Admission control and scheduling for improved QoS provisioning in transport and access networks

Xue, Daojun

2009

Xue, D. (2009). Admission control and scheduling for improved QoS provisioning in transport and access networks. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/18620>

<https://doi.org/10.32657/10356/18620>

ADMISSION CONTROL AND SCHEDULING FOR IMPROVED QOS PROVISIONING IN TRANSPORT AND ACCESS NETWORKS

XUE DAOJUN

School of Electrical and Electronic Engineering

A thesis submitted to the Nanyang Technological University
in fulfillment of the requirement for the degree of
Doctor of Philosophy

2009

Acknowledgements

First of all I would like to thank my supervisors, Dr. Qin Yang and Dr. Siew Chee Kheong, for their guidance, encouragement and support throughout my doctoral study. They have instructed me all the way from a bachelor degree holder to a Ph.D. I believe I would not be able to finish my study and this thesis without their guidance.

I would also like to thank my parents and my grandparents who care about me so much. Their encouragements have always been the inspiration for me that keeps me advancing.

Lastly I want to express my gratitude to all the other Ph.D. students in the Infocomm Research Lab. I have benefited a lot from the group discussions and it was always a pleasure to work with them.

Table of Contents

Acknowledgements.....	I
Table of Contents	II
Summary.....	V
List of Figures.....	VII
List of Tables	IX
List of Abbreviations.....	X
List of Mathematical Symbols.....	XII
Chapter 1 Introduction.....	1
1.1 Background and Motivation	1
1.2 Research Objectives	8
1.2.1 Formulation and Performance Analysis of the Scheduling Problem in TWIN	8
1.2.2 A General Framework for Guaranteed QoS in Slotted Optical Transport Networks	8
1.2.3 Admission Control and Scheduling for Guaranteed QoS in a Scalable WDM EPON Network Architecture	9
1.2.4 An Admission Control Scheme for IEEE 802.11e WLAN Supporting Deterministic QoS Provisioning	9
1.3 Thesis Contributions.....	10
1.4 Thesis Organization.....	12
Chapter 2 Literature Review	13
2.1 Scheduling in TWIN and Existing Works	13
2.2 QoS Provisioning in Single-Hop and Multi-Hop Optical Networks	15
2.3 QoS Provisioning in EPON and WDM EPON/PON.....	16
2.3.1 QoS Provisioning in EPON	16
2.3.2 WDM EPON/PON Upgrade Architecture and QoS Provisioning.....	17
2.4 QoS Provisioning in WLAN.....	18
2.5 Network Calculus and Its Applications	20
Chapter 3 Performance Analysis of a Novel Traffic Scheduling Algorithm in TWIN	25

3.1	Introduction	25
3.2	Problem Formulation.....	26
3.3	The Scheduling Policy.....	28
3.3.1	DSS Scheduling Algorithm	28
3.3.2	Complexity Analysis	30
3.3.3	Pipeline Scheduling	32
3.4	Performance Analysis.....	33
3.5	Simulation and Result Discussion	40
3.6	Conclusion.....	46
Chapter 4 Providing Deterministic Quality of Service in Slotted Optical Networks		47
4.1	Introduction	47
4.2	Framework Configuration Overview.....	47
4.3	Control Plane	48
4.3.1	Overview	48
4.3.2	Propagation Delay Compensation	49
4.3.3	Conflict Resolution Algorithm	52
4.3.4	Simulation Results	54
4.3.5	Periodic Timeslot Service.....	56
4.4	Data Plane.....	57
4.4.1	Class-based Traffic Assembly Process	57
4.4.2	Flow Classification	58
4.4.3	Class-based Scheduler	59
4.4.4	Application Examples.....	60
4.4.5	Simulation Results	63
4.5	Conclusion.....	64
Chapter 5 Admission Control and Scheduling in a Scalable WDM EPON Architecture		66
5.1	Introduction	66
5.2	The Proposed Scheme	67
5.2.1	Overview of the WDM EPON Architecture.....	67
5.2.2	Wavelength scaling process.....	68
5.2.3	Admission Control Schemes.....	71
5.2.4	Scheduling Discipline.....	80
5.3	Simulation Results.....	83
5.4	Conclusion.....	91
Chapter 6 A Scheme for IEEE 802.11e WLAN Supporting Deterministic QoS Provisioning		92

6.1	Introduction	92
6.2	Admission Control Policy and Scheduling Algorithm	93
6.2.1	Admission Control Policy.....	93
6.2.2	Service Curve Based Scheduling Algorithm (SCBS).....	96
6.3	Simulation Results and Discussion.....	99
6.4	Conclusion.....	104
Chapter 7 Conclusions and Future Work.....		105
7.1	Conclusions	105
7.2	Future Work.....	108
Author's Publications		109
Bibliography		110

Summary

The increasing popularity of new multimedia services has made it desirable to support end-to-end Quality-of-Service (QoS) provisioning across multiple network domains. It is a very challenging problem and there is little progress in an implementable solution so far. Instead of providing a full solution at this stage, our goal in this thesis work is to implement QoS guarantees in each individual network that a user connection passes (e.g. transport networks and access networks), which is the necessary condition for an overall inter-networking QoS architecture.

Time-Domain Wavelength Interleaved Network (TWIN) has been proposed by Bell Laboratories as a promising transport network architecture. TWIN performs efficient traffic grooming without resorting to electronic operations as is done today in SONET/SDH. Moreover, it eliminates the need for fast optical switches and optical buffers, which is still the main technical barrier faced by Optical Packet Switching (OPS) and Optical Burst Switching (OBS).

To facilitate the analysis of TWIN, we build an analytical model for TWIN from which delay characteristics could be derived. Further, we propose a novel scheduling algorithm for TWIN that is able to provide improved QoS. The simulation results show that our algorithm could achieve much better QoS, in terms of smaller queueing delay and queueing delay variance, than existing works without sacrificing throughput.

As a step forward, we propose a framework based on the architecture of TWIN that is able to provide guaranteed QoS. The framework uses the combination of a control plane and a data plane to solve the complex problem of capacity allocation, slot-matching and traffic scheduling. Most related works could only provide such guarantees in single-hop networks such as star-coupled networks. In contrast with them, our framework could guarantee deterministic delay bounds in arbitrary topology mesh networks.

Access networks are also studied in this thesis. Providing deterministic QoS guarantees in WDM Ethernet Passive Optical Network (EPON) is a very challenging issue. We propose a two-

level WDM EPON architecture. The upper level is responsible for capacity scaling and bandwidth sharing across multiple groups, and the lower level takes charge of admission control and traffic scheduling inside each group. The scheme is able to provide class-based deterministic delay bounds to traffic flows and maximize wavelength utilization.

The IEEE 802.11 Wireless LAN (WLAN) has been widely deployed for ease of access which can be regarded as the extension of traditional access networks. We propose an admission control and scheduling scheme for IEEE 802.11e HCCA. With the help of traffic regulators, it eliminates the impact of ill-behaving traffic flows and maintains the delay bound for each class of traffic, gaining good improvement over the reference HCCA scheduler suggested in the standard.

This thesis has proposed a series of solutions for deterministic QoS provisioning in individual autonomous networks from the core (transport networks) to the edge (access networks), as well as at the user interface (wireless networks). The results can be used as the theoretical foundation for future research of end-to-end QoS provisioning.

List of Figures

Figure 2.1: Illustration of TWIN architecture and its scheduling problem	14
Figure 2.2: A TXOP allocation example in 802.11e HCF	19
Figure 2.3 Arrival curve for Intserv flows, rate-latency service curve,	22
Figure 3.1: The DSS scheduling algorithm	29
Figure 3.2: Illustration of the upper bound of Q_j	31
Figure 3.3: Information collection in central scheduling	32
Figure 3.4: Scheduling framework of TWIN	33
Figure 3.5: Configuration inside a node	34
Figure 3.6: Decomposition of the queues inside a node	35
Figure 3.7: An example network and its destination tree	35
Figure 3.8: Delay comparison in a 6-node network	41
Figure 3.9: Delay comparison in the 14-node NSFNET	42
Figure 3.10: Queueing delay comparison in NSFNET under Poisson traffic	42
Figure 3.11: Queue length comparison in NSFNET under Poisson traffic	43
Figure 3.12: Queueing Delay comparison in NSFNET under bursty traffic	43
Figure 3.13: Queue length comparison in NSFNET under bursty traffic	44
Figure 3.14: Variance of queueing delays comparison in NSFNET	45
Figure 3.15: Throughput comparison in NSFNET	45
Figure 4.1: TWIN network with a bandwidth broker controller	48
Figure 4.2: Slot transmission in a period	49
Figure 4.3: The conflict-resolution algorithm in the control plane	53
Figure 4.4: The arrival curve and service curve of an aggregate flow I	56
Figure 4.5: The class-based traffic assembly and scheduling process	57
Figure 4.6: Flow classifications with M QoS classes and a best effort class (class 0)	59
Figure 4.7: Comparison of service rates required for two arrival curves $R_1(t):(\sigma_j, \rho_j)$	62
Figure 4.8: Delay distribution with three classes of traffic	64
Figure 5.1: An illustration of the proposed WDM EPON architecture	67
Figure 5.2: The creation of a Dynamic Group	68
Figure 5.3: Flowchart of the wavelength scaling process	70
Figure 5.4: Pseudocode of process Suspend_Flows	77
Figure 5.5: Admission priority in scheme C	78

Figure 5.6: Pseudocode of process Restore_Flows	78
Figure 5.7: Illustrative example of the transmission sequence.....	80
Figure 5.8: Arrival curve and service curve illustration.....	82
Figure 5.9: Two different cases in scheduling	83
Figure 5.10: Delay distribution of admission control scheme A, B and C.....	84
Figure 5.11: Bandwidth utilization of admission control scheme A, B and C	86
Figure 5.12: Flow loss ratio of admission control scheme A, B and C	87
Figure 5.13: Maximum packet delay for different schemes.....	89
Figure 5.14: Extra wavelengths required for different schemes.....	90
Figure 6.1: The Service Curve Based Scheduling (SCBS) algorithm.....	98
Figure 6.2: Comparison of average delay	100
Figure 6.3: Comparison of maximum delay.....	100
Figure 6.4: Delay distribution at the load of 0.8	101
Figure 6.5: Percentage of packets exceeding their delay bound	102
Figure 6.6: Session blocking probability of the two schedulers.....	103
Figure 6.7: Throughput comparison of the two schedulers.....	104

List of Tables

Table 3.1: Variables used in the DSS algorithm	29
Table 4.1: Network utilization in TWIN networks with different number of nodes.....	54
Table 4.2: Network utilization in a 200-node TWIN network for 2 slots allocation	54
Table 4.3: Network utilization in a 200-nodes TWIN network for random of 1 to 4 slots allocation.....	55
Table 5.1: Chosen token buckets for the three classes	84
Table 6.1: Variables used in the SCBS algorithm.....	97
Table 6.2: Traffic specifications of the three VBR flows	100
Table 6.3: The parameters of the three classes.....	100

List of Abbreviations

AC	Access Category
AIFS	Arbitration Inter Frame Space
AP	Access Point
APON	ATM Passive Optical Network
AWG	Arrayed Waveguide Grating
BB	Bandwidth Broker
CAP	Controlled Access Phase
CBR	Constant Bit Rate
CO	Central Office
CSMA/CD	Carrier Sense Multiple Access/Collision Detect
DBA	Dynamic Bandwidth Allocation
DCF	Distributed Coordination Function
DIFS	Distributed Inter Frame Space
DXC	Digital Cross-Connects
DSL	Digital Subscriber Line
EDCA	Enhanced Distributed Channel Access
EPON	Ethernet Passive Optical Network
HCCA	HCF-controlled Channel Access
HCF	Hybrid Coordination Function
HFC	Hybrid Fiber Coax
ILP	Integer Linear Programming
LAN	Local Area Network
MPCP	Multipoint Control Protocol
OEO	Optical-Electrical-Optical
OBS	Optical Burst Switching
OLT	Optical Line Terminal
ONU	Optical Network Unit
OPS	Optical Packet Switching
PSC	Passive Star Coupler
P2P	Peer-to-Peer
PCF	Point Coordination Function

PIFS	Point Coordination Function Inter Frame Space
PON	Passive Optical Network
QAP	QoS Access Point
QoS	Quality of Service
QSTA	QoS Station
RAM	Random-access Memory
RN	Remote Node
RSVP	Resource Reservation Protocol
RTT	Round Trip Time
SDH	Synchronous Digital Hierarchy
SONET	Synchronous Optical Network
SI	Service Interval
STA	Station
TS	Traffic Stream
TSPEC	Traffic Specification
TWIN	Time-domain Wavelength Interleaved Network
TXOP	Transmission Opportunity
UP	User Priority
VBR	Variable Bit Rate
WDM	Wavelength Division Multiplexing
WLAN	Wireless Local Area Network
WSS	Wavelength-Selective Switch

List of Mathematical Symbols

$A_{ij}(t)$	number of bursts having arrived and queued for node pair (i, j) at time t
b	number of bits in a slot
C	PHY data rate
C_{Avail}	available bandwidth of the upstream channel
C_{HCCA}	upstream bandwidth that can be used by HCCA
C_w	available upstream bandwidth in group w
$D_{ij}(t)$	number of bursts waiting to be transmitted from node i to node j at time t
D_i	the delay bound of class i
d_j	the delay requirement of flow j
F_h	set of flows in class h
g	guard time
L_{th}	queue size threshold
L_i	nominal packet size of class i traffic
L	length of a transmission cycle
M	number of classes
N	number of nodes
N_w	number of ONUs in group w
O_{max}	maximum polling overhead in a beacon
Q_j	set of time slots which have been occupied at node j
$R_i^k(t)$	arrival curve of class i traffic in node k
RTT_x	round trip time of polling ONU x
$S_i^k(t)$	service curve provided to class i traffic in node k
$S_{ij}(t)$	equal to 1 if node pair (i, j) is selected to transmit a burst at time t , and 0 otherwise
$TXOP_i^k$	TXOP assigned to class i traffic in QSTA k
$u(t)$	unit function
w	number of slots assigned to current node pair
α_i	bandwidth share of class i
Δ	a time slot in TWIN
δ_{ij}	propagation delay from node i to node j in units of time slots
ρ_j	average rate of flow j

σ_j	burstiness of flow j
ρ_{new}	average rate of the new flow
σ_{new}	burstiness of the new flow

Chapter 1 Introduction

1.1 Background and Motivation

Quality-of-Service (QoS) provisioning plays an increasingly important role in today's networks as more and more real-time multimedia applications need to be supported, such as voice over IP, online games and IP-TV. These applications are delay sensitive and require that their traffic be delivered in certain timeframes otherwise the service would fail. End-to-end QoS guarantee is highly desirable in communication networks but it is a very challenging problem. Current solutions in wired networks such as the IntServ and DiffServ models are not able to truly achieve end-to-end QoS due to scalability or inability to control acquired burstiness in shared link. So far there is little progress in an implementable solution for overall end-to-end QoS across multiple domains. Before such inter-networking QoS guarantees could be truly achieved, we first attempt to achieve guaranteed QoS in individual networks such as transport networks and access networks, because end-to-end QoS guarantee is possible if and only if all autonomous networks through which a connection passes are able to provide QoS guarantees individually.

Traditional transport networks are widely based on synchronous optical network (SONET) or synchronous digital hierarchy (SDH) rings [1] and they have certain weaknesses. These SONET-based networks are primarily designed to provide point-to-point connectivity for synchronous traffic, and are not well suited to supporting asynchronous traffic with a variety of bandwidth granularities or dynamic bandwidth requirements. Currently, services based on these point-to-point wavelength connections are generally not economical since the total traffic demand between node pairs has not scaled to the wavelength capacity. Moreover, SONET-based networks rely on digital cross-connects (DXCs) to perform the switching function. A DXC requires an optical-electrical-optical (OEO) converter at each port, which tends to incur a relatively large portion of the system cost.

Chapter 1. Introduction

Two schemes have been proposed to address these issues. One is the optical packet switching (OPS) [2]. It is a theoretically ideal solution for future transport networks due to its high-speed, fine-granularity switching, flexibility, and ability to use the resources economically. However, the technology is still in a very early stage, and several issues need to be solved, including switching architectures, synchronization, contention-resolution schemes, etc.

Contention of packets in a switch fabric is a major problem and has important impact on the performance of the network such as packet delay, packet-loss ratio, throughput, etc. Contention occurs when two or more packets try to leave the switch from the same output port at the same time. To resolve such a conflict, one of the packets is routed through the output port while others are routed elsewhere, depending on the contention-resolution scheme. In electronic switches, contention can be resolved by the store-and-forward technique where packets in contention are stored in a buffer and sent out when the port becomes available. Unfortunately, optical buffers similar to electronic counterparts (i.e., random-access memory (RAM)) do not exist. The only way to store an optical packet is to use optical delay lines (fixed length fibers). These “sequential access” buffers are less flexible than an electronic “random access” buffer because a packet entering a delay line will emerge from the other end of the line after a fixed amount of time. In such optical buffers, the delay time can not be precisely controlled.

Optical Burst Switching (OBS) [3] was proposed as another candidate for next-generation transport networks. In OBS, a control packet is sent first to reserve an appropriate amount of bandwidth and to pre-configure the switches along the path. The burst of data, which can consist of several packets, immediately follows the control packet, without waiting for an acknowledgement. OBS may lead to better resource usage compare to circuit switching because reserved resources are released after the completion of the burst. However, there are several issues that need to be addressed before OBS can be deployed, such as optical burst switch architecture, and deciding of the time gap between the control packet and the data burst (offset time). The latter issue is the major problem in OBS and it greatly affects the performance of OBS. OBS also need optical buffers to resolve burst conflicts, or it has to suffer a great packet loss because a lost burst usually consists of a number of packets.

Chapter 1. Introduction

Although OPS and OBS could theoretically improve the utilization of wavelength connections in transport network, the technological barriers they face do not seem to be easily overcome in the near future. Recently, Bell Laboratories proposed a new optical transport network called Time-Domain Wavelength Interleaved Network (TWIN) [4]. In TWIN, a number of practical difficulties in building next-generation transport networks, in particular the need for fast optical switching and optical buffers, can be effectively dealt with. TWIN also performs efficient traffic grooming without resorting to electronic cross-connects, as is done today in SONET/SDH (see [5,6]).

Briefly speaking, TWIN defines an optical network consisting of a simple core that is based on passive wavelength-selective switches (WSS) capable of routing incoming wavelengths to the appropriate output ports, and an intelligent edge using fast tunable lasers to emulate fast switching of data in the core. In TWIN each destination node j is assigned a unique wavelength λ_j . When a source node has data to send to destination j , the source tunes its laser to λ_j for the duration of the data transmission. To ensure that the optical signal of a given wavelength from a source will be routed to the intended destination, the WSS at each node is preprovisioned so that an incoming optical signal of a given wavelength is routed to the appropriate port. Once preprovisioned, each WSS performs self-routing of optical signals on-the-fly based on their wavelengths. Fast active optical switching is physically eliminated in this architecture as the WSS reconfiguration timescale is equivalent to the timescale when routers are modified, which is relatively long under normal operating conditions (e.g., no failures). Tunable lasers at the sources emulate fast switching and enable optical grooming through dynamic tuning of appropriate wavelengths at designated times. Optical bursts sent out from the sources will not be buffered in any intermediate nodes which removes the need for optical buffers. To prevent collisions at any intermediate node or at the destination, TWIN relies on scheduling to coordinate the tunable lasers. The special scheduling problem of TWIN will be described in detail in Chapter 2 together with existing scheduling algorithms.

With the initial design, TWIN has a physical limitation that the number of nodes in the network is restricted by the number of wavelengths available in a single fiber. This issue has been

Chapter 1. Introduction

addressed in a recent work called TWIN with wavelength reuse (TWIN-WR) [7]. By dividing the problem into several subproblems and solving them in order, it successfully achieved the goal of supporting N nodes with roughly $N^{1/2}$ wavelengths.

We are interested in TWIN because, aside from the fact that it is not constrained by current technology limitations (i.e. lack of fast optical switches and optical buffer) as OPS and OBS do, it bridges the gap between end-to-end demands and wavelength capacity by performing efficient traffic grooming without resorting to electronic cross-connects. It also has a simple layering and a simple core which make it highly scalable. Due to these merits, it is a good transport network platform on which QoS mechanisms could be developed. It will be demonstrated in Chapter 2 that although there are some scheduling algorithms proposed for TWIN, they focused mainly on maximizing the throughput and ignored the consideration of QoS. This has not fully exploited the potential of TWIN. We would like to propose feasible schemes that could achieve improved or even deterministic QoS in transport networks based on the architecture of TWIN.

Aside from transport networks, access networks are another main point considered in this thesis work. Access networks are often referred to as the *last mile* or *first mile*, of which the latter emphasizes its importance to customers. Digital Subscriber Line (DSL) and Hybrid Fiber Coax (HFC) are the most widely deployed access network solutions today. Compared to the transport network employing optics technology, these solutions provide much lower bandwidth and have become the bottleneck in today's network infrastructure. Passive Optical Network (PON) is a technology viewed by many as an attractive solution to the first mile problem. Compared to the DSL and HFC networks, PON is cost-effective and is able to provide much larger bandwidth. Different types of PON have appeared in the last few decades. In 1990s the Full Services Access Network (FSAN) group defined a PON-based optical access network that used ATM as its layer 2 protocol (ITU G.983). However, ATM suffers from heavy cell tax overhead and high cost of equipments (network cards, switches, etc.). Due to these weaknesses, ATM-based PON (APON) has gradually lost its attraction. The Ethernet PON (EPON), standardized by the IEEE 802.3ah Ethernet in the First Mile Task Force [8], has gained a great deal of interests in both industry and academia recently. EPON carries all data encapsulated in Ethernet frames. Given the fact that

Chapter 1. Introduction

more than 90 percent of today's data traffic originates from and terminates in Ethernet LANs, EPON appears to be a natural candidate for future first mile solutions.

Typically, EPON has a physical tree topology with the central office located at the root and the subscribers connected to the leaf nodes of the tree. The Optical Line Terminal (OLT), which resides at the central office, delivers data to multiple customer premises equipments called Optical Network Units (ONUs) through a *passive* 1:N optical splitter/combiner. Each ONU provides access service to multiple business/residential subscribers. EPON has two dedicated wavelength channels for transmitting the downstream and upstream traffic respectively. In the downstream direction, it is a broadcast network: Ethernet frames are broadcast by OLT and selectively received by each ONU. In the upstream direction, however, EPON is a multipoint-to-point network, where multiple ONUs share the same transmission channel. A proper mechanism is thus required for coordinating the upstream transmissions. The IEEE 802.3ah task force has developed the so-called multipoint control protocol (MPCP) to arbitrate the access of ONUs to the shared medium. The dynamic bandwidth allocation (DBA) algorithm module is the key component in MPCP which is responsible for assigning non-overlapping upstream transmission windows (time slots) to each ONU.

With the rapidly increasing number of users and bandwidth intensive applications, people have begun to consider upgrading current single-channel EPON to satisfy the traffic demands in the future. One possible upgrade approach is to increase the line rate of current EPON by replacing lower speed transceivers with higher speed transceivers. Although this can increase the system capacity to a higher level (e.g., from 2.5 Gb/s to 10 Gb/s), the huge bandwidth of an optical fiber is still mostly wasted. Alternatively, we could deploy multiple wavelength channels in the existing infrastructure in the upstream and/or downstream directions, resulting in wavelength division multiplexing (WDM) EPON [9]. WDM provides a cautious upgrade, wherein wavelengths can be added as needed. More importantly, the existing hardware (optical fibers) can be exploited more efficiently, bringing higher revenues to the service providers without much further investment.

Chapter 1. Introduction

QoS provisioning in EPON has been investigated thoroughly in the past few years. However, it is a rather new topic in WDM EPON/PON. To the best of our knowledge, very few research works have investigated this issue. Providing QoS guarantees in WDM EPON/PON still remains a big challenge, let alone strict QoS assurances.

Wireless technology has been largely used as a complementary means for access networks where wired connections are not available or easily accessible. Currently, IEEE 802.11 Wireless LAN (WLAN) is the most popular broadband wireless access solution. Its convenience and low cost have made it widely deployed in public and residential locations such as airports, shopping malls, hotels and apartments. Providing strict QoS guarantees in WLAN is also essential as it has become the last step to user interface. Since the transmission speed by radio is much lower than by wire, in the early stage of WLAN people focused on increasing the transmission rate or maximizing the bandwidth utilization rather than providing QoS guarantees. The IEEE 802.11 a/b/g standards did not provide any specific means to guarantee the QoS. Recently, the 802.11 Working Group realized its importance and developed a new standard protocol, namely IEEE 802.11e [10], to provide service differentiation mechanism in WLAN.

A new MAC layer function called the hybrid coordination function (HCF) is proposed in 802.11e standard. It uses both enhanced distributed channel access (EDCA) mechanism for contention-based transfer and a HCF-controlled channel access (HCCA) mechanism for contention-free transfer. They have replaced distributed coordination function (DCF) and point coordination function (PCF) in the legacy standards. The access point (AP) and stations (STAs) are called QoS access point (QAP) and QoS stations (QSTAs) to differentiate them from legacy ones. Under HCF, the basic unit of channel allocation is the transmission opportunity (TXOP), which refers to a time duration during which a QSTA is allowed to transmit a burst of data frames. The TXOP obtained by EDCA contention is called EDCA TXOP and the TXOP obtained by receiving a CF-Poll from QAP during contention period or contention-free period is called HCCA TXOP. A QSTA can transmit multiple frames within its TXOP allocation.

EDCA

The EDCA mechanism provides differentiated, distributed access to the wireless medium for QSTAs. It defines four access categories (ACs). Each data packet from the higher layer with a specific user priority (UP) value should be mapped into a corresponding AC. EDCA is similar to the legacy DCF access mechanism, but differs in the following aspects. First, each AC has its specific contention parameters ($CW_{\min}[AC]$, $CW_{\max}[AC]$, $AIFS[AC]$), while in legacy DCF the parameters are defined for all STAs. Besides, the arbitration inter frame space (AIFS) is introduced in EDCA to replace distributed inter frame space (DIFS) in DCF. It can be chosen as different values to give different ACs different priorities to access the medium. This is because the AC with a shorter AIFS could access the medium earlier than another AC with a longer AIFS, hence is given a higher priority. Second, each AC behaves as a single DCF contending entity and receives its TXOP from the AP. Third, EDCA allows for multiple frame transmission in the TXOP limit instead of one frame transmission.

HCCA

The HCCA manages access to the wireless medium using a hybrid coordinator that has higher medium access priority than non-AP STAs. HCCA is more flexible than PCF in that the latter is only allowed in a contention-free period (CFP), while a QAP can initiate HCCA whenever it wishes in the whole beacon interval (including contention period (CP)). During a beacon interval, a QAP is allowed to start several contention-free periods called controlled access phases (CAPs) using HCCA at any time after detecting the medium has been idle for a time interval of point coordination function inter frame space (PIFS). Different traffic classes called traffic streams (TSs) are introduced in HCCA, and each QSTA is allowed to have no more than eight TSs. The standard has defined a reference scheduler to assign TXOPs to different QSTAs. Its working manner will be introduced in detail in Chapter 2.

Although EDCA is an improved version of DCF, it is still incapable of providing any strict delay bounds because of its contention-based transmission manner. HCCA can precisely control the access of each QSTA, hence it is a better candidate for guaranteeing strict QoS in WLAN.

1.2 Research Objectives

1.2.1 Formulation and Performance Analysis of the Scheduling Problem in TWIN

Traditional transport networks based on SONET or SDH rings are generally not economical since the total traffic demand between node pairs has not scaled to the wavelength capacity. They have also employed expensive OEO converters in the electrical cross-connects. Time-Domain Wavelength Interleaved Network (TWIN) [4] has been proposed as cost-effective transport network architecture that could improve the utilization of end-to-end wavelength connections. It also eliminates the need for fast optical switches and optical buffers, which are the main technological barriers that obstruct OPS and OBS. QoS guarantees can be provided in TWIN by well-designed scheduling algorithms. Due to this, TWIN has supplied a good platform for achieving good QoS. Currently, several scheduling algorithms have been proposed for TWIN [11,12] but they focused mainly on maximizing throughput without any consideration of QoS. Our first objective is to develop a scheduling algorithm that could achieve improved QoS in TWIN without sacrificing throughput.

1.2.2 A General Framework for Guaranteed QoS in Slotted Optical Transport Networks

QoS provisioning in optical networks has been studied in different scenarios. Some investigated the issue in single-hop networks, such as the optical packet switch [13-15] or star coupled network [16-24]. In single-hop networks, to guarantee the total delay the scheduler only need to control the queueing delay since each node (port) has equal (negligible) propagation delay to all the other nodes (ports). A more complicated case is the multi-hop network with mesh topology, such as the OBS network. In this kind of networks, it is not easy to control the total delay because the routes are changing from time to time. A number of QoS provisioning schemes in OBS networks have been proposed [25-35] but they are unable to achieve strict QoS guarantees such as delay bounds for each burst/packet. Thanks to TWIN's special architecture, the routes in TWIN are generally fixed (unless emergencies) for each node pair which implies that the

propagation delays are fixed. Hence, we are able to control the total delay by controlling the queueing delay at source nodes, making it possible to achieve strict QoS in a mesh network. Our second objective is to propose a general framework for slot-based optical transport networks that is able to support *deterministic* QoS. This framework provides functions of resource allocation, admission control and traffic scheduling in a mesh network of arbitrary topology.

1.2.3 Admission Control and Scheduling for Guaranteed QoS in a Scalable WDM EPON Network Architecture

EPON has been viewed by many as an attractive solution to the first mile problem. Though it has rather large capacity, it will soon become insufficient because of the rapidly increasing number of bandwidth-intensive applications. Upgrading its architecture with WDM technology is a natural and cost-effective choice. However, compared to EPON which has substantial relevant works on QoS provisioning [36-47], providing QoS assurance in WDM EPON/PON is a rather new and challenging topic. There are a number of WDM EPON/PON architectures or upgrade solutions in the literature [48-51], but to the best of our knowledge, none of them could provide deterministic QoS in their schemes. On the other hand, it is also desirable that any proposed scheme is able to utilize the multiple wavelengths with maximum efficiency. Our third objective is to propose a WDM EPON architecture which could perform efficient capacity scaling with minimum number of wavelengths and provide deterministic QoS guarantees to multimedia traffic.

1.2.4 An Admission Control Scheme for IEEE 802.11e WLAN Supporting Deterministic QoS Provisioning

IEEE 802.11 Wireless LAN (WLAN) is widely used as a complimentary means for access networks and hence becomes the last step to user interface. It becomes essential to provide guaranteed QoS in WLAN for time-constrained applications. The newly developed IEEE 802.11e standard offers a very good platform for achieving such a goal. It defines EDCA mode for contention-based access and HCCA mode for contention-free access respectively. There are many schemes proposed for EDCA for enhanced QoS provisioning [52-66], however, HCCA is

more competent in providing guaranteed QoS thanks to its contention-free nature. The HCCA reference scheme in the standard has very limited capabilities and only performs well under constant bit rate (CBR) traffic. A number of enhancements have been proposed for achieving better QoS on HCCA [67-75]. Our last objective is to propose an admission control scheme in conjunction with the scheduling discipline for IEEE 802.11e HCCA that is able to handle bursty traffic and achieve deterministic delay bounds for each traffic flow.

1.3 Thesis Contributions

In this thesis, we have proposed a series of solutions for guaranteed QoS provisioning in different network domains that a user connection may pass, including optical network at the core (TWIN), optical network at the edge (WDM EPON) and wireless network (WLAN) connected to user interface. Deterministic QoS in these networks is a necessary condition for deterministic end-to-end QoS, yet more research is required to truly achieve such a goal. The contributions of the thesis are summarized as follows.

1. We have formulated the scheduling problem of TWIN as an integer linear programming (ILP) problem and presented a novel heuristic algorithm called Destination Slot Set (DSS) to solve it effectively. Existing batch scheduling algorithms proposed for TWIN have good performance on throughput, but have relatively large queueing delays because of their batch transmission behavior (a batch may include a large number of time slots). In contrast, our algorithm schedules transmissions in every time slot, hence achieving much smaller queueing delays. Due to the special architecture of TWIN, smaller queueing delay implies smaller total delay from the source to the destination. Hence such an algorithm could achieve much better QoS than existing batch scheduling algorithms. More importantly, this benefit is not gained at the expense of throughput. We have conducted simulations for our DSS algorithm and the results proved its advantages over existing batch algorithms. This work has been published in [76].

Chapter 1. Introduction

2. Based on TWIN architecture, we have proposed an efficient framework for deterministic service guarantees in slot-based WDM mesh networks. The framework uses the combination of a control plane and a data plane to solve the complex problem of capacity allocation, slot-matching and traffic scheduling. The control plane implements admission control and capacity allocation to source-destination node pairs and the data plane handles traffic aggregation, buffering, and scheduling. Through our analysis, we proved that the framework realizes Service Curve [77] assurance that is able to provide class-based queueing delay bounds to traffic flows. Our simulation results have validated our analysis. The simplicity and scalability of this framework make it suitable for implementation in mesh optical transport networks. This work has been published in [78].
3. We have proposed a two-level WDM EPON architecture which implements two main functions: efficient group capacity scaling with bandwidth sharing at the upper level, and deterministic QoS provisioning for individual flows at the lower level. By wavelength scaling each group in the network could ask for extra upstream bandwidth when it becomes saturated, and different groups could share the same extra wavelength such that its utilization could be maximized. Inside each group, an integrated scheme including the admission control policies and scheduling discipline has been developed to guarantee deterministic QoS for multiple classes of traffic. We have proposed three different admission control policies, in particular fixed-proportional, non-preemptive and preemptive admission control, and provided a complete study of their unique features. Through simulations we showed that all the three admission control schemes could guarantee deterministic delay bounds for bursty traffic. The non-preemptive admission control scheme achieves most balanced performance in terms of bandwidth utilization and flow loss ratio. Some preliminary results of this work have been published in [79].
4. We have developed a class based admission control and scheduling algorithm called SCBS for IEEE 802.11e WLAN. The reference scheduler suggested in the standard only performs well with constant bit rate (CBR) traffic and has substantial delays if the traffic becomes

bursty. In contrast, our scheme could handle variable bit rate (VBR) traffic effectively; at the same time, the impact of non-conformant traffic could be avoided. Our simulations showed that while the standard reference scheduler suffers significant delays with VBR traffic, our scheme still maintains the delay bound for traffic flows in each class. Simulation study also showed that the reference scheduler has a much larger portion of packets exceeding their delay bounds than SCBS. On the other hand, SCBS has a higher session blocking probability than the reference scheme when the traffic intensity is high, because it reserves more resource than the reference scheduler when taking account of the burstiness. This is a necessary cost for guaranteed QoS. This work has been published in [80].

1.4 Thesis Organization

The thesis is organized as follows. In Chapter 2, we provide a thorough review of related research works in the literature, including the basics of network calculus theory that we are going to apply in later chapters. In Chapter 3, we propose a heuristic algorithm for scheduling in TWIN that is able to achieve better QoS than existing scheduling algorithms. We go one step further in Chapter 4, where deterministic QoS could be achieved in slotted optical transport networks based on our proposed framework. After achieving deterministic QoS in optical core networks, we try to achieve such guarantees in optical access networks. In Chapter 5, we propose a WDM EPON architecture together with three admission control schemes, which are capable of providing deterministic QoS to traffic flows inside each user group. Finally, we place our focus on achieving deterministic QoS in wireless networks connected to user interface. In Chapter 6, we study the IEEE 802.11e WLAN and propose a scheme capable of supporting deterministic QoS provisioning. These four chapters have provided a series of solutions for deterministic QoS provisioning in optical core network, optical access network and user wireless network individually.

We present our conclusions and discuss future research directions in Chapter 7.

Chapter 2 Literature Review

In this chapter we will provide a comprehensive review of related important research works. Firstly, we will describe the scheduling problem in TWIN and its distinction from other scheduling problems. We will then discuss the pros and cons of the existing scheduling algorithms in TWIN. Secondly, we will review those schemes capable of providing QoS guarantees in single-hop and multi-hop optical networks. Thirdly, we will review DBA schemes for QoS provisioning in EPON and WDM PON. Fourthly, we will review QoS schemes in IEEE 802.11e WLAN, including the reference admission control and scheduling scheme in the standard and other QoS enhancement schemes.

2.1 Scheduling in TWIN and Existing Works

Scheduling in communication networks is a well-studied topic. Existing works can be classified into two main categories: packet scheduling in a core router and packet scheduling in a crossbar switch. In typical packet networks, scheduling takes place in a router where multiple packets compete for transmission on a common outgoing link. How to properly select one packet at a time to transmit is the problem that the scheduler needs to solve. The well-known scheduling algorithms such as Virtual Clock [81], Self-clocked Fair Queueing [82], Generalized Processor Sharing (GPS) and packetized GPS (PGPS) [83], Deficit Round-Robin [84], Worst-case Fair Weighted Fair Queueing (WF²Q) [85], and Credit Based Fair Queueing [86] belong to this category. More recent works such as [87,88] considered fair scheduling for the support of Diffserv traffic in a core router. In a crossbar switch, however, the schedulers generally set up connections for transmission of packets between a number of input and output ports. In this kind of scheduling, the scheduler should avoid conflict of multiple packets at a same output port. While early schemes focused on maximizing throughput or improving fairness [89,90], recently

proposed switch schedulers began to incorporate QoS support [91-93] for scheduling of multimedia traffic.

Figure 2.1: Illustration of TWIN architecture and its scheduling problem

Scheduling in TWIN is much different. We illustrate it by Fig. 2.1. The figure shows traffic on two different wavelengths (represented by solid and dashed line) destined for two destination nodes (D and F). The wavelength connections from various sources to a particular destination form a multipoint-to-point tree called destination tree [11]. For example, in the figure the solid and dashed lines form two different destination trees for node D and F respectively. In TWIN it is required that bursts (a burst is the amount of traffic sent in a time slot) from various sources to a destination should not collide at the destination or at any intermediate node, otherwise the transmission would fail. Since time in TWIN is slotted, it requires that no two or more bursts can arrive at the destination or at any intermediate node in a same slot on a same channel. For example, in Fig. 2.1 the boxes with solid border represent the bursts received at the destination, and boxes with dashed border represent empty slots. We can see that there is a burst from node A arriving at node D in time slot 1. If there is another burst from node E also arriving at node D in time slot 1, conflict occurs and transmission fails. Since this kind of scheduling problem has not been studied in any previous work before TWIN appears, there are no available schemes that can be applied to it directly.

Currently only a few scheduling algorithms have been proposed for TWIN. In [12] a dynamic scheduling algorithm was proposed. The basic idea is to formulate the problem into a well-studied crossbar switch scheduling problem called Birkhoff-von Neumann crossbar switch [94]. By applying the Birkhoff-von Neumann theorem [95], the algorithm decomposes the traffic

demand into a batch of transmissions. It has to wait a certain number of time slots for all bursts to reach their destinations in one batch before starting the next batch due to non-zero propagation delays between node pairs.

Another scheduling algorithm called TWIN Iterative Independent Set (TIIS) was proposed in [11]. The basic idea is to construct a bipartite graph based on the scheduling information, and use the algorithm by Kahale et, al. [96] to find the maximum-weight independent set which can be applied as a batch schedule. This algorithm performs better than the previous one with low traffic, but it has very high computation complexity. In each scheduling, it needs the computation time of $O(N^5)$ where N is the number of nodes.

Both [11] and [12] focused on maximizing the throughput without any consideration of QoS. The work of [97] achieved better QoS (in terms of smaller average queueing delay) without sacrificing throughput. However, it assumed a simple topology of equal distance on each link, which limited its scope of application.

2.2 QoS Provisioning in Single-Hop and Multi-Hop Optical Networks

The issue of QoS provisioning has been studied for both single-hop and multi-hop optical networks. A large portion of the existing works are based on single-hop networks, such as the star coupler [16-24] or optical packet switch [13-15]. Most of them derived statistical delay characteristics by queueing theory which are not strict guarantees, except works in [22-24] could provide deterministic QoS by employing max-plus algebra [98]. In a single-hop network, each node has the same propagation delay to all the other nodes without the need for routing. Therefore, these results can not be applied directly to multi-hop networks with arbitrary topology in which packets may have different propagation delays at different times due to changing of routes.

OBS networks are usually multi-hop networks. A number of QoS provisioning schemes have been proposed for OBS [25-35]. They usually rely on resource reservation or controlling the offset time to achieve better end-to-end delays. However, since routes in OBS networks changes

from time to time, these mechanisms may not always make accurate predictions (e.g., offset time) and hence are unable to achieve deterministic QoS such as delay bounds. Although there are some works claiming supporting absolute QoS differentiation in OBS [99-101], their definition of “absolute QoS” is that the loss probability of the guaranteed traffic does not exceed a certain value. To the best of our knowledge, so far there is no scheme that can guarantee deterministic end-to-end delay in OBS.

2.3 QoS Provisioning in EPON and WDM EPON/PON

2.3.1 QoS Provisioning in EPON

The existing dynamic bandwidth allocation (DBA) algorithms for EPON can be classified into two main categories: algorithms with statistical multiplexing [43,45] and algorithms with QoS support. The latter ones can be further subdivided into algorithms with absolute QoS assurances [36,41,42] and relative QoS assurances [44,46,47]. Although the number of DBA schemes in the literature is large, those capable of providing absolute QoS assurance are very limited. However, due to more and more multimedia services with strict time constraints are going to be supported in today’s network, providing such absolute QoS guarantees is critical. In the following, we discuss this kind of DBA algorithms in greater detail.

In [41], the authors proposed a new bandwidth guaranteed polling (BGP) scheme which considers dividing the ONUs in the network into two sets: one set contains the ONUs with bandwidth guaranteed services while the other set contains the ONUs with best effort services. The algorithm provides guaranteed bandwidth to premium subscribers while providing best effort services to other subscribers. In [42], a system in which ONUs and OLT employ *deterministic effective bandwidth* (DEB) admission control in conjunction with Generalized Processor Sharing (GPS) scheduling is developed. The scheme provides deterministic QoS guarantees to each QoS queue, ensuring lossless, bounded-delay service. However, the GPS scheduling in each ONU has increased complexity proportional to the number of flows. Recently, a DBA scheme with admission control capable of providing per-stream QoS in EPON was proposed in [36]. By

computing the *guaranteed rate*, it transforms the bandwidth requirement of a VBR flow into that of a CBR flow. Each ONU decides whether to admit new flows based on the comparison of the aggregate guaranteed rates of the flows to the total allocated bandwidth to this ONU. The scheduling algorithm used in conjunction with the admission control in [36] is a simple variant of the Deficit Weighted Round Robin (DWRR) algorithm [102]. It may not be strong enough to guarantee per-stream QoS since its polling is per-ONU based.

2.3.2 WDM EPON/PON Upgrade Architecture and QoS Provisioning

The research on how to upgrade current PON architecture with WDM technology has been in progress for some time. One straightforward upgrade approach is to assign a dedicated pair of upstream and downstream wavelength channels to each ONU [103]. It delivers high performance (in terms of bandwidth) but results in a poor resource utilization, because if one ONU has no data to transmit, the corresponding wavelength channels and transceivers at the OLT would be idle, and could not be exploited by other ONUs as well. It also brings high upgrade cost due to supporting a great number of wavelengths (equal to double the number of ONUs). The authors of [51] proposed a new hybrid architecture (referred to as SUCCESS), which provides a practical migration from current TDM PONs to future WDM access networks while maintaining backward compatibility with existing TDM PONs. The SUCCESS architecture is based on a collector ring and several distribution stars connecting the Central Office (CO) and the users. However, this architecture does not allow for any resource sharing across different PONs. Later, the authors of [48] proposed an improved scheme of SUCCESS called SUCCESS-DWA that employs dynamic wavelength allocation (DWA) to support bandwidth sharing across multiple physical PONs, and hence achieves both cost-effective and high-performance architecture. Another WDM EPON architecture called STARGATE was proposed in [49]. STARGATE uses the WDM Resilient Packet Ring (WDM RPR) [104] to interconnect multiple central offices among each other as well as to the Internet. Each CO on the ring is connected with a WDM EPON tree network. Different from a conventional collector ring, the central offices on the WDM RPR can communicate

directly with each other through a passive star coupler (PSC) and an arrayed waveguide grating (AWG). STARGATE provides a cost-effective all-optical WDM integration of Ethernet-based access and metro networks, with particular support of P2P/P2MP applications. The work of [50] proposed a tunable WDM PON architecture which supports incremental upgrade. The ONUs are divided into a number of user groups and a fixed wavelength is allocated to each group. The OLT is also equipped with Tunable Laser Diodes which enables it to activate extra wavelengths to be used by these user groups. However, in this architecture each extra wavelength can only be used by one user group at a time and the scheme does not support bandwidth sharing of the extra wavelength across multiple user groups.

With respect to bandwidth and resource management in WDM EPON/PON, the related works appears to be very limited, let alone service guarantees. Most of the above mentioned architectures do not have any WDM-DBA schemes that could support deterministic QoS provisioning, except some strict-priority traffic scheduling algorithms were proposed for SUCCESS-DWA in [105]. However, they could only support two types of traffic: high-priority (HP) traffic and best-effort (BE) traffic, and there is no more differentiation within the HP traffic. In general, some WDM-DBA algorithms are WDM variants [106,107] of the well-known IPACT scheme in EPON, and the others provide bandwidth allocation in WDM EPON/PON [9,108] in a similar multiplexing way. Since their scheduling is in a “best effort” manner based on collected information such as free channels and free transmitters, they could hardly achieve any strict QoS guarantees.

2.4 QoS Provisioning in WLAN

The IEEE 802.11e standard has defined two medium access mechanisms: EDCA and HCCA. EDCA is proposed as the substitute of the widely adopted DCF mode for QoS provisioning. To further improve the performance of EDCA and achieve better QoS, a number of studies have been proposed in the literature [52-66]. Although EDCA is favored for its distributed operation, it is hard to achieve deterministic QoS because no guarantee could be really achieved when

contention exists. In contrast, HCCA could precisely control the access of each QSTA and hence has stronger capabilities in providing strict QoS assurances.

The reference admission control and scheduling scheme of HCCA defined in the IEEE 802.11e standard is not mandatory. It provides a basic scheme and allows new schemes to be developed based on the standard framework for better QoS. The reference scheduler operates as follows. In order to initiate a traffic stream (TS) connection, a QSTA sends a request frame to the QAP containing a traffic specification (TSPEC). The TSPEC includes parameters such as mean/peak data rate, mean/maximum frame size, delay bound, and maximum required service interval (RSI). On receiving all the QoS requests, the QAP scheduler first determines the selected service interval (SI), which should be the highest submultiple value of the beacon interval and also no larger than all the maximum RSIs required by different TSs of different QSTAs. Then an 802.11e beacon interval is divided into an integer number of SIs, and QSTAs are polled sequentially during each SI. Lastly, the QAP scheduler computes the corresponding HCCA TXOP for different QSTAs by using their QoS requests in TSPEC. An allocation example is shown in Fig. 2.2.

Figure 2.2: A TXOP allocation example in 802.11e HCF

For the calculation of the TXOP duration for an admitted stream, the simple scheduler uses the following parameters: mean data rate (ρ) and nominal MSDU size (L) from the negotiated traffic stream, the service interval (SI) calculated above, physical transmission rate (R), size of maximum frame (2304 bytes, M), and overheads in time units (O). The physical transmission rate is the minimum PHY rate negotiated in the traffic stream. The overheads in time include inter frame spaces, ACKs, and CF polls. First, the scheduler calculates the number of frames that arrived at the mean data rate during SI: N_i , and then calculates the TXOP duration:

$$N_i = \left\lceil \frac{SI * \rho_i}{L_i} \right\rceil \quad (2.1)$$

$$TXOP_i = \max(\frac{L_i * N_i}{R_i} + O, \frac{M}{R_i} + O) \quad (2.2)$$

The reference admission control is simply as follows: the QAP scheduler checks if the new request $TXOP_{K+1}$ plus all the allocated TXOPs are lower than or equal to the maximum fraction of time that can be used by HCCA:

$$\frac{TXOP_{K+1}}{SI} + \sum_{i=1}^K \frac{TXOP_i}{SI} \leq \frac{T - T_{CP}}{T} \quad (2.3)$$

where K is the number of existing streams and $K+1$ is the index for the newly arriving stream. T is the beacon interval and T_{CP} is the time used for EDCA traffic.

It is apparent that this reference scheme computes the TXOP based on the mean data rate, which is very inaccurate when dealing with bursty traffic streams. A number of HCCA enhancement schemes have been proposed in [67-75], while not many could really achieve deterministic QoS.

2.5 Network Calculus and Its Applications

We are going to use the technique of network calculus in this thesis. Network calculus is a set of recent developments which provides a deep insight into flow problems encountered in networking. The foundation of network calculus lies in the mathematical theory of dioids, and in particular, the Min-Plus dioid (also called Min-Plus algebra [98]). With network calculus, we are able to understand some fundamental properties of integrated service networks, of window flow control, of scheduling and of buffer or delay dimensioning.

The concepts of arrival curve and service curve were first used by Cruz [109-111]. In these works, general bounds based on the concept of arrival and service curves were derived. They were later extended and simplified independently under equivalent forms by Sariowan [112]. Le Boudec developed these techniques into a complete system in [77]. In this section, we are going to introduce some basic definitions and notations of network calculus, and then summarize some basic results of this theory. More in-depth results are given in the text [77].

Chapter 2. Literature Review

Definition 2.1: The input function $x(t)$ of an arrival process is the number of bits that arrive in the interval $[0, t]$. In particular, $x(0) = 0$ and x is wide-sense increasing, i.e. $x(t_1) \leq x(t_2)$ for all $t_1 \leq t_2$.

Definition 2.2: The output function $y(t)$ of a system G is the number of bits that have left G in the interval $[0, t]$. In particular, $y(0) = 0$ and y is wide-sense increasing.

Definition 2.3 (Backlog and Delay): The backlog at time t is $w(t) = x(t) - y(t)$. The virtual delay at time t is

$$d(t) = \inf\{\tau \geq 0 : x(t) \leq y(t + \tau)\} \quad (2.4)$$

The backlog is the amount of bits that are held inside the system. The virtual delay at time t is the delay that would be experienced by a bit arriving at time t if all bits received before it are already served.

Definition 2.4 (Min-Plus Convolution): Let f and g be wide-sense increasing and $f(0) = g(0) = 0$. Then their convolution under min-plus algebra is defined as

$$(f \otimes g)(t) = \inf_{0 \leq \tau \leq t} \{f(t - \tau) + g(\tau)\} \quad (2.5)$$

We now define the arrival curve.

Definition 2.5 (Arrival Curve): Given a wide-sense increasing function R such that $R(t) = 0$ for all $t < 0$, we say that a flow x is constrained by R if and only if for all $\tau \leq t$:

$$x(t) - x(\tau) \leq R(t - \tau) \quad (2.6)$$

With the definition of Min-plus convolution, this is equivalent to imposing that for all $t \geq 0$

$$x(t) \leq \inf_{0 \leq \tau \leq t} \{R(t - \tau) + x(\tau)\} = (R \otimes x)(t) \quad (2.7)$$

The simplest arrival curve is $R(t) = Bt$. The constraint means that, on any time window of width τ , the number of bits from the flow is limited by $B\tau$. We say in that case the flow is peak rate limited. This occurs if we know that the flow is arriving on a link whose physical bit rate is limited by B bits/sec.

Figure 2.3 Arrival curve for Intserv flows, rate-latency service curve, and vertical and horizontal deviations between both curves

Another typical example of an arrival curve is give by

$$R(t) = \rho t + \sigma \quad (2.8)$$

which results from using the prominent token bucket model [113,114] as traffic regulation mechanism. Having this function as an arrival curve allows a source to send σ bits at once, but no more than ρ bits/s over the long run. Parameters σ and ρ are called the maximum burstiness (in units of data) and the long-term rate (in units of data per time unit). The Integrated services framework of the Internet (Intserv) uses arrival curves, such as

$$R(t) = \min \{M + pt, \rho t + \sigma\} \quad (2.9)$$

where M is interpreted as the maximum packet size, p as peak rate, σ as the maximum burstiness, and ρ as the sustainable rate, as shown in Fig. 2.3. In Intserv jargon, the combination of (p, M, σ, ρ) is also called TSPEC (traffic specification).

One first principle in integrated services networks is to put arrival curve constraints on flows. In order to provide reservations, network nodes in return need to offer some guarantees to flows. This is done by packet schedulers. The details of packet scheduling are abstracted using the concept of service curve. It is defined as follows.

Chapter 2. Literature Review

Definition 2.6 (Service Curve): Consider a system G and a flow through G with input and output function x and y . We say that G offers to the flow a service curve S if and only if for all $t \geq 0$, there exist some $t_0 \geq 0$ and $t_0 \leq t$, such that

$$y(t) - x(t_0) \geq S(t - t_0) \quad (2.10)$$

Again, with the definition of Min-plus convolution, we can recast this definition as

$$y(t) \geq \inf_{0 \leq \tau \leq t} \{S(t - \tau) + x(\tau)\} = (S \otimes x)(t) \quad (2.11)$$

Let us consider a few examples. A simple one is a GPS (Generalized Processor Sharing [83]) node which, by offering a service curve $S(t) = Bt$, guarantees that each flow is served at least at rate B bits/sec during a busy period.

Another example is the rate-latency service curve as shown in Fig. 2.3. The IETF assumes that routers using Resource ReReservation Protocol (RSVP) offer such a service curve in the form of

$$S(t) = B[t - T]^+ = \begin{cases} B(t - T) & \text{if } (t > T) \\ 0 & \text{otherwise} \end{cases} \quad (2.12)$$

where the notation $(a)^+$ denotes a if $a \geq 0$, and 0 otherwise.

After providing the basic definitions and notations in network calculus, we now summarize the main results of network calculus. They are all bounds for lossless systems with service guarantees. The proofs are straightforward applications of the definitions of service and arrival curves.

The first theorem states that the maximum backlog is bounded by the *vertical deviation* between the arrival and service curves:

Theorem 2.1 (Backlog Bound): Assume a flow, constrained by arrival curve R , traverses a system that offers a service curve S . The backlog $x(t) - y(t)$ for all t satisfies:

$$x(t) - y(t) \leq \sup_{\tau \geq 0} \{R(\tau) - S(\tau)\} \quad (2.13)$$

The second theorem gives a bound on maximum delay that is also called *horizontal deviation* between the arrival and service curves:

Chapter 2. Literature Review

Theorem 2.2 (Delay Bound): Assume a flow, constrained by arrival curve R , traverses a system that offers a service curve S . At any time t , the virtual delay $d(t)$ satisfies:

$$d(t) \leq \sup_{\tau \geq 0} \{ \inf_{\varepsilon \geq 0} \{ R(\tau) \leq S(\tau + \varepsilon) \} \} \quad (2.14)$$

Let us see an application of these results next. Consider a flow, defined by TSPEC(M, p, σ, ρ) and served in one node which guarantees a service curve equal to the rate-latency function $S(t) = B[t - T]^+$. This example is the standard model used in the Intserv. Let us apply Theorem 2.1 and 2.2. Assume that $B \geq \rho$ namely the reserved rate is as large as the sustainable rate of the flow. The buffer required for the flow is bounded by

$$w_{\max} = \sigma + \rho \max \left(\frac{\sigma - M}{p - \rho}, T \right) \quad (2.15)$$

The maximum delay for the flow is bounded by

$$d_{\max} = \frac{M + \frac{\sigma - M}{p - \rho} (p - B)^+}{B} + T \quad (2.16)$$

So far, we have introduced the basics of network calculus, including definitions, notations, and main results. We have also presented a simple application of network calculus. These techniques will be used in our later chapters to derive deterministic bounds on delay in various networks, including optical core network, optical access network and wireless network.

Next we introduce the applications of network calculus in the literature. Network calculus can derive deterministic bounds rather than statistical averages provided by traditional queueing theory [115], hence is favored by many researchers. Since its analysis is based on the arrival curve and service curve of a traffic flow independent of its location, it can be applied in various scenarios for guaranteed performance. In early years it was primarily used for packet scheduling [116,117] and flow control [118,119] in packet networks. Later it was used for congestion control [120] and multicasting [121]. Recently people began to employ the technique in wireless networks [122-125], while its applications in optical networks [78] are still very limited.

Chapter 3 Performance Analysis of a Novel Traffic Scheduling Algorithm in TWIN

3.1 Introduction

As we have discussed in Chapter 1, end-to-end QoS guarantee is possible if and only if all autonomous networks through which a connection passes are able to provide QoS guarantee individually. These networks may include optical network at the core, optical network at the edge, and wireless network at the user interface. We start this work from optical core networks, and use Time-Domain Wavelength Interleaved Network (TWIN) as the platform for delivering such QoS. In this chapter, we conduct an in-depth study of TWIN and propose a scheduling algorithm that could achieve better QoS than existing schemes.

Total delay and delay jitter are two of the most important QoS parameters in the transmission of real-time multimedia traffic. In a network, total delay of a packet generally consists of queueing delay and transmission delay at the source, and propagation delay from the source to the destination (including processing delay in the intermediate nodes). Total delay is hard to control in common mesh topology networks because propagation delay is unfixed due to constantly changing of routes. In TWIN, once preprovisioned, the routes remain fixed unless emergencies (e.g. link failure) [4]. TWIN does not have any processing delay in the intermediate nodes because the optical signals are routed on-the-fly based on their wavelengths without any OEO conversion. This implies that propagation delay is constant for every source-destination node pair. Transmission delay is also fixed in TWIN because each slot has the same length. This has made it possible to transform the total delay problem into a much simpler queueing delay problem at the source node, while queueing delay could be minimized by the scheduling algorithm.

Chapter 3. Performance Analysis of a Novel Traffic Scheduling Algorithm in TWIN

Delay jitter (or known as delay variance) is also hard to control in common transport networks as there are many variables (e.g., queueing delay, propagation delay, etc.) in total delay. The situation is much simplified in TWIN as queueing delay is the only variable. One can minimize delay jitter by minimizing variance of queueing delays at source nodes. Hence in TWIN, a scheduling algorithm that could minimize both queueing delay and queueing delay variance is highly desirable.

3.2 Problem Formulation

TWIN has very important characteristics which makes it distinct from other optical networks. In TWIN, each node is assigned a unique wavelength and all the traffic going to it will be transmitted on that wavelength only. As a result, a network with N nodes will have N wavelengths in each fiber link. On a specific channel (wavelength), the routes from various sources to the destination form a multipoint-to-point tree (*destination tree* [11]). The scheduling algorithm is responsible for eliminating any collision in the network. However, tracking the location of every burst in the network is a formidable task and almost impossible in practice. Fortunately we have discovered the fact that since the union of paths traversed by various sources to the destination lies on a tree, any data burst's position at a given time defines a unique arrival time at the destination. Conflicts in the network, therefore, correspond to a common arrival time at a destination. Correspondingly, if a schedule is such that the arrival times have no conflicts at any destination, then there are no conflicts elsewhere in the network. Hence we only need to check the arrival times at the destinations.

Consider a TWIN network with N nodes and let their set be denoted by $\mathcal{N} = \{1, 2, 3, \dots, N\}$. Each node is equipped with a tunable transmitter and a fixed receiver. Time is slotted in the system and at most one data burst can be transmitted or received in a slot by a node. We define a time indicator T which denotes current time slot and increases monotonically with time (i.e., $T = 1, 2, 3, \dots$). The *traffic demand vector* is denoted by a two-dimensional $N \times N$ matrix $D(T) = \{D_{ij}(T)\}$, where $D_{ij}(T)$ represents the number of bursts in the queue waiting to be transmitted from

Chapter 3. Performance Analysis of a Novel Traffic Scheduling Algorithm in TWIN

node i to node j at the beginning of time slot T . The initial traffic demand matrix is $D(0) = \{D_{ij}(0)\}$.

At the beginning of time slot T , the scheduler chooses a *service configuration vector* $S(T) = \{S_{ij}(T)\} \in \hat{S}$, where $S_{ij}(T)$ is 1 if node pair (i, j) is selected to transmit in time slot T , and 0 otherwise. Hence each service configuration vector S describes the set of traffic queues scheduled for transmission, whereas all possible S form the feasible set $\hat{S}$.

To ensure that at time T there are no conflicts at any destination, the transmission time of the bursts must satisfy:

$$\sum_{i=1}^N S_{ij}(T - \delta_{ij}) \leq 1, \forall j \in \mathcal{N} \quad (3.1)$$

which ensures that at time T , only one data burst will reach the receiver at the destination node j . δ_{ij} is the propagation delay from node i to node j and we assume that it is an integral multiple of a time slot.

Since each source has only one tunable transmitter, it can not transmit two data bursts at the same time on two different channels. The source nodes must satisfy that:

$$\sum_{j=1}^N S_{ij}(T) \leq 1, \forall i \in \mathcal{N} \quad (3.2)$$

The traffic demand vector evolves at the beginning of each time slot. Denote the arriving traffic by $A(T) = \{A_{ij}(T)\}$, where $A_{ij}(T)$ is the number of new bursts arriving and being queued during time slot T for node pair (i, j) . Thus the traffic demand evolves as follows:

$$D_{ij}(T) = D_{ij}(T - 1) - S_{ij}(T - 1) + A_{ij}(T - 1) \quad (3.3)$$

We summarize the full problem as follows:

$$\begin{aligned} &\text{Maximize} && \sum_{i=1}^N \sum_{j=1}^N S_{ij}(T) \\ &\text{Subject to} && \sum_{i=1}^N S_{ij}(T - \delta_{ij}) \leq 1, \text{ for all } j \in \mathcal{N} \end{aligned}$$

Chapter 3. Performance Analysis of a Novel Traffic Scheduling Algorithm in TWIN

$$\sum_{j=1}^N S_{ij}(T) \leq 1, \text{ for all } i \in \mathcal{N}$$

$$D_{ij}(T) = D_{ij}(T-1) - S_{ij}(T-1) + A_{ij}(T-1), \forall i, j \in \mathcal{N}$$

$$S_{ij}(T) \leq D_{ij}(T), \forall i, j \in \mathcal{N}, \quad T = 1, 2, 3, \dots \quad (3.4)$$

where the last constraint in the formulation ensures that only when there are bursts waiting in the queue can it be selected for transmission.

This formulation turns out to be an Integer Linear Programming (ILP) problem. When N is large, solving the problem is time-consuming and may not be practical in real time since it has to solve an NP-complete problem in every time slot. Next we propose a heuristic algorithm called Destination Slot Set (DSS) which can solve the problem effectively with low complexity.

3.3 The Scheduling Policy

3.3.1 DSS Scheduling Algorithm

Definition 3.1: In TWIN if node i is farther away to node j than node k , node i is called node k 's *upstream* node to node j .

For example, denote the propagation delay between node i and node j by δ_{ij} and the propagation delay between node k and node j by δ_{kj} , and if $\delta_{ij} > \delta_{kj}$ then node i is node k 's upstream node to node j in the network, no matter whether node k is on the route from node i to node j or not. We give this definition first because it is slightly different from the conventional definition of upstream node in networking. This definition will be used in the algorithm.

The DSS is a slot based scheduling algorithm. The variables in the algorithm are listed in Table 3.1:

Chapter 3. Performance Analysis of a Novel Traffic Scheduling Algorithm in TWIN

Table 3.1: Variables used in the DSS algorithm

Δ	A time slot
T	The indicator denoting current time slot (in units of Δ)
L_{th}	The queue size threshold value
δ_{ij}	The propagation delay from node i to node j in units of Δ
Q_j	Set of time slots which have been occupied at node j (i.e., there will be bursts arriving at node j in these time slots)
D	The traffic demand matrix
E	Copy of D
S	The service configuration matrix

Initialization: $T=0$, and $Q_j = \emptyset$ for every $j \in \mathcal{N}$.

(In every time slot, this algorithm executes from step 1 to step 6)

Step 1) $E \leftarrow D$, $S = 0$;

Step 2) Sort the nonzero elements of E ;

Step 3) Search in all Q_j , if there exists x belonging to Q_j such that $x < T$, remove it;

Step 4) Get the maximal element E_{ij} in E ;

IF ($T + \delta_{ij} \in Q_j$)

IF ($E_{ij} > L_{th}$)

Search for node i 's upstream nodes b such that $\delta_{bi} > \delta_{ij}$;

Set $E_{bi} = 0$ for every such b ;

ELSE

$E_{ij} = 0$;

ENDIF

ELSE

$S_{ij} = 1$;

$Q_j \leftarrow Q_j \cup \{T + \delta_{ij}\}$;

Set $E_{im} = 0$ for every $m \in \mathcal{N}$;

ENDIF

Step 5) Repeat step 4) until all the elements in E are 0;

Step 6) $T = T + 1$;

Figure 3.1: The DSS scheduling algorithm

The algorithm is shown in Fig. 3.1. The main idea of the algorithm is that we iteratively schedule a source-destination pair that has the biggest queue size to transmit in the available time slots until all the node pairs are scheduled, or no more time slot is available. In the algorithm, Q_j

Chapter 3. Performance Analysis of a Novel Traffic Scheduling Algorithm in TWIN

is used to record those future time slots which have already been taken up by previous scheduling at destination node j (these occupied time slots are actually the arrival time of the corresponding bursts at node j). Since there are N nodes, there are N such sets altogether. We delete those elements in Q_j which are smaller than time indicator T as the corresponding bursts have arrived already. This is done by Step 3. To check if a burst can be scheduled, we only need to check whether its arrival time conflicts with any element in Q_j . The burst can be scheduled if there is no same arrival time slot in Q_j , otherwise it is excluded from scheduling (Step 4).

The queue size threshold value L_{th} is used to prevent a phenomenon called *starvation*. This is due to the fact that when a burst from node i comes to a position which has the same distance to the destination j as node k (according to our definition node i is node k 's upstream node), node k has no choice but stop transmitting to node j or conflict at the destination would occur. If an upstream node continues to transmit its bursts, all its downstream nodes would be blocked (i.e., no chance to transmit). To address this, we check the queue size every time we do the scheduling. If any queue's size exceeds the threshold and still has no chance to transmit, all its upstream nodes will be stopped from sending any burst in current time slot. In the simulation, we tune L_{th} for various values and find that when L_{th} is set to be the number of time slots to traverse the average length of the links, the best performance is obtained.

3.3.2 Complexity Analysis

The complexity of our algorithm lies on the size of Q_j because the sizes of other variables are constant. Let δ_{max}^j denote the propagation delay of the *maximum* distance between any node and node j on the destination tree of node j , and δ_{min}^j denote the propagation delay of the *minimum* distance between any node and node j on the destination tree of node j (δ_{max}^j and δ_{min}^j are in units of Δ). We claim that the size of Q_j has an upper-bound δ_{max}^j . We explain it in Fig. 3.2.

Chapter 3. Performance Analysis of a Novel Traffic Scheduling Algorithm in TWIN

Figure 3.2: Illustration of the upper bound of Q_j

In Fig. 3.2, the shadowed boxes denote the occupied time slots and blank boxes denote the time slots that no one takes. Assume that t_0 is the initial time point. At t_0 the lattermost time slot the source nodes in the network can reserve at node j is $t_0 + \delta_{max}^j$, and the earliest time slot the source nodes can reserve at node j is $t_0 + \delta_{min}^j$ (as we have stated before, the reserved slot is actually the arrival time at node j). Since no two same time slots will be included in the set Q_j , at this point its size will be no more than $(\delta_{max}^j - \delta_{min}^j)$ (from A to B).

When time advances to the point $t_1 = t_0 + \delta_{min}^j$, the earliest time slot that could be reserved at node j is $t_1 + \delta_{min}^j$ and the lattermost time slot that could be reserved at node j is $t_1 + \delta_{max}^j$. There are still some reserved time slots earlier than $t_1 + \delta_{min}^j$ left in Q_j ; however, no element that is earlier than t_1 will be retained in the set (this is done by Step 3 in the algorithm). Thus the size of Q_j will be no bigger than δ_{max}^j (from C to E). As time keeps going, this relationship persists. Hence the upper bound of all Q_j ($j = 1, 2, \dots, N$) is $\Phi = \max \{\delta_{max}^j\}$ ($j = 1, 2, \dots, N$).

In the algorithm, the complexity of Step 1 is $O(N^2)$, and that of Step 2 is $O(N^2 \log N)$. The complexity of Step 3 is $O(N\Phi)$. Since Φ is usually a constant in a network of a certain area, we assume it could be replaced by N when N is large. And Step 4 and Step 5 have maximum $O(N^2)$ running time together. Thus the total complexity of the algorithm could be deemed as $O(N^2 \log N)$.

Compared to our algorithm, the dynamic scheduling algorithm in [12] has the complexity of $O(N^{4.5})$. The other scheduling algorithm TIIS in [11] normally has an complexity of $O(N^5)$ given a feasible batch length input. However, TIIS needs a binary search to find the minimum batch length, thus the total complexity can be far beyond that.

Chapter 3. Performance Analysis of a Novel Traffic Scheduling Algorithm in TWIN

3.3.3 Pipeline Scheduling

Figure 3.3: Information collection in central scheduling

Basically, our algorithm is a central scheduling algorithm for TWIN. In a central scheduling scheme, the scheduler needs to communicate with source nodes to collect traffic information and then distribute the schedule results to them after scheduling. There should be some delays in this process. Is the scheme still feasible when this factor is taken into account? We answer the question by Fig. 3.3, which handles the issue by pipeline scheme in the information collection.

We assume that in the network, the scheduler is put at a near central position, but the distance from each node to it may still differ. For the sake of simplicity, we assume that the *maximum* propagation delay from each node to the central scheduler is an integral multiple of Δ , i.e. $m\Delta$, and the processing time is negligible compared to the slot length Δ (for a reasonable sized network). At an initial time T_1 , all nodes begin to send their traffic information to the central scheduler, and after a round-trip delay of $2m\Delta$, all nodes will receive the first scheduling result at time $T_1+2m\Delta$. Hereafter, at the beginning of every Δ each node will send out their traffic information (A_{ij}) and receive the corresponding schedule $2m\Delta$ later. The bursts transmissions actually start at the time $T_1+2m\Delta$.

As a result, the scheduling incurs a $2m\Delta$ constant delay to all transmissions in the network. Constant delay is usually no big issue in QoS provisioning as it will not contribute to delay jitter.

Chapter 3. Performance Analysis of a Novel Traffic Scheduling Algorithm in TWIN

3.4 Performance Analysis

In this section, we will present a model to analyze the TWIN network by queueing theory. We would like to measure how close our model could approximate the real network. Firstly we will calculate the average scheduling time (scheduling time is the sum of queueing time and transmission time of a burst in a specific queue) of each traffic queue in a node. The result will show that, even though these traffic queues share the same server, their average scheduling time are actually different. However, by allocating a specific share of the total service to each queue, we could indeed achieve the objective that these queues can be deemed identical. This will be the basis for our next analysis. In the second step, by modeling the TWIN network into a big switch with different delays between input and output ports, we could derive the delay characteristics of TWIN by utilizing the result in [15]. With these expressions, we can calculate the delay of the bursts in the network and compare it with the simulation.

Figure 3.4: Scheduling framework of TWIN

Fig. 3.4 shows an abstraction of the TWIN network. Basically, the scheduling of TWIN consists of two parts: avoidance of collision at the destinations and avoidance of collision at the sources. But in fact, they could be done together by carefully scheduling the transmissions at the sources. As we have explained, the arrival time is only subject to the transmission time (the time when the burst is transmitted). By carefully selecting the transmission time, both collisions can be avoided. We assume that in each node i there are a number of traffic queues, with each traffic

Chapter 3. Performance Analysis of a Novel Traffic Scheduling Algorithm in TWIN

queue corresponding to a source-destination pair (i, j) ($j = 1, 2, \dots, N, j \neq i$). Since a node cannot send data to itself ($j \neq i$), the total number of queues in a node is $N-1$. Data bursts going to different destinations will be queued up in the respective queues before transmission.

Figure 3.5: Configuration inside a node

We magnify the configuration of a node in Fig. 3.5. Node i can be deemed as a server and there are $N-1$ queues in the system. Assume that the service rate of node i is r_i , thus each of these $N-1$ queues has a share of the total service. We assume that the traffic in the network is *homogeneous* for all source-destination node pairs and follows a Poisson distribution. Besides, the traffic to each destination is independent. Thus we can decompose the node into $N-1$ *independent* $M/G_{i,j}/1$ ($j = 1, 2, \dots, N, j \neq i$) queues with each queue having an independent virtual server, as shown in Fig. 3.6. Each virtual server has its own service rate r_{ij} , which should satisfy

$$\sum_{j=1,2,\dots,N,j \neq i} r_{ij} \leq r_i \quad (3.5)$$

This is a simplified model of TWIN as we treat each queue in a node as independent (i.e., one queue is not other queues in its transmission), where these queues do affect each other because they share a common transmitter. This may bring discrepancies between theoretical data and simulated results. However, without this simplification, we could not find a proper queueing model to accurately describe the behavior of the system, as serving of one queue may affect the status of all other queues.

Chapter 3. Performance Analysis of a Novel Traffic Scheduling Algorithm in TWIN

Figure 3.6: Decomposition of the queues inside a node

To determine the service rate of each virtual server in Fig. 3.6, let us analyze the average scheduling time of each queue. Due to the properties of the scheduling policy in Section 3.3, the average scheduling time of each queue is actually subject to the node's position in the network, which we will explain in greater detail.

Figure 3.7: An example network and its destination tree

To see this, let us study an example network in Fig. 3.7(a). Assume that node 4 is the destination, and its destination tree can be obtained as shown in Fig. 3.7(b). According to Definition 3.1, Fig. 3.7(b) can be transformed into Fig. 3.7(c) without losing any information. For example, in Fig. 3.7(b) node 5 is farther away to node 4 than node 3, hence in Fig. 3.7(c) node 5 is node 3's upstream node. The shape of the tree in Fig. 3.7(b) is meaningless to scheduling. We

Chapter 3. Performance Analysis of a Novel Traffic Scheduling Algorithm in TWIN

call the tree of Fig. 3.7(c) the *transformed destination tree*. On this tree, an upstream node would virtually have higher priority than its downstream nodes. Since the burst sent out could not be dropped or delayed, those downstream nodes will be blocked when a burst from an upstream node is passing them (although in the physical topology, the burst may not really “pass” the downstream node). For example, in Fig. 3.7(b), if a burst is sent from node 2 to node 4 and after some time it reaches somewhere in the network which has the same distance as from node 5 to node 4, node 5 cannot send anything then. On the transformed destination tree, this burst is “passing” node 5. If node 2 keeps sending out data bursts, its downstream nodes (node 5, 1, 3, 6) would be blocked all along. The more upstream nodes a node has, the harder it could transmit its own traffic, and the bigger its average scheduling time would be. One node’s positions on different destination trees are different. Hence even though the traffic queues in node i has the same source, they have different average scheduling times because they correspond to different destinations.

Next we calculate the average scheduling time of each traffic queue. Let Y_{ij} denote the number of nodes that include node i ’s upstream nodes and node i itself on node j ’s transformed destination tree. It is clear that for a specific node on a specific transformed destination tree, all its downstream nodes will have no impact on its average scheduling time, and only its upstream nodes will do. The average scheduling time would be the weighted mean of all possible scheduling times with the weighting factors being the probability of these times. For example, the probability that a burst gets transmitted in the first slot is $1/Y_{ij}$ (the selection will merely be made among the node’s upstream nodes and itself), and the probability that it gets transmitted in the second slot is $(1-1/Y_{ij}) \cdot (1/Y_{ij})$, and so on. Hence, the random variable of the scheduling time t_{ij} is subject to a geometric distribution with the parameter (i.e., probability) $1/Y_{ij}$. Therefore the average scheduling time of queue (i, j) can be obtained by (in units of time slots)

$$\bar{t}_{ij} = \frac{1}{Y_{ij}} + (1 - \frac{1}{Y_{ij}}) \cdot \frac{2}{Y_{ij}} + (1 - \frac{1}{Y_{ij}})^2 \cdot \frac{3}{Y_{ij}} + \dots = \sum_{k=1}^{\infty} \frac{1}{Y_{ij}} \cdot (1 - \frac{1}{Y_{ij}})^{k-1} \cdot k \quad (3.6)$$

This turns out to be the sum formula of geometric distribution with the parameter $1/Y_{ij}$. From the well-known results of the geometric distribution, the result of Eq. (3.6) is

Chapter 3. Performance Analysis of a Novel Traffic Scheduling Algorithm in TWIN

$$\overline{t_{ij}} = \frac{1}{\frac{1}{Y_{ij}}} = Y_{ij} \quad (3.7)$$

Equation (3.7) is the average scheduling time of a burst in queue (i, j) . Since different queues have different average scheduling times, some of them may need a long time to transmit a single burst. Our aim is to give different queues different transmission opportunities such that each of them could transmit the same amount of traffic in a unit time. If a queue has a larger average scheduling time, it is given more opportunity to transmit. In the algorithm, this is accomplished by using the parameter L_{th} : if a node's queue length is bigger than L_{th} , the scheduler should stop all its upstream nodes and let it transmit in current time slot, until its queue length drops below L_{th} . Let g_{ij} denote the probability of selecting queue (i, j) to transmit in node i , it should be determined by

$$g_{ij} = \frac{\overline{t_{ij}}}{\sum_{\substack{k=1 \\ k \neq i}}^N \overline{t_{ik}}} = \frac{Y_{ij}}{\sum_{\substack{k=1 \\ k \neq i}}^N Y_{ik}} \quad (3.8)$$

where the selecting probability of a queue is proportional to its average scheduling time.

If we assume the traffic amount of a burst is 1, then according to Eq. (3.7), the transmission rate of queue (i, j) is actually

$$h_{ij} = \frac{1}{\overline{t_{ij}}} = \frac{1}{Y_{ij}} \quad (3.9)$$

If we define the unit time to be 1, then queue (i, j) will be transmitting for the duration of g_{ij} in a unit time (because it is picked by the probability of g_{ij} and thus has a share of g_{ij} of the total service). Then the traffic amount that queue (i, j) transmit in a unit time is

$$z_i = h_{ij} \cdot g_{ij} = \frac{1}{Y_{ij}} \cdot \frac{Y_{ij}}{\sum_{\substack{k=1 \\ k \neq i}}^N Y_{ik}} = \frac{1}{\sum_{\substack{k=1 \\ k \neq i}}^N Y_{ik}} \quad (3.10)$$

which is equivalent for every traffic queue in node i . Hence it can be deemed that the server serves a number of *identical* and *independent* traffic queues *equally* in a round-robin manner.

Chapter 3. Performance Analysis of a Novel Traffic Scheduling Algorithm in TWIN

This is an important prerequisite for our next analysis. Now we know what the service rate of each virtual server in Fig. 3.6 is:

$$r_{ij} = g_{ij} \cdot r_i \quad (3.11)$$

After this transformation the traffic queues can be considered as $N-1$ independent $M/G/1$ queues.

In the second step, we will analyze the average delay of a burst quantitatively. TWIN can be regarded as a big switch with its ports spreading in the whole network and with different propagation delays between input and output ports. As has explained in Chapter 2, the scheduling schemes in crossbar switches cannot be applied to TWIN directly because there is no propagation delay in a crossbar switch. However, since the propagation delays in TWIN do not change, the statistical conflicting probability in TWIN is the same as in a switch in the long run. Hence the statistical analyzing method for switches can also be applied to TWIN.

Like a $N \times N$ switch, the source nodes in TWIN network are treated as the input ports and the destination nodes are treated as the output ports (see Fig. 3.4). We have assumed that the traffic for all source-destination pairs is homogeneous. In addition, we assume that the traffic is generated following a Poisson process and arrives at a node (not an individual traffic queue) with the rate of λ . Let y denote the average transmission time of a burst, the utilization of each node is given by:

$$\rho = \lambda y \quad (3.12)$$

Since the traffic is homogeneous for all node pairs, the total number of bursts waiting to be transmitted at the input ports is $N\rho$, and each burst is destined to the output ports with equal probability. Hence the average number of bursts that can be successfully transmitted in one slot is given by [15]

$$\Theta = \sum_{k=1}^{N\rho} \frac{C_N^k S(N\rho, k) k! k}{N^{N\rho}} \quad (3.13)$$

in which $S(N\rho, k)$ is the *Stirling number of the second kind* ($N\rho$ and k are nonnegative integers), which by definition, is the number of ways of distributing $N\rho$ distinct objects into k identical boxes such that no box is empty (i.e., at least k ($k \leq N$) bursts can be transmitted to different

Chapter 3. Performance Analysis of a Novel Traffic Scheduling Algorithm in TWIN

output ports). C_N^k ensures that k outputs can be chosen from any of the N output nodes, and $k!$ guarantees that the output (destination) nodes are indeed distinct.

We now consider those bursts originating from those $N\rho$ nonempty nodes. The average transmission time of a burst can be computed by Little's result [115]

$$y = \frac{\rho N}{\Theta} \quad (3.14)$$

Obviously, y is a function of λ , and a different λ will result in a different y . y can be computed using a standard iterative method from Eq. (3.12–3.14). Since $N\rho$ needs to be a nonnegative integer, in the calculation we could vary the value of ρ such that $N\rho$ is 1, 2, ..., N . We can then calculate the value of y .

The average number of bursts that can be transmitted by one node in a slot is

$$\beta = \frac{\Theta}{N} \quad (3.15)$$

Let us consider the situation inside a node. We have established the result in our previous analysis that the traffic queues in a node can be considered as $N-1$ independent and identical $M/G/1$ queues. Since there are $N-1$ traffic queues in a node, the traffic coming to each traffic queue will follow the same Poisson distribution but with the rate of $\lambda_v = \lambda/(N-1)$. Let w denotes the mean service time of a queue and W_2 denotes its second moment. Its utilization is $\rho_v = \lambda_v w$. Since these queues are served equally in a round-robin manner, the service time of a queue can be computed by $H = Fy + y$, where y is the burst transmission time. The random variable F is the number of nonempty queues other than the one selected to transmit. It is easy to see that F is binomial with parameters $(N-2, \rho_v)$. Hence, the expectation of H is

$$w = (N-2)\rho_v y + y \quad (3.16)$$

And its second moment is

$$W_2 = (N-2)Y_2(\rho_v(1-\rho_v) + (N-2)\rho_v^2) + 2(N-2)\rho_v Y_2 + Y_2 \quad (3.17)$$

where Y_2 is the second moment of y . Y_2 can be obtained similarly using the same iterative method for calculating the first moment y .

We can now compute ρ_v from the above equations

Chapter 3. Performance Analysis of a Novel Traffic Scheduling Algorithm in TWIN

$$\rho_v = \lambda_v w = \lambda_v ((N-2)\rho_v y + y) \quad (3.18)$$

Therefore, we have

$$\rho_v = \frac{\lambda_v y}{1 - \lambda_v (N-2)y} \quad (3.19)$$

According to the *Pollaczek-Khinchin mean-value formula* [115], the mean delay a burst undergoes is

$$T_{mean} = w + \frac{\lambda_v \cdot W_2}{2(1 - \rho_v)} \quad (3.20)$$

which is in the units of time slots.

Using the above equations we can calculate the average time a burst stays in a traffic queue. We can compare this result with the data from simulations to see how well our model approximates the TWIN network.

3.5 Simulation and Result Discussion

In this section, we show some numerical results obtained from the analytical model and simulations. In the simulation, the link capacity is set to 10 Gbps, and the distance of every link is randomly selected, varying from 5 km to 50 km. We choose the time of propagating 5 km (the greatest common divisor of each link) to be one time slot. The queue length threshold L_{th} is important in the simulation. Generally, a different L_{th} will give a different simulation result. Since the effect of L_{th} is already considered in the analytical model in Section 3.4 (by using L_{th} , we give different queues different services such that they can be virtually treated as $N-1$ identical and independent $M/G/1$ queues), the analytical values are deemed as ideal values. When comparing them with simulation results, we should choose the best results (minimum delay) from simulations. We find that when L_{th} is set to be the number of slots to traverse the average length of the links, the minimum delay is achieved. Hence in the following simulations, L_{th} is always selected this way. On the other hand, since we randomly choose the distances of links in the simulation, we repeat each simulation five times and compute the average value from them.

Chapter 3. Performance Analysis of a Novel Traffic Scheduling Algorithm in TWIN

Figure 3.8: Delay comparison in a 6-node network

Fig. 3.8 shows the comparison between the theoretical values and simulation results in the 6-node network of Fig. 3.7(a). We assume the traffic is homogeneous (i.e., each node has traffic to all the other nodes, and the traffic follows the same Poisson distribution). The x-axis is the traffic arrival rate to a node (i.e., λ), and y-axis is the average time a burst stays in the queue.

According to Eq. (3.15), the average number of bursts that can be successfully transmitted in one slot is β , which is about 0.66 in the example network of Fig. 3.7(a) (this number can be obtained by Eq. (3.13) and (3.15) after the number of nodes is known). Hence, the saturation input load of the network for each node is about 0.66 bursts per slot. It can be observed that the analytical results match those data obtained from simulations well when the input load is smaller than the saturation input load. However, there are still some discrepancies between the theoretical and simulation results. This can be interpreted from several aspects. First, the mathematical model derived for TWIN is a general model that is not based on any specific scheduling algorithm. This may bring discrepancies between the theoretical and simulated results. As we can see, the simulated results still match the theoretical data most of the time, which proves the correctness of our mathematical model. Second, for simplicity we treat traffic queues in a node as independent, but actually these queues do affect each other as they need to share a common transmitter. This may also bring difference between the theoretical and simulated results. Third, it is due to the approximation in calculating the Stirling number of the second kind $S(N\rho, k)$ in Eq. (3.13), which requires $N\rho$ to be an integer. In reality, $N\rho$ is not necessarily an integer.

Chapter 3. Performance Analysis of a Novel Traffic Scheduling Algorithm in TWIN

Figure 3.9: Delay comparison in the 14-node NSFNET

Fig. 3.9 shows the result in the 14-node NSFNET network. The topology is well known so we omit its figure. Again, the distances are randomly selected from 5 to 50 *km*. The traffic is homogeneous for all node pairs. According to Eq. (3.15), the average number of bursts that can be successfully transmitted in one slot, β , is about 0.64 in this network. Therefore, the saturation input load of this network for each node is 0.64 bursts per slot. The figure again demonstrates that the analytical results approximate the simulation results quite well when the input load is smaller than the saturation input load.

Figure 3.10: Queueing delay comparison in NSFNET under Poisson traffic

Chapter 3. Performance Analysis of a Novel Traffic Scheduling Algorithm in TWIN

Figure 3.11: Queue length comparison in NSFNET under Poisson traffic

Fig. 3.10 and Fig. 3.11 show the comparison by means of simulation between our DSS algorithm and the batch scheduling algorithm from [12]. We still use the NSFNET as the network topology. In the figures, “BS” denotes the batch scheduling algorithm in [12]. This algorithm has been proven to be able to achieve maximum throughput. We can see that under the same Poisson traffic pattern, the average delay of DSS is much smaller than that of BS. The same happens for the average queue length. This implies that a network deploying DSS would have smaller total delay, and in application it requires a smaller buffer, which would reduce the hardware cost.

Figure 3.12: Queueing Delay comparison in NSFNET under bursty traffic

Chapter 3. Performance Analysis of a Novel Traffic Scheduling Algorithm in TWIN

Figure 3.13: Queue length comparison in NSFNET under bursty traffic

In the same experiment network, we change the traffic to on-off bursty traffic and the result is shown in Fig. 3.12 and Fig. 3.13. The on-off bursty traffic is generated as follows: during the “on” time interval the traffic is generated at the sources following a Poisson distribution. During the “off” time interval no traffic is generated. Both the “on” and “off” time intervals are subject to an exponential distribution. The arrival rate in Fig. 3.12 and Fig. 3.13 is average arrival rate of the bursty traffic calculating from the simulation data. Again the delay and queue length of DSS is much smaller compared to BS.

The reason of this phenomenon lies on the nature of batch scheduling [11,12]. In batch scheduling, at each scheduling time the traffic demand is decomposed into a batch of transmissions. To avoid collisions between batches, no traffic can be scheduled before the current batch transmission is over. When the traffic load becomes higher, the size of the batch will increase correspondingly, resulting in more bursts queued at the sources. In contrast, DSS is a slotted scheduling algorithm. The slot length is fairly short compared to the batch length. It reacts more quickly to the traffic change, resulting in a much smaller queueing delay.

Chapter 3. Performance Analysis of a Novel Traffic Scheduling Algorithm in TWIN

Figure 3.14: Variance of queueing delays comparison in NSFNET

In Fig. 3.14, we compare the variance of queueing delays between DSS and batch scheduling algorithm (BS). For each Poisson arrival rate, we calculate the variance of the queueing delays for each source-destination node pair and then take average for the whole network. We can see that BS has a much larger queueing delay variance, which implies that its delay jitter is also very large (this has been discussed in Section 3.1). When the arrival rate increases, the variance of BS becomes larger and larger. This shows that BS has more focused on achieving the maximum throughput but considered little about the queueing delay and its variance. The batch scheduling may not be the best option when total delay and delay jitter are the key concerns.

Figure 3.15: Throughput comparison in NSFNET

Chapter 3. Performance Analysis of a Novel Traffic Scheduling Algorithm in TWIN

Fig. 3.15 shows that our algorithm can achieve the same throughput as the batch scheduling. The throughput is calculated on per-node basis and the traffic follows Poisson distribution. Since each node has only one transmitter, the throughput can not exceed 10 Gb/s which is the capacity of a transmitter. It could be noticed that the throughputs of the two algorithms are almost identical at any arrival rate. Moreover, denote the throughput of BS by a and that of DSS by b , 95% confidence intervals are found where a is within $\pm 2.5\%$ of b 's values.

3.6 Conclusion

In this chapter, we use the Time-Domain Wavelength Interleaved Network (TWIN) as the platform for QoS provisioning in optical transport networks. We first formulate the scheduling problem in TWIN into an integer linear programming (ILP) problem and propose a heuristic algorithm called Destination Slot Set (DSS) to solve it effectively. To study the delay characteristics of TWIN, we derive an analytical model by queueing theory and compare the analytical values with simulation data. The analytical values track the simulation results quite well when the input load is smaller than the saturation input load, which proves the correctness of the model. Further simulations have shown that DSS achieves much smaller queueing delay and queueing delay variance than the existing batch scheduling algorithm, which could guarantee smaller total delay and delay jitter in TWIN. On the other hand, DSS can also maximize the throughput as the batch algorithm does.

Chapter 4 Providing Deterministic Quality of Service in Slotted Optical Networks

4.1 Introduction

In the previous chapter, we have studied the delay characteristics of TWIN and proposed a new scheduling algorithm for it. The algorithm is able to achieve better QoS than existing batch scheduling schemes, but still incapable of providing any *deterministic* QoS assurance. In this chapter, we propose a general framework for slotted optical networks based on the architecture of TWIN that is able to provide such deterministic assurances. This framework uses the combination of a control plane and a data plane to solve the complex problem of capacity allocation, slot-matching and traffic scheduling, as published in [78]. The data plane employs a class based admission control scheme to guarantee the delay bounds, which has been previously proposed in [126] but developed into a more general form here.

Once deterministic QoS has been achieved in TWIN, applications should have guaranteed performance in optical core networks, which is one of the necessary conditions for ensuring deterministic end-to-end QoS.

4.2 Framework Configuration Overview

Fig. 4.1 shows the main structure of our proposed framework. It consists of a control plane and a data plane. Most of the control plane function is implemented in the Bandwidth Broker (BB) shown in Fig. 4.1. The flow of information in Fig. 4.1 shows how aggregated requests propagate to the BB of TWIN, considered as an autonomous region [127]. The BB [127,128] stores state information of local connections that become inputs to the admission control function of the control plane. Each ingress node acts as the external interface of the control plane. On receiving a

Chapter 4. Providing Deterministic Quality of Service in Slotted Optical Networks

flow request, the control plane verifies whether there is available capacity in source-destination node pairs to meet the requests before admitting the flow. Details are discussed in Section 4.3.

Figure 4.1: TWIN network with a bandwidth broker controller

The data plane function is implemented in every edge node. An edge node is equipped with a slot assembly and scheduling mechanism. We will consider how to handle multiple classes in the mechanism, and single class case can be deemed as its subset. In the mechanism, traffic is channeled into queues according to their delay budgets. Just before a slot transmission, the node assembles traffic from these queues into a burst and transmits into its assigned slot. Details of the process are discussed in Section 4.4.

4.3 Control Plane

4.3.1 Overview

Capacity allocation and matching of conflict-free source-destination time slots are two of the main functions in the control plane. Capacity in terms of periodic slots is allocated for traffic between a source-destination node pair according to specific flow requests during admission control. While TWIN architecture aggregates traffic from different source nodes for a destination node using a particular wavelength, it assumes that the source nodes have perfect knowledge of which slot to transmit. Ross et al. [11] proposed a centralized controller for slot scheduling. In contrast, we propose to allocate periodic slots to edge nodes according to traffic requirements. In so doing, we encounter the problem of slot contention due to different propagation delays

Chapter 4. Providing Deterministic Quality of Service in Slotted Optical Networks

between different source-destination pairs for a particular destination node. To overcome timeslot conflict at source node, the control plane performs timeslot synchronization among all source nodes for each of the destination nodes.

Fig. 4.2 shows an illustration of different propagation delays for a particular destination node. For node N , the propagation delays from nodes 1, 2, ..., $N-1$ are $t_1, t_2, \dots, t_{N-1}$ respectively. The problem is more challenging than the zero delay case (Birkhoff-von Neumann switch) or the equal delay case (star topology), which have been discussed in [94] and [129], respectively. We begin our analysis for a TWIN with zero and equal source-destination delay first and extent the results to that of arbitrary delays.

Figure 4.2: Slot transmission in a period

4.3.2 Propagation Delay Compensation

For a TWIN network of N nodes with zero or equal propagation delay among all node pairs, we assume that traffic arrivals at all edge nodes are uniformly distributed. The utilization of the network is given by the following lemma (we define *utilization* as the ratio of used slots to the total number of slots on the average):

Lemma 4.1: In a TWIN network of N nodes with zero or equal propagation delay between each source-destination node pair, the number of slots in a period for full reachability and 100% utilization is $N-1$.

Chapter 4. Providing Deterministic Quality of Service in Slotted Optical Networks*Proof:*

It is obvious that when every node has traffic to other $N-1$ nodes, there must be $N-1$ transmission in a period, with one slot transmission to each destination. In any slot, there should be $N-1$ transmission from N nodes for full reachability. Next, we show that $N-1$ slots are sufficient for 100% utilization. With no propagation delay, a burst sent in the m -th slot in the period will be received in the same slot. Hence we can give such a schedule: in the first slot, let each node select any of the $N-1$ destinations with no two nodes selecting the same destination, e.g. node 1 selects node 2, node 2 selects node 3, ..., and node N selects node 1. In fact, there are totally $(N-1)!$ combinations with no source-destination conflict for all source nodes in the first slot.

In the second slot, let each node selects any of the remaining $N-2$ destinations without overlapping (except the destination already selected in the previous slot), e.g. node 1 selects node 3, node 2 selects node 4, ..., and node N selects node 2. This procedure is repeated in the remaining $N-3$ slots. In the last $(N-1)$ th slot, there will be just one destination left for each node. In particular, each node sends exactly one conflict-free burst to each of the other $N-1$ destination nodes utilizing every slot in every period thus achieving 100% network utilization. This is also true for all source-destination node pairs with equal propagation delay because the delay will result in a constant shift of the arrival time at the destination without causing any collision among transmissions, a time-shifted version of the zero-propagation delay case. ■

Zero propagation delay or equal propagation delay in the network are ideal cases that simplifies the matching of input and output slots without the difficulty of slot collisions. In a practical network with arbitrary topology, the random distribution of these propagation delays is one of the main reasons for performance degradation. An intuitive idea is to equalize all propagation delays to that of the longest link among all node-pairs. But, such a solution will increase propagation delay between node-pairs unnecessarily.

We solve this synchronization problem efficiently using periodic property of slot allocation. To satisfy Lemma 4.1, a small compensation delay can be added to the delay of each node pair so that a source node will transmit into a particular slot of the $N-1$ slots at the destination node. We

Chapter 4. Providing Deterministic Quality of Service in Slotted Optical Networks

have thus compensated for the difference in propagation delays *virtually* using the property of the periodic $N-1$ slots. This novel synchronization scheme is stated in the following Theorem.

Theorem 4.1: In a TWIN network of N nodes, denoting the propagation delay between node i and node j by δ_{ij} (δ_{ij} is expressed in number of slots), if $(\delta_{ij} \bmod (N-1))$ ($\forall i, j$) is an integer constant, the minimum number of slots in a period to achieve 100% utilization is $N-1$.

Proof:

We prove that when the condition “ $(\delta_{ij} \bmod (N-1))$ is an integer constant” is satisfied, the conclusion is the same as zero or equal propagation delay case in Lemma 4.1.

Assume for any node i and j ,

$$\delta_{ij} \bmod (N-1) = p \quad (0 \leq p < N-1, p \text{ is a constant}) \quad (4.1)$$

It is clear that a burst sent in m -th slot of current period will arrive at the destination in slot $q = (m + \delta_{ij}) \bmod (N-1)$ in a later period. Consider together the assumption of (1), $q = (m + \delta_{ij}) \bmod (N-1) = (m + p) \bmod (N-1)$. Hence, q is only subject to m . Applying the property of periodicity, an assignment with different m implies different q , and thus there will be no slot overlapping in any period. Hence, the requirement in Lemma 4.1 is satisfied, fulfilling the condition for 100% utilization. ■

Theorem 4.1 fulfils the synchronization condition required in Lemma 4.1 using the properties of periodic $N-1$ slots. To satisfy condition (4.1), a small compensating delay is required to be added to each node pair (i, j) . Denoting the *largest* compensating delay added to each source-destination link (i, j) by β_{ij} (in number of slots), it is easy to know that β_{ij} will be smaller than the period τ . This can be derived as follows: denote δ_{ij}' the original propagation delay of link (i, j) in the network and δ_{ij} the propagation delay after compensation, then $\delta_{ij} = \delta_{ij}' + \beta_{ij}$. According to (1), $(\delta_{ij}' + \beta_{ij}) \bmod (N-1) = p$ (p is an integer constant and $0 \leq p < N-1$). If $(\delta_{ij}' \bmod (N-1)) \leq p$, $\beta_{ij} = p - (\delta_{ij}' \bmod (N-1)) < N-1$; if $(\delta_{ij}' \bmod (N-1)) > p$, $\beta_{ij} = (N-1) + p - (\delta_{ij}' \bmod (N-1)) < N-1$. Therefore, $\beta_{ij} < N-1$. Since the period τ is quite small compared to the longest propagation delay, this is a reasonable scheme.

Chapter 4. Providing Deterministic Quality of Service in Slotted Optical Networks

Electronic Domain Compensation

To fulfill the requirement of Theorem 4.1, a total of $N-1$ specific optical delays are needed at each source node for the synchronization to $N-1$ destination nodes, which will be impractical to implement. To overcome this problem, we add compensating delays in the electrical domain and show that this is a feasible but suboptimal solution. Adding a compensating delay in the electrical domain implies queueing a burst in the buffer for the delay before transmission. Due to this delay, the ideal condition in Theorem 4.1 will be violated. While slots at destination nodes remain conflict-free, conflicts will arise at source nodes with traffic to different destination nodes. For example, consider a traffic burst between a node pair (i, j) designated for transmission in slot t_{ij} according to Theorem 4.1. The electrical delay compensation will shift the transmission slot from t_{ij} to $((t_{ij} + \beta_{ij}) \bmod (N-1))$. A conflict will arise if $((t_{ij} + \beta_{ij}) \bmod (N-1))$ maps to a slot for more than one node pair for source i .

In other words, we have applied electrical domain compensation at source nodes that can lead to some conflicts. We will propose an algorithm to resolve these conflicts.

4.3.3 Conflict Resolution Algorithm

The theoretical estimation of the network utilization due to such slot conflicts at a source is given by the following lemma:

Lemma 4.2: In an N -node TWIN network, there are $N-1$ slots in each period. Let δ_{ij} denote the propagation delay between node i and node j , if $(\delta_{ij} \bmod (N-1)) (\forall i, j)$ is uniformly distributed in $[1, N-1]$ and electrical domain delay compensation is used, the average utilization of the network is given by [15]

$$\Phi = \sum_{k=1}^{N-1} \frac{C_{N-1}^k \text{Str}(N-1, k) k! k}{(N-1)^N} \quad (4.2)$$

where $\text{Str}(N-1, k)$ is the *Stirling number of the second kind*.

Proof: The $N-1$ slots in a period can be imagined as $N-1$ boxes. If two or more transmission times (after delay compensation) fall into the same box, only one can be successful. Assume $(\delta_{ij}$

Chapter 4. Providing Deterministic Quality of Service in Slotted Optical Networks

$\text{mod } (N-1)$) is uniformly distributed in $[1, N-1]$, then the transmission slots to each destination in the period after electrical compensation are also uniformly distributed in $[1, N-1]$. Hence, the problem becomes calculating the average throughput of each source node under uniform arrival process, equivalent to the problem in [15]. ■

Lemma 4.2 provides a theoretical estimation of the number of successful slots in a period for a network with random uniform source-destination propagation delays. Notably, some available slots remain unused. We propose a simple Conflict-Resolution Algorithm (CRA) for performance enhancement as shown in Fig. 4.3.

```

WHILE not the end of free slots in source node
    Search for the next free slot at source node
    Compute the corresponding destination slot based on the propagation delay
    IF conflict-free source-destination slots are available
        Assign source-destination slots
        Set Successful
    ENDIF
ENDWHILE
IF Successful is set
    Accept connection request
ELSE
    Reject connection request
ENDIF

```

Figure 4.3: The conflict-resolution algorithm in the control plane

Since the BB runs the CRA algorithm, it is necessary to assess whether the computational complexity is reasonable. To allocate the timeslot, the CRA algorithm searches for a free time slot at the source node and a corresponding free time slot at the destination node. In the event that the destination slot is already occupied, the algorithm repeats the search for the next free timeslot at the source and computes the destination slot. As there is a clear one-to-one mapping between the timeslots in the source-destination node pair after electrical delay compensation, the worst case complexity for the search is a reasonable $O(N)$.

In periodic slot assignment, the control plane may assign one or multiple slots in a single-period or over multiple periods to a source-destination node pair. A further question is how to fix the period τ ? The period τ defines the maximum delay for a service-slot or a multiple service-slot. Therefore, the period τ will define the minimum delay bound. Higher delay bounds of integer multiples of τ can also be provided. It appears that τ should be as small as possible subject to the minimum slot size and the number of slots in each period.

Chapter 4. Providing Deterministic Quality of Service in Slotted Optical Networks

In summary, the CRA algorithm searches for conflict-free source-destination slots for capacity allocation between the node pair. If such a slot pair is not available, no further capacity can be allocated to the source-destination node pair. We apply Network Simulator 2 (ns-2) simulations to assess the performance of the combination of our delay compensation scheme and the CRA algorithm in the next section.

4.3.4 Simulation Results

We conduct the first simulation for a TWIN of 100 nodes with 99 slots in each period τ . The network topology is mesh-connected and randomly generated. The modulus of propagation delay between any source-destination node pair is uniformly distributed between 1 and 99 according to the condition in Lemma 4.2. In the experiment, we assume that the aggregate traffic from each source would require one slot per period and two cases are considered. In the first case, conflicts will be deemed as failures with no further action. In the second case, a conflict will trigger the CRA algorithm to search for alternative conflict-free source-destination slots in the period. If such slots are unavailable, it rejects the request.

Table 4.1: Network utilization in TWIN networks with different number of nodes

Network Load	Utilization (100 Nodes)		Utilization (200 Nodes)		Utilization (400 nodes)	
	No CRA	With CRA	No CRA	With CRA	No CRA	With CRA
1.00	0.63	0.92	0.63	0.94	0.63	0.95
0.90	0.60	0.88	0.59	0.88	0.59	0.89
0.80	0.55	0.79	0.55	0.79	0.55	0.79
0.70	0.50	0.70	0.50	0.70	0.50	0.70

Table 4.2: Network utilization in a 200-node TWIN network for 2 slots allocation

Network Load	Utilization	
	No CRA	With CRA
1.00	0.63	0.96
0.90	0.60	0.89
0.80	0.55	0.80
0.70	0.50	0.70

Chapter 4. Providing Deterministic Quality of Service in Slotted Optical Networks

Table 4.3: Network utilization in a 200-nodes TWIN network for random of 1 to 4 slots allocation

Network Load	Utilization (CRA)
1.00	0.96
0.90	0.88
0.80	0.80
0.70	0.70

Table 4.1 shows the results of our first simulation (the column with 100 nodes). In the table, the network load is a normalized network load. It is normalized to $N-1$ where N is the number of nodes in the network. In this case, the conflicts reduce the utilization to 0.63 which is close to the result of Lemma 4.2 and recent results in [130]. After applying the CRA algorithm, the resulting utilizations improve about 40% to 50% for network loads from 0.7 to 1.

In experiment 2, we repeat the simulation for a TWIN network with 200 and 400 nodes. Other conditions are the same as experiment 1. Table 4.1 shows the results (the columns with 200 nodes and 400 nodes). The results of experiment 2 show a similar improvement of utilization for CRA algorithm. The performance is slightly better for a network with more nodes. The reason is due to better choices for conflict resolution with more available slots.

To determine the effect of increasing the number of slots per period τ , we conducted the third experiment: for a network of N ($N = 200$) nodes, we increase the number of slots to $2(N-1)$ in each period. We conduct this experiment to assess the effect of having double the number of slots in a period. In this experiment, each node requests for two slots and CRA can allocate non-consecutive slots in the period. The results are shown in Table 4.2. Small improvements can be perceived when CRA is used.

In our final experiment, we assume that an aggregate flow could request a random number of slots from 1 to 4 in a 200 nodes TWIN. This experiment is to assess the impact of traffic demand on network utilization. Table 4.3 shows the simulation results. There is little difference in the results as compared to experiment 2. Experiments 2 and 3 imply that there is little advantage for increasing the number of slots per period after taking into consideration the additional overheads.

Chapter 4. Providing Deterministic Quality of Service in Slotted Optical Networks

4.3.5 Periodic Timeslot Service

We apply the service curves (i.e. network calculus) theory to quantify the service provided by period timeslots assigned to a source-destination node pair. For basic definitions and applications of service curves, please refer to Section 2.5. The service offered by a stream of periodic slots is defined as follows:

Definition 4.1: The service offered by a stream of periodic slots to a aggregate flow I between a source-destination pair is given by a *slotted service curve*

$$S_I(t) = \sum_{k=1}^{\infty} w_I b u(t - z_I k \tau), \quad \text{where } u(t) = \begin{cases} 0, & t < 0 \\ 1, & t \geq 0 \end{cases} \quad (4.3)$$

where w_I is the number of slots assigned to the aggregate flow I , and b is the number of bits in a timeslot, and $z_I \tau$ is the period in the assignment for aggregate flow I .

Figure 4.4: The arrival curve and service curve of an aggregate flow I

The timeslot duration and period are defined by parameters w_I and z_I . By varying these two parameters, we could construct a *slotted service curve* with specific timeslot duration and period. Fig. 4.4 shows a *slotted service curve* for $w_I = 1$ and $z_I = 1$. For the sake of exposition, the following notations will be used in the rest of this chapter: $R_I(t)$ denotes the arrival curve of the aggregate flow I , $S_I(t)$ denotes its service curve, $A_I(t)$ denotes its actual traffic envelope, and d_i denotes its *queueing delay budget*. To represent an individual flow i between this source-destination pair, we replace the subscript by a lowercase letter i . The *slotted service curve* offers a “staircase service” with step size $w_I b$ after each period $z_I \tau$ with no service between the periods.

Chapter 4. Providing Deterministic Quality of Service in Slotted Optical Networks

The proposed *slotted service curve* has useful properties that provide deterministic QoS guarantee in relation to admission control and slot scheduling in the data plane in Section 4.4.

4.4 Data Plane

The keys functions of the data plane include traffic policing and scheduling mechanism. The main purpose of traffic policing is to separate conformant traffic from non-conformant traffic in a flow. This prevents non-conformant traffic in an aggressive flow from degrading the performance of conformant flows. We assume that a token bucket parameterized by individual flow's traffic specifications is used to identify conformant and non-conformant traffic. The scheduling mechanism will schedule according to QoS requirements and treat non-conformant traffic as best-effort traffic in the network.

In the next section, we describe how class-based traffic is assembled into a slot before transmission.

4.4.1 Class-based Traffic Assembly Process

Fig. 4.5 depicts the class-based buffering and scheduling mechanism. The mechanism consists of three subprocesses: 1) class-based buffering; 2) class-based traffic assembly; and 3) a periodic slot service of $w_i b$ bits per $z_i \tau$ seconds. Fig. 4.5 shows $M+1$ buffers for M real-time traffic classes and one best-effort traffic class. Also shown in the figure are flow specific traffic monitors $R_i^*, \forall i$ that identify and forward conformant and non-conformant traffic to their buffers.

Figure 4.5: The class-based traffic assembly and scheduling process

Chapter 4. Providing Deterministic Quality of Service in Slotted Optical Networks

While traffic arrived are buffered continuously, traffic for a transmission slot is assembled just before the transmission. The highest priority traffic buffered is given precedence over the rest. In other words, the slot is assembled by backlogged traffic starting from the highest priority class, followed by the next and so on. A slot will be transmitted periodically independent of whether the slot is full. Hence, the presence of best-effort traffic will allow more efficient utilization of the allocated capacity. To implement class-based buffering, classification scheme that provides a quantifiable QoS-to-class mapping is required. To address this issue, we apply the traffic classification scheme in [126]. A brief review of the scheme is presented next.

4.4.2 Flow Classification

There are limited number of works on flow classification in the literature [131-133]. It is probably due to the difficulties in mapping specific deterministic QoS guarantees to a particular class. The typical parameters for flow classification include type of application, end-to-end delay, ingress-egress addresses and/or port numbers and precedence. In most cases, the main purpose of flow classification is to provide class-based service differentiation to admitted flows.

In [126], a simple flow classification procedure that provides a quantitative QoS-to-class mapping is proposed. In the proposal, flows are classified into classes according to their delay budget. We introduce the idea of queueing delay bound next. We define a queueing delay vector $D = \{D_M, D_{M-1}, \dots, D_1\}$, where $D_M < D_{M-1} < \dots < D_1$ are the delay bounds of the classes from M to 1 respectively. The basic idea is that flows classified into a particular class will have the same class-based delay bound. The following definition specifies how a flow is classified into a particular class shown in Fig. 4.6.

Definition 4.2: A flow i is classified into class k if the queueing delay budget d_i is

$$D_k \leq d_i < D_{k-1}, \quad \text{for } 1 \leq k \leq M \quad (4.4)$$

By classifying flows into various classes with different delay bounds, we can compute the aggregate arrival curve for various traffic classes and provide the minimum aggregate service

Chapter 4. Providing Deterministic Quality of Service in Slotted Optical Networks

curves to these classes for queueing delay control. We provide a detailed analysis to show how the scheduler provides class-based queueing delay bound next.

Figure 4.6: Flow classifications with M QoS classes and a best effort class (class 0)

4.4.3 Class-based Scheduler

We assume the control plane allocates w slots per period τ to a source-destination node pair. The switch (Fig. 4.5) will close for w slots transmission every period τ . Just before the switch is closed, traffic is assembled into the slot according to their priority class as described in Section 4.4.1.

We apply the idea of capacity allocation in [126] to M traffic classes. We define a capacity allocation vector $\alpha = \{\alpha_M, \alpha_{M-1}, \dots, \alpha_1\}$, where $\alpha_M < \alpha_{M-1} < \dots < \alpha_1$ and $\sum_k \alpha_k = 1$, reserving capacity for various classes of traffic. It does not preclude a network operator from allocating the full capacity to one class of traffic, a special case of class-based scheduling. Using the allocation vector, we propose the following theorem for per-flow service guarantees by means of class-based traffic treatment.

Theorem 4.2: For a *slotted service curve* of wb bits per τ second allocated to an ingress-egress node pair with M classes and a capacity allocation vector $\alpha = \{\alpha_M, \alpha_{M-1}, \dots, \alpha_1\}$, a flow j with arrival curve $R_j(t)$ could be admitted to class h ($1 \leq h \leq M$) by this optical slot scheduling mechanism if

$$R_j(t - D_h) + \sum_{k=1}^M \sum_{i \in F_k} R_i(t - D_k) \leq \sum_{k=1}^{\infty} wbu(t - k\tau) \quad \text{and} \quad (4.5)$$

$$S_j(t) + \sum_{i \in F_h} S_i(t) \leq \alpha_h wb, \quad \text{for } 0 \leq t \leq \tau \text{ and } 1 \leq h \leq M \quad (4.6)$$

Chapter 4. Providing Deterministic Quality of Service in Slotted Optical Networks

where F_k (F_h) is the number of flows already accepted in class k (h).

Proof:

For M classes of traffic, we apply flow classification to all QoS flows according to their delay budgets. For a flow to be admitted to class M , we must ensure that the arrival traffic of class M shifted by D_M is upper bounded by the *slotted service curve* $S(t)$. That is,

$$R_M(t - D_M) \leq S(t) \quad \text{and} \quad (4.7)$$

$$\sum_{i \in F_M} S_i(t) \leq \alpha_M wb, \quad \text{for } 0 \leq t \leq \tau \quad (4.8)$$

If we extend to all existing flows in M classes, the following conditions are required:

$$\sum_{k=1}^M \sum_{i \in F_k} R_i(t - D_k) \leq \sum_{k=1}^{\infty} wbu(t - k\tau) \quad \text{and} \quad (4.9)$$

$$\sum_{i \in F_h} S_i(t) \leq \alpha_k wb, \quad \text{for } 0 \leq t \leq \tau \text{ and } 1 \leq k \leq M \quad (4.10)$$

If a new flow j arrives, it could be admitted into class h if and only if the service curve required is still upper bounded by the *slotted service curve*. Hence the results in the theorem are obvious. This completes the proof. ■

4.4.4 Application Examples

In this section, we consider two different scenarios in which the traffic characteristics of flow j are modeled as either the token bucket TB (σ_j, ρ_j) or $TSPEC$ ($M_j, p_j, \sigma_j, \rho_j$) in IETF specifications (In $TSPEC$, M_j stands for the maximum packet size of flow j and p_j stands for the peak rate of flow j). While the token bucket provides a neat solution, the $TSPEC$ traffic characteristic provides a more realistic model of the real traffic.

First we consider the case when the in-profile traffic of a flow j could be characterized by the token bucket TB (σ_j, ρ_j), where σ_j and ρ_j denote the burstiness and long term rate of flow j respectively. In this context, it is not difficult to show that either of these parameters determines whether this flow could be admitted. For ease of exposition, we neglect all out-of-profile traffic

Chapter 4. Providing Deterministic Quality of Service in Slotted Optical Networks

as they are not protected by the scheduling mechanism. We apply Theorem 4.2 to define a set of simple and concise constraints for admission control in the following proposition.

Proposition 4.1: For a *slotted service curve* of wb bits per τ second allocated to an ingress-egress node pair with M classes and a capacity allocation vector $\alpha = \{\alpha_M, \alpha_{M-1}, \dots, \alpha_1\}$, a flow j with requirement (σ_j, ρ_j, d_j) could be admitted to class h ($1 \leq h \leq M, D_h \leq d_j \leq D_{h-1}$) by this optical slot scheduling mechanism if

$$\frac{\sigma_j + \sum_{k=h}^M \sum_{i \in F_k} \sigma_i^k}{D_h} \leq \frac{wb}{\tau} \left(1 - \sum_{k=h+1}^M \alpha_k\right) \quad \text{and} \quad (4.11)$$

$$\rho_j + \sum_{i \in F_h} \rho_i^h \leq \alpha_h \frac{wb}{\tau} \quad (4.12)$$

where σ_i^k denotes the burstiness of flow i in class k and ρ_i^h denotes the long-term rate of flow i in class h and F_h is the set of existing flows (already admitted) in class h .

Proof:

To admit a flow j into class M , we must ensure that the traffic of the flow j is upper bounded by the envelope of the arrival curve deduced from its service curve. That is,

$$R_M(t) \leq S_M(t + D_M) \quad (4.13)$$

To meet the delay bound, the following conditions must hold [78,126]:

$$\frac{\sigma_j + \sum_{i \in F_M} \sigma_i^M}{D_M} \leq \frac{wb}{\tau} \quad \text{and} \quad (4.14)$$

$$\rho_j + \sum_{i \in F_M} \rho_i^M \leq \alpha_M \frac{wb}{\tau} \quad (4.15)$$

where σ_i^M and ρ_i^M represent the burstiness and long-term rate of flow i in class M and F_M is the set of existing flows in class M . Inequalities (4.14) and (4.15) ensure that the class M traffic arrival curve is upper bounded by the *slotted service curve* shifted left by τ and thus bound the delay and limit the capacity allocated to class M traffic. When $\alpha_M = 1$, all the capacity is allocated to class M and we have the scenario of single-class case.

Chapter 4. Providing Deterministic Quality of Service in Slotted Optical Networks

Similarly, a flow j could be admitted in class $M-1$ (delay bound D_{M-1}) if

$$\frac{\sigma_j + \sum_{k=M-1}^M \sum_{i \in F_k} \sigma_i^k}{D_{M-1}} \leq \frac{wb}{\tau} (1 - \alpha_M) \quad \text{and} \quad (4.16)$$

$$\rho_j + \sum_{i \in F_{M-1}} \rho_i^{M-1} \leq \alpha_{M-1} \frac{wb}{\tau} \quad (4.17)$$

If we extend this to class h , the results follow. This completes the proof. ■

It is worth noticing that Proposition 4.1 separates the requirements into two constraints. The first constraint deals with the burstiness of the flows while the second deals with the long term rate. Whether a flow j could be admitted depends on whether its burstiness or its long-term rate causes the resulting traffic curve to exceed the guaranteed arrival curve, which is deduced from the *slotted service curve*. By means of class-based traffic assembly and scheduling, the queueing delay bound achieved for higher priority class traffic is smaller than that of lower priority class traffic independent of capacity allocation.

Figure 4.7: Comparison of service rates required for two arrival curves $R_1(t):(\sigma_j, \rho_j)$ and $R_2(t):(M_j, p_j, \sigma_j, \rho_j)$

Let us consider a more realistic IETF RSVP type of traffic of a flow j whose characteristics are given by the quadruple $(M_j, p_j, \sigma_j, \rho_j)$ and a queueing delay budget d_j . We sketch a simple graphical solution in Fig. 4.7. In this figure, we plot the arrival curves of the flow based on *TB* (σ_j, ρ_j) and *TSPEC* $(M_j, p_j, \sigma_j, \rho_j)$ as $R_1(t)$ and $R_2(t)$. We sketch the rates r_1 and r_2 which are needed to

Chapter 4. Providing Deterministic Quality of Service in Slotted Optical Networks

provide the deterministic guarantees according to $R_1(t)$ and $R_2(t)$. The graphical solution shows that for a given queueing delay budget d_j , $r_1 \geq r_2$. The tight arrival curve based on IETF specification allows a lower minimum rate for the same queueing delay budget. It allows a more accurate computation of the service curve based on a more realistic arrival curve. This result shows that by means of the realistic IETF specification, it could admit more flows for a given capacity, leading to a more efficient utilization of resources.

4.4.5 Simulation Results

To assess the effectiveness of the admission scheme proposed in Section 4.4.4, we designed an experiment to test its performance. Assume that the capacity allocated to this source-destination node pair is 4Mb/s for three classes of traffic at the source node. For the sake of clarity, we assume that all the traffic flows have the same token bucket parameter (ρ, σ) , in which $\rho = 0.5\text{Mb/s}$ and $\sigma = 40\text{kbit}$. The delay bound of the three classes are 20ms, 40ms, 80ms respectively and the capacity allocation vector $\{a_3, a_2, a_1\} = \{0.3, 0.3, 0.4\}$. We compute the number of flows admitted into each class without violating the service contract.

According to Proposition 4.1, the number of flows n_3 that can be admitted in class 3 (highest priority) must satisfy:

$$\frac{n_3 \times 40 \times 10^3}{20 \times 10^{-3}} \leq 4 \times 10^6 \quad \text{and} \quad n_3 \times 0.5 \times 10^6 \leq 0.3 \times 4 \times 10^6 \Rightarrow n_3 = 2$$

Similarly, the number of flows n_2 that can be admitted in class 2 must satisfy:

$$\frac{n_2 \times 40 \times 10^3}{40 \times 10^{-3}} \leq 4 \times 10^6 \times (1 - 0.3) \quad \text{and} \quad n_2 \times 0.5 \times 10^6 \leq 0.3 \times 4 \times 10^6 \Rightarrow n_2 = 2$$

And finally the number of flows n_1 that can be admitted in class 1 must satisfy:

$$\frac{n_1 \times 40 \times 10^3}{80 \times 10^{-3}} \leq 4 \times 10^6 \times (1 - 0.3 - 0.3) \quad \text{and} \quad n_1 \times 0.5 \times 10^6 \leq 0.4 \times 4 \times 10^6 \Rightarrow n_1 = 3$$

Chapter 4. Providing Deterministic Quality of Service in Slotted Optical Networks

Figure 4.8: Delay distribution with three classes of traffic

If numbers of flows admitted to various classes are limited by these numbers, the queueing delay of flows in the three classes should be bounded by 20ms, 40ms and 80ms respectively. We simulated the scenario in Network Simulator 2 (ns-2) and obtained the delay of admitted flows in each of the three classes. The simulation has been repeated five times and the results are similar. They are plotted in Fig. 4.8. We observe that class 3, the highest class, has the delay bound by 20 ms, and we could see that 90% of the traffic's delay are within 15ms. The class 2 has larger delay bound, 40ms and it is observed that 90% of the traffic delay is bound within 35ms. The case is also similar for class 3 traffic. The result has proved that our scheme provides strict QoS guarantees to flows if the admission control policy is observed.

4.5 Conclusion

Providing deterministic QoS guarantees in optical networks is a challenging issue. Few available results on QoS guarantees in star-coupler networks, assumed equal end-to-end delays, are not applicable to networks with arbitrary topologies. In this chapter, we propose an efficient framework for deterministic service guarantees in slot-based optical transport networks with arbitrary topology. The framework uses the combination of a control plane and a data plane to solve the complex problem of capacity allocation, slot-matching and traffic scheduling. We have applied admission control and capacity allocation to source-destination node pairs in the control

Chapter 4. Providing Deterministic Quality of Service in Slotted Optical Networks

plane successfully. It uses electrical domain delay compensations and a collision resolution algorithm of $O(N)$ for a N edge nodes optical network. We have proposed a comprehensive aggregation, buffering, and scheduling mechanism to provide deterministic QoS guarantees using class-based traffic treatment. Through our analysis, we prove that the framework realizes Service Curves assurance that provides class-based queueing-delay bounds. Simulation results have validated our analysis and shown that the class-based delay bounds are maintained. The simplicity and scalability of this framework make it suitable for implementation in slotted optical transport networks.

Chapter 5 Admission Control and Scheduling in a Scalable WDM EPON Architecture

5.1 Introduction

In previous sections, we have managed to achieve deterministic QoS in optical core networks using TWIN as the platform. Our next step is to achieve deterministic QoS in optical access networks, which is also one of the necessary conditions for achieving deterministic end-to-end QoS guarantees.

Ethernet passive optical network (EPON) is viewed by many as an attractive solution to the first mile problem as it can provide much larger bandwidth than current DSL and CABLE MODEM. With the Wavelength Division Multiplexing (WDM) technology, it can offer even larger bandwidth capable of satisfying the insatiable demand of future bandwidth-intensive applications. However, providing QoS guarantees in WDM EPON is challenging and to the best of our knowledge, very few works have offered feasible solutions.

To solve this problem, we propose a two-level WDM EPON architecture with each level accomplishing one main function. The upper level is the wavelength level responsible for capacity scaling and bandwidth sharing among different stations, and the lower level is the EPON level which supports deterministic QoS provisioning by efficient class based admission control policies. Similar class based admission control schemes have been proposed in general IP networks [126] and slotted optical transport networks [78]. In these works, each class of traffic has a fixed share of the total bandwidth. Though the delay bound of each class of traffic is guaranteed, it is not very flexible since a saturated class can not take advantage of the free capacity of other classes (if there is any) other than its own share. As a step forward, we propose two new admission control schemes, namely non-preemptive and preemptive admission control, and discuss their pros and cons over the fixed-proportional admission control policy.

Chapter 5. Admission Control and Scheduling in a Scalable WDM EPON Architecture

5.2 The Proposed Scheme

5.2.1 Overview of the WDM EPON Architecture

Figure 5.1: An illustration of the proposed WDM EPON architecture

The proposed WDM EPON architecture is shown in Fig. 5.1. In the architecture, the Optical Network Units (ONUs) are divided into a number of groups, and each group is assigned a pair of upstream and downstream wavelengths initially. For example in Fig. 1, Group 1 is assigned an upstream wavelength λ_1 and a downstream wavelength λ_1' respectively, and so is Group 2, Group 3. The optical couplers in the figure are used as optical combiners and splitters in the upstream and downstream direction respectively. We mark those upstream wavelengths in Fig. 5.1 to facilitate our explanations about the wavelength scaling process in the following sections.

The situation is simple in the downstream direction. Within each group, it is a broadcast network without any collision like an ordinary EPON. Qos provisioning in such a scenario could be handled by many well-known packet schedulers such as the Weighted Fair Queueing (WFQ) algorithm. The downstream is not our focus as it poses no big challenges. We assume that the downstream bandwidth is enough for each group.

In the upstream direction, however, coordination is required for ONUs in each group to ensure certain QoS, as they are sharing the same upstream wavelength for transmission. We rely

Chapter 5. Admission Control and Scheduling in a Scalable WDM EPON Architecture

on the admission control and scheduling algorithm to achieve the goal, which will be introduced in Section 5.2.3 and 5.2.4. When the bandwidths of some groups are used up, they could ask for more and the Optical Line Terminal (OLT) should allocate extra upstream wavelengths. Instead of allocating one extra upstream wavelength for each group, we consider sharing the added wavelength across multiple groups to improve the resource utilization. To do this, each ONU that intends to take advantage of the new wavelength should be connected to a common optical coupler and then connected to a new AWG port in Remote Node (RN). For example in Fig. 5.1, λ_4 is the new upstream wavelength that could be shared by all ONUs. The excessive traffic is carried on λ_4 and combined by the coupler before entering the AWG at the RN. In the RN, different wavelengths are multiplexed onto a single optical fiber by the AWG and sent to the OLT. After reaching the OLT, these wavelengths are demultiplexed by the AWG and the signals of λ_4 are fed to the corresponding receiver. The detailed scaling process is presented in the next section. By repeating the same process more extra wavelengths could be supported in the architecture.

5.2.2 Wavelength scaling process

Figure 5.2: The creation of a Dynamic Group

Chapter 5. Admission Control and Scheduling in a Scalable WDM EPON Architecture

Each group operates like an EPON. The OLT runs multiple instances of the admission control and scheduling algorithm, one for each group, to ensure its QoS. In case that a specific group has used up all its upstream bandwidth and new incoming flows are rejected, the OLT will assign a new upstream wavelength to it, and this new wavelength could also be shared by other overflowing ONUs in different groups for better resource utilization. For example in Fig. 5.1, initially there are three groups: Group 1, 2, and 3. Suppose at sometime Group 1 and Group 3 are saturated and there are still new flows coming to ONU1 and ONU6. The detail is shown in Fig. 5.2. Suppose Flow 1 ~ 6 and Flow 7 ~ 12 are the already admitted flows in Group 1 and Group 3 respectively, and Flow 13 and Flow 14 are the new flows coming to ONU1 and ONU6 respectively. Since Group 1 and Group 3 are already saturated, the OLT will allocate a dynamic wavelength, namely λ_4 , to carry the excess traffic. Hence the existing traffic of Group 1 and Group 3 is carried by λ_1 and λ_3 , and the new traffic of Flow 13 and Flow 14 is carried by λ_4 . A virtual group is thus formed, namely Group 4. We call it a *Dynamic Group* to distinguish it from the *Fixed Groups* that are created by initially dividing the ONUs (e.g., Group 1, 2, 3). A Dynamic Group can be dynamically created and deleted. New flows from other saturated Fixed Groups could also join in the Dynamic Group if needed. The detailed flowchart of wavelength scaling process is drawn in Fig. 5.3.

Let us look at the left half of the flowchart first. When there is a new flow coming to ONU i of Group j (which could only be a Fixed Group), the OLT checks whether this flow could be admitted in Group j (the admission control policy for a group will be introduced in the next section). If yes the flow will be registered in Group j (including the recording of the flow's parameters and allocation of resources (i.e. transmission time slots)). If Group j is saturated, the OLT will try to admit it in the Dynamic Group, or create one if there is none. However, if the Dynamic Group is also saturated, the flow will be dropped.

Chapter 5. Admission Control and Scheduling in a Scalable WDM EPON Architecture

Figure 5.3: Flowchart of the wavelength scaling process

The right half of the chart shows the process when there is a flow terminating in ONU i of Group j . There could be two cases: if Group j is a Dynamic Group, OLT just deregisters the flow and frees the resources; if Group j is a Fixed Group, it implies that there is surplus capacity in it and OLT should check whether any flows in the Dynamic Group *originating from ONUs of Group j* could be shifted back. This is to ensure that the capacity of the fixed wavelengths is utilized furthest. Finally if the flows in the Dynamic Group are all shifted back into the Fixed Groups, the Dynamic Group should be deleted to release the corresponding wavelength. For the sake of simplicity, we only show the creation and deletion of a single Dynamic Group in Fig. 5.3. If multiple Dynamic Groups are employed, a few loops are needed to check the contents in all Dynamic Groups. This can be easily done on the basis of Fig. 5.3.

To be able to join a Dynamic Group, an ONU needs to be upgraded by installing a tunable transmitter that enables it to transmit on the new wavelength. More tunable transmitters could be installed depending on the upgrading requirement. Note that it is not necessary to upgrade a

Chapter 5. Admission Control and Scheduling in a Scalable WDM EPON Architecture

whole group of ONUs if only a few of them are requesting extra bandwidth. In the example of Fig. 5.2, only ONU1 and ONU6 need to be upgraded. The OLT will check whether a given ONU has a free tunable transmitter before creating or adding it into a Dynamic Group. Since the dynamic wavelengths can be reused after they are released, it is not necessary to install a great number of tunable transmitters when performing the upgrade.

Compared to the work of [50], our scheme is more efficient as the dynamic wavelengths can be shared by ONUs in multiple groups, and hence less number of wavelengths is needed for network upgrade. We will see this difference more clearly in the simulation.

5.2.3 Admission Control Schemes

In this section we present the admission control schemes to be used in each group for deterministic QoS provisioning. The admission control schemes are derived by Network Calculus [77]. The basic definitions and general results of network calculus have been introduced in Section 2.5.

We start from a simple case: consider a group that contains only one ONU. Assume that the ONU contains M classes of traffic and each class i has a delay bound requirement D_i . Each class of traffic is backlogged in its queue waiting to be transmitted. It can be seen that in this particular group, admission control and scheduling process is exactly like that in a switch/router, if we could neglect the overhead (polling and guard time between consecutive transmissions). This overhead usually does not exist in a switch/router since the accessing time of buffers is extremely small. However, it must be taken into account in a network scenario. We will discuss how to minimize this overhead in Section 5.2.4. Presently we only assume that the physical line rate of a wavelength channel is C and the available bandwidth excluding the overhead in a group is C_w , where w denotes the upstream wavelength allocated to this group. Their quantitative relationship will be analyzed and revealed when we discuss the scheduling discipline.

If there are more than one ONU in the group (multi-ONU case), the above conclusion is still valid as long as some assumptions are made. If the class definition (delay bound requirement for

Chapter 5. Admission Control and Scheduling in a Scalable WDM EPON Architecture

each class) is identical for each ONU, the same class of traffic from all ONUs could be regarded as a whole, as if it came from a single ONU. Hence a group can be considered as a big switch/router if certain conditions are met. A theorem will be proposed next for our first admission control scheme.

Before we proceed to the theorem, some preliminaries about traffic regulations should be provided first. Traffic regulation is a necessary condition for achieving strict QoS guarantees in packet networks. Without traffic regulation the network resources may be completely exhausted by a few ill-behaving traffic flows. We adopt the well-known token bucket traffic regulator in our scheme. The token bucket regulator has two parameters: the token arrival rate ρ (or the long-term rate of the constrained traffic) and the bucket depth σ (or maximum burst size of the constrained traffic). In our scheme, each individual flow is constrained by its own token bucket traffic regulator. The following theorem provides a valid token bucket allocation of each class from [78,126]:

Theorem 5.1: Let M be the total number of classes such that class 1 is the highest class and class M is the lowest class. Each class i has a bandwidth share α_i that satisfies $\sum_{i=1}^M \alpha_i \leq 1$ and a delay bound $D_i (i = 1, \dots, M, D_i < D_j \text{ if } i < j)$ that should be strictly observed. To achieve class i 's delay bound and to ensure that each class of traffic obtains its bandwidth share α_i , the classes' token buckets should be chosen as

$$\rho_i = \alpha_i C_w \quad (5.1)$$

$$\sigma_i = D_i (C_w - \sum_{j=1}^{i-1} \rho_j) - \sum_{j=1}^{i-1} \sigma_j \quad (5.2)$$

for $i = 1, \dots, M$, where (ρ_i, σ_i) are the allocated token bucket for class i , and C_w is the available bandwidth in group w .

Theorem 5.1 allocates the *aggregate* token bucket for each class of traffic based on its bandwidth share. After acquiring this information, the admission of new flows is simple and easy. The following admission control scheme is a direct result of Theorem 5.1:

Chapter 5. Admission Control and Scheduling in a Scalable WDM EPON Architecture

Admission Control Scheme A (Fixed Proportional Admission Control)

1. A new incoming flow $(\rho^{new}, \sigma^{new})$ with delay requirement d^{new} is firstly examined by finding the most appropriate class i :

$$D_i \leq d^{new} < D_{i+1} \quad (5.3)$$

(Note: We assume $D_{M+1} = \infty$ since there are only M classes)

2. It could be admitted into class i if and only if

$$\rho^{new} + \sum_{k=1}^{N_w} \rho_i^k \leq \rho_i \quad (5.4)$$

$$\sigma^{new} + \sum_{k=1}^{N_w} \sigma_i^k \leq \sigma_i \quad (5.5)$$

where (ρ_i^k, σ_i^k) are the aggregate token bucket of already admitted flows in class i of ONU k , and N_w is the total number of ONUs in group w .

Eq. (5.3) finds the most appropriate class for the new flow. After the classification, the admission process is simple and swift. The OLT only needs to check whether the sum of the token buckets of all traffic flows belonging to class i (including the new flow) has exceeded the token bucket allocated to this class (the information collection for admission control will be explained later in this section). However, one possible drawback of this scheme is that the bandwidth share vector α is fixed and can not be changed dynamically during operation. This may cause bandwidth underutilization as one class can not take advantage of the free bandwidth of other classes when its own bandwidth is used up. Of course, this scheme does have advantages: the fixed bandwidth allocation protects the interest of each class as the free bandwidth of each class is reserved for future traffic flows. Next we would like to propose a new scheme that does not employ fixed bandwidth shares as an improvement. The following theorem provides the theoretical basis for the new admission control scheme:

Theorem 5.2: Let M be the total number of classes such that class 1 is the highest class and class M is the lowest class. Each class i has a delay bound requirement $D_i (i = 1, \dots, M, D_i < D_j \text{ if } i$

Chapter 5. Admission Control and Scheduling in a Scalable WDM EPON Architecture

$< j$) that should be guaranteed. A new incoming flow $(\rho^{new}, \sigma^{new})$ with delay requirement d^{new} could be admitted to class i if all the inequalities are true:

$$D_i \leq d^{new} < D_{i+1} \quad (5.6)$$

$$\rho^{new} + \sum_{h=1}^M \sum_{k=1}^{N_w} \rho_h^k \leq C_w \quad (5.7)$$

$$\sigma^{new} + \sum_{h=1}^M \sum_{k=1}^{N_w} \sigma_h^k \leq D_M (C_w - \sum_{h=1}^{M-1} \sum_{k=1}^{N_w} \rho_h^k) \quad (5.8)$$

where (ρ_h^k, σ_h^k) are the *aggregate* token bucket of *already admitted* flows in class h of ONU k , and N_w is the total number of ONUs in group w .

Proof:

As has been mentioned, Eq. (5.6) finds the most appropriate class for the new flow. It can be seen that putting the new flow into any class higher than i would result in wastage of resources (over-provisioning).

Eq. (5.7) is a necessary condition as it states the constraint that the sum of average rate of the new flow and all the accepted flows in the group should not exceed the total available capacity.

Eq. (5.8) is another necessary condition which states the constraint about the bucket depths. Since fixed bandwidth shares are not used, the admission of a new flow could affect all the existing flows. According to the general conclusion of Eq. (5.2) from Theorem 5.1, the bucket depth of class i depends on the average rates and bucket depths of all the classes higher than i . To ensure the admission conditions are not violated we only need to verify the lowest class M , since this checking includes all the flows from all the classes:

$$\begin{aligned} \sigma^{new} + \sum_{k=1}^{N_w} \sigma_M^k &\leq D_M (C_w - \sum_{h=1}^{M-1} \sum_{k=1}^{N_w} \rho_h^k) - \sum_{h=1}^{M-1} \sum_{k=1}^{N_w} \sigma_h^k \\ \Leftrightarrow \sigma^{new} + \sum_{h=1}^M \sum_{k=1}^{N_w} \sigma_h^k &\leq D_M (C_w - \sum_{h=1}^{M-1} \sum_{k=1}^{N_w} \rho_h^k) \end{aligned} \quad (5.9)$$

This has sufficiently shown that the system capacity could support all the traffic flows. ■

Chapter 5. Admission Control and Scheduling in a Scalable WDM EPON Architecture

Theorem 5.2 has defined a new admission policy in which each traffic flow could take advantage of the whole available bandwidth rather than the fixed share of the class it belongs to. This would lead to a higher overall bandwidth utilization, especially when the traffic from each class is not uniform. The following admission control scheme is a direct result of Theorem 5.2:

Admission Control Scheme B (non-preemptive admission control)

1. A new incoming flow $(\rho^{new}, \sigma^{new})$ with delay requirement d^{new} is firstly examined by finding the most appropriate class i according to Eq. (5.6).
2. It could be admitted into class i if and only if Eq. (5.7) and Eq. (5.8) are all true.

This is a non-preemptive admission control scheme as a new flow could only be admitted when Eq. (5.7) and Eq. (5.8) are all true, otherwise it is rejected. Non-preemptive admission control will not accept a new flow if it detects that the available capacity is insufficient, even if this flow is classified into the highest class. Alternatively, we could have a preemptive admission control scheme in which high class flows can occupy the resources of lower classes flows. The new flow could be admitted into high class i (which it has been classified into) without considering preserving the flows in classes lower than i . The following theorem defines the conditions for such admissions:

Theorem 5.3: Let M be the total number of classes such that class 1 is the highest class and class M is the lowest class. Each class i has a delay bound requirement D_i ($i = 1, \dots, M, D_i < D_j$ if $i < j$) that should be guaranteed. A new incoming flow $(\rho^{new}, \sigma^{new})$ with delay requirement d^{new} could be admitted to class i if

$$D_i \leq d^{new} < D_{i+1} \quad (5.10)$$

$$\rho^{new} + \sum_{h=1}^i \sum_{k=1}^{N_w} \rho_h^k \leq C_w \quad (5.11)$$

$$\sigma^{new} + \sum_{h=1}^i \sum_{k=1}^{N_w} \sigma_h^k \leq D_i (C_w - \sum_{h=1}^{i-1} \sum_{k=1}^{N_w} \rho_h^k) \quad (5.12)$$

Further, to maintain the delay bound of each class the following conditions must be satisfied:

Chapter 5. Admission Control and Scheduling in a Scalable WDM EPON Architecture

$$\rho^{new} + \sum_{h=1}^q \sum_{k=1}^{N_w} \rho_h^k \leq C_w, \quad \text{for } q = i+1, i+2, \dots, M \quad (5.13)$$

$$\sigma^{new} + \sum_{h=1}^q \sum_{k=1}^{N_w} \sigma_h^k \leq D_q (C_w - \sum_{h=1}^{q-1} \sum_{k=1}^{N_w} \rho_h^k), \quad \text{for } q = i+1, i+2, \dots, M \quad (5.14)$$

where (ρ_h^k, σ_h^k) are the *aggregate* token bucket of *already admitted* flows in class h of ONU k , and N_w is the total number of ONUs in group w .

Proof:

As usual, Eq. (5.10) finds the most appropriate class for the new flow. Eq. (5.11) and (5.12) are evident as they are the necessary conditions to admit the new flow into class i . If they are true, the new flow will be admitted. Note that this admission may cause the capacity of lower classes become insufficient (classes higher than i should not be affected). To maintain the system functionality, all the classes lower than i should be examined. As a result, for class $i+1$ the following conditions must not be violated taking the new flow into account:

$$\rho^{new} + \sum_{h=1}^{i+1} \sum_{k=1}^{N_w} \rho_h^k \leq C_w \quad (5.15)$$

$$\sigma^{new} + \sum_{h=1}^{i+1} \sum_{k=1}^{N_w} \sigma_h^k \leq D_{i+1} (C_w - \sum_{h=1}^i \sum_{k=1}^{N_w} \rho_h^k) \quad (5.16)$$

For class $i+2$ the conditions are

$$\rho^{new} + \sum_{h=1}^{i+2} \sum_{k=1}^{N_w} \rho_h^k \leq C_w \quad (5.17)$$

$$\sigma^{new} + \sum_{h=1}^{i+2} \sum_{k=1}^{N_w} \sigma_h^k \leq D_{i+2} (C_w - \sum_{h=1}^{i+1} \sum_{k=1}^{N_w} \rho_h^k) \quad (5.18)$$

and so on. Lastly for class M the conditions are

$$\rho^{new} + \sum_{h=1}^M \sum_{k=1}^{N_w} \rho_h^k \leq C_w \quad (5.19)$$

$$\sigma^{new} + \sum_{h=1}^M \sum_{k=1}^{N_w} \sigma_h^k \leq D_M (C_w - \sum_{h=1}^{M-1} \sum_{k=1}^{N_w} \rho_h^k) \quad (5.20)$$

This has proved Eq. (5.13) and (5.14) and completed the proof. ■

Chapter 5. Admission Control and Scheduling in a Scalable WDM EPON Architecture

Theorem 5.3 has provided the conditions for the preemptive admission control. Since the admission of new flows does not care whether the capacity is still enough for lower classes, all the flows in lower classes must be readjusted if Eq. (5.13) and (5.14) do not hold true. In this process, a number of flows in lower classes may be suspended. The detail of the new admission control scheme is given below:

Admission Control Scheme C (preemptive admission control)

1. A new incoming flow $(\rho^{new}, \sigma^{new})$ with delay requirement d^{new} is firstly examined by finding the most appropriate class i according to Eq. (5.10).
2. Check whether Eq. (5.11) and (5.12) are valid; if yes then admit the flow and jump to step 3, otherwise reject the flow and exit.
3. $k \leftarrow i$ and call *Suspend_Flows*(k).

```

Suspend_Flows( $k$ )
{
     $j = M$ ;
    while ( $j > k$ ) {
        while (Eq. (5.13) and (5.14) are false when  $q = j$ ) and (class  $j$  is not empty) {
            randomly select a flow in class  $j$ ;
            suspend this flow and put it into suspending queue  $Q_j$ ;
        }
         $--j$ ;
    }
}

```

Figure 5.4: Pseudocode of process *Suspend_Flows*

The process *Suspend_Flows* is used to suspend flows in lower classes whose capacity becomes insufficient after the admission of the new flow in class i . It starts from the lowest class M and terminates after checking class $i+1$. Flows selected are suspended and put temporarily in the suspending queue Q_j if they belong to class j . The OLT keeps $M-1$ such suspending queues, one for each class (the highest class does not need such a queue since the flows in the highest class will never be suspended). This storing, however, does not mean queueing of the packets; rather, the flow's identifier and the time it has waited in the queue are recorded. This information is used for restoration of these flows when the bandwidth becomes enough for accommodating them. For example, the ending of an existing flow will vacate certain bandwidth which may accommodate some suspended flows. However, there may be new flows requesting for admission

Chapter 5. Admission Control and Scheduling in a Scalable WDM EPON Architecture

at the same time. If they belong to the same class, it is obvious that the suspended flows should be considered first as they have already “started”. This can be done by assigning them higher priority. More specifically, the priority in the admission is shown in Fig. 5.5.

Figure 5.5: Admission priority in scheme C

Since class 1 is the highest class, its flows will not be suspended by any other classes once they are admitted and hence there are no “class 1 suspended flows” in the figure. In the same class, the order implies that suspended flows can not push out existing flows, and suspended flows should join admission before new flows once there is available capacity. However, new high class flows will join the admission before low class flows, even suspended low class flows exist. New high class flows can also suspend existing low class flows.

```

Restore_Flows( $k$ )
{
  for  $j = k$  to  $M$  {
    while (not all flows have been checked in suspending queue  $Q_j$ ) {
      select one flow in  $Q_j$ ;
       $d$  = the time the flow has waited in the queue;
      while ( $d > D_j$ ) and (not all flows in  $Q_j$  have been checked) {
        drop this flow and delete it from  $Q_j$ ;
        select the next flow in  $Q_j$ ;
         $d$  = the time the flow has waited in the queue;
      }
      if ( $Q_j$  is not empty) {
        retrieve the token bucket of currently selected flow;
        treat the flow as a new flow and initiate the admission process;
        if successful {
          re-admit the flow and delete it from  $Q_j$ ;
          Suspend_Flows( $j+1$ );
        }
      }
    }
  }
}

```

Figure 5.6: Pseudocode of process Restore_Flows

Chapter 5. Admission Control and Scheduling in a Scalable WDM EPON Architecture

When an existing flow ends (assuming it belongs to class i) and there are no new flows of higher classes coming at the same time (otherwise the admission process of these higher class flows is initiated), we let $k = i$ and call *Restore_Flows*(k). The pseudocode is shown in Fig. 5.6. It starts to check from class i to the lowest class to see whether any suspended flows could be restored. Only those flows that have not violated their service agreement (i.e., the time waited in the queue is smaller than the delay bound of that class) are considered. These eligible flows will join the admission again before they can be restored from suspended status. If any flow is readmitted, the process *Suspend_Flows* would be called to ensure that lower classes do not violate their admission conditions, if there are any.

The preemptive admission control scheme provides better service to higher classes than to lower classes. This is because a new flow, which is classified into a high class, could make room for itself by suspending flows in lower classes. This scheme provides strict priority based service, but it also involves much more complicated operations. We will compare its performance with the non-preemptive admission control scheme in Section 5.3.

Next we introduce the information collection for the admission control. The admission control is implemented at the OLT. As defined in the Multipoint Control Protocol (MPCP) of the EPON standard [8], the ONUs use REPORT message to report their bandwidth requirements (typically in the form of queue occupancies) to the OLT. Our scheme inherits the same mechanism but the REPORT message contains different information. An ONU reports to the OLT only when a new flow (ρ^{new}, σ^{new}) with delay requirement d^{new} requests to be admitted. After receiving the request, the OLT will decide whether to admit the flow based on the admission control scheme it uses.

The advantage of this approach is that it does not rely on the REPORT messages to determine the grants (transmission windows). Traditional DBA schemes which determine the grants based on REPORT messages have a delay of one cycle because the information retrieved in current cycle can only be used in the next cycle (MPCP defined that REPORT message should be attached at the end of each transmission). There are some improved schemes in the literature trying to avoid this impact by using estimates of the traffic, but obviously they could not be

Chapter 5. Admission Control and Scheduling in a Scalable WDM EPON Architecture

always accurate. In our scheme, once having received the REPORT message the OLT will determine the grants without requesting further information (such as queue lengths) from the ONUs, which is much more convenient and accurate.

If a new flow is admitted, its token bucket will be recorded by the OLT to calculate the appropriate grants. The corresponding ONU will be informed and assigned increased grants by the GATE message in the next cycle. The implementation of the preemptive admission control scheme is more complicated than the other two. The OLT must inform the ONUs about those flows which have been suspended after the admission of high class flows, and let the corresponding ONUs store the packets from these flows in a separate set of priority queues (backup queues). When any of these suspended flows is resumed, the OLT sends the message to the corresponding ONU and the ONU transfers the packets of the suspended flow back to the normal priority queue. If any of these suspended flows is chosen to be dropped, the packets stored in the backup queues of the ONU are deleted.

5.2.4 Scheduling Discipline

Figure 5.7: Illustrative example of the transmission sequence

The admission control in the previous section has provided the guarantee that all admitted flows could be served within their delay bounds. Now we need to find a means to schedule the transmission. The well-known round-robin algorithm is adopted here. Firstly, the scheduling should strictly follow the order of the classes, serving from the highest class to the lowest class. Secondly, inside each class, the scheduling serves the ONUs in a fixed round robin manner. The

Chapter 5. Admission Control and Scheduling in a Scalable WDM EPON Architecture

transmission time assigned to each ONU should be constrained by the token buckets negotiated in the admission control. This mechanism could ensure that all admitted flows maintain their delay bounds as desired.

Fig. 5.7 shows the transmission order in Group 2 of Fig. 5.1. The group has two ONUs: ONU4 and ONU5. We assume that both of them have three classes of traffic. An arrow line from OLT to ONU represents a polling (GRANT message), and an arrow line from ONU to OLT represents the start of a data transmission. The dark parallelograms represent the data transmissions. To minimize the overhead, the Round Trip Time (RTT) should be taken into account to determine the time of next polling. Suppose that the finishing time of current transmission is t_1 , the time to send the next polling should be $(t_1 + g - RTT_x)$, where x is the next ONU to be polled and g is the guard time. Such a mechanism could eliminate the waiting time due to round trip transmission. Recall that previously we have defined the total usable bandwidth to be C_w in group w . Denoting the length of a transmission cycle by L , we should have

$$\frac{C_w}{C} = \frac{L - N_w M g}{L} \Leftrightarrow C_w = \frac{L - N_w M g}{L} C \quad (5.21)$$

where C is the physical line rate of a wavelength channel, N_w is the total number of ONUs in group w and M is the total number of classes.

Since each flow is constrained by its own traffic shaper, the overall traffic of each class in each ONU is computable. First we define the following terms:

$R_i^k(t)$: the arrival curve of class i traffic in ONU k

$S_i^k(t)$: the service curve provided to class i traffic in ONU k

Denote the start of a cycle by τ , the traffic arrival of class i in ONU k within any cycle $[\tau, \tau + L]$ is constrained by

$$R_i^k(\tau, \tau + L) = \rho_i^k L + \sigma_i^k \quad (5.22)$$

To fulfill the requirement of the admission control, the service provided must follow a certain rate. To be more specific, we plot the relation of arrival curve and service curve in Fig. 5.8.

Chapter 5. Admission Control and Scheduling in a Scalable WDM EPON Architecture

In Fig. 5.8, $R_i^k(t)$ and $S_i^k(t)$ are the theoretical arrival curve and service curve of class i traffic in ONU k respectively. Note that $A_i^k(t)$ is the *real traffic arrival* of class i in ONU k . $R_i^k(t)$ is essentially the theoretical traffic arrival; in other words, $R_i^k(t)$ is the *upper bound* of the real traffic arrival. Because of the traffic regulation by token bucket, $A_i^k(t)$ can never go beyond $R_i^k(t)$. The maximum deviation between $R_i^k(t)$ and $S_i^k(t)$ is the maximum delay (i.e., d_i^k). As this scheduling discipline follows the principles of the admission control, it is guaranteed that $d_i^k \leq D_i$ (D_i is the delay bound of class i) as long as the service could clear the accumulated traffic arrival.

Figure 5.8: Arrival curve and service curve illustration

As can be seen from the figure, to fulfill this requirement the *service rate* should be

$$y_i^k = \frac{R_i^k(\tau, \tau+L)}{L} = \rho_i^k + \frac{\sigma_i^k}{L} \quad (5.23)$$

The maximum transmission time assigned to class i of ONU k can be computed by

$$Z_i^k = \frac{R_i^k(\tau, \tau+L)}{C} = \frac{\rho_i^k L + \sigma_i^k}{C} \quad (5.24)$$

in which C is the physical line rate of the channel.

Based on this knowledge, we can predict the latest finishing time of current transmission. Due to the uncertainty of bursty traffic, the transmission may or may not finish at the predicted time. For example, in Fig. 5.7 all the ONUs finish their transmissions at expected times (the extreme case). However, they may also finish earlier if the traffic is not so “bursty”. According to

Chapter 5. Admission Control and Scheduling in a Scalable WDM EPON Architecture

the actual finishing time, there could be two cases as illustrated in Fig. 5.9. We will show how the scheduling handles these cases with different strategies.

Figure 5.9: Two different cases in scheduling

Suppose the OLT is serving class i . At time P_i^k the OLT polls current ONU k and the OLT begins to receive data at T_1 ($T_1 = P_i^k + RTT_k$). We predict that current transmission will end at T_3 ($T_3 = T_1 + Z_i^k$). According to it, the time of polling next ONU (namely ONU $k+1$) can be computed as T_2 ($T_2 = T_3 + g - RTT_{k+1}$), and the OLT is expected to receive the data at T_4 ($T_4 = T_3 + g$). Suppose T_f is the actual finishing time of current transmission, the actual polling time of next ONU (P_i^{k+1}) is

$$P_i^{k+1} = \begin{cases} T_f & \text{if } (T_f < T_2) \\ T_2 & \text{otherwise} \end{cases} \quad (5.25)$$

In the first case (case a), when the OLT knows that the data transmission has ended before expected time and then sends the next polling ahead of schedule, it saves a little bandwidth which could be allocated to best effort (BE) traffic (if it exists) after all the priority classes have been served within the cycle. In the second case (case b), since the next polling is already sent out at time T_2 , no additional actions are needed.

5.3 Simulation Results

In the simulation our objectives are, firstly, to test the capabilities of the admission control schemes; secondly, to evaluate the effectiveness of the wavelength scaling mechanism. In these two parts, the simulation scenarios are slightly different: the network consists of 16 ONUs initially without grouping, and the three admission control schemes (i.e., fixed-proportional, non-

Chapter 5. Admission Control and Scheduling in a Scalable WDM EPON Architecture

preemptive and preemptive) are tested by analyzing their respective delays, bandwidth utilization and flow loss ratio (defined later). The 16 ONUs share the same upstream wavelength channel and no extra wavelengths are employed. Later, the number of ONUs is increased to 64 and they are equally divided 4 Fixed Groups. Each Fixed Group is assigned a dedicated upstream wavelength channel and extra wavelengths can be deployed at this time. We compare our scheme with other WDM PON solutions in terms of maximum delay and number of extra wavelengths needed.

Table 5.1: Chosen token buckets for the three classes

Class i	α_i (%)	D_i (ms)	ρ_i (Mbps)	σ_i (kb)
1	20	1	200	1000
2	30	2	300	600
3	50	5	500	900

Figure 5.10: Delay distribution of admission control scheme A, B and C

In the simulation, each ONU has a uniformly distributed propagation delay over $[50 \mu\text{s}, 100 \mu\text{s}]$, which corresponds to OLT-ONU distances of 10 ~ 20 km. The available upstream bandwidth (C_w) can be obtained by Eq. (5.21) once the physical line rate is known. For ease of calculation,

Chapter 5. Admission Control and Scheduling in a Scalable WDM EPON Architecture

we assume $C_w = 1000 \text{ Mbps}$ in the simulation. We define three different classes in our scheme, and their bandwidth shares (used in scheme A only) and delay bound requirements are shown in Table 5.1. The aggregate token bucket of each class (used in scheme A only) can be acquired according to Theorem 5.1. To ensure the correctness of the simulation, each simulation is repeated at least five times to eliminate any accidental results.

In the first part of the simulation, 16 ONUs share the same upstream wavelength and no extra wavelengths are employed. VBR traffic flows are randomly generated at each ONU. Each flow is constrained by a token bucket with ρ randomly chosen from $[1 \text{ Mbps}, 5 \text{ Mbps}]$ and σ randomly chosen from $[10 \text{ kb}, 50 \text{ kb}]$. The random selection process has a uniform distribution. The delay requirements of these flows are chosen such that they fall randomly into any of the three classes. We keep generating new flows at each ONU until the network is fully saturated. Fig. 5.10 plots the Cumulative Distribution Function (CDF) of the delays of the three classes in the three schemes. The Cumulative distribution function is able to provide complete information on how delays are distributed. As can be seen, none of the classes violates its service agreement (delay bound) in scheme A (Fig. 5.10(a)). Similar results can be observed in scheme B (Fig. 5.10(b)) and scheme C (Fig. 5.10(c)), though lower classes (i.e., class 2 and class 3) in scheme C show relatively higher delays than in other schemes. This is because that in scheme C lower class traffic flows are often suspended by high class flows when the network is saturated, and delay is eventually increased even some of these suspended flows are restored later. However, since the suspended flows that have broken service agreement will be dropped, this mechanism still maintains good delay bounds. The results have demonstrated that the network calculus based admission schemes achieve good control over delays.

Chapter 5. Admission Control and Scheduling in a Scalable WDM EPON Architecture

Figure 5.11: Bandwidth utilization of admission control scheme A, B and C

Next we compare the bandwidth utilization among the three schemes. To obtain comparable results for the three schemes, every incoming VBR flow has the same token bucket of $\rho = 2 \text{ Mbps}$ and $\sigma = 100 \text{ kb}$ so different schemes can be compared under the same flow arrival rate. Each flow has a lifetime exponentially distributed with a mean value of 60 s . Fig. 5.11 plots the bandwidth utilization of the three schemes. The x-axis is the average number of flows arriving at each ONU per second. As can be seen from the figure, scheme B and scheme C have very close bandwidth utilization at all flow arrival rates. The preemptive admission control does not show any noticeable superiority over non-preemptive admission control in this respect. However, scheme A achieves lower bandwidth utilization than others when flow arrival rate becomes high. This can be explained as follows. Recall that the flows generated at each ONU fall randomly into any of the three classes. This implies that one third of the flows go into each class averagely. In the bandwidth pre-allocation of scheme A (see Table 5.1), 20%, 30% and 50% of the bandwidth are allocated to class 1, 2 and 3 respectively. It is easy to infer that with the coming of new flows class 1 will be the first to exhaust its capacity, followed by class 2 and 3. When there is no more capacity in class 1, new flows coming to it will be rejected. With the increase of load, class 2 will be the next to reject flows and class 3 will be the last to do so. As a contrast, scheme B and scheme C will not reject flows until the bandwidth of the whole wavelength channel is exhausted (To be more accurate, scheme C does not reject high class flows but suspends flows in lower classes when the capacity becomes tight. This leads to approximately the same bandwidth utilization as scheme B). It can be seen that with very high flow arrival rate, scheme A will

Chapter 5. Admission Control and Scheduling in a Scalable WDM EPON Architecture

eventually achieve the same bandwidth utilization as scheme B and C (i.e., when the network is saturated), but in the middle scheme A suffers lower bandwidth utilization because of fixed bandwidth shares. This has led to the conclusion that in scheme B if the bandwidth shares allocation is not in accord with the distribution of the real traffic, it will result in bandwidth underutilization. The bandwidth shares can surely be adjusted to acquire a better performance in this simulation; however, in the real network the traffic may be always changing and a fixed allocation will never apply to all cases. We will discuss another impact of this mechanism next.

Figure 5.12: Flow loss ratio of admission control scheme A, B and C

In the following experiment we examine the loss of flows in each scheme. Due to the properties of preemptive admission control, comparison of the flow blocking probability is not appropriate, as the preemptive admission control is always easier to accept new flows. This acceptance, however, is at the expense of suspending lower classes flows. We define a new term called *flow loss ratio*, which is the total number of flows lost divided by the total number of flows generated. For scheme A and scheme B, this flow loss ratio is exactly flow blocking probability; for scheme C, the lost flows should also include those suspended flows which are dropped finally

Chapter 5. Admission Control and Scheduling in a Scalable WDM EPON Architecture

because they can not keep their delay bounds. The other settings of the simulation are the same as the previous one.

Fig. 5.12 shows the flow loss ratio of the three schemes. In Fig. 5.12(a), we can see that class 1 in scheme A begins to lose (reject) flows earliest due to its smallest bandwidth share. Class 2 in scheme A has larger capacity so it begins to lose flows later than class 1. And so does class 3. This result has again shown that fixed bandwidth shares are inefficient in utilizing resources though the delay bound of admitted flows can still be guaranteed. In Fig. 5.12(b), scheme B has a moderate and similar flow loss ratio for all the three classes, as the whole bandwidth is shared by the three classes of traffic without priority. In Fig. 5.12(c), however, the flow loss ratio of class 1 always maintains 0, but that of class 2 and class 3 is much higher than scheme B. This is because that the acceptance of high class flows pushes the flows in low classes to suspended status, and a lot of suspended flows are finally dropped because they have waited too long in the queue. *From the results of the above two experiments we can conclude that scheme A can be used when the service provider wants to guarantee the resource of each class without any sharing, and scheme C can be used when the service provider wants to provide the service quality strictly in the order of the classes; scheme B can be used when there are no specific requirements and balanced performance is the first concern.*

Next we plan to compare the performance of our scheme with other WDM DBA schemes. As has been introduced earlier, many proposed WDM DBA schemes are WDM extensions of the well-known IPACT [45] algorithm and the work of [107] is a good representative. To be fair, in the simulation the total number of wavelengths in the network is identical for both our scheme and WDM IPACT [107]. In our scheme, four wavelengths are fixedly assigned (one for each group) and two extra wavelengths can be used to create Dynamic Groups. In the WDM IPACT, however, each ONU can transmit on any of the six wavelengths as there is no grouping. VBR traffic flows as defined before are continuously generated at each ONU resulting in increase of load. We compare the maximum packet delay between WDM IPACT and our scheme in Fig. 5.13. Our scheme used in the comparison is admission control scheme B (non-preemptive admission control) as it shows very balanced performance in the previous simulations.

Chapter 5. Admission Control and Scheduling in a Scalable WDM EPON Architecture

Figure 5.13: Maximum packet delay for different schemes

In Fig. 5.13, the load is the average load of each wavelength channel normalized to its capacity (also an indication of the network load in this experiment). The maximum delay for our scheme is calculated as follows: we first acquire the maximum delay of each group (Fixed and Dynamic) at different loads, and then compute the average at the same load. We have defined three classes in our scheme (same for all groups) hence there are three curves plotted in the figure, one for each class. The WDM IPACT [107] does not have any class differentiation so there is only one curve to show its maximum packet delay. As can be seen, the curves of three guaranteed classes in our scheme are quite smooth even when the load becomes very high. The maximum delay of each class does not exceed its bound at any load. As a contrast, the maximum delay of the WDM IPACT is low when the load is light but increases fast with the increase of load. It appears that the WDM IPACT lacks the ability to constrain the delays at high traffic loads.

The work of [50] proposed a WDM PON upgrade architecture that is similar to ours. The difference between [50] and our scheme is that [50] does not allow wavelength sharing by ONUs in different groups. This is not advantageous when each group only asks for very little extra bandwidth. The OLT may have to allocate a new wavelength for each of these bandwidth requests (i.e., one for each group) since the wavelengths can not be shared. Our scheme could accommodate all these requests by allocating a single wavelength as long as the aggregate bandwidth requested does not exceed a wavelength's capacity.

Chapter 5. Admission Control and Scheduling in a Scalable WDM EPON Architecture

Figure 5.14: Extra wavelengths required for different schemes

As usual, we choose admission control scheme B as the representative to compare with [50]. VBR traffic flows are continuously generated at each ONU resulting in increase of load but this time we do not limit how many extra wavelengths are used. In other words, extra wavelengths can be requested once any group is saturated. Note that the load in the figure is the average load of each wavelength channel normalized to its capacity *without counting in later requested extra wavelengths* (that is why it can go above 1). The number of extra wavelengths needed for the two different schemes at different loads is shown in Fig. 5.14. We denote by WDM-PON IU (WDM-PON supporting Incremental Upgrade) the work of [50]. As can be seen, the number of extra wavelengths needed in our scheme increases almost linearly with the increase of load. This is because that each new wavelength can be shared by all ONUs in the network. The WDM-PON IU [50], however, has several jumps on its curve. The bandwidth utilization at these jump points is obviously low, as some extra wavelengths have to be allocated to satisfy small bandwidth requests from different groups. Overall the WDM-PON IU [50] requires a larger number of extra wavelengths than ours due to lacking the capability of supporting wavelength sharing among groups.

*Chapter 5. Admission Control and Scheduling in a Scalable WDM EPON Architecture***5.4 Conclusion**

In this chapter, we propose a cost-effective and scalable WDM EPON architecture in which deterministic QoS guarantees could be achieved. We have presented the work in detail including the wavelength scaling process, the admission control policies and the scheduling discipline. With wavelength scaling, each group in the network could request extra bandwidth when it becomes saturated, and different groups could share the same extra wavelength such that its utilization is maximized. Those extra wavelengths could be released once the traffic load drops. The admission control schemes are capable of providing deterministic QoS guarantees for bursty traffic in each group. Admission control scheme A allocates fixed share of bandwidth for each class of traffic, while scheme B and scheme C are more flexible in allocating resources without employing fixed bandwidth shares. By means of extensive simulations we show that all the three admission control schemes could guarantee deterministic delay bounds for bursty traffic; at the same time, both scheme B and scheme C could achieve higher bandwidth utilization than scheme A. On the other hand, although the preemptive admission control scheme (scheme C) provides better service to high classes, it suffers higher flow loss ratio (especially in low classes) than the other two. The non-preemptive admission control scheme (scheme B) achieves balanced performance in terms of bandwidth utilization and flow loss ratio. When comparing our work with another WDM PON upgrade solution, our scheme achieves smaller delay, and requires fewer extra wavelengths for scaling at the same traffic load.

Chapter 6 A Scheme for IEEE 802.11e WLAN Supporting Deterministic QoS Provisioning

6.1 Introduction

In previous chapters, we have managed to achieve deterministic QoS in both optical core network (TWIN) and optical access network (WDM EPON). Wireless networks such as Wireless LAN (WLAN) are widely used as a complementary method where wired connections are not available or easily accessible. In such cases, it connects the users to the wired network. Therefore, providing deterministic QoS in WLAN is also considered one of the necessary conditions for end-to-end QoS provisioning.

As we have introduced in Chapter 1, in the new 802.11e standard, EDCA is not capable of providing strict QoS guarantees because of its contention based operation. Simulations in [58] have shown that when the channel load is high, only the traffic from the highest class is protected in low delay. HCCA does not own this weakness because the TXOP allocation for each QSTA is solely controlled by the QAP. However, a major issue with the reference scheme in the standard is that the TXOPs are calculated based on mean data rates, which would work fine with CBR traffic but would perform poorly if the traffic is a little “bursty” (e.g. VBR). Even if the mean sending rate of the VBR application is lower than that submitted in its QoS request, peaks of its traffic will not be absorbed by the TXOPs allocated by the reference scheduler. Simulations in [58] showed that VBR traffic is completely out of control when the HCCA reference scheduler is used. To address this issue, we propose a new scheme for the scheduling of upstream traffic in 802.11e WLAN. Similar ideas have been proposed in [78] and [126] for strict QoS provisioning in slotted optical networks and general IP networks. In these works, the scheduler handles the scheduling in a single node. The case is different in WLAN as the scheduler needs to deal with multiple nodes.

Chapter 6. A Scheme for IEEE 802.11e WLAN Supporting Deterministic QoS Provisioning

6.2 Admission Control Policy and Scheduling Algorithm

6.2.1 Admission Control Policy

First let us assume that there is only one QSTA in the network. It has M classes of traffic numbered from 1 to M , with M being the highest priority class and 1 being the lowest priority class. Each class i has a delay bound requirement D_i , and a bandwidth share α_i of the total available bandwidth C_{Avail} . We define a delay vector D , where $D_M < D_{M-1} < \dots < D_1$, and an allocation vector $\alpha = \{\alpha_M, \alpha_{M-1}, \dots, \alpha_1\}$, where $\sum_k \alpha_k \leq 1$. We assume that there are new traffic flows requesting to be admitted to these classes and each traffic flow is regulated by a token bucket. The token bucket is a well-known bursty traffic model which has two parameters: the mean data rate ρ and the bucket depth σ (burstiness). Based on the results of [78,126], we have the following theorem for the admission control in a WLAN when the polling overhead can be neglected (i.e., only one QSTA in the network):

Theorem 6.1: If a QSTA has M classes of traffic, and each class i has a delay bound requirement D_i and bandwidth share α_i , a new flow j with token bucket (ρ_j, σ_j) and delay requirement d_j could be admitted to class h if all the inequalities are true:

$$D_h \leq d_j \leq D_{h-1} \quad (6.1)$$

$$\frac{\sigma_j + \sum_{k=h}^M \sum_{i \in F_k} \sigma_i^k}{D_h} \leq C_{Avail} \left(1 - \sum_{k=h+1}^M \alpha_k\right) \quad (6.2)$$

$$\rho_j + \sum_{i \in F_h} \rho_i^h \leq \alpha_h C_{Avail} \quad (6.3)$$

where σ_i^k and ρ_i^k denote the burstiness and long-term rate of flow i in class k respectively, and F_h is the set of existing flows (already admitted) in class h .

When there is only one QSTA in the network, the polling overhead among different QSTAs does not exist. It becomes the same single node problem as studied in [78,126] and their conclusions are also true here.

Theorem 6.1 is the general form of the admission control (it is a single node case and we

Chapter 6. *A Scheme for IEEE 802.11e WLAN Supporting Deterministic QoS Provisioning*

will extend it to multiple node case later), but its form is not straightforward in application. We propose the following proposition for easy admission operations and it is easy to prove that it is equivalent to Theorem 6.1.

Proposition 6.1: If a QSTA has M classes of traffic, and each class i has a delay bound requirement D_i and bandwidth share α_i , to achieve each class's delay bound and to ensure each class obtains its bandwidth share, the aggregate token buckets (ρ_i, σ_i) for each class i need to be chosen as:

$$\rho_i = \alpha_i C_{Avail} \quad (6.4)$$

$$\sigma_i = D_i (C_{Avail} - \sum_{j=i+1}^M \rho_j) - \sum_{j=i+1}^M \sigma_j \quad (6.5)$$

A new flow j with token bucket (ρ_j, σ_j) and delay requirement d_j could be admitted to class h if:

$$D_h \leq d_j < D_{h-1} \quad (6.6)$$

$$\rho_j + \sum_{k \in F_h} \rho_k^h \leq \rho_h \quad (6.7)$$

$$\sigma_j + \sum_{k \in F_h} \sigma_k^h \leq \sigma_h \quad (6.8)$$

where ρ_k^h and σ_k^h are the long term rate and burstiness of flow k in class h respectively, and (ρ_h, σ_h) is the aggregate token bucket of class h , and F_h is the set of already admitted flows in class h .

When using Proposition 6.1 for admission control, we first calculate the aggregate token bucket for each class, then do a simple check according to Eq. (6.6 - 6.8) to verify whether a new flow could be admitted.

Theorem 6.1 has addressed the admission problem when there is only one QSTA in the network. How could it be extended to the multiple QSTAs case? A natural idea is that, if there are N identical QSTAs in the network with each QSTA having the same number of classes of traffic, we could treat the *same* priority class of traffic in different QSTAs as a whole and the admission policy of Theorem 6.1 can still hold.

To be more particular, we assume that each QSTA has the same number of classes and the

Chapter 6. A Scheme for IEEE 802.11e WLAN Supporting Deterministic QoS Provisioning

same class of traffic in different QSTAs has the same delay bound. In other words, the partition of classes for all QSTAs is identical. Therefore, traffic flows in the same class from different QSTAs could join the admission under the same criteria. On the other hand, to fulfill the requirements of Theorem 6.1, the QAP should serve the QSTAs strictly following the order of the classes, which implies that the accumulated traffic of a higher class in all QSTAs must be cleared before the traffic in a lower class could be transmitted.

We assume that in the WLAN, the PHY rate is C , and the available capacity that can be used by HCCA is C_{HCCA} . According to the 802.11e standard [10], they should satisfy

$$C_{HCCA} = C \frac{T - T_{CP}}{T} \quad (6.9)$$

where T is the beacon interval and T_{CP} is the time used for EDCA traffic.

If the overhead for one time polling is O , the maximum polling overhead in a beacon should be no more than

$$O_{\max} \leq MNO \quad (6.10)$$

where M is the maximum number of classes and N is the number of QSTAs. This is because we must poll the same class in all the QSTAs before we can move to the next class. The right-hand side of (6.16) is only the worst case because each QSTA may not have traffic in all classes.

It is obvious that the available capacity for the admission control is

$$C_{Avail} = C_{HCCA} - O_{\max} \quad (6.11)$$

The admission control policy works as follows. When there is a new flow coming to a QSTA, the QSTA sends its traffic information (including its token bucket parameters and delay requirement) to the QAP. This can be done by sending a request frame or by piggybacking the information in the data packets (since it is a very small fraction of data). The QAP will try to admit the new flow according to Proposition 6.1. If there is enough capacity, the flow will be admitted and new TXOP value will be computed accordingly (how to compute it will be discussed in the next section), and the QSTA will be notified in the next round poll. The QAP maintains the information of all admitted flows of each class, thus could make decisions rapidly when there is a new request.

Chapter 6. A Scheme for IEEE 802.11e WLAN Supporting Deterministic QoS Provisioning

On the other hand, when a QSTA knows that one flow is leaving (e.g., the real-time session has ended), it should also notify the QAP so that new TXOP could be computed accordingly, releasing the extra capacity for future use.

To ensure that the admission control works as expected, we need a scheduling algorithm to control the transmissions of the QSTAs. It is called Service Curve Based Scheduling algorithm (SCBS), which will be introduced in the next section.

6.2.2 Service Curve Based Scheduling Algorithm (SCBS)

According to our admission control, the incoming traffic flows in each QSTA should not violate their token bucket constraints which they have negotiated with the QAP. However, unlike in the packet networks, implementing traffic regulators in a wireless network could be very expensive, especially for some small devices lacking strong computing capabilities such as the Pocket PC. Therefore, we implement a *virtual* regulator in the QAP for each class of each QSTA, since the QAP is usually much more powerful in hardware. The procedures are as follows: first we calculate the selected service interval (SI), which should be the highest submultiple value of the beacon interval and also no larger than all the maximum RSIs required by different classes of different QSTAs, as specified in [10]. Then the QAP will calculate the corresponding TXOP of each class for each QSTA in SI. We define the following variables to be used in the scheduling algorithm:

$S_i^k(t)$: the service curve provided to class i in QSTA k

$R_i^k(t)$: the arrival curve of class i traffic in QSTA k

Since we use token bucket as the traffic model, the traffic arrival could be obtained by

$$R_i^k(t) = \rho_i^k t + \sigma_i^k \quad (6.12)$$

where ρ_i^k is the aggregate mean data rate of all flows belong to class i in QSTA k , and σ_i^k is the aggregate burstiness of all flows belong to class i in QSTA k .

The service $S_i^k(t)$ is updated whenever new packets in class i of QSTA k get transmitted.

According to the Service Curve theory, the backlog of class i traffic in QSTA k is the vertical

Chapter 6. A Scheme for IEEE 802.11e WLAN Supporting Deterministic QoS Provisioning

deviation between the arrival curve and service curve, i.e. $R_i^k(t) - S_i^k(t)$. This could be deemed as the upper bound of traffic amount that the QAP could schedule to transmit if the delay bound is to be maintained. On the other hand, based on the characteristics of token buckets, the maximum traffic that could be accumulated and transmitted for class i in QSTA k each time should be no more than σ_i^k . Let L_i be the nominal packet size in class i and Z_i^k be the number of packets that can be transmitted for class i traffic in QSTA k , the corresponding TXOP could be computed as follows:

$$Z_i^k = \left\lfloor \min\left(\frac{R_i^k(t) - S_i^k(t)}{L_i}, \frac{\sigma_i^k}{L_i}\right) \right\rfloor \quad (6.13)$$

$$TXOP_i^k = \frac{L_i * Z_i^k}{C} + O \quad (6.14)$$

where C is the PHY data rate.

Table 6.1: Variables used in the SCBS algorithm

t	System time starting from 0
Y	Number of classes
N	Number of QSTAs
C	PHY data rate
R_i^k	Total arrival (in bits) of class i traffic in QSTA k
S_i^k	Total service (in bits) received by class i in QSTA k
ρ_i^k	Aggregate mean data rate of class i traffic in QSTA k
σ_i^k	Aggregate burstiness of class i traffic in QSTA k
L_i	Nominal packet size of class i traffic
Z_i^k	Number of packets that class i traffic in QSTA k can transmit
$TXOP_i^k$	TXOP assigned to class i traffic in QSTA k
Total_TXOP	Total TXOP allocated so far in current SI
Avail_Time	Time that can be used by HCCA in each SI

Chapter 6. A Scheme for IEEE 802.11e WLAN Supporting Deterministic QoS Provisioning

Initialization: $t = 0$; $S_i^k = 0$ for $i = \{1, \dots, Y\}$ and $k = \{1, \dots, N\}$.

At the start of every SI:

Total_TXOP = 0;

Avail_Time = SI * $(T - T_{CP}) / T$;

FOR $i = 1$ to Y

FOR $k = 1$ to N

IF QSTA k has class i traffic

$$R_i^k = \rho_i^k t + \sigma_i^k;$$

$$Z_i^k = \left\lfloor \min\left(\frac{R_i^k - S_i^k}{L_i}, \frac{\sigma_i^k}{L_i}\right) \right\rfloor$$

$$TXOP_i^k = \frac{L_i * Z_i^k}{C} + O$$

IF (Total_TXOP + TXOP $_i^k$) > Avail_Time

EXIT;

ELSE

 Poll QSTA k and assign TXOP $_i^k$ for class i traffic;

 Wait until the transmission finishes;

 Update S_i^k by adding actual transmitted amount of traffic;

 Update Total_TXOP by adding actual transmission time used by QSTA k ;

ENDIF

ENDIF

ENDFOR

ENDFOR

Figure 6.1: The Service Curve Based Scheduling (SCBS) algorithm

The SCBS algorithm is shown in Fig. 6.1 and it works as follows. In each SI after computing the TXOPs, the QAP begins to poll each QSTA in a fixed sequence, starting from the highest class. When the QAP polls a certain QSTA, say, class i in QSTA k , TXOP $_i^k$ is granted. If it is not enough for transmitting all the accumulated traffic, only the time of TXOP $_i^k$ could be used, and the rest of the packets (non-conformant traffic) in the QSTA should be dropped. If the time is more than enough, the control will be returned to QAP after the transmission finishes. Then the QAP polls the next QSTA of the same class. After the highest class has been polled in all QSTAs, the QAP begins to poll the next priority class in the same sequence.

In the algorithm, when the sum of already allocated TXOPs in current SI is bigger than the portion that can be used by HCCA, the algorithm stops. We say that this may occur and is reasonable because only the traffic of highest class should be transmitted once in each SI. Traffic of lower classes only needs to be transmitted once in the duration of multiple SIs (they have larger delay bound than the highest class). The admission control policy ensures that the delay

Chapter 6. A Scheme for IEEE 802.11e WLAN Supporting Deterministic QoS Provisioning

bound of each class is still satisfied. This would be better than the reference scheduling algorithm [10] which tries to transmit all classes of traffic in the smallest delay bound negotiated.

Admission control is critical in achieving guaranteed performance. Without proper admission control, the fairness of the system would be easily impaired by ill-behaving traffic flows, and the delay bounds of the traffic would be breached too. We will show in the following section that our admission control scheme could effectively guarantee the delay of bursty traffic.

6.3 Simulation Results and Discussion

In this section, we compare the performance of the reference scheduler and our scheduler by means of extensive simulations. We implement SCBS and the reference scheduler of 802.11e in Network Simulator 2 (ns-2). The PHY data rate is set to 11Mb/s. An ideal, error-free wireless channel is assumed, as our focus is on the scheduling mechanism. Since the transmission time in the wireless channel is trivial compared to queueing time, the delay bound here refers to queueing delay bound. The simulation topology contains one fixed QAP and 6 QSTAs. We set the maximum number of classes to 3. There are three types of VBR flows and their traffic specifications are shown in Table 6.2. Their parameters are selected such that we have three different kinds of flows: large mean data rate but small burstiness (flow 1), small mean data rate but large burstiness (flow 3) and ordinary mean data rate and burstiness (flow 2). All traffic flows use UDP as the transport layer protocol. The beacon interval is 600 ms. In this implementation, we assume the maximum RSI and delay bound have the same value. The required delay bound of each predefined class is shown in Table 6.3, as well as the bandwidth shares. According to it, SI is chosen as 60 ms. The bandwidth shares are determined by the service provider depending on its policy. For example, it may allocate more bandwidth to high classes since they have paid more. With the parameters in Table 6.3, the aggregate token bucket for each class could be calculated by Proposition 6.1. For the sake of simplicity, we assume the maximum capacity that can be used by HCCA excluding all polling overheads is 7.7Mb/s (i.e., $C_{Avail} = 7.7\text{Mb/s}$). In the following simulations, each test is repeated at least five times to eliminate any accidental results.

Chapter 6. A Scheme for IEEE 802.11e WLAN Supporting Deterministic QoS Provisioning

Table 6.2: Traffic specifications of the three VBR flows

VBR flow	Mean Data Rate	Bucket Depth
1	360 kb/s	26 kb
2	126 kb /s	40 kb
3	60 kb/s	60 kb

Table 6.3: The parameters of the three classes

Class i	α_i (%)	D_i (ms)	ρ_i (kb/s)	σ_i (kb)
1	60	200	3860	164
2	30	100	2310	231
3	20	60	1640	386

Figure 6.2: Comparison of average delay

Figure 6.3: Comparison of maximum delay

Chapter 6. A Scheme for IEEE 802.11e WLAN Supporting Deterministic QoS Provisioning

We gradually increase the number of flows at each QSTA to evaluate the performance of the two schemes under different loads. In the process, one VBR flow is generated each time (randomly selected from one of the three kinds) and requests to be admitted into one of the three classes (also randomly selected). Fig. 6.2 and Fig. 6.3 show the average delay and the maximum delay of the three classes with two different schedulers. It is apparent that with the reference scheduler, the average delay and maximum delay are both much larger than with SCBS. In addition, we notice a very interesting phenomenon: the class 3 traffic (highest class), which is supposed to have the minimum delay in the three classes, has the biggest delay instead with the reference scheme. To inspect it further, we plot the cumulative distribution function of the delays at the load of 0.8 in Fig. 6.4.

Figure 6.4: Delay distribution at the load of 0.8

In Fig. 6.4, we could notice two things: first, the delays of the three classes' traffic are kept within their bounds when using SCBS, while the delays are totally uncontrolled with the reference scheme; second, when using the reference scheme the highest class (class 3) traffic has the biggest delay compared to other two classes, just the same as we have observed in Fig. 6.2 and Fig. 6.3. To explain this phenomenon, recall the parameters defined in Table 6.3. Class 3 has the biggest bucket (σ_i), which implies that the admission control allows the biggest burstiness in class 3. Since the reference scheduler only allocates a fixed TXOP to each QSTA based on the mean data rate, a lot of class 3 packets are queued at the QSTAs when the bursts come, and delay

Chapter 6. A Scheme for IEEE 802.11e WLAN Supporting Deterministic QoS Provisioning

is consequently increased. The situations for class 2 and class 1 are the same but relatively better because they have smaller burstiness. This effect is undesirable because it is expected that class 3 should receive the best service. It has proved that the reference scheduler is very weak in handling bursty traffic.

As to the SCBS scheme, we can see that the delays of the three classes are kept within their bounds. Since class 3 traffic is given the highest priority during the transmission, its delay is lowest. Class 2 traffic has larger delay but it is still kept in its bound. And so does class 1. This is a desirable improvement upon the 802.11e, as the reference scheduler is trying to provide different classes *the same and most stringent delay bound* (the minimum delay bound of all), which is usually not necessary. There is actually no service differentiation by this means. Our scheme ensures that each class only obtains the service it requests, hence increases the bandwidth utilization, and avoids the impact of non-conformant traffic as well.

Figure 6.5: Percentage of packets exceeding their delay bound

We would like to examine how well the two schemes can control the delay of packets. In the following experiment, we compute the percentage of packets that exceeded their delay bound at different loads when different schemes are used (all the three classes are taken into account). The result is plotted in Fig. 6.6. It can be seen that the reference scheduler is very weak in controlling delays and it becomes worse and worse when traffic load rises. Though those packets exceeding delay bounds will be transmitted eventually, the large delay incurred is disadvantageous for QoS

Chapter 6. A Scheme for IEEE 802.11e WLAN Supporting Deterministic QoS Provisioning

provisioning. As a contrast, SCBS maintains a low percentage of packets exceeding delay bounds at different loads. This has rendered SCBS a much better choice for QoS provisioning.

Figure 6.6: Session blocking probability of the two schedulers

Next we analyze the session blocking probability of the two schemes, since both of them have admission control. The session blocking probability is defined as the probability that a new flow is rejected by the admission control. In the simulation, flows are randomly generated at each QSTA, one at a time, with the average arrival interval varying from 6 s to 14 s. The class they request to join is also random. The lifetime of a flow is exponentially distributed with a mean value of 60 s. The average blocking probability is computed and plotted in Fig. 6.6. It can be seen that our proposed scheme has a higher blocking probability than the reference scheme. This is because the SCBS admission control takes the burstiness into account and reserves more resources, hence it admits fewer flows than the reference scheme, especially when the system is busy. This is the cost we need to pay for guaranteed QoS. It may happen when the reference scheme could still admit new flows while SCBS could not. But this only occurs when the capacity is almost fully occupied. Admitting new flows at this time will only bring service degradation because more bursts will join in. As we will see later in the throughput comparison, admitting more flows does not necessarily guarantee higher throughput.

To examine the throughput of the two schemes at different loads, we conduct the final experiment. Fig. 6.7 plots the throughputs of the two schedulers. There is not too much difference

Chapter 6. A Scheme for IEEE 802.11e WLAN Supporting Deterministic QoS Provisioning

between them, but SCBS always manages to get a little higher throughput. It is because SCBS could dynamically adjust the length of TXOP and thus absorb the bursts in the traffic. However, the reference scheduler always allocates fixed length TXOP based on the mean data rate, hence it cannot assign more TXOP even though there is still capacity unused.

Figure 6.7: Throughput comparison of the two schedulers

6.4 Conclusion

In this chapter we propose an admission control scheme capable of providing deterministic QoS guarantees for bursty traffic in IEEE 802.11e WLAN. We have presented the admission control policy and the SCBS algorithm in detail, and conducted extensive simulations using ns-2 simulator. The results show that while the reference scheduler suffers great delays under bursty traffic, SCBS still maintains the delay bound for each class of traffic. SCBS also has a low percentage of packets exceeding delay bounds, which is favored by real-time applications. On the other hand, SCBS has a higher session blocking probability than the reference scheme when the traffic intensity is high. Though this may look like a disadvantage compared to the reference scheduler, it is a necessary cost for providing deterministic QoS. As we have pointed out, the reference scheme's lower session blocking probability does not necessarily guarantee higher throughput in this case, but only degrades the QoS it could achieve.

Chapter 7 Conclusions and Future Work

7.1 Conclusions

In this thesis, we have managed to provide deterministic QoS guarantees in optical core network (TWIN), optical access network (WDM EPON), and wireless network (WLAN) connected to user interface. Such deterministic QoS in individual networks may not be sufficient for guaranteeing deterministic end-to-end QoS, but they are the necessary conditions for this ultimate goal. As deterministic end-to-end QoS provisioning across multiple network domains is a very challenging problem and there is no implementable solution so far, our contributions are essential in finding a feasible solution for the end-to-end QoS problem.

We start from the investigation of TWIN, a promising candidate for future transport networks. In Chapter 3, we build an analytical model for TWIN from which delay properties could be derived. We compare the analytical values with simulation data and find that the analytical values track the simulation results quite well when the input load is smaller than the saturation input load, which proves the correctness of our model. Further, we formulate the scheduling problem into an integer linear programming (ILP) problem and propose our Destination Slot Set (DSS) algorithm. Simulation shows that DSS achieves much smaller queueing delay and queueing delay variance than existing batch scheduling algorithm and hence could guarantee smaller total delay and delay jitter in TWIN. Besides, DSS achieves the same maximum throughput as the batch algorithm does. This has demonstrated that our algorithm could provide better QoS than existing works without sacrificing performance in other aspects.

As a step forward, we would like to achieve deterministic QoS in optical networks. Providing such QoS guarantees in an optical network is a challenging task. Few available works on QoS guarantees in star-coupled networks assumed equal end-to-end delays which are not applicable to networks with arbitrary topologies. In Chapter 4, we propose an efficient

Chapter 7. Conclusions and Future Work

framework for deterministic service guarantees in slot-based optical networks with arbitrary topology based on the architecture of TWIN. The framework uses the combination of a control plane and a data plane to solve the complex problem of capacity allocation, slot-matching and traffic scheduling. We have applied admission control and capacity allocation to source-destination node pairs in the control plane successfully. It uses electrical domain delay compensations and the collision resolution algorithm has reasonable complexity of $O(N)$ for a N edge nodes optical network. We have proposed a comprehensive aggregation, buffering, and scheduling mechanism to provide deterministic QoS guarantees using class-based traffic treatment. Through our analysis, we prove that the framework realizes Service Curves assurance that provides class-based queueing delay bounds. Our simulations have validated our analysis and shown that the class-based delay bounds are maintained. The simplicity and scalability of this framework make it suitable for implementation in future transport networks.

Our next step is to achieve deterministic QoS in optical access networks. EPON is viewed by many as a promising access network solution. As more and more bandwidth-consuming applications emerge in the network, upgrading its architecture with WDM technology becomes a natural and cost-effective choice. However, there are very few works that are able to provide deterministic QoS in WDM EPON/PON. In Chapter 6, we propose a scalable WDM EPON architecture in which deterministic QoS guarantees could be achieved. We present the work in detail including the wavelength scaling process, the admission control policies and the scheduling discipline. With wavelength scaling, each group in the network could request extra bandwidth when it becomes saturated, and ONUs in different groups could share the same extra wavelength such that the utilization is maximized. The extra wavelengths could be released once the traffic load drops and reused later. The admission control schemes are capable of providing deterministic QoS guarantees for bursty traffic in each group. Among the three admission control schemes, scheme A allocates fixed share of bandwidth to each class of traffic, while scheme B and scheme C do it more flexibly by eliminating the use of bandwidth shares. Simulations have proved that all the three admission control schemes could guarantee deterministic delay bounds for bursty traffic; at the same time, both scheme B and scheme C can achieve higher bandwidth

Chapter 7. Conclusions and Future Work

utilization than scheme A. On the other hand, although scheme C provides better service to high classes, it suffers higher flow loss ratio in low classes than the other two. Scheme B achieves balanced performance in terms of bandwidth utilization and flow loss ratio. When comparing with another WDM PON upgrade solution, the simulation results show that our work achieves lower maximum delay and requires fewer extra wavelengths for scaling at the same traffic load.

Providing deterministic QoS in wireless networks is the last step in this thesis, as wireless networks such as WLANs have served as the last segment connecting edge networks and users in many occasions. IEEE 802.11e is a newly developed standard which has incorporated service differentiation mechanism, but the reference scheduling algorithm of HCCA only performs well with CBR traffic. In Chapter 5, we propose an admission control and scheduling scheme for IEEE 802.11e HCCA which could achieve deterministic delay bounds for bursty traffic. We present our SCBS algorithm and conduct extensive simulations. The results show that while the reference scheduler suffers great delays under bursty traffic, SCBS still maintains the delay bound for each class of traffic. SCBS also has a low percentage of packets exceeding their delay bounds, which is favored by real-time applications. On the other hand, our SCBS has a higher session blocking probability than the reference scheme when the traffic intensity is high, which may look like a disadvantage. However, this is a necessary cost for providing deterministic QoS. In this case, the lower session blocking probability of the reference scheme does not necessarily guarantee higher throughput, but only degrades the QoS.

7.2 Future Work

In this thesis work, we studied the issue of end-to-end deterministic QoS provisioning. This is a very challenging problem and so far there is little progress in an implementable solution. Instead of providing a full solution achieving deterministic end-to-end QoS at this stage, we managed to achieve such guarantees in individual networks including optical core network, optical access network and user wireless network. Of course, concatenation of individual autonomous networks that provide deterministic QoS may not necessarily guarantee end-to-end deterministic QoS. Hence our next step research will be concentrated on this topic.

One important reason that makes end-to-end deterministic QoS so hard to implement is that there are many types of networks and different networks exhibit different behaviors when dealing with traffic. To solve this problem effectively, a unified network architecture will surely help. Based on these analyses, we consider the following targets as our future research:

- A unified network architecture combining both transport network and access network. In fact, [49] has done some preliminary work by connecting multiple EPONs with Resilient Packet Ring (RPR) which could serve as a transport network. However, whether ring is a good choice for future transport network is still uncertain. Is tree based network such as TWIN more effective in such unified network architecture? Before we could answer this question, extensive investigations much be done to evaluate their performances, especially when multiple wavelengths are employed and efficiency is the primary concern.
- A deterministic QoS provisioning scheme for such an architecture. If we could provide such guarantees in this architecture, we have actually managed to achieve end-to-end deterministic QoS in a limited form of WAN (we say it is a limited form because it only include transport network and access network), which is a big step from current research. However, this will not be easy, as the scheme must be scalable, and handle other issues such as admission control and scheduling well.

Author's Publications

- D. J. Xue, Y. Qin, and C. K. Siew, "Supporting deterministic quality-of-service in WDM EPON networks," submitted to *IEEE/OSA Journal of Lightwave Technology*
- D. J. Xue, Y. Qin, and C. K. Siew, "Performance analysis of a novel traffic scheduling algorithm in slotted optical networks," *Computer Communications*, vol. 30, no. 18, pp. 3559-3571, Dec. 2007.
- Y. Qin, D. J. Xue, and C. K. Siew, "Virtual topology-based traffic scheduling algorithm for slotted optical networks," *Journal of optical networking*, vol. 6, no. 2, pp. 166-179, Feb. 2007.
- C. K. Siew, D. J. Xue, and Y. Qin, "Providing deterministic quality of service in slotted optical networks," *Optics Express*, vol. 14, no. 26, pp. 12679-12692, Dec. 2006.
- D. J. Xue, Y. Qin and C. K. Siew, "Deterministic QoS provisioning with network calculus based admission control in WDM EPON networks," *Proc. ICC 09*, Jun. 2009.
- D. J. Xue, Y. Qin, and C. K. Siew, "A service curve based scheduling algorithm for QoS support in 802.11e wireless LANs," *Proc. GLOBECOM 06*, Nov. 2006.
- C. K. Siew, D. J. Xue, and Y. Qin, "A novel framework for guaranteed service provisioning in optical network," *Proc. IEEE International Conference on Communication Systems (ICCS)*, Oct. 2006.
- D. J. Xue, Y. Qin, and C. K. Siew, "A slotted dynamic traffic scheduling algorithm in time-domain wavelength interleaved networks," *Proc. IEEE International Conference on Networks (ICON)*, Nov. 2005.
- Y. Qin, D. J. Xue, C. K. Siew, I. Saniee, and I. Widjaja, "A topology based dynamic traffic scheduling in time-domain wavelength interleaved networks," *Proc. IEEE International Conference on Networks (ICON)*, Nov. 2004.

Bibliography

- [1] W. J. Goralski, *SONET, 2nd Edition* McGraw-Hill, New York, 2000.
- [2] Shun Yao, B. Mukherjee, and S. Dixit, "Advances in photonic packet switching: an overview," *IEEE Communications Magazine*, vol. 38, no. 2, pp. 84-94, 2000.
- [3] C. Qiao and M. Yoo, "Optical burst switching (OBS) - a new paradigm for an optical internet," *Journal of High Speed Networks*, vol. 8, no. 1, pp. 69-84, 1999.
- [4] I. Widjaja, I. Saniee, R. Giles, and D. Mitra, "Light core and intelligent edge for a flexible, thin-layered, and cost-effective optical transport network," *IEEE Communications Magazine*, vol. 41, no. 5, p. S30-S36, 2003.
- [5] O. Gerstel, P. Lin, and G. Sasaki, "Combined WDM and SONET network design," *Proc. INFOCOM'99*, 1999.
- [6] A. L. Chiu and E. H. Modiano, "Traffic grooming algorithms for reducing electronic multiplexing costs in WDM ring networks," *Journal of Lightwave Technology*, vol. 18, no. 1, pp. 2-12, 2000.
- [7] C. Nuzman and I. Widjaja, "Time-domain wavelength interleaved networking with wavelength reuse," *Proc. INFOCOM'06*, 2006.
- [8] IEEE 802.3 WG, "IEEE Standard for Information technology- Telecommunications and information exchange between systems- Local and metropolitan area networks- Specific requirements Part 3: Carrier Sense Multiple Access with Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications Amendment: Media Access Control Parameters, Physical Layers, and Management Parameters for Subscriber Access Networks," *IEEE Std 802. 3ah-2004*, 2004.
- [9] M. P. McGarry, M. Maier, and M. Reisslein, "WDM ethernet passive optical networks," *IEEE Communications Magazine*, vol. 44, no. 2, pp. 15-22, 2006.
- [10] IEEE 802.11 WG, "IEEE Standard for Information technology - Telecommunications and information exchange between systems - Local and metropolitan area networks - Specific requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications Amendment 8: Medium Access Control (MAC) Quality of Service Enhancements," *IEEE Std 802. 11e-2005*, 2005.
- [11] K. Ross, N. Bambos, K. Kumaran, I. Saniee, and I. Widjaja, "Scheduling bursts in time-domain wavelength interleaved networks," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 9, pp. 1441-1451, 2003.
- [12] K. Ross, N. Bambos, K. Kumaran, I. Saniee, and I. Widjaja, "Dynamic scheduling of optical data bursts in time-domain wavelength interleaved networks," *Proc. 11th Symposium on High Performance Interconnects*, 2003.

- [13] Z. Zhang and Y. Yang, "Distributed QoS-aware scheduling algorithm for WDM optical interconnects with arbitrary wavelength conversion capability," *Proc. International Conference on Parallel Processing*, 2004.
- [14] F. Callegati, G. Corazza, and C. Raffaelli, "Exploitation of DWDM for optical packet switching with quality of service guarantees," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 1, pp. 190-201, 2002.
- [15] B. Li and Y. Qin, "Traffic scheduling in a photonic packet switching system with QoS guarantee," *Journal of Lightwave Technology*, vol. 16, no. 12, pp. 2281-2295, 1998.
- [16] X. Huang and M. Ma, "Efficient scheduling with reduced message delay for passive star coupled WDM optical networks," *IEE Proceedings Communications*, vol. 152, no. 6, pp. 765-770, 2005.
- [17] X. Huang and M. Ma, "An efficient scheduling algorithm for real-time traffic on WDM passive star optical networks," *Journal of Lightwave Technology*, vol. 23, no. 11, pp. 3683-3695, 2005.
- [18] C.-F. Hsu, T.-L. Liu, and N.-F. Huang, "Multicast traffic scheduling in single-hop WDM networks with arbitrary tuning latencies," *IEEE Transactions on Communications*, vol. 52, no. 10, pp. 1747-1757, 2004.
- [19] M. Ma, "A cost-based scheduling algorithm for differentiated service on WDM optical networks," *IEEE Communications Letters*, vol. 7, no. 9, pp. 460-462, 2003.
- [20] M. Ma and M. Hamdi, "An adaptive scheduling algorithm for differentiated services on WDM optical networks," *Proc. GLOBECOM'01*, 2001.
- [21] H.-Y. Tyan, J. C. Hou, B. Wang, and C.-C. Han, "On supporting temporal quality of service in WDMA-based star-coupled optical networks," *IEEE Transactions on Computers*, vol. 50, no. 3, pp. 197-214, 2001.
- [22] M. Ma and M. Hamdi, "Providing deterministic quality-of-service guarantees on WDM optical networks," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 10, pp. 2072-2083, 2000.
- [23] M. Ma and M. Hamdi, "Providing guaranteed deterministic performance service to multimedia applications on WDM optical networks," *Proc. GLOBECOM'00*, 2000.
- [24] M. Ma and M. Hamdi, "A systematic scheme for guaranteed deterministic performance service on WDM optical networks," *Proc. ICC'00*, 2000.
- [25] M. Klinkowski, D. Careglio, D. Morato, and J. Sole-Pareta, "Effective burst preemption in OBS network," *2006 Workshop on High Performance Switching and Routing*, 2006.
- [26] W. Liao and C.-H. Loi, "Providing service differentiation for optical-burst-switched networks," *Journal of Lightwave Technology*, vol. 22, no. 7, pp. 1651-1660, 2004.
- [27] H. C. Cankaya, S. Charcranoon, and T. S. El-Bawab, "A preemptive scheduling technique for OBS networks with service differentiation," *Proc. GLOBECOM'03*, 2003.
- [28] L. Yang, Y. Jiang, and S. Jiang, "A probabilistic preemptive scheme for providing service differentiation in OBS networks," *Proc. GLOBECOM'03*, 2003.

- [29] F. Farahmand and J. P. Jue, "Supporting QoS with look-ahead window contention resolution in optical burst switched networks," *Proc. GLOBECOM'03*, 2003.
- [30] C.-H. Loi, W. Liao, and D.-N. Yang, "Service differentiation in optical burst switched networks," *Proc. GLOBECOM'02*, 2002.
- [31] J. Liu and N. Ansari, "Forward resource reservation for QoS provisioning in OBS systems," *Proc. GLOBECOM'02*, 2002.
- [32] M. Yoo, C. Qiao, and S. Dixit, "Optical burst switching for service differentiation in the next-generation optical Internet," *IEEE Communications Magazine*, vol. 39, no. 2, pp. 98-104, 2001.
- [33] Y. Chen, M. Hamdi, and D. H. K. Tsang, "Proportional QoS over OBS networks," *Proc. GLOBECOM'01*, 2001.
- [34] M. Yang, S. Q. Zheng, and D. Verchere, "A QoS supporting scheduling algorithm for optical burst switching DWDM networks," *Proc. GLOBECOM'01*, 2001.
- [35] M. Yoo, C. Qiao, and S. Dixit, "QoS performance of optical burst switching in IP-Over-WDM networks," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 10, pp. 2062-2071, 2000.
- [36] A. R. Dhaini, C. M. Assi, M. Maier, and A. Shami, "Per-stream QoS and admission control in ethernet passive optical networks (EPONs)," *Journal of Lightwave Technology*, vol. 25, no. 7, pp. 1659-1669, 2007.
- [37] Y. Zhu, M. Ma, and T. H. Cheng, "An urgency fair queuing scheduling to support differentiated services in EPONs," *Proc. GLOBECOM'05*, 2005.
- [38] M. Ma, Y. Zhu, and T. H. Chen, "A systematic scheme for multiple access in ethernet passive optical access networks," *Journal of Lightwave Technology*, vol. 23, no. 11, pp. 3671-3682, 2005.
- [39] Y. Zhu, M. Ma, and T. H. Chen, "An urgency fair queuing scheduling to support differentiated services in EPONs," *Proc. GLOBECOM'05*, 2005.
- [40] L. Zhang and G.-S. Poo, "Delay constraint dynamic bandwidth allocation in ethernet passive optical networks," *Proc. ICCS'04*, 2004.
- [41] M. Ma, Y. Zhu, and T. H. Cheng, "A bandwidth guaranteed polling MAC protocol for ethernet passive optical networks," *Proc. INFOCOM'03*, 2003.
- [42] L. Zhang, Eung-Suk An, Chan-Hyun Youn, Hwan-Geun Yeo, and Sunhee Yang, "Dual DEB-GPS scheduler for delay-constraint application in ethernet passive optical networks," *IEICE Transactions on Communications*, vol. E86-B, no. 5, pp. 1575-1584, 2003.
- [43] H.-J. Byun, J.-M. Nho, and J.-T. Lim, "Dynamic bandwidth allocation algorithm in Ethernet passive optical networks," *Electronics Letters*, vol. 39, no. 13, pp. 1001-1002, 2003.
- [44] C. M. Assi, Y. Ye, S. Dixit, and M. A. Ali, "Dynamic bandwidth allocation for quality-of-service over Ethernet PONs," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 9, pp. 1467-1477, 2003.

- [45] G. Kramer, B. Mukherjee, and G. Pesavento, "IPACT: a dynamic protocol for an Ethernet PON (EPON)," *IEEE Communications Magazine*, vol. 40, no. 2, pp. 74-80, 2002.
- [46] S.-I. Choi and J.-D. Huh, "Dynamic bandwidth allocation algorithm for multimedia services over Ethernet PONs," *ETRI Journal*, vol. 24, no. 6, pp. 465-468, 2002.
- [47] G. Kramer, B. Mukherjee, Y. Ye, S. Dixit, and R. Hirth, "Supporting differentiated classes of service in Ethernet passive optical networks," *Journal of Optical Networking*, vol. 1, no. 8, pp. 280-298, 2002.
- [48] Yu-Li Hsueh, M. S. Rogge, S. Yamamoto, and L. G. Kazovsky, "A highly flexible and efficient passive optical network employing dynamic wavelength allocation," *Journal of Lightwave Technology*, vol. 23, no. 1, pp. 277-286, 2005.
- [49] M. Maier, M. Herzog, and M. Reisslein, "STARGATE: The next evolutionary step toward unleashing the potential of WDM EPONs," *IEEE Communications Magazine*, vol. 45, no. 5, pp. 50-56, 2007.
- [50] Huan Song, A. Banerjee, and B. Mukherjee, "A protocol for efficient tunable laser utilization to support incremental upgrade in a WDM-PON," *Proc. OFC'07*, 2007.
- [51] Fu-Tai An, Kyeong Soo Kim, D. Gutierrez, S. Yam, E. Hu, K. Shrikhande, and L. G. Kazovsky, "SUCCESS: a next-generation hybrid WDM/TDM optical access network architecture," *Journal of Lightwave Technology*, vol. 22, no. 11, pp. 2557-2569, 2004.
- [52] Ho Young Hwang, Seong Joon Kim, Dan Keun Sung, and Nah-Oak Song, "Performance Analysis of IEEE 802.11e EDCA With a Virtual Collision Handler," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 2, pp. 1293-1297, 2008.
- [53] D. Wu, G. Zhu, G. Liu, W. Yuan, and M. Chen, "Resolving collision and delay deterioration by a novel channel collision based adaptive backoff (CCBAB) scheme in wireless LANs," *Proc. 32th IEEE Conference on Local Computer Networks*, 2007.
- [54] K. A. Meerja and A. Shami, "A Collision Avoidance Mechanism for IEEE 802.11e EDCA Protocol to Improve Voice Transmissions in Wireless Local Area Networks," *Proc. GLOBECOM'07*, 2007.
- [55] R. G. Garroppo, S. Giordano, S. Lucetti, and L. Tavanti, "A model-based admission control for IEEE 802.11e networks," *Proc. ICC'07*, 2007.
- [56] J. Liu and Z. Niu, "A dynamic admission control scheme for QoS supporting in IEEE 802.11e EDCA," *Proc. WCNC'07*, 2007.
- [57] X. Chen, H. Zhai, X. Tian, and Y. Fang, "Supporting QoS in IEEE 802.11e wireless LANs," *IEEE Transactions on Wireless Communications*, vol. 5, no. 8, pp. 2217-2227, 2006.
- [58] Q. Ni, "Performance analysis and enhancements for IEEE 802.11e wireless networks," *IEEE Network*, vol. 19, no. 4, pp. 21-27, 2005.
- [59] Hyunmin Kyung, Sangho Seo, and Sin-Chong Park, "Performance evaluation of enhanced distributed contention control (EDCC) mechanism in IEEE 802.11e," *Proc. ICC'05*, 2005.

- [60] A. Nafaa, A. Ksentini, and A. Mehaoua, "Sliding contention window for efficient service differentiation in IEEE 802.11 networks," *Proc. WCNC'05*, 2005.
- [61] Y. Xiao and H. Li, "Evaluation of distributed admission control for the IEEE 802.11e EDCA," *IEEE Communications Magazine*, vol. 42, no. 9, p. S20-S24, 2004.
- [62] Kyoung-Ju Noh, Woo-Yong Choi, and Sok-kyu Lee, "Adaptive and dynamic tuning of the operation parameter value for QoS and fairness in wireless LAN," *Proc. VTC'04*, 2004.
- [63] V. Scarpa, G. Convertino, A. Palmieri, and J. C. De Martin, "Adaptive techniques to guarantee QoS in a IEEE 802.11 wireless LAN," *Proc. VTC'04*, 2004.
- [64] J. Wall and J. Y. Khan, "An adaptive ARQ enhancement to support multimedia traffic using 802.11 wireless LANs," *Proc. GLOBECOM'04*, 2004.
- [65] R. Romdhani, Q. Ni, and T. Turletti, "Adaptive EDCF: enhanced service differentiation for IEEE 802.11 wireless ad hoc networks," *Proc. WCNC'03*, 2003.
- [66] D. Pong and T. Moors, "Call admission control for IEEE 802.11 contention access mechanism," *Proc. GLOBECOM'03*, 2003.
- [67] M. M. Rashid, E. Hossain, and V. K. Bhargava, "Controlled channel access scheduling for guaranteed QoS in 802.11e-based WLANs," *IEEE Transactions on Wireless Communications*, vol. 7, no. 4, pp. 1287-1297, 2008.
- [68] D. Gao, J. Cai, and C. W. Chen, "Admission control based on rate-variance envelop for VBR traffic Over IEEE 802.11e HCCA WLANs," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 3, pp. 1778-1788, 2008.
- [69] G. Boggia, P. Camarda, L. A. Grieco, and S. Mascolo, "Feedback-based control for providing real-time services with the 802.11e MAC," *IEEE/ACM Transactions on Networking*, vol. 15, no. 2, pp. 323-333, 2007.
- [70] Y. Higuchi, A. Foronda, C. Ohta, M. Yoshimoto, and Y. Okada, "Delay guarantee and service interval optimization for HCCA in IEEE 802.11e WLANs," *Proc. WCNC'07*, 2007.
- [71] Y.-W. Huang, T.-H. Lee, and J.-R. Hsieh, "Gaussian approximation based admission control for variable bit rate traffic in IEEE 802.11e WLANs," *Proc. WCNC'07*, 2007.
- [72] I. Inan, F. Keceli, and E. Ayanoglu, "An adaptive multimedia QoS scheduler for 802.11e wireless LANs," *Proc. ICC'06*, 2006.
- [73] P. Ansel, Q. Ni, and T. Turletti, "An efficient scheduling scheme for IEEE 802.11e," *Proc. Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks*, 2004.
- [74] A. Annese, G. Boggia, P. Camarda, L. A. Grieco, and S. Mascolo, "Providing delay guarantees in IEEE 802.11e networks," *Proc. IEEE VTC Spring*, 2004.
- [75] A. Grilo, M. Macedo, and M. Nunes, "A scheduling algorithm for QoS support in IEEE 802.11e networks," *IEEE Wireless Communications*, vol. 10, no. 3, pp. 36-43, 2003.
- [76] D. J. Xue, Y. Qin, and C. K. Siew, "Performance analysis of a novel traffic scheduling algorithm in slotted optical networks," *Computer Communications*, vol. 30, no. 18, pp. 3559-3571, 2007.

- [77] J.-Y. Le Boudec and P. Thiran, *Network Calculus - A Theory of Deterministic Queueing Systems for the Internet*, Springer Verlag, LNCS 2050, 2004.
- [78] C. K. Siew, D. J. Xue, and Y. Qin, "Providing deterministic quality of service in slotted optical networks," *Optics Express*, vol. 14, no. 26, pp. 12679-12692, 2006.
- [79] D. J. Xue, Y. Qin, and C. K. Siew, "Deterministic QoS provisioning with network calculus based admission control in WDM EPON networks," *Proc. ICC'09*, 2009.
- [80] D. J. Xue, Y. Qin, and C. K. Siew, "A service curve based scheduling algorithm for QoS support in 802.11e wireless LANs," *Proc. GLOBECOM'06*, 2006.
- [81] L. Zhang, "Virtual clock: a new traffic control algorithm for packet switching networks," *Proc. ACM SIGCOMM*, 1990.
- [82] S. J. Golestani, "A self-clocked fair queueing scheme for broadband applications," *Proc. INFOCOM'94*, 1994.
- [83] A. K. Parekh and R. G. Gallager, "A generalized processor sharing approach to flow control: the single node case," *Proc. INFOCOM'92*, 1992.
- [84] M. Shreedhar and G. Varghese, "Efficient fair queueing using deficit round-robin," *IEEE/ACM Transactions on Networking*, vol. 4, no. 3, pp. 375-385, 1996.
- [85] J. C. R. Bennett and Hui Zhang, "WF²Q: worst-case fair weighted fair queueing," *Proc. INFOCOM'96*, 1996.
- [86] D. Pan and Y. Yang, "Credit based fair scheduling for packet switched networks," *Proc. INFOCOM'05*, 2005.
- [87] A. G. P. Rahbar and O. Yang, "Loan-grant based round robin scheduling," *Proc. Communication Networks and Services Research Conference (CNSR)*, 2006.
- [88] A. G. P. Rahbar and O. Yang, "OCGRR: A new scheduling algorithm for differentiated services networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 18, no. 5, pp. 697-709, 2007.
- [89] W. Wang, Libin Dong, and W. Wolf, "iSKIP: a fair and efficient scheduling algorithm for input-queued crossbar switches," *Proc. GLOBECOM'02*, 2002.
- [90] N. McKeown, "The iSLIP scheduling algorithm for input-queued switches," *IEEE/ACM Transactions on Networking*, vol. 7, no. 2, pp. 188-201, 1999.
- [91] K. Ross and N. Bambos, "Dynamic quality of service control in packet switch scheduling," *Proc. ICC'05*, 2005.
- [92] K. Ross and N. Bambos, "Optimizing quality of service in packet switch scheduling," *Proc. ICC'04*, 2004.
- [93] T. H. Szymanski, "A QoS switch scheduling algorithm based on recursive fair stochastic matrix decomposition," *Workshop on High Performance Switching and Routing*, 2006.
- [94] C. Chang, W. Chen, and H. Huang, "Birkhoff-von Neumann input buffered crossbar switches," *Proc. INFOCOM'00*, 2000.

- [95] J. Von Neumann, "A certain zero-sum two-person game equivalent to the optimal assignment problem," *Contributions to the Theory of Games*, vol. 2, no. 28, pp. 5-12, 1953.
- [96] N. Kahale and P. Wright, "Dynamic global packet routing in wireless networks," *Proc. INFOCOM'97*, 1997.
- [97] Y. Qin, D. J. Xue, C. K. Siew, I. Saniee, and I. Widjaja, "A topology based dynamic traffic scheduling in time-domain wavelength interleaved networks," *Proc. 12th IEEE International Conference on Networks*, 2004.
- [98] F. Baccelli, G. Cohen, G. Oslder, and J. Quadrat, *Synchronization and linearity: an algebra for discrete event systems*, New York: Wiley, 1992.
- [99] H. Lui and H. T. Mouftah, "A new absolute QoS differentiation scheme supporting best-effort class in OBS networks," *Proc. 9th International Conference on Transparent Optical Networks*, 2007.
- [100] S. Kim, J. S. Choi, and M. Kang, "Providing absolute differentiated services for optical burst switching networks: loss differentiation," *IEE Proc. Communications*, vol. 152, no. 4, pp. 439-446, 2005.
- [101] Q. Zhang, V. M. Vokkarane, J. P. Jue, and B. Chen, "Absolute QoS differentiation in optical burst-switched networks," *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 9, pp. 1781-1795, 2004.
- [102] C. Semeria, *Supporting differentiated service classes: queue scheduling disciplines*, Juniper Networks, Inc. white paper, 2002.
- [103] J. Kani, M. Teshima, K. Akimoto, N. Takachio, H. Suzuki, K. Iwatsuki, and M. Ishii, "A WDM-based optical access network for wide-area gigabit access services," *IEEE Communications Magazine*, vol. 42, no. 2, p. S43-S48, 2003.
- [104] F. Davik et al., "IEEE 802.17 resilient packet ring tutorial," *IEEE Communications Magazine*, vol. 42, no. 3, pp. 112-118, 2004.
- [105] Yu-Li Hsueh, M. S. Rogge, Wei-Tao Shaw, S. Yamamoto, and L. G. Kazovsky, "Quality of service support over SUCCESS-DWA: a highly evolutionary and cost-effective optical access network," *Proc. OFC'05*, 2005.
- [106] Kae Hsiang Kwong, D. Harle, and I. Andonovic, "Dynamic bandwidth allocation algorithm for differentiated services over WDM EPONs," *Proc. 9th International Conference on Communications Systems*, 2004.
- [107] F. Clarke, S. Sarkar, and B. Mukherjee, "Simultaneous and interleaved polling: an upstream protocol for WDM-PON," *Proc. OFC'06*, 2006.
- [108] A. R. Dhaini, C. M. Assi, M. Maier, and A. Shami, "Dynamic wavelength and bandwidth allocation in hybrid TDM/WDM EPON networks," *Journal of Lightwave Technology*, vol. 25, no. 1, pp. 277-286, 2007.
- [109] R. L. Cruz, "Quality of service guarantees in virtual circuit switched networks," *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 6, pp. 1048-1056, 1995.
- [110] R. L. Cruz, "A calculus for network delay. I. Network elements in isolation," *IEEE Transaction on Information Theory*, vol. 37, no. 1, pp. 114-131, 1991.

- [111] R. L. Cruz, "A calculus for network delay. II. Network analysis," *IEEE Transaction on Information Theory*, vol. 37, no. 1, pp. 132-141, 1991.
- [112] H. Sariowan, "A service curve approach to performance guarantees in integrated service networks." Ph.D. Thesis, Univ. Calif. San Diego, 1996.
- [113] S. Shenker and J. Wroclawski, *General Characterization Parameters for Integrated Services Network Elements*, RFC 2215, 1997.
- [114] P. T. Puqi and T.-Y. C. Tai, "Network traffic characterization using token bucket model," *Proc. INFOCOM'99*, 1999.
- [115] L. Kleinrock, *Queueing systems volume I: theory*, New York: Wiley, 1976, pp. 189-191.
- [116] H. Sariowan, R. L. Cruz, and G. C. Polyzos, "SCED: a generalized scheduling policy for guaranteeing quality-of-service," *IEEE/ACM Transactions on Networking*, vol. 7, no. 5, pp. 669-684, 1999.
- [117] I. Stoica, H. Zhang, and T. S. E. Ng, "A hierarchical fair service curve algorithm for link-sharing, real-time, and priority services," *IEEE/ACM Transactions on Networking*, vol. 8, no. 2, pp. 185-199, 2000.
- [118] J.-Y. Le Boudec and O. Verscheure, "Optimal smoothing for guaranteed service," *IEEE/ACM Transactions on Networking*, vol. 8, no. 6, pp. 689-696, 2000.
- [119] R. Agrawal, R. L. Cruz, C. Okino, and R. Rajan, "Performance bounds for flow control protocols," *IEEE/ACM Transactions on Networking*, vol. 7, no. 3, pp. 310-323, 1999.
- [120] C. K. Siew, G. Feng, F. Long, and M. H. Er, "Congestion control based on flow-state-dependent dynamic priority scheduling," *IEE Proc. Communications*, vol. 152, no. 5, pp. 548-558, 2005.
- [121] C. K. Siew, T. Cui, and M. Zukerman, "Adaptive video multicast using the FDPS algorithm," *Proc. International Conference on Information, Communications & Signal Processing*, 2007.
- [122] Lun Tong and P. Ramanathan, "Adaptive power and rate allocation for service curve assurance in DS-CDMA network," *IEEE Transactions on Wireless Communications*, vol. 3, no. 2, pp. 555-564, 2004.
- [123] R. Leelahakriengkrai and R. Agrawal, "Scheduling in multimedia CDMA wireless networks," *IEEE Transactions on Vehicular Technology*, vol. 52, no. 1, pp. 226-239, 2003.
- [124] J. Tan, G. Feng, C. K. Siew, and L. Zhang, "Providing differentiated services over shared wireless downlink through buffer management," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 1, pp. 548-555, 2008.
- [125] D. J. Xue, Y. Qin, and C. K. Siew, "Virtual topology-based traffic scheduling algorithm for slotted optical networks," *Journal of Optical Networking*, vol. 6, no. 2, pp. 166-179, 2007.
- [126] C. K. Siew and M. H. Er, "Multiservice provisioning mechanism with service curves assurance for per-class scheduling delay guarantees," *IEE Proceedings Communications*, vol. 153, no. 6, pp. 846-855, 2006.

- [127] B. Teitelbaum et al., "QBone architecture (v1.0)," *Internet 2 QoS Working Group Draft*, 1999.
- [128] H. A. Mantar, Junseok Hwang, I. T. Okumus, and S. J. Chapin, "A scalable model for interbandwidth broker resource reservation and provisioning," *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 10, pp. 2019-2034, 2004.
- [129] M. Chen and T.-S. Yum, "A conflict-free protocol for optical WDMA networks," *Proc. GLOBECOM'91*, 1991.
- [130] I. Saniee and I. Widjaja, "Simplified layering and flexible bandwidth with TWIN," *Proc. Workshop on Future Directions in Network Architecture, SIGComm*, 2004.
- [131] J. Zheng, V. O. K. Li, and X.-C. Yuan, "An adaptive flow classification algorithm for IP switching," *Proc. GLOBECOM'99*, 1999.
- [132] K. Yasukawa, K. Baba, and K. Yamaoka, "Dynamic class assignment for stream flows consideration characteristics of non-stream flow classes," *IEICE Transactions on Communications*, vol. E87-B, no. 11, pp. 3242-3254, 2004.
- [133] W. Wang and C. C. Shen, "An adaptive flow classification scheme for data-driven label switching networks," *Proc. ICC'01*, 2001.