQ) provides isolation between flows, thus protecting the well-behaved flows against the ill-behaved ones. At the same time it distributes bandwidth fairly among the flows. In particular, the criteria used to distribute bandwidth, is max-min fairness. These algorithms are mainly used in the multimedia integrated services (IntServ) networks for their fairness and delay boundedness. Some important variation of Fair Queueing Algorithms are briefly described below:

* Round-Robin and Weighted Round-Robin Algorithm: Round-Robin Scheduling (RR) distributes each request sequentially around the pool of real servers. Using this algorithm,

all real servers are treated as equals without any regard to capacity or load. This scheduling model resembles round-robin DNS (Domain Name System/Service/Server) but is more granular due to the fact that it is network-connection based and not host-based.

Weighted Round-Robin Scheduling (WRR) distributes each request sequentially around the pool of real servers but gives more jobs to servers with greater capacity. Capacity is indicated by a user-assigned weight factor, which is then adjusted upward or downward by dynamic load information. Weighted round-robin scheduling is a better choice if there are significant differences in the capacity of certain real servers in the pool. However, if the requested load varies dramatically, the more heavily weighted server may answer more than its share of requests.

* Weighted Fair Queueing Algorithms: Weighted Fair Queueing (WFQ) provides traffic priority management that automatically sorts among individual traffic streams without requiring that you first define access lists. WFQ can also manage duplex data streams such as those between pairs of applications, and simplex data streams such as voice or video. There are two categories of WFQ sessions: high bandwidth and low bandwidth.

Low-bandwidth traffic has effective priority over high-bandwidth traffic, and high bandwidth traffic shares the transmission service proportionally according to assigned weights. When WFQ is enabled for an interface, new messages for high-bandwidth traffic streams are discarded after the configured or default congestive messages threshold has been met.

However, low-bandwidth conversations, which include control message conversations, continue to enqueue data. As a result, the fair queue may occasionally contain more messages than its configured threshold number specifies. With standard WFQ, packets are classified by flow. Packets with the same source IP address, destination IP address, source Transmission Control Protocol (TCP) or User Datagram Protocol (UDP) port, or destination TCP or UDP port belong to the same flow. WFQ allocates an equal share of the bandwidth to each flow. Flow-based WFQ is also called fair queueing because all flows are equally weighted.

* Core Stateless Fair Queueing Algorithm: An alternate approach to achieve congestion control is to provide router support for fair bandwidth allocations. In this way we achieve protection against ill-behaved flows. In addition, each flow is ensured to receive its fair share no matter what control algorithms, if any, the end hosts implement. The major disadvantage of this approach is that the known algorithms to achieve fair share are complex, as they require routers to perform per flow management. More precisely, a router needs to perform:

(1) per-packet classification,

(2) per-flow buffer management, and eventually

(3) per-flow scheduling.

This complexity may prevent them from being cost-effectively deployed at high speeds. To address this problem, a network architecture and an algorithm, called Core-Stateless Fair Queueing (CSFQ), has been proposed. This significantly reduces the implementation complexity yet still achieves approximately fair allocations. The architecture differentiates between edge and core nodes. While edge nodes do perform per flow management, core nodes do not perform per flow management, and therefore can be efficiently implemented at high speeds. In addition, edge nodes themselves are simpler than regular fair queueing nodes.

- Priority Queueing Algorithms: Priority Queueing (PQ) allows you to define how traffic is prioritized in the network. You can configure various traffic priorities. You can define a series of filters based on packet characteristics to cause the router to place traffic into these four queues; the queue with the highest priority is serviced first until it is empty, then the lower queues are serviced in sequence. During transmission, PQ gives priority queues absolute preferential treatment over low priority queues; important traffic, given the highest priority, always takes precedence over less important traffic. Packets are classified based on user-specified criteria and placed into one of the output queues high, medium, normal, and low based on the assigned priority. Packets that are not classified by priority fall into the normal queue.

- BLUE and Stochastic Fair Blue Algorithms: In order to stem the increasing packet loss rates caused by an exponential increase in network traffic, we may consider the deployment of active queue management techniques such as RED. While active queue management can potentially reduce packet loss rates in the Internet, one can show that current techniques are ineffective in preventing high loss rates. The inherent problem with these queue management algorithms is that they all use queue lengths as the indicator of the severity of congestion. In light of this observation, a fundamentally different active queue management algorithm called BLUE is proposed. BLUE uses packet loss and link idle events to manage congestion. Using simulation and controlled experiments, BLUE is shown to perform significantly better than RED both in terms of packet loss rates and buffer size requirements in the network.

Stochastic Fair Blue (SFB) Algorithm is a novel technique for enforcing fairness among a large number of flows. SFB scalably detects and rate-limits non-responsive flows through the use of a marking probability derived from the BLUE queue management algorithm and a Bloom filter. Using analysis and simulation, SFB is shown to effectively handle non-responsive flows using an extremely small amount of state information. Next a review of queueing theory is provided.

Overview of Queueing Theory

In this section we use the following notation for simplicity:

Kendall-Lee Notation:

$$A/B/x/y/m\text{- } Z \tag{A.1}$$

- A: letter for arrival distribution,

"M" = exponential interarrival distribution (M=Markovian, memoryless); Poisson process

"D" = deterministic, constant interarrival times

"G" = general (unspecified)

"Ek" = Erlang-k distribution

"PH" = phase distribution

"Cox" = Cox distribution

- B: letter for service distribution,

"M" = exponential service time distribution

"G" = general service

- x: number of service channels,

- y: number allowed in queue + servers (default: ∞),

- m: total job population (default: ∞),

- Z : queue discipline (default: FIFO).

The M/M/1 Queueing System

In this part we will focus on M/M/1 queueing system. Note that M/M/1 refers to negative exponential arrivals and service times with a single server. This is the most widely used queueing system in analysis as pretty much everything is known about it. M/M/1 is a good approximation for a large number of queueing systems. M/M/1 queueing systems assume a Poisson arrival process. This assumption is a very good approximation for arrival process in real systems that meet the following rules:

- The number of customers in the system is very large.

- Impact of a single customer on the performance of the system is very small, i.e. a single customer consumes a very small percentage of the system resources.

- All customers are independent, i.e. their decisions to use the system are independent of other users. It is known that M/M/1 can be applied to systems that meet certain criteria. But if the system under consideration can be modelled as an M/M/1 queueing system, it is a straight forward process.

First we define ρ, the traffic intensity (sometimes called occupancy). It is defined as the average arrival rate (λ) divided by the average service rate (μ). For a stable system the average service rate should always be higher than the average arrival rate. (Otherwise the queue would rapidly race towards infinity). Thus ρ should always be less than one. Also note that we are talking about average rates here, instantaneous arrival rate may exceed the service rate. Over a longer time period, the service rate should always exceed arrival rate.

$$\rho = \frac{\lambda}{\mu} \tag{A.2}$$

Lastly we obtain the total waiting time (including the service time):

$$T = \frac{1}{\mu - \lambda} \tag{A.3}$$

Again we see that as mean arrival rate (λ) approaches mean service rate (μ), the waiting time becomes very large.