Multiple access schemes for Ethernet Passive Optical Networks (EPONS)

Zhu, Yongqing

2006

Zhu, Y. (2006). Multiple access schemes for Ethernet Passive Optical Networks (EPONS).
Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/3435>

<https://doi.org/10.32657/10356/3435>

Nanyang Technological University

Downloaded on 26 Apr 2023 17:46:27 SGT

Multiple Access Schemes for Ethernet Passive Optical Networks (EPONs)

Zhu Yongqing

School of Electrical & Electronic Engineering

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

2006

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research and has not been submitted for a higher degree to any other University or Institution.

Date

Zhu Yongqing

Acknowledgement

I would like to express my deepest and sincerest appreciation to my supervisor, A/P Ma Maode, for his thorough supervision of this dissertation as well as the continuous guidance, support, and encouragement throughout my studies at Nanyang Technological University. I also gratefully acknowledge many other professors and researchers who have encouraged and helped me to complete this work, especially A/P Cheng Tee Hiang, A/P Shum Ping, A/P Lu Chao, and Ms. Xie Jing.

Moreover, I greatly appreciate all members of Communications Lab IV, with whom I have spent all the pleasant time during my PhD study. Special thanks I owe to Huang Xiaohong, Hou Zhihua, Leng Supeng and Zhang Yan for their technical help and endless encouragement, and to Luo Xingfang, Xiao Shi, Wang Wen, Zhang Keli, and Wang Haibo for their warmest friendship.

My boundless gratitude is devoted to my beloved husband for his everlasting support and accompaniment during the PhD study, and my parents and parents-in-law who take care of my little son in their tireless efforts. I am also highly thankful to my brother and sisters-in-law for their generous help and constant encouragement.

Finally I would like to dedicate this dissertation to my sweetest lovely son, whom I have been missing so much.

Table of Contents

Acknowledgement.....	i
Table of Contents	ii
Summary.....	v
List of Figures.....	vii
List of Tables	xi
List of Abbreviations	xii
List of Symbols	xvii
Chapter 1 . Introduction	1
1.1 Background.....	1
1.2 Motivation.....	4
1.3 Objectives	6
1.4 Methodology	7
1.5 Main Contribution of the Thesis.....	8
1.6 Organization of the Thesis	12
Chapter 2 . Literature Review	14
2.1 Passive Optical Networks (PONs).....	15
2.1.1 PON Topologies.....	15
2.1.2 Primary PON Solutions.....	16
2.1.3 ATM PON (APON)	17
2.1.4 Ethernet PON (EPON).....	21
2.1.5 Multi-Point Control Protocol (MPCP).....	27
2.2 Survey of the Centralized DBA Schemes for EPONs	30
2.2.1 DBA Schemes with Statistical Multiplexing	32
2.2.2 DBA Schemes with QoS Assurances.....	36
2.3 Summary.....	48
Chapter 3 . A Proposal to Enhance the IPACT Scheme.....	49
3.1 System Model	50

3.2 Review of The IPACT Scheme	52
3.2.1 Interleaved Polling Algorithm	52
3.2.2 Supporting of the Differentiated Services (DiffServ)	56
3.3 The Proposed IPACT with Grant Estimation (IPACT-GE) Scheme.....	57
3.3.1 Weakness of the IPACT Scheme	58
3.3.2 Principle of the IPACT-GE Scheme	61
3.3.3 Advantages of the IPACT-GE Scheme.....	65
3.4 Simulation Experiments.....	71
3.4.1 Simulation Model.....	71
3.4.2 Comparison of the IPACT-GE, IPACT, and EB-DBA Scheme.....	72
3.4.3 Comparison of IPACT-GE and IPACT Combined with the SPQ Mechanism.....	79
3.5 Summary	82
Chapter 4 . A Systematic Multiple Access Scheme.....	84
4.1 Network Architecture and System Model	86
4.2 Bandwidth Guaranteed Polling (BGP) Scheme	87
4.2.1 Principle of the BGP Scheduling Scheme	88
4.2.2 Control Message Transmissions	90
4.2.3 Operations of the OLT and ONUs	91
4.2.4 Implementation of the BGP Scheme.....	94
4.2.5 Even Distribution Algorithm (EDA)	95
4.3 Issues of the Bandwidth Utilization.....	97
4.3.1. Control Message Overhead.....	98
4.3.2. Guard Time Overhead.....	100
4.3.3. Remainder Window Overhead.....	101
4.3.4. Discussion	102
4.4 Performance Estimation.....	103
4.4.1 Average Cycle Time	104
4.4.2 Average Waiting Delay $E(W_1)$	107
4.4.3 Average Waiting Delay $E(W_2)$	109
4.4.4 Summary	111
4.5 A Continuous-Time Markov Chains (CTMC) Model	112
4.5.1 Model Description	113
4.5.2 Markov Chain State Probabilities	116
4.5.3 Conditional Cycle Time Approximation.....	118
4.5.4 Recursive Algorithm for the CTMC Model.....	119
4.6 BGP-Based MAC Protocol.....	122
4.6.1 Parameter-based CAC Mechanism	123

4.6.2 Implementation of the BGP-Based MAC Protocol.....	126
4.7 Simulation Experiments.....	128
4.7.1 Experiments on the BGP Scheme	129
4.7.2 Experiments on the BGP-Based MAC Protocol	139
4.8 Summary	145
Chapter 5 . Proposed Solutions to Support DiffServ	147
5.1 Hierarchical Scheduling.....	148
5.1.1 Inter-ONU Scheduling	149
5.1.2 Intra-ONU Scheduling	154
5.2 Solution with the SPQ Scheme.....	156
5.2.1 SPQ Algorithm.....	156
5.2.2 Benefits to the Light Traffic Load	159
5.2.3 Simulation Experiments.....	161
5.3 Solution with the TSQ Scheme.....	166
5.3.1 TSQ Algorithm	167
5.3.2 Benefits to the Lower-Priority Traffic	169
5.3.3 Simulation Experiments.....	170
5.4 Solution with the UFQ Scheme	178
5.4.1 UFQ Algorithm.....	179
5.4.2 Simulation Experiments.....	185
5.5 Summary	193
Chapter 6 . Conclusions.....	195
6.1 Summary.....	195
6.2 Recommendation to the Future Research	199
6.2.1 New DBA Schemes for EPON Networks.....	199
6.2.2 Network Management for EPON Networks	200
6.2.3 Adapting EPON to Wireless Access Networks	202
Author's Publications	204
Bibliography	206

Summary

The long-haul backbone networks have experienced substantial growth through the use of Wavelength Division Multiplexing (WDM) and other technologies. However, the access networks have little changes in recent years, resulting in the bottleneck between the backbone networks and subscribers. Ethernet Passive Optical Network (EPON) is regarded as the best candidate for the next-generation access network to overcome the bottleneck, because it represents the convergence of the inexpensive Ethernet equipment and the high-speed fiber infrastructure. This dissertation is mainly concerned about the multiple access issue in EPON networks, where the upstream bandwidth is allocated to multiple Optical Network Units (ONUs) for transmission to the Optical Line Terminal (OLT).

We have first proposed an enhancement to the famous Dynamic Bandwidth Allocation (DBA) scheme, Interleaved Polling with Adaptive Cycle Time (IPACT). Our proposed scheme, IPACT with Grant Estimation (IPACT-GE), considers the estimated amount of new packets arriving between pollings when deciding the grant size. Since the grant size with estimation is closer to the instant buffer occupancy when the ONU is polled, the proposal can achieve better performance than IPACT with regard to the average waiting delay and buffer occupancy. Besides, IPACT-GE can greatly mitigate the light-load penalty when it is combined with the Strict Priority Queuing (SPQ) mechanism to support the Differentiated Service (Diffserv), while the penalty is an obvious phenomenon in the solution combining IPACT with SPQ.

Then a novel polling-based DBA scheme, Bandwidth Guaranteed Polling (BGP), has been proposed to arbitrate EPONs upstream transmissions efficiently. According to the bandwidth requirement specified in the Service Level Agreement (SLA) contracts, ONUs are

allocated with different bandwidth units and are polled by the OLT with different times in a cycle. The BGP scheme can guarantee the absolute Quality of Service (QoS) to high-demand users, while the best-effort service is provided to the low-demand users. An Even Distribution Algorithm (EDA) is devised to evenly distribute the bandwidth units belonging to the same ONU and decide the polling sequence for the BGP scheme. Based on BGP, we have further presented a parameter-based Call Admission Control (CAC) mechanism for EPONs to admit the transmission requests with various QoS requirements. Then a systematic Medium Access Control (MAC) protocol is formed consisting of the CAC, EDA, and BGP schemes that are cooperating closely.

Next, several solutions have been proposed for EPONs to support Diffserv effectively. Deploying the hierarchical scheduling architecture, all solutions adopt the systematic BGP-based MAC protocol for the inter-ONU scheduling. Various mechanisms are exploited in the intra-ONU scheduling to provide different levels of fairness in addition to the Diffserv support. First, the default SPQ mechanism has been suggested in the intra-ONU scheduling. The solution with SPQ can provide efficient Diffserv for different traffic flows, while it cannot achieve any fairness between flows. The solution with the Two-Stage Queuing (TSQ) mechanism is then proposed, which can improve the performance of the lower-priority traffic by diminishing the performance polarization, and obtain the fairness to some extent for traffic flows. At last, we have designed an efficient Urgency Fair Queuing (UFQ) scheme for the intra-ONU scheduling. The best-effort traffic in UFQ can obtain the maximum bandwidth allocation without violating the performance of the QoS traffic. The solution combining UFQ with the BGP-based MAC scheme can achieve the better fairness between the QoS traffic and best-effort traffic than the solutions with SPQ and TSQ.

Finally, we have performed the mathematical estimation and developed a Continuous-Time Markov Chains (CTMC) model to numerically evaluate the BGP scheme separately.

List of Figures

Figure 1.1 Deployment of the Optical Fiber in the Access Networks	4
Figure 2.1 Topologies of Passive Optical Networks	16
Figure 2.2 Downstream Transmissions in EPON Networks	22
Figure 2.3 Upstream Transmissions in EPON Networks	22
Figure 2.4 MPCP Auto-Discovery Procedures	28
Figure 2.5 MPCP GATE Operations	29
Figure 2.6 Individual Polling	30
Figure 2.7 Joint Polling.....	30
Figure 2.8 Category of DBA Schemes for EPONs.....	31
Figure 3.1 A typical EPON network based on tree topology	51
Figure 3.2 Interleaved Polling Algorithm in IPACT	53
Figure 3.3 Illustration of the IPACT Transmission	59
Figure 3.4 Flowchart of the ONU Operations in IPACT-GE	64
Figure 3.5 Flowchart of the OLT Operations in IPACT-GE	65
Figure 3.6 Illustration of the Packet Waiting Delay	67
Figure 3.7 Comparison of Average Delay Between IPACT-GE, IPACT and EB-DBA.....	75
Figure 3.8 Comparison of Average Delay Between IPACT-GE, IPACT and EB-DBA at the Light Load.....	75
Figure 3.9 Comparison of Average Buffer Occupancy Between IPACT-GE, IPACT and EB- DBA	76
Figure 3.10 Comparison of Average Buffer Occupancy Between IPACT-GE, IPACT and EB-DBA at the Light Load	77
Figure 3.11 Comparison of Average Loss Rate Between IPACT-GE, IPACT and EB-DBA	78

Figure 3.12 Comparison of Upstream Channel Utilization Between IPACT-GE, IPACT and EB-DBA.....	78
Figure 3.13 Comparison of Average Delay Between Solutions with IPACT-GE and IPACT	81
Figure 3.14 Comparison of Maximum Delay Between Solutions with IPACT-GE and IPACT	81
Figure 4.1 Upstream Transmissions in the Tree-Based EPON Networks.....	86
Figure 4.2 Flowchart of the ONU Operations in BGP.....	93
Figure 4.3 Flowchart of the OLT Operations in BGP	93
Figure 4.4 An Example of Entry Table with 100 Entries	95
Figure 4.5 Flowchart of the EDA Algorithm.....	97
Figure 4.6 Basic Model of the Multi-Queue Polling System with Finite Buffers	113
Figure 4.7 Flowchart of the CAC Operations During the Initialization	125
Figure 4.8 Entry Table with 100 Entries for the Simulation Experiments	130
Figure 4.9 Average Delay of Different ONUs in BGP and IPACT.....	132
Figure 4.10 Throughput of Different ONUs in BGP and IPACT.....	133
Figure 4.11 Average Buffer Occupancy of Different ONUs in BGP and IPACT.....	134
Figure 4.12 Average Loss Rate of Different ONUs in BGP and IPACT	135
Figure 4.13 Comparison of Average Delay by Estimation and Simulation for BGP	138
Figure 4.14 Results of the EDA Algorithm	143
Figure 4.15 Average Delay of the BGP-Based MAC Protocol	144
Figure 4.16 Average Loss Rate of the BGP-Based MAC Protocol.....	145
Figure 5.1 Hierarchical Scheduling Architecture in EPONs	148
Figure 5.2 Flowchart of the CAC Operations with Consideration of Individual Traffic	153
Figure 5.3 Mapping of the Traffic Classes into Priority Queues.....	155

Figure 5.4 The ONU Structure in the SPQ Scheme.....	157
Figure 5.5 Flowchart of the ONU Operations in the SPQ Scheme	158
Figure 5.6 Illustration of the Light-Load Penalty	160
Figure 5.7 Entry Table with 20 Entries for the Simulation Experiments	163
Figure 5.8 Average Delay of ONUs with 5 and 3 Entries in the SPQ Solution	164
Figure 5.9 Maximum Delay of ONUs with 5 and 3 Entries in the SPQ Solution	164
Figure 5.10 Average Delay of ONUs with 2 and 1 Entry in the SPQ Solution.....	165
Figure 5.11 Maximum Delay of ONUs with 2 and 1 Entry in the SPQ Solution.....	165
Figure 5.12 The ONU Structure in the TSQ Scheme	168
Figure 5.13 Flowchart of the ONU Operations in the TSQ Scheme	168
Figure 5.14 Average Delay of ONUs with 5 and 3 Entries in the TSQ Solution	171
Figure 5.15 Maximum Delay of ONUs with 5 and 3 Entries in the TSQ Solution	172
Figure 5.16 Average Delay of ONUs with 2 and 1 Entry in the TSQ Solution.....	172
Figure 5.17 Maximum Delay of ONUs with 2 and 1 Entry in the TSQ Solution	173
Figure 5.18 Loss Rate of BG ONUs in the TSQ Solution	173
Figure 5.19 Comparison of Average Delay for 3-Entry ONUs in TSQ and SPQ Solutions	175
Figure 5.20 Comparison of Loss Rates for 3-Entry ONUs in TSQ and SPQ Solutions.....	175
Figure 5.21 Comparison of Average Delay for 2-Entry ONUs in TSQ and SPQ Solutions	176
Figure 5.22 Comparison of Loss Rates for 2-Entry ONUs in TSQ and SPQ Solutions.....	176
Figure 5.23 Comparison of Average Delay for 1-Entry ONUs in TSQ and SPQ Solutions	177
Figure 5.24 Comparison of Loss Rates for 1-Entry ONUs in TSQ and SPQ Solutions.....	177
Figure 5.25 The ONU Structure in the UFQ Scheme	179
Figure 5.26 Comparison of Average Delay for 2-Entry ONUs in UFQ and TSQ Solutions	187
Figure 5.27 Comparison of Average Delay for 1-Entry ONUs in UFQ and TSQ Solutions	188
Figure 5.28 Comparison of Loss Rate 1 for 5-Entry ONUs in UFQ and TSQ Solutions.....	189

Figure 5.29 Comparison of Loss Rate 1 for 3-Entry ONUs in UFQ and TSQ Solutions.....	189
Figure 5.30 Comparison of Loss Rate 1 for 2-Entry ONUs in UFQ and TSQ Solutions.....	190
Figure 5.31 Comparison of Integrated Loss Rate for 2-Entry ONUs in UFQ and TSQ Solutions	191
Figure 5.32 Comparison of Loss Rate 1 for 1-Entry ONUs in UFQ and TSQ Solutions.....	192
Figure 5.33 Comparison of Integrated Loss Rate for 1-Entry ONUs in UFQ and TSQ Solutions	192

List of Tables

Table 2.1 Summary of centralized TDMA-based DBA Schemes in EPONs	47
Table 4.1 Original ONUs Parameters for the BGP-Based MAC Protocol	139
Table 4.2 Results of the CAC Mechanism.....	141

List of Abbreviations

AF	Assured Forwarding
APON	Asynchronous Transfer Mode Passive Optical Network
ATM	Asynchronous Transfer Mode
BE	Best Effort
BP	Burst-Polling
BG	Bandwidth Guaranteed
BGP	Bandwidth Guaranteed Polling
BPON	Broadband Passive Optical Network
CAC	Call Admission Control
CBR	Constant Bit Rate
CBRC	Constant-Bit-Rate Credit
CDMA	Code Division Multiple Access
CDV	Cell Delay Variation
CO	Central Office
CPE	Customer Premise Equipment
CSMA/CD	Carrier Sense Multiple Access with Collision Detection
CTMC	Continuous-Time Markov Chains
DBA	Dynamic Bandwidth Allocation
D-CRED	Dynamic Credit Distribution
DDBA	Delta Dynamic Bandwidth Allocation
DEB	Deterministic Effective Bandwidth
DiffServ	Differentiated Services

DPOA	Dynamic Polling Order Arrangement
DSL	Digital Subscriber Line
DSLAM	DSL Access Multiplexer
EB-DBA	Estimation-Based DBA
EDA	Even Distribution Algorithm
EF	Expedited Forwarding
EFM	Ethernet in the First Mile
EMS	Element Management System
EOT	End of Transmission
EPON	Ethernet Passive Optical Network
ESC	Escape Code
FCAPS	Fault Configuration Accounting Performance and Security
FIFO	First In First Out
FPS	Full Priority Scheduling
FQSE	Fair Queuing with Service Envelopes
FSAN	Full Service Access Network
FTTB	Fiber-To-The-Building
FTTC	Fiber-To-The-Curb
FTTH	Fiber-To-The-Home
FULL-RCMA	Full Utilization Local Loop Request Contention Multiple Access
GEM	GPON Encapsulation Method
GPON	Gigabit Passive Optical Network
GPS	Generalized Processor Sharing
HFC	Hybrid Fiber Coax
HOL	Head-Of-Line

HSDC	Hybrid-Slot Decentralized Control
HSSR	Hybrid Slot Size/Rate
ID	Identification
IP	Internet Protocol
IPACT	Interleaved Polling with Adaptive Cycle Time
IPACT-GE	IPACT with Grant Estimation
IPS	Interval Priority Scheduling
ITU-T	International Telecommunication Union-Telecommunication
LAN	Local Area Network
LEC	Local Exchange Carrier
LRD	Long-Range Dependencies
LST	Laplace Stieltjes Transform
MAC	Medium Access Control
MAN	Metropolitan Area Network
MPCP	Multi-Point Control Protocol
MTW	Maximum Transmission Window
MQ-FIFO	Multiple Queue First In First Out
NID	Node Identification
NMS	Network Management System
Non-BG	Bandwidth Non-Guaranteed
OLT	Optical Line Terminal
ONU	Optical Network Unit
OOC	Optical Orthogonal Code
OSI	Open Systems Interconnection
PDF	Probability Distribution Function

PHB	Per-Hop Behavior
PIS	Priority with Insertion Scheduling
PON	Passive Optical Network
POTS	Plain Old Telephone Service
QoS	Quality of Service
QR	Queue Report
RTT	Round-Trip Time
SC	Star Coupler
SE	Service Envelope
SFQ	Start-time Fair Queuing
SLA	Service Level Agreement
SMS	Service Management System
SPQ	Strict Priority Queuing
T-Cont	Traffic Container
TDMA	Time Division Multiple Access
TLBA	Two-Layer Bandwidth Allocation
TSQ	Two-Stage Queuing
TUR	Transmission Upon Reception
UFQ	Urgency Fair Queuing
VLAN	Virtual Local Area Network
VoD	Video on Demand
WAN	Wide Area Network
WDM	Wavelength Division Multiplexing
WDM IPACT-ST	WDM IPACT with a Single polling Table
WFBA	Weighted Fair Buffer Allocation

WFQ	Weighted Fair Queuing
WG	Working Group
WRR	Weighted Round Robin
WS	Window Size

List of Symbols

A	packet access delay
a_k	packet arrival time
$a_m(t)$	arrival process with Poisson distribution
B	minimum peak bandwidth requirement
B^*	bandwidth required to guarantee delay bound requirement
B_i	minimum bandwidth requirement for ONU i
B_i^*	bandwidth required to guarantee delay bound for ONU i
b_j	block probability of queue j
b_m', b_m''	conditional arrival probability with m arrivals
$b_{m,j}', b_{m,j}''$	conditional arrival probability with m arrivals of queue j
$b_{m,j}^{**}, b_{m,j}^{''*}$	conditional arrival probability with m arrivals of queue j during the backward recurrence time
C	channel transmission rate of the EPON system
$c_{c,j}', c_{c,j}''$	coefficient of variation for conditional cycle time of queue j
D	maximum waiting delay
D_i	maximum waiting delay accepted by ONU i
$D(n)$	difference between the grant size and ONU's buffer occupancy at the n^{th} cycle
d	packet waiting delay in the ONU
$d1, d2, d3$	different components of packet waiting delay
$\bar{d}$	average packet waiting delay in the ONU

$\overline{d1}, \overline{d2}, \overline{d3}$	average value of different components of packet waiting delay
$\overline{d}_{diff}$	difference of ONU's average delays in IPACT and IPACT-GE
d_k	delay bound requirement of packet k
$E(n)$	estimated amount of packets arriving during the n^{th} cycle
$E_i[j]$	the j^{th} entry number of BG ONU i
e	convergence indicator
$F(t)$	Probability Distribution Function
$f_c(t), f_{c''}(t)$	probability distribution of conditional cycle time
$f(x)$	probability distribution of ONU's transmission amount
G	granted window size
$G(n)$	granted window size at the n^{th} cycle
G_i	granted window size to ONU i
$G_i(n)$	granted window size to ONU i in the n^{th} cycle
G_T	guard time adopted in the UFQ scheme
g	number of queues in the model of multi-queue polling system
L_B	ONU buffer size
L_{diff}	difference of ONU's buffer occupancies in IPACT and IPACT-GE
L_{IPACT}	buffer length of an ONU in IPACT
$L_{IPACT-GE}$	buffer length of an ONU in IPACT-GE
L_k	buffer occupancy when ONU k sends out the REPORT message
L_k'	buffer occupancy when ONU k receives the GATE message
L_o	ONU's buffer occupancy
l_i	priority queue i occupancy in the ONU
M	number of ONUs in the EPON system
m'	number of non-BG ONUs

m_n	number of BG ONUs with n entries
N	total number of entries in the Entry Table
N'	number of polling times in a cycle
N_i	total number of entries for BG ONU i
n	number of entries assigned to a BG ONU
$\bar{n}$	average number of entries assigned to a non-BG ONU
O_{guard}	overhead of guard time
$O_{remainder}$	overhead of remainder transmission window
O_{up}	overhead of upstream message transmission
P	number of entries used to poll a second ONU
$\bar{P}$	average number of entries used to poll a second ONU
P_0	traffic stream with the highest priority
P_i	traffic stream with priority i
P_n	traffic stream with the lowest priority
P_{jk}	Markov chain transition probability from state j to state k
$P_{k,j}^{(n)}$	Markov chain state probability of queue j at state k in the n^{th} polling epoch
$P_{k,j}$	Markov chain steady state probabilities of queue j at state k
$P_k^{(n)}(t)$	Markov chain state probability at state k in the n^{th} polling epoch expressed by Poisson arrival distribution
$P_k^{(n)}$	Markov chain state probability at state k in the n^{th} polling epoch expressed by the conditional arrival probabilities
$P_{k,j}^*$	arbitrary time state probabilities of queue j at state k
p_l	pointer to implement BGP and point to entries in the Entry Table

p_2	pointer to implement BGP and point to non-BG ONUs in the List
p_n	probability that one entry can be used to poll two ONUs
Q_0	priority queue in the buffer with the highest priority
Q_i	priority queue in the buffer with priority i
Q_n	priority queue in the buffer with the lowest priority
$q(n)$	queue length in the n^{th} polling cycle
$R1$	average remainder size of the entry used to poll only one ONU
$R2$	average remainder size of the entry used to poll two ONUs
$R(n)$	request size of ONU i (equal to the instant buffer occupancy)
$R_i(n)$	buffer occupancy of ONU i reported by the REQUEST message in the n^{th} cycle
$r(n)$	estimated real-time arrival rate for the n^{th} cycle
$\overline{r(n)}$	average arrival rate of the ONU in the n^{th} cycle
$S(t)$	system current time
$S_{\Delta}(n)$	packet size arriving during a small interval ΔT in the n^{th} cycle
S_j	buffer size of queue j in the model of multi-queue polling system
s_k	size of packet k
T	threshold determining whether or not to use the superfluous transmission window to poll a second ONU
$T[N]$	array of the Entry Table in the EDA algorithm
$T_c(n)$	scan time in the n^{th} polling cycle
$\overline{T_c}$	average scan time of a polling cycle
$T_{c,j}$	the cycle time of queue j
$T'_{c,j}$	conditional cycle time of queue j when it is empty in the previous

	scanning period
$T_{c,j}^*$	conditional cycle time of queue j when it is not empty in the previous scanning period
$T_{E,j}$	time interval between the polling epochs of queue j and queue $(j+1)$
$T_i^G(n)$	GRANT message arriving time of ONU i in the n^{th} cycle
$T_i^R(n)$	REQUEST message generation time of ONU i in the n^{th} cycle
T_{hj}	message service time in the model of multi-queue polling system
T_{uj}	switchover time in the model of multi-queue polling system
$\bar{T}_{w,j}$	mean waiting delay of a packet in queue j
ΔT	small time interval to estimate the arrival rate in IPACT-GE
t	total data transmission time in a polling cycle
$\bar{t}$	average transmission time for a packet
t_c	polling cycle time
$\bar{t}_c$	average cycle time
t_g	guard time between any pair of ONUs
t_{msg}	transmission time of a control message
t_n	time of the n^{th} polling epoch
TC_i	time period between two consecutive times that ONU i is polled
U	maximum bandwidth the user can afford to pay
$U_k^j(t)$	Urgency Parameter of packet k in QoS traffic j
$v_k(t)$	lingering time of packet k in the ONU
W_1	packet waiting delay while other ONUs are being polled
W_2	packet waiting delay while the ONU is being polled and

	transmitting earlier packets
$W[i]$	transmission window size requested by ONU i
W_{max}	maximum transmission window size
W_{non}	all window size for non-BG ONUs
$\overline{W_{non}}$	average window size for a non-BG ONU
WT	walk time for transmitting the control messages in a polling cycle
WT_{down}	downstream walk time, the propagation time for polling message traveling from OLT to the ONU
WT_{up}	upstream walk time, the propagation time for polling message traveling from the ONU to OLT
$X^{(n)}$	system state of the queue in the n^{th} polling epoch
α	estimation factor adopted in the IPACT-GE scheme
β	control gain adopted in the EB-DBA scheme
λ	system arrival rate
λ_i	arrival rate of ONU i
$\lambda(n)$	arrival packet size during the $[n, n+1)$ cycle
μ	channel capacity in the EPON system
ρ	system traffic load
ρ'	traffic load for one kind of ONUs
ρ_e	effective traffic load
ρ_i	traffic load of ONU i
ρ_s	saturate traffic load
$\overline{\tau}$	average propagation delay from ONU to OLT

Chapter 1 . Introduction

1.1 Background

Although the access network has experienced little changes in recent years, bandwidth in the backbone network is increasing dramatically through the use of Wavelength Division Multiplexing (WDM) and other new technologies. At the same time, Local Area Networks (LANs) have grown up from 10Mb/s to 100Mb/s then to 1Gb/s; even 10Gb/s speed is available for the residential subscribers now. The result is a growing gulf between high-capacity LANs and backbone networks with the bottleneck of access networks. Once called the “last mile”, the access network between the backbone and residential subscribers has been renamed as the “first mile” to express its importance. Because the subscribers have an increasing demand for the Internet traffic with a variety of types, a powerful technology is needed for the “first mile” to provide broadband access to the backbone. It is also expected to be inexpensive, simple, scalable, and capable of delivering integrated voice, data, and video services to end users over a single network.

Two most widely deployed broadband access solutions today are Digital Subscriber Line (DSL) and Hybrid Fiber Coax (HFC) networks [1], which depend on the reuse of the existing infrastructure. Using the twisted pairs as the transmission medium, DSL is deployed

mainly by the traditional Plain Old Telephone Service (POTS) providers. The point-to-point configuration requires a DSL modem at the customer premises and a DSL Access Multiplexer (DSLAM) in the Central Office (CO). However, with the typical data rate of 128 kb/s~1.5 Mb/s, DSL is hardly to be considered broadband to support integrated voice, data and video applications. In addition, the distance that a CO can cover with DSL is limited to less than 5.5 km, which only covers approximately 60% [1] of the potential subscribers.

HFC networks are preferred by cable television providers to deliver the data service together with the television signals over the existing CATV infrastructure. HFC combines the optical fiber and coaxial cable in the transmission path with the point-to-multipoint configuration. In the curbside optical node, the downstream optical signal is converted into electrical signal and is further transmitted on the coaxial tree. For the upstream transmission, multiple Customer Premise Equipments (CPEs) are connected to the same HFC tree to share the common sections of coaxial cable. The drawback is that each optical node has less than 36Mbps effective data throughput in the HFC network, which is shared by up to 2000 users [1]. The resulting slow speed is unable to provide enough bandwidth for the emerging services like Video on Demand (VoD), interactive gaming and two-way video conferencing.

Optical fiber is an ideal transmission medium that can deliver bandwidth-intensive, integrated voice, data and video services at distances around 20 km in the subscriber access network. In order to alleviate the bandwidth bottleneck at the access network, it is required to deploy the optical fiber and optical nodes with deeper penetration. Especially, the optical fiber should be deployed throughout the “first mile” to meet the bandwidth requirements. A simple way to deploy the optical fiber in the access network is to use a point-to-point topology with a dedicated fiber from the CO to each end-user subscriber [2]. Although a simple architecture, the point-to-point topology is cost prohibitive because it requires

significant fiber deployment and connector termination space in the local exchange. Considering M subscribers at an average distance L km from the CO in Figure 1.1(a), a point-to-point design requires $2M$ transceivers and $M*L$ fiber length totally (assuming that a single fiber is used for bi-directional transmissions).

A remote curb switch can be deployed close to the neighborhood in order to reduce the fiber deployment (Figure 1.1(b)). Such architecture will reduce the fiber consumption to only L km (assuming negligible distance between the switch and customers). However, because there is one more link added to the network, the number of transceivers will increase to $2M+2$. In addition, the curb-switched network architecture requires electrical power as well as back-up power at the curb switch, which will increase the cost of Local Exchange Carriers (LECs).

Hence, it is reasonable and logical to replace the active curbside switch with an inexpensive passive optical splitter, which results in the solution of Passive Optical Network (PON). A PON is a point-to-multipoint optical network with no active elements in the signal path from the source to destination. The only interior elements used in a PON are passive optical components, such as optical fiber, splices and splitters. PONs can minimize the number of optical transceivers, CO terminations and fiber deployment. An access network based on a single-fiber PON only requires $M+1$ transceivers and L km fiber length as shown in Figure 1.1 (c).

PON has been viewed by many as an attractive solution for the “first mile” access networks, because it can support gigabit speed at low cost comparable to the DSL and HFC solutions. With PON, the subscriber access network can be effectively implemented into the Fiber-To-The-Home (FTTH), Fiber-To-The-Building (FTTB) and Fiber-To-The-Curb (FTTC), which is one of the objectives of next-generation access networks. Three primary types of PON solution are Asynchronous Transfer Mode (ATM) PON, Gigabit PON, and

Ethernet PON. Among these PON solutions, Ethernet PON (EPON) is highly regarded as the best candidature for the next-generation access networks since it represents the convergence of the inexpensive Ethernet equipments and the high-speed fiber infrastructure. So our research will mainly focus on EPON networks in this thesis.

Figure 1.1 Deployment of the Optical Fiber in the Access Networks

1.2 Motivation

In an EPON network, multiple Optical Network Units (ONUs) access to the shared fiber channel to reach the Optical Line Terminal (OLT) through a passive optical splitter. To arbitrate the multiple access from ONUs, an effective bandwidth allocation scheme is

required for the upstream transmission. Because the traffic arriving in the access network is quite bursty, the bandwidth requirements of the traffic vary greatly from time to time. Therefore, static assignments of bandwidth to the subscribers or ONUs are very inefficient in EPONs. It is much more efficient to employ a Dynamic Bandwidth Allocation (DBA) scheme adapting to the instantaneous bandwidth requirements. Furthermore, although the IEEE 802.3ah Ethernet in the First Mile (EFM) Task Force is standardizing for EPONs to ensure the interoperability of products from different vendors, it has not specified the particular bandwidth allocation scheme. Instead, designing and developing multiple access schemes for upstream transmissions is an open issue in EPONs.

In the literature, the main categories of DBA schemes for EPONs include WDM-based schemes, contention-based schemes and TDMA-based schemes. With a WDM-based scheme, each ONU operates at a different wavelength to avoid conflicts in the upstream transmission. Although simple to implement, this kind of solutions is cost prohibitive and cannot be accepted widely. A contention-based scheme is essentially a distributed access control scheme where multiple ONUs perform the Carrier Sense Multiple Access with Collision Detection (CSMA/CD). Such a scheme is difficult to implement in EPONs because ONUs cannot detect collisions at the OLT due to the optical splitter's directional properties. In a TDMA-based scheme, each ONU is allocated a timeslot to transmit data upstream without collision. Including both decentralized and centralized schemes, TDMA-based DBA schemes are more popular and cost-effective than WDM-based and contention-based schemes. Since the decentralized schemes require the connectivity between ONUs, this imposed some constraints on EPON topologies. Comparatively, a centralized TDMA-based DBA scheme is an ideal choice for EPONs because it ensures only the connectivity between the OLT and each ONU, which can be employed in all EPON topologies. Being robust and scalable, the centralized TDMA-based DBA schemes have attracted more and more interests

from research and industry. The research in this thesis will also concentrate on the centralized TDMA-based DBA schemes for EPON networks.

Although next-generation access networks will mainly carry the IP data traffic, traditional services, such as T1/E1s, ISDN, POTS, analog video, etc, will still exist in the foreseeable future. Hence, it is crucial for EPONs to provide both IP-based services and jitter-sensitive and time-critical legacy services that have not been the focus of Ethernet traditionally. That means, EPONs should be designed to transmit the best-effort data together with the time-critical voice and video traffic and fulfill their transmission requirements. Therefore, it is imperative to devise DBA schemes for EPONs that can support the Differentiated Services (DiffServ) and provide Quality of Service (QoS) guarantees to traffic flows with specific service requirements.

1.3 Objectives

This thesis aims to investigate the multiple access issue among different ONUs in EPONs, to design and analyze novel bandwidth allocation schemes with QoS and DiffServ support for the upstream transmission in EPON networks.

The general objective of this thesis is to design the centralized TDMA-based DBA schemes with the consideration of providing absolute QoS guarantees. With such a DBA scheme, end users having absolute QoS requirements will be satisfied with the specified transmission quality such as the minimum bandwidth and maximum waiting delay. For users with more consideration on budget than transmission quality, best effort services will be provided by the DBA scheme. Then we will develop the Call Admission Control (CAC) mechanism for EPONs based on the DBA scheme. It will decide which kind of subscribers can be admitted to the EPON network with required QoS guarantees. The results of the CAC

mechanism can be further used to facilitate the implementation of the DBA scheme, thus the admitted subscribers will be provided with the corresponding QoS guarantees.

The further objective is to support DiffServ for different traffic flows in EPONs. It is quite necessary because an end user may have various traffic flows to transmit, with different service requirements. The thesis will work out efficient solutions based on the DBA scheme to provide DiffServ for traffic streams from the same user. Fairness is another objective we will pursue for different classes of traffic in EPONs. With the fair scheduling mechanism, no traffic flow will get the bandwidth more than it requests. The excess bandwidth unconsumed by the light-load traffic will be distributed to all backlogged traffic flows in a predefined manner.

Then we need to develop appropriate mathematic models to evaluate the performance of polling-based bandwidth allocation schemes. In EPON networks, the OLT polls multiple ONUs in an adaptive order, inviting them transmit data packets over the upstream channel. Such a mechanism can be modeled as a multi-queue polling system, where the OLT and ONUs are regarded as the central server and queues respectively.

An additional objective is to enhance the performance of a famous pioneer DBA scheme, the Interleaved Polling with Adaptive Cycle Time (IPACT), to obtain lower waiting delay and buffer occupancy.

1.4 Methodology

In order to work out the DBA scheme to guarantee the absolute QoS, service requirements of subscribers will be taken into account in the thesis. Generally, service requirements are represented as the Service Level Agreement (SLA) contracts between service providers and subscribers. SLA contracts define the service classes that will be provided to subscribers, and the penalties that will be charged if the service provider cannot meet the established

goals. By incorporating SLA contracts into the bandwidth allocation mechanism design, the related DBA schemes can provide absolute QoS guarantees to subscribers with definite transmission requirements.

With regard to supporting DiffServ for traffic flows, an EPON network is required to classify all flows into a limited number of classes according to their characteristics, and provide the differentiated service for each class. Thus an ONU will be equipped with multiple queues to accommodate different classes of traffic from the subscribers. In order to support DiffServ, the bandwidth allocation in EPONs will be divided into two levels. While the first level of scheduling is responsible for allocating the upstream transmission capacity among multiple ONUs, the second level of scheduling arbitrates the transmissions from different traffic queues within an ONU. Accordingly, the bandwidth allocation solutions devised for EPONs will include these two levels of scheduling functions respectively.

To improve the performances of IPACT regarding the waiting delay and buffer occupancy, the thesis will deploy an efficient method to estimate the transmission window size being granted to ONUs. Only when the granted window size approximates to the instant buffer occupancy, the bandwidth waste will be controlled, and the better performances can be obtained with lower delay and buffer occupancy.

1.5 Main Contribution of the Thesis

1. An enhanced proposal has been presented to improve the performances of a pioneer DBA scheme, the Interleaved Polling with Adaptive Cycle Time (IPACT). The newly proposed scheme, IPACT with Grant Estimation (IPACT-GE), estimates the amount of new packets arriving between two consecutive pollings, and decides the grant size to the ONU based on this estimated amount as well as the request size. By additionally considering the newly arriving packets, the grant size in IPACT-GE is more approximate than that in IPACT, to the

instantaneous buffer occupancy when the ONU is polled. Comparing to IPACT, the IPACT-GE scheme can achieve the shorter waiting delay and smaller buffer occupancy at the light load. Furthermore, when combined with the Strict Priority Queuing (SPQ) mechanism to support DiffServ, the IPACT-GE scheme can greatly mitigate the light-load penalty, which is an obvious phenomenon in the solution of IPACT combined with SPQ.

2. Next, a novel bandwidth allocation scheme, Bandwidth Guaranteed Polling (BGP), has been proposed for EPONs to provide the absolute QoS guarantees as well as arbitrate the upstream transmissions from multiple ONUs without collisions. As a brand new DBA scheme proposed by us, the polling-based BGP scheme takes the SLA contracts into consideration, which represent each ONU's bandwidth requirement. According to SLA, ONUs can be divided into different groups with relative bandwidth allocations. In every polling cycle, the OLT polls different ONUs in an adaptive order predefined by an Even Distribution Algorithm (EDA). The EDA algorithm is responsible for preparing the polling sequence for BGP. It evenly distributes the bandwidth units of an ONU so that the burst from that ONU can be controlled at the low level. With the BGP scheme, some ONUs with high bandwidth requirements can be polled more than one time in a cycle, while other ONUs without any requirement may not be polled once in several cycles. Thus, the BGP scheme can guarantee the absolute QoS to users with high demands, and provide the best-effort service to users with low demand.

Then a parameter-based CAC mechanism has been designed for EPONs based on the BGP scheme. The CAC mechanism admits the transmission requests from ONUs in both the initialization period and operation period, based on the QoS requirements of new requests and the network bandwidth utilization. The merit of the proposed CAC mechanism is that it can accept all transmission requests from the ONUs without any rejection. If the transmission cannot be admitted with the required QoS, the CAC mechanism will suggest

providing it with the downgraded service. Combined the admission control and bandwidth allocation, a systematic Medium Access Control (MAC) protocol have been developed for EPONs in this thesis, which contains the proposed CAC mechanism, EDA algorithm, and BGP scheme. These three parts operate autonomously but are related to each other closely. The parameter-based CAC mechanism processes the users' requirements expressed in SLA parameters, decides which group each ONU belongs to and how many bandwidth units each ONU is allocated. Based on the results of CAC, the EDA algorithm determines the polling sequence for the BGP scheme. Then by BGP, the OLT polls ONUs one after another based on the adaptive order, inviting ONUs transmit data over the upstream channel.

3. Several effective solutions have been proposed to support the differentiated services to different traffic flows in EPONs. We have employed a hierarchical scheduling architecture for EPONs including the inter-ONU and intra-ONU scheduling, to accomplish the two levels of bandwidth allocation among the ONUs and within an ONU respectively. While the newly designed BGP-based MAC protocol is modified and adopted for the inter-ONU scheduling in all solutions, various mechanisms are suggested for the intra-ONU scheduling. All the proposed solutions can provide efficient DiffServ to traffic flows, with the higher-priority traffic obtaining better performance than the lower-priority traffic. Besides, solutions with different mechanisms in the intra-ONU scheduling can achieve different levels of fairness for traffic flows.

1) First we have suggested the default SPQ mechanism as the intra-ONU scheduling scheme. In SPQ, different types of traffic are grouped into multiple queues with different priorities when arriving at the ONU. Packets are transmitted strictly according to the priorities when the ONU is polled. Besides supporting DiffServ, the solution combining SPQ with the BGP-based MAC scheme can greatly reduce the light-load penalty that is obvious

in the solution of IPACT and SPQ. This solution cannot provide the relative fairness to traffic flows in the same ONU.

2) Then the Two-Stage Queuing (TSQ) mechanism has been recommended in the intra-ONU scheduling. When the ONU is polled for transmission in TSQ, packets are advanced from the priority queues to the second stage of First-In-First-Out (FIFO) queue, and then are delivered to the upstream channel. The solution with TSQ can diminish the performance polarization between the traffic flows, and provide better performance to the lower-priority traffic than the solution with SPQ. Thus it can achieve the fairness to some extent for different traffic flows.

3) Last we have proposed an efficient Urgency Fair Queuing (UFQ) scheme for the intra-ONU scheduling. Considering the specific QoS requirements, the newly devised UFQ scheme schedules the packets from different queues within an ONU according to the urgency with regard to the delay bound requirements. Packets from the best-effort traffic can be transmitted first if all packets in the QoS traffic are not urgent for transmission. By this means, the bandwidth can be allocated fairly between the QoS traffic and the best-effort traffic in the same ONU. The best-effort traffic can be provided with better performance than in SPQ and TSQ, without violating the performance of the QoS traffic. The solution combining UFQ with the BGP-based MAC scheme can not only provide differentiated services for EPONs, but also achieve the better fairness between the QoS traffic and best-effort traffic than solutions with SPQ and TSQ.

4. Finally, we have developed a mathematical estimation model and a Continuous-Time Markov Chains (CTMC) model to numerically evaluate the polling-based BGP scheme. Both models have taken some approximations to facilitate the adaptation to the EPON network with BGP. The models can be used to testify the rationality of the proposed BGP schemes, and verify the models themselves.

1.6 Organization of the Thesis

The thesis includes six chapters totally. Chapter 2 first introduces the characteristics of PON technologies. Then three primary PON solutions are presented, with the detailed description of MAC protocols used in APON and EPON solutions. The Multi-Point Control Protocol (MPCP) specified by the IEEE 802.3ah EFM Task Force is introduced briefly, followed by the comprehensive survey of the DBA schemes proposed for EPONs so far.

In chapter 3, the IPACT scheme is reviewed in detail first, followed by its weakness. Then we propose the novel IPACT-GE scheme to enhance the IPACT performance. The experimental simulation shows the performance of IPACT-GE surpassing IPACT. It also shows that IPACT-GE can achieve better improvement to IPACT than another scheme that is also aiming to enhance the IPACT scheme.

Chapter 4 is focused on the novel BGP scheme that we design for EPON upstream transmissions. The principle, operation and efficiency of BGP are fully presented, followed by the description of the EDA algorithm. Then we develop the mathematical estimation model and CTMC model for the BGP scheme separately. A parameter-based CAC mechanism is further proposed based on BGP, and the systematic MAC protocol is presented to contain the CAC mechanism, EDA algorithm, and BGP scheme as a whole. Last, the numerical experiments show the efficiency of the proposed BGP scheme and BGP-based MAC protocol by comparing with the IPACT scheme. The results of the mathematical estimation model are also compared with the simulation results of BGP to testify the rationality of the scheme.

In chapter 5, three solutions are separately presented to support DiffServ for different traffic flows in EPONs. The hierarchical scheduling architecture is introduced. Then the BGP-based MAC protocol proposed in chapter 4 is modified to be adapted for the inter-ONU scheduling. The solution with the SPQ mechanism in the intra-ONU scheduling is first

Chapter 1. Introduction

presented. Following is the solution with TSQ that can provide better performance to lower-priority traffic. The performance of TSQ solution is compared with that of the SPQ solution by the experimental simulation. Finally the UFQ scheme is proposed for the intra-ONU scheduling. The solution with UFQ can achieve better fairness than the TSQ solution, which can be shown by the performance evaluation.

At last, we conclude the whole thesis in chapter 6, and recommend several future directions for the potential research on the EPON area.

Chapter 2 . Literature Review

Deploying the high-capacity optical fiber, a Passive Optical Network (PON) can provide very large bandwidth to meet the subscribers increasing demand of transmitting various types of traffic. At the same time, PONs can also reduce the cost of building the access networks and free the operators from maintaining the active components in the transmission path. Hence, the PON solutions are highly regarded as the best choices for the next-generation access networks. The most famous solutions of PON are Asynchronous Transfer Mode PON (APON), Ethernet PON (EPON), and Gigabit PON (GPON). A PON deploys a point-to-multipoint architecture where an Optical Line Terminal (OLT) is connected with multiple Optical Network Units (ONUs) through a passive optical component. In the upstream direction, multiple ONUs transmit data to the OLT over the shared channel, where the specific Medium Access Control (MAC) protocols should be adopted to arbitrate the multiple access from ONUs effectively. This chapter will give an overview for PONs including the PON topologies and primary solutions, especially presenting the detailed description for the significant EPON solutions. Then a comprehensive survey will be provided for the Dynamic Bandwidth Allocation (DBA) schemes for the upstream transmissions in EPONs.

2.1 Passive Optical Networks (PONs)

Being a point-to-multipoint optical network, a PON employs only passive optical components, such as passive couplers and splitters in the transmission path. It minimizes the amount of optical transceivers, central office terminals, and fiber deployment in both local exchange office and local loop, thus reducing the cost of building the access network [1]. By using passive components deployed as part of the optical fiber cable plant, PONs eliminate the necessity to install active multiplexers and de-multiplexers at splitting locations, thus relieving network operators from maintaining and providing power to them.

Besides the low cost and maintenance due to deploying passive components in the transmission path, PON has many other advantages making it an attractive choice of access networks. It allows around 20 km long distance transmissions between the central office and customer premises. PON can provide very high bandwidth due to the deep fiber penetration, offering solutions with gigabit per second. In the downstream direction, PON operates as a broadcast network, thus allowing for video broadcasting and other applications. In addition, PON allows upgrades to higher bit rates or additional wavelengths because of the optical end-to-end transparency.

2.1.1 PON Topologies

Logically, the access network deploys the point-to-multipoint architecture, with a central office serving multiple users. There are several topologies suitable for the point-to-multipoint access network, including tree, tree-and-branch, ring, and bus. Using 1:2 optical tap couplers and 1: M optical splitters, PONs can be flexibly configured in any of these topologies (Figure 2.1). Additionally, a PON can be deployed into the redundant configuration, i.e. double rings or double trees. The redundancy may be added only to a part of the PON, such as the trunk of the tree [2].

In a PON, all transmissions are performed between an OLT and ONUs, which are located at the local exchange (central office) and end user locations respectively. The OLT is responsible for connecting the optical access network to the backbone. The ONU takes charge of conveying the broadband voice, data, and video services between the end users and the OLT. In the downstream direction (from the OLT to ONUs), a PON is a point-to-multipoint network that can easily operate by broadcasting. In the upstream direction (from ONUs to the OLT), a PON is a multipoint-to-point network where different ONUs share the upstream channels to transmit data.

Figure 2.1 Topologies of Passive Optical Networks

2.1.2 Primary PON Solutions

There are three main solutions for the PON technology: APON, EPON, and GPON, which are developed by different standard organizations respectively. The main technological difference between these solutions is how to encapsulate the upper layer data packets in the layer 2 transmissions. In the APON solutions, packets are carried in the Asynchronous

Transfer Mode (ATM) cells for transmission. While in EPON, Ethernet frames are the carrier in layer 2 to take the data packets. Finally, GPON Encapsulation Method (GEM) is used to encapsulate packets in GPON. APON and GPON permit fragmentation of data packets, which can achieve strict service requirements for the applications. While, it is also required to reconstruct the fixed frames at the destination, which leads to the extra complexity. No segmentation is allowed in the EPON solutions, the packet reconstruction is neither needed. Hence, the individual frame length in EPON is variable.

Since Ethernet can achieve Gigabit and even 10-Gigabit capacity, an EPON network can provide very high transmission rate with Gigabit per second for both the upstream and downstream transmissions [2][3][4]. Using the standard line rates, APON can support the symmetric transmission with 155.52 Mbps for both directions and the asymmetric transmission with 622.08 Mbps for downstream and 155.52 Mbps for upstream [5][6]. GPON has been optimized for higher rate applications than APON with the support of ATM. It provides several combinations of asymmetric or symmetric line rate from 155.52 Mbps to 2.48 Gbps to fit various operational situations [7][8]. In the following sections, the APON and EPON solutions will be presented in further detail.

2.1.3 ATM PON (APON)

2.1.3.1 Context of APON

In the mid-1990s, the Full Service Access Network (FSAN) defined a PON-based optical access network that uses ATM as its layer 2 protocol, which is the origin of APON. At that time, ATM was highly expected to become the prevalent technology in the Local Area Networks (LANs), Metropolitan Area Networks (MANs), and Wide Area Networks (WANs) with the decreasing cost. The APON format has been accepted as an International Telecommunication Union-Telecommunication (ITU-T) standard series G.983 [6].

However, ATM has many shortcomings that prevent it from an ideal technology for PONs to transmit the predominant Internet Protocol (IP) data traffic in the future networks. Since the ATM cell has the fixed size of 53 bytes, the IP packet with variable length should be fragmentized to segments in order to be carried by ATM cells. That will impose a high overhead over IP packet transmissions. In addition, a dropped or corrupted ATM cell will invalidate the entire IP packet, however other ATM cells carrying the same IP packet will propagate further to the destination, thus consuming network resources unnecessarily [2]. Moreover, the cost of ATM equipment has not declined as expected, which inhibits the deployment of the APON solutions.

Due to the technological and economic considerations, APON has lost the dominant position for next-generation access networks. The FSAN developed another Broadband PON (BPON) technology, GPON, which has been standardized as ITU-T G.984 series [8]. GPON supports transport of various native protocols including ATM and Ethernet. Furthermore, it expands APON's transmission capacity from Megabit to Gigabit. Although achieving the comparative bandwidth with the EPON solutions, GPON needs more complex operations and frame structure to provide backward compatibility with the legacy technologies including APON. Moreover, the necessities to segment and reconstruct packets [7] also result in the additional complexity of GPON. Comparatively, supported by the IEEE community, EPON has many advantages that help it to attract more and more attentions in recent years, which will be addressed in detail later.

2.1.3.2 MAC Protocols for APONs

A great deal of efforts has been offered to APON networks by research from the outset. Since different ONUs in an APON share the upstream channel to deliver data to the OLT, many research works are focused on the MAC protocols used for the upstream data

transmission. The most prevalent MAC protocols that have been proposed and demonstrated are based on the Time Division Multiple Access (TDMA) method.

If the APON networks mainly carry the Constant Bit Rate (CBR) service, static or semi-static MAC protocols, which are based on the connection setup information, will be very efficient. However, this type of protocol becomes inefficient in the presence of burst and heavy traffic (e.g. IP service) because it is difficult to calculate an appropriate equivalent bandwidth for the burst traffic.

An efficient APON operation in the presence of burst traffic requires the implementation of a dynamic MAC protocol. Dynamic MAC protocols allow ONUs to report their instantaneous bandwidth requests to the OLT. It is very important to transport these requests as efficiently as possible from ONUs to the OLT by the dynamic MAC algorithm. In the literature, mainly two types of upstream channels are proposed: reservation-based upstream channels and piggybacking channels. For reservation-based channels, dedicated timeslots carrying bandwidth requests are sent per group of ONUs at regular polling intervals. One such timeslot is used by multiple consecutive ONUs to report one request per mini-slot. For a piggybacking channel, an ONU uses a one-byte field for bandwidth request in each data cell. ONUs can request new permits while sending an upstream data cell. The reservation-based TDMA scheme is considered to be most promising and sufficient for a standard APON network, while the piggybacking TDMA scheme is an additional way for APONs with large number of ONUs.

Because the static or semi-static MAC protocols are relative simple and inefficient for APONs with bursty traffic, we only review some dynamic MAC protocols for APONs containing the piggybacking TDMA schemes ([9][10]) and reservation-based TDMA schemes ([11-16]) as follows.

[9] presents a dynamic resource allocation algorithm called Multiple Queue First-In-First-Out (MQ-FIFO) for the home network environment. It combines piggybacking with the mini-slot polling mechanism. Bandwidth allocation is based on two levels of multiplexing: multiplexing at the OLT and multiplexing at the ONU. An ONU sends to the OLT a request asking the permit to transmit, which is contained in a tag added to each upstream ATM cell. In each ONU, a Weighted Fair Buffer Allocation (WFBA) algorithm guarantees the fair distribution of buffer resources among multiple queues. A dynamic MAC protocol is described in [10] that allows implementing upstream requests by piggybacking or dedicated polling slots. The downstream grants can also be processed per ONU or subdivided per Traffic Container (T-Cont) class. The authors also configure a demonstrator platform to evaluate the MAC protocol.

[11] proposes a reservation based TDMA method for channel sharing in the APON network. Focusing on the service transparency, the proposed allocation method is based on requests made by ONUs in the special MAC frame header. The OLT responds by sending “tickets” that indicate the number of cells each ONU is permitted to insert. The order of transmissions is determined by the management procedures. To address the performance of Cell Delay Variation (CDV), [12] focuses on the design and implementation of a MAC mechanism emphasizing a very low CDV. It relies on a meticulous recording of the cells arrival time that enables a subsequent allocation of time slots, thus producing almost the same output and CDV performance as the centralized FIFO multiplexer. The bandwidth allocation algorithm uses the interim storage belt and the permit scheduling belt, to implement the allocation of time slots and achieve timing correlation of arrivals at different terminals. Other reservation-based dynamic MAC protocols [13-16] are proposed to achieve the good CDV performance or have other Quality of Service (QoS) considerations.

2.1.4 Ethernet PON (EPON)

EPON is a PON-based network where data packets are encapsulated into Ethernet frames for transmission. EPON has been regarded as the best candidature for the next-generation access networks because it merges the virtues of Ethernet and PON. Being a low-cost technology, Ethernet has been universally accepted and can be interoperable with a great deal of legacy equipments. It can achieve very high speed, and even 10 Gigabit Ethernet products are available today. Besides, newly adopted QoS techniques have made Ethernet networks capable of supporting the integrated voice, data, and video services efficiently. These techniques include the full-duplex transmission mode, prioritization (P802.1p), and Virtual LAN (VLAN) tagging (P802.1Q) [2]. Ethernet becomes a perfect choice of PONs for delivering IP packets.

Compared to the mature APON solutions that have been developed for almost twenty years, EPON has just been proposed and developed in the couple of years. Although the IEEE 802.3ah Ethernet in the First Mile (EFM) Task Force [4] has made the related standardizations for EPON, there are many open issues need to be investigated in the EPON area, especially the MAC protocols for the multiple access of ONUs, which is generally referring to the bandwidth allocation schemes, for upstream transmissions in EPON networks.

2.1.4.1 Transmission Principle

In the downstream direction (from the OLT to ONUs) of an EPON network, Ethernet frames transmitted by the OLT pass through a $1:M$ passive splitter and reach each ONU. Ethernet fits perfectly with the EPON architecture due to its broadcast property (Figure 2.2). Frames are broadcast by the OLT, carrying the destined ONU's MAC address. If the destination

address matches with the ONU's MAC address, it will extract the frames and deliver to the end users; otherwise the ONU will ignore and discard the frames.

Figure 2.2 Downstream Transmissions in EPON Networks

In the upstream direction (from ONUs to the OLT), data frames from any ONU will only reach the OLT due to directional properties of the passive combiner (optical splitter). Since data frames from different ONUs being transmitted simultaneously may collide over the channel, an efficient MAC scheme is required for EPONs to perform the bandwidth allocation among multiple ONUs. Figure 2.3 illustrates the upstream data transmissions from different ONUs to the OLT.

Figure 2.3 Upstream Transmissions in EPON Networks

2.1.4.2 MAC Protocols for EPONs

In the upstream direction of EPON networks, effective MAC protocols are required to allocate the bandwidth to multiple ONUs for access the shared trunk fiber channel. The main

functionality of a MAC protocol should be avoiding collisions of packets from different users. In addition, a MAC protocol should also be capable of introducing low overhead, making efficient use of the resources and guaranteeing the requested QoS for all classes of service.

Because of the traffic bursty nature in the access networks, the bandwidth requirements change greatly from time to time. Hence, it is inefficient to allocate bandwidth to ONUs or individual traffic queues in a static manner. A Dynamic Bandwidth Allocation (DBA) scheme is more efficient for EPONs where the instantaneous requirements are considered in the bandwidth allocation. Hereafter we will focus on the DBA schemes deployed in the EPON networks for the upstream transmissions. Besides the widely accepted TDMA-based DBA schemes, there are also some schemes deploying the Wavelength Division Multiplexing (WDM) and Carrier Sense Multiple Access with Collision Detection (CSMA/CD), which will be described respectively as follows.

2.1.4.2.1 WDM-Based Schemes

One possible way to share the upstream bandwidth in EPONs is using WDM, in which each ONU operates at a different wavelength. Although a simple solution, it is cost prohibitive because tunable transmitters are required in ONUs and either a tunable receiver or a receiver array is needed in the OLT. A more serious problem is that multiple types of ONUs based on their laser wavelength would be required. Some schemes are proposed to combine WDM with other methods such as TDMA and Code Division Multiple Access (CDMA). By the combination, EPONs can achieve the high capacity without employing one wavelength for each ONU. However, the cost of a WDM-based EPON is still too high to be accepted widely.

In [17], the TDMA mechanism is combined into a WDM-based EPON system where multiple wavelength channels are established in both upstream and downstream directions.

The system employs a WDM IPACT with a single polling table (WDM IPACT-ST) scheme based on the interleaved polling. The OLT can predict when transmissions will finish on all upstream channels, thus to schedule the next ONU to transmit packets over the first available channel. A Weighted Fair Queuing (WFQ) scheme is adopted to support the Differentiated Services (DiffServ) by reserving different weighted proportion for various types of traffic from cycle to cycle.

[18] evaluates the efficiency of the hybrid time/wavelength/code division approach for EPONs. The EPON system employs N wavelength channels for transmission, where the optical CDMA is combined in order to increase the network capacity. All ONUs are divided into N groups, each with one wavelength. Within the group, each ONU randomly picks an Optical Orthogonal Code (OOC) from the multiple OOCs to encode the data packets for transmission. The ONU should make an announcement on the control channel about the selected wavelength, OOC sequence, and the amount of packets to be sent.

2.1.4.2.2 Contention-Based Schemes

According to [2], the contention-based media access (similar to CSMA/CD) is difficult to implement in EPONs, because ONUs cannot detect a collision at the OLT due to directional properties of the optical splitter. A collision can only be detected by the OLT and be informed to ONUs by the contention signal. However, propagation delays in EPONs greatly reduce the efficiency of such a scheme. In order to achieve the prompt collision detection in a contention-based scheme, several proposals consider including additional components in the EPON architecture. Accordingly, the cost of the access network will increase and the extra maintenance should be taken for the components.

[19] and [20] proposes an implementation of optical CSMA/CD scheme for EPONs employing a $3 \times N$ Star Coupler (SC). A redirection mechanism is introduced where a portion

of the optical power transmitted upstream is redirected back and distributed to all ONUs for carrier sensing and collision detecting. If an ONU has sensed that the upstream wavelength is unoccupied, it can transmit the packets. Otherwise, the packets are backed off. If a collision is detected, transmission is promptly aborted and the collided packets are backed off. The OLT is removed from the implementation of the scheme. [21] extends the scheme to EPON networks with the $N \times N$ SC architecture.

A Hybrid-Slot Decentralized Control (HSDC) scheme is presented in [22] to provide QoS for various types of traffic and simplify the network control mechanism. HSDC divides the frame into the fixed time slot part and the contention part. At the beginning of each cycle, ONUs send the high priority traffic in the respective fixed time slots whose sizes are statically set according to the subscription rates. In the remaining contention part of the frame, multiple ONUs contend to transmit the low priority traffic using the CSMA/CD scheme. In HSDC, it is unnecessary to sent the control messages between the OLT and ONUs for data packet transmissions.

2.1.4.2.3 TDMA-Based Schemes

Time-sharing is a more popular and attractive method for optical channel sharing in an EPON network. It allows for a single upstream wavelength and a single transceiver in the OLT, resulting in a cost-effective solution. In the TDMA-based schemes, each ONU is allocated a timeslot, and each timeslot is capable of carrying multiple Ethernet frames. An ONU should buffer frames received from the users until its timeslot. When its timeslot arrives, the ONU would “burst” out frames at full channel capacity.

Recently, numerous schemes based on TDMA have been proposed for EPONs including both decentralized and centralized schemes. In a decentralized scheme, the OLT is removed from the bandwidth allocation decision. All ONUs performs the scheme to share the

upstream transmission link. While in a centralized scheme, the OLT is the central controller to assign the bandwidth for all ONUs. Following are some decentralized TDMA-based schemes proposed for EPONs in the literature.

The Full Utilization Local Loop Request Contention Multiple Access (FULL-RCMA) scheme is proposed for EPONs in [23] and [24] as a decentralized time-sharing scheme. The EPON network requires that the splitter reflect the upstream data back to all ONUs for collision detection, which needs two fibers per ONU. FULL-RCMA contains a contention-based request period and a contention-free data period for each cycle. ONUs contend to submit requests during the request period in the random time slots. Each ONU knows the results of the request period by monitoring the echoes from the splitter, and transmits data packets accordingly in the data period without collisions.

[25] introduces another collision-free scheme for the distributed EPON networks. It also requires part of the upstream optical power be redirected back to all ONUs. Each ONU runs the same cycle-based algorithm with the identical bandwidth allocation results. The cycle is divided into three periods: the static update period for each ONU transmitting the control message, the fixed waiting period to process the control messages and get the ONUs transmission assignments, and the dynamic transmission period for ONUs sending data packets over the upstream channel following the assignments without collisions. The order of ONUs transmissions is dynamic for different cycles based on the traffic demand.

The decentralized approaches require the connectivity/communicability between ONUs. This imposes some constraints on EPON topologies that only a ring or a broadcasting star can be deployed [2]. Since a preferred scheme should support any point-to-multipoint EPON topology, the decentralized approach is not a good choice for TDMA-based DBA schemes.

Comparatively, a centralized scheme is an ideal choice because it ensures only the connectivity between the OLT and each ONU, which can be employed in all EPON

topologies. The OLT knows the state of the entire network and can switch from one allocation scheme to another according to the information. The entire EPON network will be more robust and scalable, and the ONU can be very cheap and simple without any intelligence. There are much more centralized schemes than decentralized schemes proposed for EPONs to date. Before surveying the centralized DBA schemes based on TDMA in section 2.2, we will first introduce the Multi-Point Control Protocol (MPCP) arbitration mechanism that is developed by the IEEE 802.3ah EFM Task Force [4] to support the centralized timeslot allocation for EPONs in next section.

2.1.5 Multi-Point Control Protocol (MPCP)

Generally, MPCP specifies a control mechanism between an OLT and multiple ONUs connected through a point-to-multipoint EPON network to allow efficient data transmissions in a centralized time-sharing approach. MPCP operates the auto-discovery mode to initialize the EPON system and detect newly connected ONUs, getting the Round-Trip Time (RTT) and MAC address of each ONU. Besides, MPCP operates the normal mode by exchanging the control messages to arbitrate the upstream data transmission from multiple ONUs to the OLT.

2.1.5.1 MPCP Auto-Discovery Mode

The OLT periodically reserves the discovery window for auto-discovery. It broadcasts the discovery GATE message to detect whether there is any new ONU connected to the EPON network. Only the un-initialized ONUs respond to the discovery GATE message and set the local time according to the timestamp contained in the arriving message. An un-initialized ONU will transmit a REGISTER_REQUEST message to request registration, which includes the ONU's address and local time. The OLT will calculate the ONU's RTT upon

receiving the message. After the relative message procedures (Figure 2.4), the ONU will be initialized and the channel between the ONU and OLT is established.

Figure 2.4 MPCP Auto-Discovery Procedures

Since there will be multiple ONUs requesting initialization at one time, auto-discovery is a contention-based procedure. The ONU whose REGISTER_REQUEST message collides with others is considered failure during the auto-discovery procedure. It can attempt to request initialization again in every following discovery window or skip random number of discovery windows (i.e. using exponential backoff) before next request.

2.1.5.2 MPCP Normal Mode

In the normal mode, MPCP controls the upstream data transmission. Two control messages, GATE and REPORT, are defined in MPCP and transmitted between the OLT and ONUs. The OLT performs the bandwidth allocation algorithm to get the transmission grants for ONUs. A GATE message is generated by the OLT and sent to an ONU granting it to transmit data over the upstream channel at an appropriate time. Illustrated in Figure 2.5 , the GATE message contains the information of timestamp when GATE is sent out, the start time that the ONU is granted for transmission, and the stop time of the transmission. On receiving the GATE message from the OLT, an ONU will update its local clock according to the

timestamp, and wait until the start time for transmission. The transmission may include multiple Ethernet frames, depending on the size of the granted transmission window and the number of waiting packets at the ONU. The ONU should ensure that no fragmentation is allowed during the data transmission. The frame that is too large to be accommodated in the assigned timeslot will be deferred to the next timeslot.

Figure 2.5 MPCP GATE Operations

ONUs can send REPORT messages to the OLT automatically or on-demand to request bandwidth allocation. REPORT is transmitted together with the upstream data frames in the assigned timeslot. It can be either transmitted at the beginning or at the end of the timeslot, depending on the bandwidth request approach implemented by the ONU. A REPORT message must contain the timestamp used by the OLT to adjust the RTT for the ONU. It may contain the desired size of the next timeslot based on the ONUs buffer occupancy. For the ONU with multiple traffic queues, it can report the entire buffer occupancy in the REPORT message, or report each queue's status to request individual grant. Depending on the bandwidth allocation scheme deployed, the OLT can choose to issue one grant for an ONU, or issue multiple grants in the same GATE message for different queues.

It is important to notice that MPCP does not specify any particular bandwidth allocation scheme for EPON upstream transmissions. Rather, it supports the implementation of various schemes as long as they conform to the MPCP framework. Recently a great number of bandwidth allocation schemes have been proposed compatible with MPCP, which will be investigated in the next section.

2.2 Survey of the Centralized DBA Schemes for EPONs

In the literature, numerous centralized DBA schemes have been proposed for EPON networks so far. Most of these schemes are based on TDMA and compatible with the MPCP protocol. In this section they will be surveyed in much detail. Considering the large number of the related DBA schemes, a proper classification is necessary for the investigation.

Figure 2.6 Individual Polling

Figure 2.7 Joint Polling

Since most of the DBA schemes are based on polling, we can simply classify them by the ways of polling. One classification is to divide the DBA schemes into groups of individual polling and joint polling. These two kinds of polling differentiate from each other by which ONUs requests are exploited for the bandwidth allocation. In the individual polling, an ONU's individual request alone can decide the transmission grant to itself (Figure 2.6). The OLT determines the grant for each ONU merely upon receiving its REPORT message. In contrast, in the joint polling shown in Figure 2.7 , the OLT collects REPORT messages from all ONUs in each cycle. Then it decides the transmission grants jointly by all ONUs requests instead of one ONU's single request.

Figure 2.8 Category of DBA Schemes for EPONs

A better way for classification has been presented in [26], where the DBA schemes are categorized into schemes with statistical multiplexing and schemes with QoS assurances. Then the schemes with QoS assurances are grouped into absolute assurances and relative assurances. We decide to follow this classification because it can be extended to the comprehensive category, where each bandwidth allocation scheme will be grouped

appropriately. Based on this method, we further divide the schemes with statistical multiplexing into interleaved polling schemes and non-interleaved schemes. Besides, the schemes with relative QoS assurances are grouped into OLT-ONU decisions and OLT decisions (Figure 2.8).

In EPON networks, multiple ONUs can statistically share the upstream transmission link without any privilege, or the individual ONU can request for the specific bandwidth according to its transmission requirements. Accordingly, the DBA schemes for EPONs are classified into schemes with statistical multiplexing and schemes with QoS assurances.

2.2.1 DBA Schemes with Statistical Multiplexing

According to the transmission mechanism of the downstream GATE messages, the DBA schemes with statistical multiplexing can be further divided into interleaved polling schemes and non-interleaved schemes. In the interleaved schemes, the GATE message transmission is overlapped with the upstream data transmission. It will not take extra time to transmit the downstream polling messages in such a scheme. Contrarily, there is no interleaved transmission for GATE messages in the non-interleaved scheme. The OLT sends out the GATE message simply upon receiving the REPORT message from the corresponding ONU. So the ONUs will have some idle time to wait for the polling messages transmitted from the OLT.

2.2.1.1 Interleaved Polling Schemes

A typical interleaved scheme called Interleaved Polling with Adaptive Cycle Time (IPACT) is proposed in [27] and [28] for the dynamic bandwidth allocation in EPONs. This OLT-based polling scheme is similar to the hub polling, where next ONU is polled by a GRANT message before data packets from the previous ONU have arrived. The OLT distributes time slots with dynamic window size according to the instantaneous amount of packets buffered

in ONUs reported by the REQUEST messages. All ONUs in IPACT share the upstream transmission link statistically.

In the IPACT scheme, the OLT maintains a polling table containing each ONU's buffer length and the RTT. Upon receiving the GRANT message from the OLT, the ONU starts sending its data up to the size of the granted window. At the end of its transmission window, the ONU generates a REQUEST message and transmits it together with the data packets to the OLT. When the REQUEST message arrives at the OLT, it will be used to update the polling table. By keeping track of the time when GRANT messages are sent out and data packets are received, the OLT can constantly update the items of buffer length for the corresponding ONUs in the polling table, and can poll each ONU without collisions. In IPACT, the OLT grants for the ONUs based on their requests in the previous polling cycle. This leads to the undesirable extra waiting delay experienced by packets arriving between the two successive report times.

In order to improve the performance of IPACT regarding to the waiting delay, [29] presents an Estimation-Based DBA (EB-DBA) algorithm. In this algorithm, the amount of packets arriving between the two consecutive requests generation time is estimated by the ONU and reported to the OLT. Then the OLT decides the granted transmission size for the ONU in the next cycle based on this estimation. By this approach, the grant size approximates to the ONU's buffer occupancy when the ONU is polled for transmission on receiving the GRANT message. Packets can experience a lower waiting delay in the EB-DBA algorithm than in IPACT. While, it is difficult to achieve a quick response to the traffic variation, especially for the highly variable traffic in EPON networks. So the accuracy of the estimation is also difficult to obtain. Moreover, it comes to the additional cost for the ONUs to keep the information of multiple variables for the estimation.

To achieve the higher utilization and support the max-min fairness under the non-uniform traffic, [30] proposes a scheme to allocate the timeslot in consideration of other ONUs queue occupancy. The authors develop two algorithms accordingly. The first algorithm makes use of the recently granted timeslot of other ONUs. It allows the OLT to grant the unused timeslot of other ONUs to the current ONU to meet its request. This algorithm relaxes the maximum timeslot restriction to improve the utilization. While, this algorithm may lead to the fairness problem because the new coming ONU will get much less bandwidth than existing ONUs. The second algorithm aims to achieve the max-min fairness for ONUs. It uses the latest requested queue length of other ONUs and allows the OLT to grant before all requests in the current cycle are collected. This algorithm assumes that there is little difference of the queue information between two successive cycles. This assumption is quite doubtful in EPONs with burst traffic or highly variable traffic.

[31] introduces a simple work for the multiple access control in EPONs. Similar to IPACT, the GRANT message transmission is interleaved with the data packet transmission. The fixed Maximum Transmission Window (MTW) allocation is provided as a simple example. Besides, it builds the detailed models for all components in EPONs for the experimental simulation.

2.2.1.2 Non-Interleaved Schemes

A polling mechanism with threshold is proposed in [32] for the Fiber-To-The-Home (FTTH) EPON networks. Since the propagation time is relatively low due to the short physical distance ($\leq 1km$) between the OLT and ONUs, the authors suggest that there is unnecessary to deploy the interleaved scheme in the FTTH EPONs. In the proposed mechanism, the OLT sends a GATE frame to the ONU after receiving the data and End of Transmission (EOT) frame from the previous ONU. Each ONU uses the threshold to limit the transmission

amount of packet in each cycle. This mechanism is very simple to be implemented in FTTH EPONs. While considering the distance between the OLT and ONU in EPONs can reach 20 km according to the IEEE 802.3ah EFM Task Force, such a mechanism is not suitable for EPON networks with multiple ONUs and long distances.

In the star topology in [33], a Transmission Upon Reception (TUR) scheme is presented to ensure both downstream and upstream transmissions in EPONs. It allows the ONU to transmit data packets with the amount proportional to the amount it receives from the OLT. In TUR, each ONU keeps a start point indicator to ensure the collision-free access to the upstream transmission link. Thus the control of the OLT is loose in TUR compared to other centralized DBA schemes. Although TUR takes some actions for the OLT to periodically detect the buffer status of ONUs, it still creates the unavoidable fairness issue because the ONUs may have more data to transmit than to receive from the OLT. Moreover, TUR is not compatible with the MPCP protocol because it does not support synchronization between the OLT and ONUs.

In conclusion, the interleaved schemes can effectively improve the channel utilization by reducing the ONUs waiting time for the GATE messages. While they require the OLT to carefully set the time for sending out the GATE message, so that the ONU can be activated for transmission by the message without idle time. For a non-interleaved scheme, the operation of the OLT is relatively simple because no interleaving is required for the GATE message transmission. However, such schemes cannot achieve the high utilization because it will take the additional time to transmit downstream GATE messages to ONUs.

Consisting of both the interleaved and non-interleaved schemes, the DBA schemes with statistical multiplexing provide the common service to all ONUs without difference. This group of DBA schemes is easy to be designed and implemented for they don't require meeting the different transmission requirements of end users. However, such schemes are

not suitable for the future emerging access networks, where different ONUs request for differentiated service requirements.

2.2.2 DBA Schemes with QoS Assurances

In order to meet the different transmission requirements, EPON networks should deploy the DBA schemes that can guarantee QoS for different end users. Recently, a variety of DBA schemes with QoS assurances are proposed in the literature, some of which can guarantee the absolute service for ONUs or users. While most of the schemes only consider providing the relative QoS assurances.

The schemes with absolute QoS assurances usually consider the Service Level Agreement (SLA) contracts between the service provider and end users. They can support the accurate service just as the contracts specify; in other words, end users will be satisfied with the specified transmission requirements, such as the minimum bandwidth allocation and the maximum tolerable waiting delay. In the contrary, the schemes with relative QoS assurances can only provide end users with the probable service without the quantitative guarantee, because they seldom take SLA into account for the bandwidth allocation.

2.2.2.1 Schemes with Absolute QoS Assurances

The authors propose a Bandwidth Guaranteed Polling (BGP) scheme for EPONs in [34] and [35], which takes the SLA contracts into consideration in the bandwidth allocation scheme design. In BGP, the ONUs are divided into two disjoint groups as bandwidth guaranteed (BG) ONUs and bandwidth non-guaranteed (non-BG) ONUs according to the SLA between the service provider and users. The OLT maintains a scheduling Entry Table and a List to determine the polling sequence of entries and the polling order of non-BG ONUs respectively. Table entries are similar to time slots in a TDM system, which will be either allocated to BG ONUs or dynamically assigned to non-BG ONUs. The bandwidth guarantee

is achieved for a BG ONU by allocating one or more bandwidth units/entries to it based on the SLA contracts. While, the non-BG ONUs can only be provided with the best-effort service without bandwidth guarantee. Entries in the Table that are not occupied by BG ONUs can be assigned to non-BG ONUs dynamically. In addition, a superfluous transmission window (unused portion of the bandwidth unit/entry) that is occupied yet unconsumed by an ONU may also be assigned to a non-BG ONU dynamically.

BGP deploys the individual polling in the bandwidth allocation, where the OLT grants the transmission window to an ONU without waiting for the REPLY messages from all ONUs. The BGP scheme ensures the BG ONUs receiving the bandwidth guarantee specified by the SLA contracts. It can also achieve the statistical multiplexing for traffic of non-BG ONUs in the unallocated bandwidth units/entries and the superfluous transmission windows. By combining the priority-based mechanisms employed by the ONUs, [36] and [37] extend the BGP scheme to support the DiffServ for EPONs.

[38] and [39] develop the duel DEB-GPS scheduler for EPONs where the OLT and ONUs employ the Deterministic Effective Bandwidth (DEB) admission control and resource allocation together with the Generalized Processor Sharing (GPS) scheduling. There are multiple queues including QoS and best-effort queues in each ONU. The incoming QoS traffic is constrained by leaky bucket parameters. By using the GPS scheme, each QoS traffic achieves a minimum service rate DEB with delay-constraint and without loss. The surplus bandwidth can be assigned to the best-effort traffic based on the weight. DEB-GPS uses the joint polling, where the master scheduler in the OLT collects requests from all ONUs to calculate the fixed bandwidth assigned to all QoS traffic and the dynamic bandwidth to the best-effort traffic in each ONU. The slave scheduler in the ONU further decides the bandwidth allocated to each queue according to the weight.

In DEB-GPS, the individual QoS traffic can get the deterministic guaranteed service with bounded-delay and lossless transmission. The surplus bandwidth can also be utilized by the best-effort traffic efficiently. While there is the increased complexity to conduct the admission control and update the weight of the traffic's DEB value, leading to the extra overhead to EPON networks.

In [40], the OLT allocates bandwidth to each ONU according to the contracted rate. If an ONU requests less bandwidth than its contracted rate, its remainder bandwidth will be allocated to other ONUs proportional to their contracted rates. In this mechanism, the ONU's REQUEST transmission is separate from its data transmission, which is not compatible with MPCP. Furthermore, the separate transmission of REQUEST and data packets leads to the low upstream utilization due to the double propagation time per ONU.

By incorporating the SLA contracts into bandwidth allocation decisions, the DBA schemes with absolute QoS assurances provide the deterministic services to ONUs that can satisfy their specific transmission requirements. In the future access networks, both the service provider and end users will prefer such DBA schemes to ensure the contracted services. However, these schemes may lead to the implementation complexity with regard to the accurate scheduling design. In the following section, the DBA schemes with relative QoS assurances will be addressed whose implementation is generally simpler compared to the schemes with absolute QoS assurances, but they will only provide the rough QoS guarantee to end users.

2.2.2.2 Schemes with Relative QoS Assurances

In EPONs with multiple types of traffic, the ONU receives traffic from the connected users and transmits different types of traffic when it is polled by the OLT. The OLT can make the bandwidth allocation among the multiple types of traffic for the ONU; either, the ONU can

schedule the transmission for its own traffic. Hence, there are two kinds of DBA schemes with relative QoS assurances: the schemes with OLT-ONU decisions and the schemes with OLT decisions. In the schemes with OLT-ONU decisions, both the OLT and ONUs participate to decide the bandwidth allocated to each type of traffic. The OLT only needs to issue one grant per ONU in each polling cycle and the ONU further assigns the bandwidth to its traffic. In the schemes with OLT decisions, ONUs are not the allocation determinant and the OLT alone decides the bandwidth allocated to each ONU as well as to each individual traffic. The OLT will issue multiple grants to each ONU corresponding to the multiple types of traffic. Actually, some DBA schemes allow applying both the OLT-ONU and OLT decisions for the bandwidth allocation. Such schemes are difficult to be grouped. Here we only classify the relative DBA schemes into the two groups roughly.

2.2.2.2.1 Schemes with OLT-ONU Decisions

[41] extends the IPACT scheme to support differentiated classes of service in EPONs. It focuses on how the message transmission mechanism (MPCP) and the bandwidth allocation scheme (IPACT) can be combined with the priority scheduling. The MAC mechanism is divided into inter-ONU scheduling and intra-ONU scheduling, where IPACT is adopted as the inter-ONU scheduling scheme. The Strict Priority Queuing (SPQ) scheme is deployed for the intra-ONU scheduling in each ONU. By SPQ, packets in each ONU are transmitted strictly according to their priorities. The lower-priority traffic will only be served after the higher-priority traffic being transmitted. Besides, a newly arriving higher-priority packet can preempt lower-priority packets when the finite buffer is not enough to accommodate it. SPQ will result in the performance polarization between different classes of traffic in the same ONU, where the higher-priority traffic gets better-than-required services while the lower-priority traffic starves at the high load. Furthermore, the unexpected network performance of

light-load penalty occurs, where the delays for some classes of traffic increase with the decrease of the network load.

To overcome the light-load penalty, the authors suggest two mechanisms named the Two-Stage Queuing (TSQ) scheme and the Constant-Bit-Rate Credit (CBRC) scheme for the intra-ONU scheduling. In TSQ, packets are sequenced in the first stage of priority queues when arriving at the ONU. When it is the turn for transmission, the ONU advances packets from the first stage to the second stage of First-In-First-Out (FIFO) queue before transmitting them over the upstream link. The new arriving packet is not allowed to replace the existing packets in the FIFO queue. The CBRC scheme predicts the number of higher-priority packets arriving between the consecutive REPORT and GATE message, so that the granted window will be large enough to carry all reported packets in the buffer. Both schemes can avoid the light-load penalty effectively.

To enhance the performance of the extended IPACT scheme described in [41], [42] proposes the Dynamic Polling Order Arrangement (DPOA) scheme for the inter-ONU scheduling and Priority with Insertion Scheduling (PIS) scheme for the intra-ONU scheduling respectively. In DPOA, the polling order is variable for different cycles, which is decided by the current queue length at each ONU. The OLT arranges the polling order at the end of each cycle. The variable polling order can achieve efficient bandwidth utilization under the unbalanced traffic from different ONUs. In PIS, the ONU will not schedule transmission among the traffic flows merely based on the strict priority. Being polled by the OLT, the ONU first sends all the highest priority packets. Then it considers transmitting some packets of the non-real-time traffic before packets of the real-time traffic, as long as the real-time traffic can be satisfied with the QoS requirements. PIS can achieve the fairness to some extent by improving the performance of the non-real-time traffic. While, PIS should

consider carefully about the transmission amount of non-real-time packets in the ONU per cycle, in case that the performance of real-time traffic might be impaired.

[43] considers the OLT-level and ONU-level bandwidth allocation corresponding to the inter-ONU and intra-ONU scheduling respectively. The OLT-level allocation deploys a frame-based Weighted Round-Robin (WRR) scheduling scheme where the frame sizes are fixed. The idle capacity of the under-loaded ONUs can be shared in a weighted manner among all over-loaded ONUs. The ONU-level allocation is based on the Start-time Fair Queuing (SFQ) scheme where packets are transmitted in the increasing order of the time-stamps. It reduces the implementation complexity by only maintaining the time-stamp for the Head-Of-Line (HOL) packet in each queue. This scheme can provide the fair scheduling between different traffic flows in the same ONU, while it does not take the traffic's specific QoS requirements into consideration.

Combined with MPCP, [44] and [45] present a bandwidth allocation algorithm by using the threshold reporting method for the individual queue in ONUs. Several threshold values, which can be assigned statically or dynamically, are associated to each queue to determine the Queue Report (QR) fields in the REPORT message. Upon receiving REPORT messages from ONUs, the OLT updates the information table according to the QR values and calculates the bandwidth allocated to each ONU by the relative algorithm. The ONU deploys the Full Priority Scheduling (FPS) and Interval Priority Scheduling (IPS) to transmit traffic from different queues, which are similar to the SPQ and TSQ respectively. However, this method will result in the implementation complexity for ONUs to assign multiple threshold values when generating the REPORT message. Furthermore, the OLT should maintain the additional table to keep the information of the individual queue's request of the ONU, which will cause the extra cost.

A hierarchical DBA scheme is proposed for the OLT's scheduler and ONU's scheduler in [46]. Four classes of priority traffic are considered in the scheme. For the OLT's scheduler, bandwidth is first guaranteed for class $P0$, $P1$, and $P2$. Then it performs the dynamic allocation for class $P1$, $P2$ and $P3$ with the consideration of queue length and priority. The OLT only sends the aggregated bandwidth to the ONU through the control message. Then the ONU's scheduler performs the similar algorithm with that used by the OLT's scheduler, to allocate the bandwidth among different classes of traffic. This scheme leads to the expense that both the OLT and ONUs perform the algorithm to calculate the bandwidth allocation.

Different algorithms are exploited for the inter-ONU scheduling and intra-ONU scheduling respectively in all the above DBA schemes. In [47] and [48], both levels of scheduling are considered as a whole and only one scheme named the Fair Queuing with Service Envelopes (FQSE) is adopted for the hierarchical scheduling. A Service Envelope (SE) represents the time slot size given to the node. The ONU generates a SE from the SEs of its users and sends to the OLT in the REPORT message. Then in a hierarchical manner, the OLT calculates the time slot start time for each ONU and the ONU decides the transmission start time for its users. FQSE can achieve the fairness among all end users, which is called the cousin-fair by the authors.

2.2.2.2.2 Schemes with OLT Decisions

[49] and [50] first propose a priority-based algorithm for the intra-ONU scheduling to overcome the light-load penalty in SPQ. Being polled by the OLT, an ONU first transmits packets arriving before it sent out the REPORT message based on their priorities. If all such packets are served and the current transmission window can carry more packets, then packets arriving after the report time can be transmitted. This is actually the SPQ scheme with the

gated service. It can improve the fairness among all classes of traffic by allowing them access to the channel as report to the OLT.

Then the authors present a dynamic bandwidth allocation scheme where the OLT assigns the guaranteed bandwidth to ONUs in proportion to the SLA requirements. The excess bandwidth is further distributed among the heavy-loaded ONUs proportional to their requests. In this scheme, the OLT should decide the allocation to all ONUs after getting REPORT messages from them, which will result in some idle time when the ONUs are waiting for the GRANT messages. In order to compensate the idle time, the authors deploy a gate-ahead mechanism where the light-loaded ONUs are scheduled for transmission early with the request size, without waiting until all REPORT messages reach the OLT. Within the allocated time slot, the ONU can request the OLT to decide the bandwidth for each class of traffic, which may lead to the additional burden on the OLT. An ONU should report individual queue status to the OLT, and the OLT will generate multiple transmission grants to the ONU for different queues. That also results in the excess cost of multiple requests and multiple grants for an ONU in each polling cycle.

A cyclic polling scheme is suggested in [51] and [52] that is typically a joint polling scheme. The OLT grants transmission for each of the three priority queues in ONUs at the beginning of each cycle. The high-priority traffic is first assigned the fixed bandwidth in the scheme. Then the remainder bandwidth is allocated to the medium-priority traffic flows trying to meet their requests. If there is still bandwidth left, it is distributed between all low-priority traffic flows in proportion to their requests. The OLT has the extra burden to control both the scheduling between ONUs and the scheduling within an ONU. It also leads to the expense that ONUs report each queue's status to the OLT and the OLT issues multiple grants to each ONU in a cycle. Finally, the ONUs with only low-priority traffic will starve for transmission at the high system load by the scheme.

[53] introduces a Hybrid Slot Size/Rate (HSSR) scheme to minimize the traffic's delay variation. HSSR divides a frame into two parts, the steady part and the dynamic part, for the transmission of two classes of traffic. The steady part of the frame is used to carry the high-priority traffic, where the fixed time slot is allocated to each ONU for the high-priority traffic transmission according to the user's subscription rate. Thus the minimum delay variation can be achieved for the high-priority traffic. The dynamic part of the frame is allocated to transmit the best-effort traffic in ONUs. If the high-priority traffic cannot consume the fixed time slot, the surplus bandwidth can be assigned to the best-effort traffic in the same ONU. This is a drawback that other ONUs cannot share the surplus bandwidth. In the case that the rate of high-priority traffic exceeds its subscription rate so that the fixed time slot is not enough for transmission, the excess high-priority traffic will be redirected into the best-effort queue and be transmitted in the dynamic part of the frame. This redirection will definitely degrade the quality of service received by the high-priority traffic.

Based on the similar idea with [53], authors of [54] and [55] propose a Two-Layer Bandwidth Allocation (TLBA) scheme including the class-layer and ONU-layer allocation for EPONs. The class-layer allocation divides the frame into three sub frames for three classes of traffic. The sub frame sizes are variable according to the requests of the corresponding service classes. Using the weight, a bandwidth threshold is set for each class of traffic to guarantee the minimum bandwidth. The surplus bandwidth of the light-loaded traffic can be distributed to other classes of traffic proportional to their weights. In the ONU-layer allocation, the bandwidth of each traffic class is distributed to all ONUs following the max-min fairness principle: the allocation complies with the increasing order of the ONUs requests and no ONU can be allocated more bandwidth than its request. The remained bandwidth is evenly shared among the unsatisfied ONUs.

For HSSR and TLBA, an ONU should separately transmit different classes of traffic by multiple segments in one polling cycle. Each segment of transmission experiences one propagation time. Moreover, a guard time is required for two consecutive transmission segments from different ONUs. Consequently, HSSR and TLBA will increase the overhead of the upstream propagation time and lead to the extra guard time, resulting in very low channel utilization, especially for EPON networks with many classes of traffic.

In [56], a Burst-Polling (BP) based Delta Dynamic Bandwidth Allocation (DDBA) scheme is proposed for the inter-ONU scheduling. There are three classes of traffic, Expedited Forwarding (EF), Assured Forwarding (AF) and Best Effort (BE), in the system. The OLT sends multiple GATE messages to ONUs in a burst manner after receiving REPORT messages from all ONUs. In DDBA, the requested bandwidth of EF and AF traffic can be adjusted by the request difference between two consecutive cycles. In the priority-based intra-ONU scheduling method, two different sets of bandwidth weight, for normal and high EF profile respectively, are applied to get the grant size for each class of traffic in the same ONU. However, it results in the additional expense because each GATE and REPORT message consists of three grants and reports corresponding to each class of traffic in ONUs.

A two-step scheduling scheme is presented in [57] to reduce the scheduling complexity. It separates the process of the transmission start-time decision from the process of the GATE message generation. In the first step, the grant scheduler converges four kinds of GATE to produce the grant size for each ONU. Then in the second step, the scheme decides the grant start-time considering the grant size and RTT of each ONU. The OLT's configuration will be very complex and the extra burden is exerting on the OLT to maintain the multiple GATE queues.

[58] introduces a Dynamic Credit Distribution (D-CRED) scheme in order to improve the utilization of EPONs. It uses two separate requests in one REPORT message including

the buffer occupancy and the MAC frame boundary. The OLT also has two grants for one ONU per cycle: one is for data packet transmission and another is for the REPORT message transmission. According to the requests of ONUs, the OLT dynamically changes the credit value that represents the assigned bandwidth to each ONU for data transmission. In order to support QoS in EPONs, the authors apply a satisfaction-based credit allocation and a weighted-fair credit allocation method for the fair allocation based on D-CRED. While, it is questionable that D-CRED can achieve the higher utilization than other schemes, because the separate requests and grants will lead to the extra overhead of control message transmissions.

Generally, in the schemes with OLT-ONU decisions, the OLT's operation can be simplified by appointing ONUs to schedule transmissions among different types of traffic. The GATE and REPORT message only need to contain one grant and one request for each ONU. However, the operation of ONUs will be complicated accordingly. In addition, because the OLT may not consider the status of each queue when making the decision, the bandwidth allocation results may be unfair to the individual traffic. For the schemes with OLT decisions, an ONU's implementation can be very simple because it only needs to transmit traffic according to the OLT's decision. Each type of traffic is more possible to obtain the fairness compared to the schemes with OLT-ONU decisions. While there is extra expense because an ONU is required to report each queue's status to the OLT, and the OLT issues multiple grants to the ONU in each polling cycle. Furthermore, the OLT is exerted additional burden to decide the bandwidth assignment for each individual queue in the ONU.

Table 2.1 gives a summary for all categories of existing centralized TDMA-based DBA schemes for the EPON networks.

Chapter 2. Literature Review

Table 2.1 Summary of centralized TDMA-based DBA Schemes in EPONs

Category		Strength	Weakness
Schemes with statistical multiplexing	Category common features	<ul style="list-style-type: none"> - multiple ONUs statistically share the upstream transmission link without any privilege - provide the common service to all ONUs without difference - easy to design and implement 	<ul style="list-style-type: none"> - cannot meet different transmission requirements of end users - not suitable for the future emerging access networks
	Interleaved polling Schemes	<ul style="list-style-type: none"> - the downstream polling message transmission is overlapped with the upstream data transmission - will not take extra time to transmit the downstream polling messages - effectively improve the channel utilization 	<ul style="list-style-type: none"> - the OLT is required to carefully set the time for sending out the downstream polling messages
	Non-interleaved schemes	<ul style="list-style-type: none"> - no interleaved transmission for downstream polling messages - the operation of the OLT is relatively simple 	<ul style="list-style-type: none"> - the ONUs have to wait for the polling messages transmitted from the OLT - the channel utilization is low
Schemes with QoS assurances	Category common features	<ul style="list-style-type: none"> - individual ONU can request specific bandwidth to meet its transmission requirements - can guarantee QoS for different end users 	<ul style="list-style-type: none"> - relatively complicated to design and implement the schemes
	Schemes with absolute QoS assurances	<ul style="list-style-type: none"> - can support the accurate service as the SLA contracts specify - provide the deterministic services to ONUs that can satisfy their specific transmission requirements 	<ul style="list-style-type: none"> - may lead to the implementation complexity
	Schemes with Relative QoS assurances	Category common features	<ul style="list-style-type: none"> - only provide the probable QoS without quantitative guarantee
		OLT-ONU decisions	<ul style="list-style-type: none"> - both the OLT and ONUs participate to decide the bandwidth allocation - the OLT only needs to issue one grant per ONU and the ONU further assigns the bandwidth to its traffic - the OLT's operation is relatively simple - the control messages only need to contain one grant and one request for each ONU
		OLT decisions	<ul style="list-style-type: none"> - extra expense for ONUs to report each queue's status to the OLT - the OLT is required to issue multiple grants to an ONU - the OLT needs to decide bandwidth assignment for each individual queue

2.3 Summary

Firstly, this chapter has reviewed the PON technology that is considered to be the best choice for the next-generation access networks. Among the three chief solutions of PON, EPON has gained more and more attentions due to its attractive advantages by combining the low-cost Ethernet with the high-capacity PON. Numerous researches have been focused on the MAC protocol design for the EPON upstream transmission, among which the centralized TDMA-based DBA schemes are the most promising ones.

Then we have thoroughly investigated the DBA schemes proposed for EPONs up to now. A classification has been presented to divide the numerous DBA schemes into corresponding groups. Besides the description and comments for each individual scheme, this chapter has also provided the common features as well as the merits and shortcomings for each category of DBA schemes. According to the comprehensive survey of these DBA schemes, an enhancement to the famous IPACT scheme will be proposed for the EPON upstream transmissions in the next chapter.

Chapter 3 . A Proposal to Enhance the IPACT Scheme

Proposed in [27] and [28], the Interleaved Polling with Adaptive Cycle Time (IPACT) scheme is widely regarded as a pioneer to dynamically allocate bandwidth to different ONUs in EPON networks. It combines Time Division Multiple Access (TDMA) and statistical multiplexing for the upstream channel sharing among multiple ONUs. By means of the control message transmission, the IPACT scheme introduces an interleaved polling approach where the next ONU is polled before the transmission from the previous one has arrived, which can provide efficient upstream utilization for EPONs. IPACT has built a foundation for the research of bandwidth allocation mechanisms, based on which lots of other DBA schemes ([29-58]) have been developed for EPONs.

Although the IPACT scheme has the great significance in EPONs, it has the weakness of prolonging the traffic waiting delay due to the discrepancy between the grant size and the instant buffer occupancy. In the IPACT scheme, the OLT grants the transmission window to an ONU based on its request in the previous polling cycle, instead of the instantaneous buffer occupancy when the ONU is polled. Hence the new packets arriving after the request

time will be delayed for transmission at least to the next cycle, leading to the inevitable extra waiting delay experienced by the packets, especially at the light load.

In this chapter, we will first have a detailed review on the IPACT scheme. Besides the full description of the interleaved polling algorithm, we will also present the combination of IPACT and the Strict Priority Queuing (SPQ) mechanism supporting the Differentiated Services (DiffServ) in EPONs, and the undesirable phenomenon of light-load penalty accompanying with the combined solution. Then we will present a proposal to enhance the performance of IPACT, named as the IPACT with Grant Estimation (IPACT-GE). By estimating the amount of new packets arriving between the two consecutive pollings, the OLT will grant to the ONU with the additional estimated size besides the ONU's request size in the previous cycle. Thus the discrepancy between the grant size and the ONU's instant buffer length will be diminished in IPACT-GE, and the performance can be efficiently improved comparing to IPACT. Another advantage of the IPACT-GE scheme is that it can greatly mitigate the light-load penalty when combined with the SPQ mechanism to support DiffServ for EPONs. The performance evaluation shows that, the IPACT-GE scheme can obviously shorten the average waiting delay and buffer occupancy of IPACT at the light load. It also demonstrates the reduced light-load penalty phenomenon achieved by the combined solution of IPACT-GE and SPQ.

3.1 System Model

In this chapter, we consider a tree-based topology (Figure 3.1) for the EPON system as that deployed in [27]. All transmissions in EPON are performed between the OLT and ONUs. Therefore, in the downstream direction (from OLT to ONUs), EPON is a point-to-multipoint network, and in the upstream direction (from ONUs to OLT) it is a multipoint-to-point network. Because Ethernet is broadcast by nature, it fits perfectly with the EPON

architecture in the downstream direction: packets are broadcast by the OLT and extracted by their destination ONU based on the Media Access Control (MAC) address. In the upstream direction, the ONUs should share the channel capacity and resources by an efficient MAC scheme to avoid collision.

Figure 3.1 A typical EPON network based on tree topology

The TDMA approach is a preferred method of channel sharing in an access network since it allows using a single upstream wavelength and results in a very cost-effective solution. However, the lack of statistical multiplexing in TDMA approaches will result in unexpected situations where some time slots overflow under very light load, and other time slots remain underutilized even if the traffic load is very high. This will lead to the EPON bandwidth being underutilized. A dynamic bandwidth allocation scheme is necessary for the upstream transmission to reduce the time slot size when there is no data for the ONU, and the excess bandwidth can be used by other ONUs. IPACT is such a DBA scheme that provides statistical multiplexing for ONUs and results in efficient upstream channel utilization in EPONs.

3.2 Review of The IPACT Scheme

IPACT uses an interleaved polling approach similar to the hub polling, where the next ONU is polled by a GRANT message before data packets from the previous ONU have arrived. The OLT distributes time slots with dynamic window size according to the instantaneous frame number buffered in ONUs reported by the REQUEST messages. All ONUs in IPACT share the upstream transmission link statistically.

3.2.1 Interleaved Polling Algorithm

In IPACT, the OLT maintains a polling table containing the buffer occupancy and Round-Trip Time (RTT) for each ONU [27][28]. Figure 3.2 illustrates the principle of the interleaved polling algorithm. Assume that the OLT knows exactly how many bytes are waiting in each ONU's buffer and the RTT to each ONU from the polling table. At some time, the OLT sends a GRANT message to ONU1 allowing it to send 6000 bytes as the polling table indicates (Figure 3.2 (1)). Because the OLT broadcasts to all ONUs in the downstream direction, the GRANT message should contain the Identification (ID) of the destination ONU as well as the size of the granted window (in bytes).

Upon receiving the GRANT message from the OLT, ONU1 starts sending its data packets up to the size of the granted window (6000 bytes). At the same time, ONU1 keeps receiving new data packets from the end users. From Figure 3.2 (2), ONU1 will generate a REQUEST message at the end of its transmission window, to tell the OLT its instantaneous buffer length when the REQUEST message is generated (550 bytes in this case).

The OLT knows when the transmission of ONU1 will finish even before it receives the reply from ONU1. Since the OLT knows how many bytes it has granted ONU1 to send data, it knows when the last bit from ONU1 will arrive. In addition, it also knows the RTT for the next ONU (that is, ONU2) from the polling table. So the OLT can schedule a GRANT

Chapter 3. A Proposal to Enhance the IPACT Scheme

message to ONU2 such that the first bit from ONU2 will arrive soon after the last bit from ONU1 arrives at the OLT, with only a small guard time in between (Figure 3.2 (2)). The guard time provides the protection for the fluctuations of RTT, the processing time of ONUs control messages, and the adjustment time of OLT to fit the different distance of each ONU.

Figure 3.2 Interleaved Polling Algorithm in IPACT

After the propagation time and transmission time, the data packets from ONU1 arrive at the OLT, together with a new REQUEST message containing the information of the data amount (in bytes) remained in ONU1's buffer when the REQUEST is generated. The OLT uses this information to update its polling table (Figure 3.2 (3)). By keeping track of the time when GRANT messages are sent out and data packets are received, the OLT can continuously update the items in the polling table regarding the buffer occupancy of the corresponding ONUs, and poll each ONU without collisions (Figure 3.2 (4)).

Similarly, the OLT can calculate the time when ONU2 will finish the transmission. And it will know when to send the GRANT message to the next ONU so that its data packets will be transmitted after the last packet of ONU2. The OLT will update its polling table regarding the entry for ONU2 after the data and REQUEST message from ONU2 arrive (Figure 3.2 (4)). If an ONU completely emptied its buffer, it will report 0 byte to the OLT in the REQUEST message. Correspondingly, the ONU will be granted 0 byte in the next cycle, that is, it will be allowed to send a new REQUEST message without data.

It is clear from the above description that, every ONU executes the same procedure in IPACT driven by GRANT messages received from the OLT. The entire scheduling and bandwidth allocation algorithm is located in the OLT. Thus, it is easy to adaptively change the scheduling at run-time based on some network conditions; the ONUs don't need to negotiate or acknowledge new parameters, nor do they need to switch to new settings synchronously.

3.2.1.1 Control Message Transmission

In the IPACT scheme, the GRANT and REQUEST messages are generated and transferred between the OLT and ONUs, in order to achieve a situation that data transmissions from multiple ONUs arrive at the OLT orderly and without collisions.

These two kinds of control messages only contain two pieces of information: ONU's Node Identification (NID) and requested/granted Window Size (WS). From [27], either the GRANT or REQUEST message in IPACT will be 4 bytes long and will consist of 1 byte Escape Code (ESC), 1 byte NID value, and 2 bytes WS. The downstream GRANT messages can be inserted in the middle of an Ethernet frame or between the frames. The ONU's receiver will recognize the beginning of the embedded control message by reading the ESC code. It then extracts the 3 bytes that follow the ESC byte before passing the rest of the

received data to a standard Ethernet MAC address. The upstream REQUEST messages use the same 4-byte format and they will be sent at the end of the transmission from the corresponding ONU.

The GRANT message to ONU $i+1$ is scheduled one cycle ahead in IPACT, so that its REQUEST message arrives at the OLT after the guard time following the end of the transmission window from ONU i . The interleaved polling can be achieved by this mechanism without collisions. The GRANT message cannot be sent before the previous REQUEST message from the same ONU is received, because the GRANT requires the information contained in the previous REQUEST.

In the IPACT scheme, the order of the REQUEST messages from ONUs is the same as round robin in each polling cycle. The GRANT messages are scheduled with regard to the corresponding RTT and granted window sizes. As a result, the order of GRANT messages may be different from cycle to cycle. Since the order of GRANT messages is determined in a cycle prior to when they are transmitted, it is possible for the OLT to schedule the GRANT to ONU $i+1$ ahead of the GRANT to ONU i . However, that may result in the GRANT scheduling conflict. The conflict occurs when two GRANT messages are scheduled less than a GRANT transmission time apart. To resolve such a conflict, the GRANT that will be scheduled next should be delayed till the end of transmission of the previous GRANT. The delay of the GRANT message transmission has no significant effect on system performance. The only consequence is the increased guard time before the transmission from the corresponding ONU.

3.2.1.2 Granted Transmission Window

If the OLT allows each ONU to send all data packets waiting in its buffer in one transmission, ONUs with high data volume could monopolize the entire bandwidth. In order

to prevent the possible upstream channel monopolization by ONUs with high traffic load, the OLT should limit the transmission size assigned to each ONU: the ONU gets a GRANT to send as many packets as it has requested, but no more than the maximum transmission window size. In IPACT, all ONUs have the same maximum transmission window size W_{max} . The value of W_{max} determines the maximum polling cycle time of the EPON networks under heavy load conditions. A large W_{max} value will result in the increased delay for all packets due to the prolonged polling cycle, while making W_{max} too small will lead to more bandwidth being wasted by guard times.

At the condition of requested window size $W[i] < W_{max}$, the OLT should determine the granted window size to the ONU. Several approaches have been considered in [27], including the fixed, limited, and gated service, etc. The limited service is regarded as the best choice because it can achieve the shortest polling cycle. Being the most conservative scheme, the limited service discipline grants the ONU with the size of the requested data packets as long as it is no more than W_{max} . In the case when the request size exceeds the maximum transmission window, the ONU will only be granted W_{max} .

3.2.2 Supporting of the Differentiated Services (DiffServ)

The IPACT scheme can be combined with the Strict Priority Queuing (SPQ) mechanism to support the Differentiated Services (DiffServ) in EPONs. [41] presents that the MAC layer mechanisms include the inter-ONU scheduling and intra-ONU scheduling in an EPON network with the DiffServ support. While the inter-ONU scheduling is used to allocate bandwidth to multiple ONUs and facilitate the transmission from different ONUs without collisions, the intra-ONU scheduling is responsible for arbitrating the transmission from multiple traffic flows within an ONU. The IPACT scheme is adopted as the inter-ONU scheduling scheme in [41], where the OLT acts as the central controller to schedule the

ONUs one after another with the interleaved polling. The SPQ mechanism is performed by each ONU to control the intra-ONU scheduling.

With the SPQ mechanism, each ONU's buffer is equipped with n queues serving n priority classes (denoted $P_0, P_1, \dots, P_n$) in correspondence with different traffic flows from the end users. Packets with P_0 have the highest priority and those with P_n have the lowest. When a packet is received from a user, the ONU will first classify according to its characteristics and place it in the corresponding queue. The multiple queues in each ONU share the common memory of the buffer. Packets with higher-priority can preempt packets with lower one in SPQ. If a packet arriving with priority P_i finds the buffer is too saturated to accommodate it, it will preempt one or more lower-priority packets P_k ($k > i$) from their queues, so that this new packet can itself be placed into the P_i queue. Between the pollings, an ONU stores all packets received from the users in their respective queues. When polled by the OLT, the ONU will transmit the packets strictly according to their priorities, which means, it cannot serve a lower-priority queue unless all the higher-priority queues are exhausted for transmission.

The combined solution of IPACT and SPQ can effectively provide DiffServ for different traffic flows in EPONs, with higher-priority traffic getting better performance than lower-priority one. However, it also results in the undesirable phenomenon of the light-load penalty, where, for lower-priority traffic, the queuing delay increases with the decrease of the network load. This issue will be addressed in more detail later.

3.3 The Proposed IPACT with Grant Estimation (IPACT-GE) Scheme

As the pioneer of the DBA schemes for EPONs, the IPACT scheme has attracted many attentions since it was proposed originally. Thereafter many other DBA schemes have been

developed for the upstream transmission in EPONs. Though a significant DBA scheme in EPONs, IPACT has the weakness of prolonging the average waiting delay experienced by packets. From the interleaved polling algorithm in the previous section, the OLT grants to the ONU according to its request in the previous cycle. However, when the ONU is polled by the OLT in the current cycle, many new packets have arrived in the ONU and the buffer occupancy exceeds the ONU's request size. The new arriving packets will be delayed for transmission at least to the next cycle, because they have not been reported to the OLT for transmission in the previous REQUEST message. Due to the discrepancy between the grant size and the instant buffer occupancy when the ONU is polled, the average waiting delay experienced by the packets is unnecessarily increased, especially when the load is light.

In order to overcome the weakness of the IPACT scheme, the transmission window granted to an ONU should take into account the amount of new packets arriving during the two consecutive pollings. The amount of new arriving packets can be estimated according to the packet arrival rate at the ONU and the time interval between the two pollings. Based on this idea, an IPACT with Grant Estimation (IPACT-GE) scheme is proposed in this section to improve the performance of the average waiting delay in IPACT.

3.3.1 Weakness of the IPACT Scheme

In the IPACT scheme, the transmission window granted to the ONU is based on its request size in the previous polling cycle. Assume ONU i sends the OLT a REQUEST message informing the entire buffer occupancy $R_i(n)$ in the n^{th} cycle. According to the limited service discipline, the OLT grants to this ONU through the next GRANT message with the granted window size:

$$G_i(n+1) = \min(R_i(n), W_{\max}) \quad (3-1)$$

Figure 3.3 illustrates the transmission of ONU i with the light load in IPACT, where the request amount in the n^{th} cycle is always less than the maximum transmission window $R_i(n) \leq W_{\max}$. So the grant size for the next cycle $G_i(n+1)$ will equal to $R_i(n)$, and ONU i will transmit packets with the exact size $R_i(n)$ in the $(n+1)^{th}$ cycle. However, new packets keep on arriving at ONU i after it sends out the REQUEST message and before it receives the GRANT message. These new arriving packets cannot be transmitted in the $(n+1)^{th}$ cycle because they have not been reported by the ONU in the n^{th} cycle. Instead, these new packets will be reported in the $(n+1)^{th}$ cycle as amount $R_i(n+1)$, and are delayed for transmission until the $(n+2)^{th}$ cycle. The average waiting delay experienced by the packets is unnecessarily prolonged in IPACT because of the specific report/grant mechanism.

Figure 3.3 Illustration of the IPACT Transmission

Let $T_i^G(n)$ indicating the GRANT message arriving time and $T_i^R(n)$ the REQUEST message generation time of ONU i in the n^{th} cycle. The scan time in the n^{th} cycle can be defined as the time interval between the two consecutive REQUEST messages generation time: $T_c(n) = T_i^R(n+1) - T_i^R(n)$. We can see from Figure 3.3 that, in the best case, a packet arrives just before the ONU sends out the REQUEST message at $T_i^R(n)$. It can be reported to the OLT in the n^{th} cycle and will be transmitted after the ONU receives the GRANT

Chapter 3. A Proposal to Enhance the IPACT Scheme

message at $T_i^G(n+1)$ in the $(n+1)^{th}$ cycle. The minimum waiting delay in this case is: $T_i^G(n+1) - T_i^R(n)$, which is a little less than the scan time $T_c(n)$. The worst case is that a packet arrives just after the ONU sends out the REQUEST message at $T_i^R(n)$. Then it should be reported to the OLT in the $(n+1)^{th}$ cycle at $T_i^R(n+1)$, because it cannot catch the time $T_i^R(n)$ for report. So this packet has to be transmitted until the $(n+2)^{th}$ cycle after the time $T_i^G(n+2)$, even be deferred to the time just before the request generation time $T_i^R(n+2)$. The maximum waiting delay in the worst case is: $T_i^R(n+2) - T_i^R(n)$, which is exactly the sum of two scan times: $T_c(n) + T_c(n+1)$.

At the light load, the average packet waiting delay in the IPACT scheme is between $\overline{T_c}$ and $2\overline{T_c}$ according to the above analysis, where $\overline{T_c}$ is the average scan time. The mechanism of reporting the buffer occupancy in advance and granting based on the report size leads to the undesirable extra waiting delay experienced by packets at the light load. If the amount of new packets can be estimated between the two consecutive pollings and the estimation is considered into the grant size to the ONU, then the waiting delay can be shortened at the light load. The reason is that, with the additional grant size, the new arriving packets will have more chances to be transmitted in the same cycle when they arrive at the buffer, without waiting for another cycle.

However, the delay improvement achieved by estimation will not be applicable to the heavy load in IPACT. The grant size to the ONU is always equal to the maximum transmission window W_{max} when the load is high. Hence, although the estimation can be made for the amount of new packets arriving between the two REQUEST messages generation time, the estimated size cannot be added to the grant size because a limited service discipline is used in IPACT.

So far, there is almost no investigation addressing this issue to enhance IPACT except the scheme proposed in [29]. In this Estimation-Based DBA (EB-DBA) scheme, the OLT grants the transmission size to the ONU based on the buffer occupancy and grant size at the previous cycle. Denoting $q(n)$ and $G(n)$ to be the queue length and the grant size at the n^{th} cycle, and $\lambda(n)$ the arrival packet size during the $[n, n+1)$ cycle, then the difference between the grant size and buffer occupancy at the n^{th} cycle can be expressed as: $D(n) = G(n) - (q(n-1) + \lambda(n-1))$. The EB-DBA scheme allocates the ONU with the grant size for the $(n+1)^{th}$ cycle:

$$G(n+1) = \min(G(n) - \beta \cdot D(n), W_{\max}) \quad (3-2)$$

where β is the control gain with $0 < \beta < 2$ for the asymptotically stable equilibrium point. By EB-DBA, packets can experience the lower waiting delay and shorter buffer occupancy than by IPACT. However, the estimation in the EB-DBA scheme is only based on the history values of the grant size and queue length, which may not be so effective to approximate the real buffer occupancy when the ONU is polled. In the following section, we will propose a new scheme to enhance the performance of IPACT with a more powerful estimation method.

3.3.2 Principle of the IPACT-GE Scheme

The weakness of the prolonged waiting delay at the light load in IPACT can be compensated if the grant size to the ONU is more approximate to the instantaneous buffer occupancy when it is polled. Instead of simply granting the request size, the OLT should grant the ONU with the additional size to carry the new packets arriving between the two consecutive REQUEST messages generation time. However, the OLT cannot get the exact amount of

Chapter 3. A Proposal to Enhance the IPACT Scheme

new arriving packets from the REQUEST message generated in the previous cycle. This amount can only be estimated accordingly.

The proposed IPACT with Grant Estimation (IPACT-GE) scheme can reduce the undesirable extended waiting delay experienced by the packets in IPACT. By IPACT-GE, the amount of packets arriving between the two pollings can be estimated for the ONU, and the OLT decides the granted transmission size for the ONU with the consideration of the estimated packet amount and the request amount in the previous polling cycle. By estimating the amount of the new arriving packets and granting the additional window size, the grant size to the ONU will be close to the real buffer occupancy at the moment when it is polled. When the traffic load is low, packets are liable to be transmitted in the same polling cycle when they arrive without waiting for another cycle. The average waiting delay experienced by the packets can be shortened in the IPACT-GE scheme at the light load.

Since the amount of packets arriving within a specific time period can be calculated by the arrival rate multiplied with the time interval, the amount of new packets between the two successive pollings can be estimated for the ONU if the values of arrival rate and scan time are available.

In the IPACT-GE scheme, the ONU is responsible for obtaining the arrival rate and scan time in each cycle for the estimation. The scan time of each polling cycle can be easily recorded by an ONU. It initiates a timer each time when receiving a GRANT message from the OLT. When the next GRANT message arrives at the ONU, it will stop the current timer and obtain the scan time for the previous cycle. At the same time, a new timer is triggered to record the scan time for the next polling cycle. Additionally, the IPACT-GE scheme tries to get the real-time arrival rate for the estimation, which is also obtained by the ONU in each polling cycle. Upon receiving the GRANT message at the n^{th} cycle, the ONU starts to count the size $S_{\Delta}(n)$ of packets arriving during a small time interval ΔT . Then the real-time

Chapter 3. A Proposal to Enhance the IPACT Scheme

arrival rate can be estimated for the n^{th} cycle by: $r(n) = S_{\Delta}(n) / \Delta T$. The smaller is the time interval ΔT , the more is $r(n)$ approximate to the real-time traffic arrival rate, and the more accurate will be the estimated packet amount. In IPACT-GE, each ONU records the scan time $T_c(n)$ and estimate the arrival rate $r(n)$ for the n^{th} cycle, whose values will be used to estimate the amount of packets arriving during the $(n+1)^{th}$ cycle.

By the values of the traffic arrival rate and scan time obtained in the n^{th} cycle, the amount of packets arriving during the $(n+1)^{th}$ cycle can be estimated as: $E(n+1) = r(n) \cdot T_c(n)$. Then the ONU will report this estimated amount $E(n+1)$ to the OLT, together with the request size that equals to the instant buffer length at the n^{th} cycle $R(n)$. The OLT will decide the grant size to the ONU for the $(n+1)^{th}$ cycle taking these two parts into account:

$$G(n+1) = \min(R(n) + \alpha \cdot E(n+1), W_{\max}) \quad (3-3)$$

where α is the estimation factor. The factor α is used to adjust the impact of the estimation to the grant size. If α is set to be too low, the grant size $G(n+1)$ will be decided mainly by the request size $R(n)$ in the previous cycle. Newly arriving packets can hardly be considered in $G(n+1)$ and have to wait for the excess cycle for transmission. Then the performance of IPACT-GE will be close to that of IPACT with very small improvement. In the contrary, if α is too large, the granted transmission window to the ONU will be large enough to carry all packets including the new arriving packets in each cycle. However, a large grant size will result in the low channel utilization, because there will be more unused time slots in the transmission windows when the load is light. Hence, α should be set carefully in the real EPON environment to achieve the balance between the shorter packet waiting delay and higher channel utilization.

Chapter 3. A Proposal to Enhance the IPACT Scheme

At the ideal state, the amount of new arriving packets can be exactly estimated and all packets in the buffer will be transmitted when the ONU is polled. So the buffer occupancy should be zero at the moment when the REQUEST message is generated. That means the request size of ONU in each cycle to be: $R(n) = 0$. In this case, the grant size to the ONU will be determined only by the estimated amount of new packets arriving between the two consecutive pollings: $G(n+1) = \min(\alpha \cdot E(n+1), W_{\max})$.

The process of the IPACT-GE scheme is shown in Figure 3.4 and Figure 3.5, including both operations of the ONUs and OLT.

Figure 3.4 Flowchart of the ONU Operations in IPACT-GE

Figure 3.5 Flowchart of the OLT Operations in IPACT-GE

Other operations of the IPACT-GE scheme are the same with those of IPACT. It uses the interleaved polling approach where the next ONU is polled by the OLT before transmissions from the previous ONU finish. The GRANT and REQUEST messages are generated and transmitted by the OLT and ONUs respectively to facilitate the immediate data transmission of the next ONU following the last bit of the previous ONU. The limited service discipline is adopted for the granted transmission window, where the OLT grants the ONU with the requested packet size plus the estimated new packet amount, as long as it is no more than the maximum window size W_{max} .

3.3.3 Advantages of the IPACT-GE Scheme

The main difference between the IPACT-GE and IPACT scheme is the dynamic time slots distributed to ONUs. In IPACT, the grant size to the ONU is based on the request amount that is the instantaneous buffer occupancy when the ONU sends out the REQUEST message in the previous polling cycle. While in IPACT-GE, the OLT grants the transmission window to the ONU considering the new packets arriving between the two consecutive pollings

besides the instant buffer occupancy in the previous polling cycle. Comparing to the IPACT scheme, the grant size in IPACT-GE is more approximate to the real buffer occupancy when the ONU is polled by the GRANT message. When the load is light, much more packets in IPACT-GE than IPACT can be transmitted in the same polling cycle when they arrive at the ONU. Thus packets can experience a shorter waiting delay in IPACT-GE than IPACT at the light load.

Although the EB-DBA scheme in [29] is also aiming to improve the performance of IPACT by estimation, it only considers the history values of the grant size and queue length. Without estimating the amount of new packets arriving between the pollings and including this amount into the grant size, EB-DBA cannot allocate the transmission window size close to the instant buffer occupancy when the ONU is polled. So the improvement of EB-DBA to the IPACT scheme is very limited. Comparatively, the estimation in our proposed IPACT-GE scheme is based on the real-time arrival rate of the ONU. Taking the amount of new packets into consideration, the estimation is more accurate than that in EB-DBA. The IPACT-GE scheme can enhance the performance of IPACT at the light load more effectively than EB-DBA.

3.3.3.1 Reducing the Average Waiting Delay and Buffer Occupancy

The IPACT-GE scheme can enhance the performance of IPACT with the shortened waiting delay and decreased buffer occupancy at the light load. Figure 3.6 shows an example of the waiting delay that a packet experiences in the EPON network with interleaved polling. The delay d of packet p waiting in the ONU i consists of three parts: $d = d1 + d2 + d3$. The part $d1$ is the time interval between packet p arrival time and the next GRANT arrival time at the ONU. Packet p is reported to the OLT in the next REQUEST message, which can be called REQUEST- p . The part $d2$ indicates the waiting delay from the time the ONU receives the

next GRANT till the time it receives GRANT- p , upon which packet p will be transmitted within the following transmission window. Finally, $d3$ is the delay between the GRANT- p arrival and the departure of packet p , which is insignificant comparing to $d1$ and $d2$.

Figure 3.6 Illustration of the Packet Waiting Delay

Denoting $\overline{T_c}$ to be the average scan time, and ρ_i the traffic load of ONU i , then the average transmission time of ONU i in a cycle is: $\rho_i \overline{T_c}$. The remaining part of the average polling cycle is: $(1 - \rho_i) \overline{T_c}$, during which ONU i is idle and waiting for a poll. So the average time between a packet's arrival and the next GRANT message is: $\overline{d1} = (1 - \rho_i) \overline{T_c} / 2$.

As to the part of delay $d2$, it may span for multiple polling cycles, depending on the packet amount stored in the buffer when packet p arrives. Let L_o to be the buffer occupancy at packet p arrival time (including the size of packet p), $W_{\max}$ the maximum transmission window size. Then the average delay of $d2$ can be expressed as: $\overline{d2} = \left\lceil \frac{L_o}{W_{\max}} \right\rceil \overline{T_c}$.

During the waiting time $d3$, packet p waits to reach the head of the buffer after ONU i is polled for transmission. In order to get the average value of $d3$, the whole EPON system can be approximated to a M/G/1 queue, with the assumption that the packet size follows the arbitrary probability distribution [61][62]. Denoting ρ to be the traffic load of the whole

Chapter 3. A Proposal to Enhance the IPACT Scheme

system, λ the system arrival rate, and $\bar{t}$ the average transmission time for a packet, the

average waiting delay of $d3$ is: $\bar{d3} = \frac{\lambda E(\tau^2)}{2(1-\rho)} = \frac{\rho \bar{t}}{2(1-\rho)}$.

Assuming the total M ONUs averagely separate the system traffic load: $\rho_i = \rho / M$, then the average time that a packet waits in the buffer for transmission is expressed as:

$$\begin{aligned} \bar{d} &= \bar{d1} + \bar{d2} + \bar{d3} \\ &= \frac{\bar{T}_c}{2}(1 - \rho / M) + \left\lceil \frac{L_o}{W_{\max}} \right\rceil \cdot \bar{T}_c + \frac{\rho \bar{t}}{2(1-\rho)} \end{aligned} \quad (3-4)$$

In the IPACT-GE scheme, $\bar{d1}$ and $\bar{d3}$ are the same to those in the IPACT scheme at any traffic load. When the load is heavy, the buffer is saturated and is always full of waiting packets. The average delay $\bar{d2}$ is also same for both IPACT-GE and IPACT scheme:

$\bar{d2} = \left\lceil \frac{L_B}{W_{\max}} \right\rceil \cdot \bar{T}_c$, where L_B is the finite buffer size. So the average waiting delay and buffer

length in IPACT-GE is similar to those in the IPACT scheme at the heavy load.

At the light load, since the IPACT-GE scheme takes into account the amount of new packets arriving during the two pollings, the grant size is more approximate to the instant buffer occupancy when the ONU is polled. The transmission window can carry more packets in each cycle of polling with IPACT-GE than with IPACT. Thus there are fewer packets accumulated in the ONU's buffer, leading to the shorter buffer occupancy in IPACT-GE than that in IPACT at the light load. At the REQUEST generation time in the n^{th} cycle, the buffer length of an ONU in IPACT is: $L_{IPACT} = \overline{r(n)} \cdot T_c(n)$, which is the exact amount of new packets arriving between the pollings. According to Formula (3-3), the buffer length in IPACT-GE is: $L_{IPACT-GE} = \max(\overline{r(n)} \cdot T_c(n) - \alpha \cdot r(n-1) \cdot T_c(n-1), 0)$. Note that $\overline{r(n)}$ is the average arrival rate of the ONU in the n^{th} cycle, and $r(n-1)$ is the estimated arrival rate in

the $(n-1)^{th}$ cycle. Then the difference between the buffer occupancies in these two schemes will be:

$$\begin{aligned} L_{diff} &= L_{IPACT} - L_{IPACT-GE} \\ &= \min\left(\alpha \cdot r(n-1) \cdot T_c(n-1), \overline{r(n)} \cdot T_c(n)\right) \end{aligned} \quad (3-5)$$

According to $\overline{d2} = \left\lceil \frac{L_o}{W_{\max}} \right\rceil \cdot \overline{T_c}$, IPACT-GE will have the smaller delay $\overline{d2}$ than the IPACT scheme at the light load accordingly. Since $\overline{d1}$ and $\overline{d3}$ are same in IPACT-GE and IPACT, the delay difference between the two schemes at the light load will be:

$$\begin{aligned} \overline{d}_{diff} &= \overline{d}_{IPACT} - \overline{d}_{IPACT-GE} \\ &= \left\lceil \frac{\overline{r} \cdot \overline{T_c}}{W_{\max}} \right\rceil \cdot \overline{T_c} - \left\lceil \frac{\max\left((1-\alpha) \cdot \overline{r} \cdot \overline{T_c}, 0\right)}{W_{\max}} \right\rceil \cdot \overline{T_c} \end{aligned} \quad (3-6)$$

where $\overline{r}$ is the average arrival rate of the ONU for all cycles. At the ideal state of the IPACT-GE scheme, the buffer occupancy will be zero when the REQUEST message is generated. Then the difference of the buffer occupancy will be: $L_{diff} = \overline{r(n)} \cdot T_c(n)$, and the delay difference: $\overline{d}_{diff} = \left\lceil \frac{\overline{r} \cdot \overline{T_c}}{W_{\max}} \right\rceil \cdot \overline{T_c}$ for the IPACT and IPACT-GE schemes. Consequently, besides the average buffer occupancy, the average waiting delay experienced by the packets in IPACT-GE is also less than that in IPACT at the light load.

3.3.3.2 Mitigating the Light-Load Penalty

Another great virtue of IPACT-GE is that, it can greatly mitigate the light-load penalty when combined with the SPQ mechanism to support DiffServ in EPONs. The penalty is obvious in the combination of IPACT and SPQ solution. This improvement is achieved mainly by the

Chapter 3. A Proposal to Enhance the IPACT Scheme

IPACT-GE process of granting the additional size to the ONU with estimation besides the request size.

The cause of the light-load penalty with IPACT lies in the discrepancy between the grant size and the instant buffer occupancy. In the IPACT scheme, the OLT grants the ONU with the request size obtained in the previous polling cycle. However, packets keep on arriving between the ONU's REQUEST generation time and the next GRANT arrival time. Newly arrived packets may have higher priority than some packets already stored in the buffer, and they will be transmitted before the lower-priority packets according to the SPQ mechanism. Since these new packets have not been reported to the OLT in the previous REQUEST, the granted transmission window cannot accommodate all packets in the buffer. Hence, some lower-priority packets have to be left in the buffer. This situation will repeat more times at the light load, causing some lower-priority packets to be delayed for more cycles than at the heavy load. That is why the delay for the lower-priority traffic grows with the decrease of the load when the network load is light.

Comparatively, the IPACT-GE scheme takes into account the new packets arriving between the two pollings. Besides considering the request size from the ONU in the previous cycle, the grant size also reflects the estimated amount of the new arriving packets. By IPACT-GE, the granted transmission window is more approximate to the instant buffer occupancy when the ONU is polled for data transmission than by IPACT. So the lower-priority packets are liable to be served within the cycle when they arrive, leading to the shorter waiting time. As a result, the light-load penalty will be greatly reduced in the IPACT-GE scheme.

3.4 Simulation Experiments

Using a powerful simulation tool — OPNET Modeler, an experimental model have been built for the EPON system to evaluate the performance of the proposed IPACT-GE scheme comparing to IPACT. We have conduct two groups of experimental simulations for the performance comparison. The first one compares IPACT-GE with IPACT and EB-DBA with regard to the improvement of average waiting delay and buffer occupancy achieved by IPACT-GE. In the latter group of simulations, the SPQ mechanism is combined with IPACT-GE and IPACT respectively, to show how the light-load penalty phenomenon is greatly mitigated by the IPACT-GE scheme.

OPNET Modeler is running on the Microsoft Windows 2000 platform. To prove the correctness of simulations, we have first simulated the EPON system with IPACT scheme and compared the results with previous related papers. After getting the similar results, we then conducted simulations for IPACT-GE and IPACT accordingly. For each data point in the experiments, we have collected results from multiple simulation trials and average to get the mean value. It is verified that the mean value of results falls within the 95% confidence interval.

3.4.1 Simulation Model

The two groups of experiment deploy the same tree-based simulation model, which consists of one OLT and M ONUs that are connected through a passive splitter. All links in the model are duplex. This model provides the independent downstream channel and upstream channel. The functions of the passive splitter are splitting the traffic stream from OLT to multiple ONUs through the downstream channel, and combining streams from multiple ONUs to the OLT through the upstream channel. The model aims to evaluate the performance of the DBA schemes adopted in the upstream transmission. Since the upstream

channel transmits data independent on the downstream channel, we will only consider the data transmission from ONUs to OLT. The downstream channel is only used to transmit control messages in the simulation.

In view of the upstream transmission, the OLT model in the simulation consists of a processor, sink, transmitter and receiver. The receiver receives the upstream traffic from each ONU and delivers to the processor. If the OLT receives data packets, the processor will send them to the sink. If it receives a REQUEST message from an ONU, the processor will decide the next grant size to this ONU according to the information contained in the message. Besides, the processor is responsible for generating GRANT messages to poll ONUs at proper time according to the adopted DBA scheme. The processor sends GRANT messages to the transmitter, through which they will be further sent to the downstream channel to reach the destination ONUs.

Each ONU model contains data sources, buffer, processor, transmitter and receiver. The sources generate data packets with the inter-arrival time based on the predefined distribution. Those data packets are sequenced in the buffer, waiting for being transmitted to the upstream channel. Once a GRANT message is received by the receiver, it will be delivered to the processor. The processor then starts to transmit data packets up to the grant size. At the end of the data transmission, it will generate a REQUEST message to report the useful information to the OLT. The data packets and REQUEST are both sent to the transmitter, through which they are further sent to the upstream channel to reach the OLT.

3.4.2 Comparison of the IPACT-GE, IPACT, and EB-DBA Scheme

With the same experiments design and system parameters, the first group of simulation will compare the performance of the proposed IPACT-GE scheme to that of IPACT and EB-

DBA in [29]. Since the grant size to the ONU in IPACT is solely based on the request size in the previous cycle, the packets have to wait for the excess cycle for transmission. Although the EB-DBA scheme considers granting based on the estimation, it is based on the history grant size instead of the amount of packets arriving between the pollings. So the improvement of EB-DBA to the IPACT scheme is limited. Contrarily, the proposed IPACT-GE scheme considers the new packets arriving between the consecutive pollings and grants the ONU based on the estimated amount and the request size. So it can obtain more improvement to IPACT than EB-DBA with shorter average waiting delay and buffer occupancy at the light load.

3.4.2.1 Experiment Designs

In the experiments, we design our simulation environment according to the real EPON network and system parameters. A system with 16 ONUs is considered and the transmission speed of the EPON is 1Gb/s. The propagation delay for each ONU is assumed to be randomly (uniformly) over the interval $[50 \mu s, 100 \mu s]$, which corresponds to the distance between the OLT and ONU ranging from 10 to 20 km. The maximum transmission window size W_{max} is set to 15000 bytes. Every ONU has a finite memory size of the buffer. In our simulation, the size is set to 10 Mbytes.

According to IEEE 802.3, the Ethernet frame can be variable from 64 bytes to 1518 bytes. In the simulation, we design packets with size uniformly selected between 64 bytes and 1518 bytes. Each ONU has its own source generating data packets with variable sizes. The inter-arrival time of the packet generation is exponentially distributed with the mean value ranging from 0.101ms to 1.01ms, making total traffic load of the entire network changing from 0.1 to 1.0 accordingly.

For the IPACT-GE scheme, the estimation factor is randomly set to 1. So the grant size to the ONU is: $G(n+1) = \min(R(n) + E(n+1), W_{\max})$. For the EB-DBA scheme, the control gain should be set between 0 and 2 according to [29]. It is set to 1.9 because the best performance can be obtained for EB-DBA with this value of control gain. Then the allocated size is: $G(n+1) = \min(G(n) - 1.9 D(n), W_{\max})$.

3.4.2.2 Simulation Results

Figure 3.7 and Figure 3.8 compare the average waiting delay experienced by the packets in IPACT-GE, IPACT, and EB-DBA schemes respectively. Figure 3.7 shows that, the average delays in all schemes increase rapidly after the system load 0.5 and tend to the balanced high values at the load 0.6. All curves have the similar balanced delay values. That is because after that load, the finite buffer of each ONU is almost full in the IPACT-GE, IPACT, and EB-DBA schemes. As long as an arriving packet is accommodated in the buffer, it has to wait until it becomes the head of the queue for transmission. Since the buffer is always saturated, and it takes almost the fixed time to serve the whole buffer with multiple cycles, thus the average waiting delay for a packet is also fixed after the load 0.6.

From the system load 0.1 to 0.5, average delays in all schemes are much lower than those after the load 0.6, because the buffer of each ONU has not been filled with packets yet. From the two figures, especially Figure 3.8, the average waiting delay in IPACT-GE is much lower than those in IPACT and EB-DBA obviously. The reason is that the IPACT-GE scheme has estimated the amount of packets arriving between two pollings. So the grant size in IPACT-GE is more approximate to the instant buffer occupancy when the ONU is polled for transmission. Packets are not required to wait for additional cycle as in the IPACT scheme, thus they can obtain shorter waiting delay in IPACT-GE at the light load. The EB-DBA scheme can only improve the average delay of IPACT before the load 0.4, and the

Chapter 3. A Proposal to Enhance the IPACT Scheme

improvement is very limited because it has not considered the amount of new packets arriving between the pollings.

Figure 3.7 Comparison of Average Delay Between IPACT-GE, IPACT and EB-DBA

Figure 3.8 Comparison of Average Delay Between IPACT-GE, IPACT and EB-DBA at the Light Load

Figure 3.9 and Figure 3.10 show the comparison of the average buffer occupancy in the IPACT-GE, IPACT, and EB-DBA schemes. In Figure 3.9, the three curves in all schemes

grow fast after the system load 0.5 and tend to the finite buffer size (10M bytes) at the load 0.6. That means the buffer of each ONU is almost saturated after the load 0.6 in all schemes. Between the system loads 0.1 and 0.5, average buffer occupancy in all schemes is much lower than that after the load 0.6.

Figure 3.9 Comparison of Average Buffer Occupancy Between IPACT-GE, IPACT and EB-DBA

Figure 3.10 emphatically shows the improved performance achieved by IPACT-GE. The buffers in IPACT-GE have much shorter average length than those in IPACT and EB-DBA. That is still because the IPACT-GE scheme considers the amount of packets arriving between two polling cycles. Since in IPACT-GE, the grant size is closer to the buffer occupancy at the moment the ONU is polled, more packets can be carried in each transmission window, resulting in the less average buffer length in IPACT-GE than in IPACT and EB-DBA. Because the estimation in the EB-DBA scheme is based on the history value of the grant size and request size, it can only achieve the restricted improvement to IPACT before the load 0.4 as the figures show.

Figure 3.10 Comparison of Average Buffer Occupancy Between IPACT-GE, IPACT and EB-DBA at the Light Load

Figure 3.11 is the average loss rate of all ONUs in IPACT-GE vs. IPACT and EB-DBA. The performances in all schemes are similar to each other. There is no data drop before the system load 0.5, because the buffer of each ONU is not full at the light load. After the load 0.6, the buffer is filled with packets and some of the new arriving packets have to be dropped due to the saturation of the buffer. The loss rate curves gradually grow with the increase of the system load.

Figure 3.12 shows the upstream channel utilization in the IPACT-GE, IPACT, and EB-DBA schemes. The upstream utilization is measured as the ratio of the throughput to the EPON network capacity, where the throughput is a measure of how many data the network can transmit over the upstream channel in a unit of time. Again, the performances in all schemes are similar to each other. Before the system load 0.6, the utilization keeps on growing with the increase of the system load. Then after that load, the curves reach to the highest utilization of the EPON network. That is because, when the buffers are saturated

Chapter 3. A Proposal to Enhance the IPACT Scheme

after the load 0.6, the throughput of the network keeps in the highest value without any increase.

Figure 3.11 Comparison of Average Loss Rate Between IPACT-GE, IPACT and EB-DBA

Figure 3.12 Comparison of Upstream Channel Utilization Between IPACT-GE, IPACT and EB-DBA

From the simulation results, the IPACT-GE scheme can obtain better performance than IPACT and EB-DBA with the less average waiting delay and buffer occupancy at the light

load, which is the chief advantage of the IPACT-GE scheme. The improvement is achieved by the IPACT-GE process of granting the additional amount with estimation to the ONU besides the request size. Additionally, IPACT-GE can obtain more improvement to IPACT than EB-DBA because it takes the better estimation method to consider the amount of new packets arriving between two consecutive pollings. The performances of average loss rate and upstream channel utilization in all schemes are similar.

3.4.3 Comparison of IPACT-GE and IPACT Combined with the SPQ Mechanism

In the second group of experimental simulations, the SPQ mechanism will be combined with the IPACT-GE and IPACT scheme to support DiffServ in the EPON network. While each ONU performs the SPQ mechanism to arbitrate the data transmission from different traffic flows, the OLT will act complying to IPACT-GE and IPACT respectively. In the solution with IPACT, there will be the obvious phenomenon of light-load penalty due to the discrepancy between the grant size and the ONU's instant buffer occupancy. However, the solution with the IPACT-GE scheme can greatly mitigate the light-load penalty with the additional grant size based on the estimation.

3.4.3.1 Simulation Designs

Three classes of traffic are defined in the experiments, which have priority P_0 , P_1 and P_2 respectively. Modeled as the traffic with the highest priority, packets of traffic P_0 have the fixed size of 70 bytes, and the constant inter-arrival time of 125 μs . So traffic P_0 has the constant data rate at 4.48Mb/s. Traffic P_1 and P_2 with lower priorities comply with the exponentially distributed inter-arrival time with packet size ranging from 64 bytes to 1518 bytes. End users generate data packets of traffic P_0 , P_1 and P_2 independently to each ONU. Besides the constant load of traffic P_0 , traffic P_1 and P_2 are supposed to occupy 50% of the

remaining network load. Other system parameters are designed to be the same with the first group of simulations.

3.4.3.2 Simulation Results

The light-load penalty can occur in both the average waiting delay and maximum waiting delay. We have collected the simulation results for both delays in order to show how the light-load penalty, which is obvious in the IPACT solution, is almost eliminated in the IPACT-GE solution.

Figure 3.13 compares the average waiting delay for all classes of traffic in the solutions with IPACT-GE and IPACT respectively. Both solutions can provide the differentiated services effectively, where the traffic with higher priority can obtain the better performance than traffic with lower priority. For traffic P_0 and P_1 , the average delay with both solutions are similar to each other. However, the delays for traffic P_2 with both solutions are different from each other apparently. When the system load is less than 0.6, traffic P_2 with the IPACT-GE solution experiences much lower waiting delay than the corresponding traffic with the IPACT solution. The light-load penalty is obvious in the IPACT solution, where the average delay of traffic P_2 decreases between the load 0.1 and 0.3. While in the IPACT-GE solution, the light-load penalty is almost removed for traffic P_2 .

Figure 3.14 shows the comparison of the maximum waiting delay between the solutions with IPACT-GE and IPACT. The phenomenon of light-load penalty is more apparent in the IPACT solution, with delays for both traffic P_1 and P_2 declining between the load 0.1 and 0.4. Again, there is no such penalty in the IPACT-GE solution for all classes of traffic. The reason of the obvious light-load penalty in the IPACT solution is the discrepancy between the grant size and the real buffer occupancy when the ONU is polled. The IPACT-GE solution greatly mitigates the light-load penalty, because IPACT-GE allocates the ONU with

Chapter 3. A Proposal to Enhance the IPACT Scheme

the additional amount for new arriving packets to diminish the discrepancy between the grant size and the buffer occupancy. So the lower-priority packets in the IPACT-GE solution have better chances to be transmitted earlier than in the IPACT solution, and the light-load penalty is greatly reduced in the IPACT-GE solution.

Figure 3.13 Comparison of Average Delay Between Solutions with IPACT-GE and IPACT

Figure 3.14 Comparison of Maximum Delay Between Solutions with IPACT-GE and

IPACT

The simulation results demonstrate that, the IPACT-GE scheme can greatly mitigate the light-load penalty when combined with the SPQ mechanism to support DiffServ in EPONs. This is another great virtue of IPACT-GE comparing to the IPACT scheme. Additionally, comparing to the IPACT solution, the lower-priority traffic can get better performance with less average and maximum waiting delays at the light load in the IPACT-GE solution.

3.5 Summary

In this chapter, we have reviewed the significant IPACT scheme that is regarded as the fundamental of many other DBA schemes for EPONs. Then its weakness is pointed out by fully analyzing the principle of IPACT, that is, the average waiting delay experienced by packets are unnecessarily prolonged because of the discrepancy between the grant size and the instant buffer occupancy when the ONU is polled for transmission. In order to overcome the IPACT weakness, we have proposed the estimation-based IPACT-GE scheme to improve the performances of average waiting delay and buffer occupancy at the light load. IPACT-GE takes into account the amount of new packets arriving between the two consecutive polling cycles, and decides the grant size to the ONU with the additional estimated amount. Due to the effective estimation, the IPACT-GE scheme can improve the performance of IPACT more efficiently than another scheme, EB-DBA, which is also aiming to enhance the IPACT scheme. Additionally, when combined with the SPQ mechanism, the IPACT-GE scheme can greatly mitigate the light-load penalty that is obvious in the IPACT and SPQ solution.

The extensive experimental simulations have demonstrated that, the IPACT-GE scheme can achieve the shorter average waiting delay and smaller buffer length at the light load, which is the chief enhancement to the IPACT scheme. Moreover, the improvement of IPACT-GE is more obvious than that obtained by the EB-DBA scheme. Finally, the

Chapter 3. A Proposal to Enhance the IPACT Scheme

simulations have also shown the light-load penalty phenomenon being greatly mitigated by IPACT-GE when it is combined with SPQ to support DiffServ in EPONs. This is another great improvement to the IPACT scheme.

Chapter 4 . A Systematic Multiple Access Scheme

We have enhanced the famous DBA scheme IPACT in chapter 3, by reducing the variance between the OLT's grant size and the ONU's instant buffer occupancy to achieve the better performance for EPON networks. However, through the comprehensive survey of the existing Dynamic Bandwidth Allocation (DBA) schemes for EPONs in chapter 2, we understand that a powerful DBA scheme should have the absolute QoS assurances in order to provide the deterministic service to users according to their transmission requirements. Both service providers and end users can benefit from such a DBA scheme because the definite service facilitates the explicit payments between the service providers and subscribers.

Generally, a DBA scheme with the absolute QoS assurances incorporates the Service Level Agreement (SLA) to support the accurate service just as the contracts specify. A SLA is a contract between the service provider and subscribers specifying the QoS level that can be expected [59]. It defines what services the provider will provide to customers and what penalties will be charged if the service provider cannot meet the established goals, including the expected behavior of the service and the parameters for QoS. According to [60], the twin

Chapter 4. A Systematic Multiple Access Scheme

drivers for the SLA management scheme are, first, revenue, i.e. payments from the customers to the service provider, which is generated by carrying traffic with the network infrastructure; and, second, penalties, i.e. payments from the service provider to the customer from SLA violations.

In this chapter, we will propose a novel DBA scheme with absolute QoS assurances named Bandwidth Guaranteed Polling (BGP) within the MPCP framework for EPON upstream transmissions, which takes SLA into consideration for the bandwidth allocation. It can achieve bandwidth guarantee for high-demand users and provide best-effort service to low-demand users. An Even Distribution Algorithm (EDA) will be developed to facilitate the implementation of BGP. We will also analyze the bandwidth utilization for the EPON network employing the BGP scheme. Next a mathematical estimation and Continuous-Time Markov Chains (CTMC) model will be presented separately to numerically analyze the BGP scheme. Based on BGP, we will further propose a parameter-based Call Admission Control (CAC) mechanism to admit customers in the EPON networks. Then we will combine the CAC, EDA and BGP scheme into a novel systematic MAC protocol, where the CAC mechanism processes the requirements of end users in consideration of SLA, the EDA algorithm determines the polling sequence of ONUs, and the BGP scheme performs the scheduling function. Finally, the simulation evaluation will be provided to the BGP scheme and the novel BGP-based MAC protocol by the extensive experiments. The performances of BGP will be compared with those of IPACT to show the efficiency of the proposed BGP scheme. Additionally, the simulation results of BGP will be compared with the results of the mathematical estimation model to verify the rationality of the BGP scheme.

4.1 Network Architecture and System Model

Typically, a tree-based topology is deployed in EPON networks as shown in Figure 4.1. In this architecture, the only OLT connects to M ONUs through a $1:M$ passive splitter. The OLT resides in the local exchange (central office), connecting the optical access network to an IP, ATM, or SONET backbone. The ONUs are located at subscriber premises in the FTTH and FTTB scenarios to provide data, voice, and video services to end users. All transmissions in EPON networks are performed between the OLT and ONUs. There is no direct transmission between ONUs. The system operates in a point-to-multipoint way in the downstream direction. It has multipoint-to-point transmissions in the upstream direction. In the downstream directions, packets are broadcast by the OLT and extracted by their destination ONUs based on the MAC addresses. In the upstream direction, the channel capacity is shared by M ONUs and an efficient bandwidth allocation scheme is required to arbitrate transmissions from multiple ONUs to the OLT. Figure 4.1 exemplifies the upstream transmissions in an EPON network.

Figure 4.1 Upstream Transmissions in the Tree-Based EPON Networks

In our proposal, the ONUs in the EPON network are divided into two disjoint ONU sets. One set contains a number of bandwidth guaranteed ONUs while another contains ONUs that do not require bandwidth-guaranteed service. In the implementation and

deployment, a subscriber premise that supports both types of services can be served with both types of ONUs, which could be two separate pieces of equipment or implemented as two separate blades plugged to the same module, rack or chassis. In the later, only one branch of EPON needs to be terminated at the subscriber premise because the optical signal power can be split internally within the equipment for the two ONU blades in the same equipment.

The system is designed to ensure that bandwidth guaranteed ONUs are provided with the absolute QoS assurances with the sufficient transmission time agreed upon between subscribers and the service provider in the SLA contract, while making use of any unused bandwidth to provide the best-effort service to bandwidth non-guaranteed ONUs. In the next section, we will propose and present a novel dynamic bandwidth allocation scheme to provide both bandwidth guaranteed service to high-demand users and best-effort service to low-demand users based on the architecture we presented in this section.

4.2 Bandwidth Guaranteed Polling (BGP) Scheme

The major challenge of designing a MAC protocol to provide bandwidth guaranteed service is, how to explicitly incorporate the bandwidth requirements from users into the design of the scheduling scheme so that all of the users can share the transmission link with almost the same bandwidth as their SLA contracts specify. The Time Division Multiple Access (TDMA) scheme is proved to be the most pragmatic way to share the transmission link. Taking into the consideration of the various characteristics of EPON networks, we believe that polling schemes have the potential and advantages to dynamically schedule upstream traffic generated by ONUs. Hence, we have the ideas to combine the advantages of both schemes to achieve an approximation of a TDM system with the dynamic scheduling ability. By this combination, the inflexibility and poor efficiency of TDM scheme can be overcome by the dynamic scheduling that allocates bandwidth to ONUs according to their instant

requirements. Most importantly, the SLA and instantaneous bandwidth requirements of OUNs can be explicitly incorporated into the scheduling mechanism to ensure that the SLA contracts are met while any excess bandwidth can be utilized to the fullest for bandwidth non-guaranteed services. Our proposed scheme is aptly named as Bandwidth Guaranteed Polling (BGP) because it is a roll-call polling system in which the OLT polls ONUs one after another in the adaptive order to allow ONUs sending data to the OLT.

4.2.1 Principle of the BGP Scheduling Scheme

In the novel BGP scheme, the OLT is the central controller to poll ONUs in the adaptive order by sending a polling message to each ONU. The polling message grants a transmission window to invite an ONU to send the proper amount of data packets to the OLT. The ONUs are divided into two disjoint groups according to the SLA contracts between the service provider and end users: bandwidth guaranteed ONUs (for high-demand users) and bandwidth non-guaranteed ONUs (for low-demand users). We call them BG ONUs and non-BG ONUs respectively.

The total upstream bandwidth is divided into equivalent small bandwidth units. Similar to the time slot in the TDMA method, each bandwidth unit in BGP is capable of carrying multiple packets. The OLT maintains an Entry Table that keeps the sequence of entries being polled. Each entry possesses one bandwidth unit that will either be allocated to a fixed BG ONU or be dynamically assigned to a non-BG ONU. The OLT polls ONUs for upstream transmissions in the order of the Entry Table cyclically; hence, multiple ONUs can share the upstream channel without collisions. There are two sections of each entry. The first section of an entry stores the ID number of the ONU that occupies this entry. The second section keeps the propagation delay of the occupying ONU to the OLT. The total number of the entries in the Table is the maximum number of bandwidth units of the upstream channel

Chapter 4. A Systematic Multiple Access Scheme

capacity. For example, the total 1Gb/s upstream bandwidth can be divided into 100 bandwidth units, each with 10 Mb/s, in this case the total entry number of the Entry Table is 100.

The OLT also maintains a non-BG ONUs' List that determines the polling sequence of non-BG ONUs. The non-BG ONUs' List has the similar structure as the Entry Table. Each element of the list has two sections too. The first section of an element keeps the ID number of the non-BG ONU. The second section keeps the propagation delay of the ONU to the OLT. The total number of the elements in the List is dynamic.

A BG ONU's bandwidth guarantee is achieved by allocating one or more bandwidth units/entries in a polling cycle, according to the SLA contracts about their bandwidth requirements, which is further corresponding to the payment of end users. BG ONUs with more than one entry will take more than one bandwidth unit for transmission. And they will be polled more than one time by the OLT in a round of polling. If a BG ONU occupies more than one entry, the entries of this ONU will be evenly distributed among all entries in the Entry Table by the EDA algorithm, which will be described later in detail.

Best-effort service without bandwidth guarantee will be provided to the non-BG ONUs. Entries in the Table that are not occupied by BG ONUs can be assigned to non-BG ONUs dynamically. The OLT polls non-BG ONUs in these available entries in the order presented in the List. In addition, the superfluous transmission window in one entry that is occupied yet unconsumed by an ONU may also be assigned to a non-BG ONU dynamically, according to the sequence of the List. In this case, there should be a predefined threshold to determine whether to use the superfluous transmission window or give it up. If the entire transmission window consumed by an ONU in one entry is less than the threshold, then the superfluous window can be used to poll a non-BG ONU; otherwise, the superfluous transmission window will be given up.

Each ONU buffers data packets received from the end users until its turn to transmit. When its turn arrives, the ONU will “burst” out packets to the OLT at the full channel capacity. The OLT allocates bandwidth to ONUs complying with the limited service. That is, each ONU is allowed to send packets up to a limited amount confined by the maximum transmission window. It represents the maximum amount of packets (in byte) that each entry can accommodate. The maximum transmission window size is a system parameter and can be adjusted according to the network environment.

4.2.2 Control Message Transmissions

In the BGP scheme, data transmissions are conducted by the control message transferred between the OLT and ONUs. Compatible with the Multi-Point Control Protocol (MPCP) [4], BGP uses two kinds of control message for data transmissions: GRANT and REPLY, which are generated and delivered by the OLT and ONUs respectively. The OLT sends GRANT messages to ONUs based on the adaptive order determined by the Entry Table and the non-BG ONU List, informing ONUs to transmit the granted amount of packets at the appropriate time. Upon receiving a GRANT message, an ONU reports its buffer information by sending a REPLY message to the OLT and transmits data packets up to the duration of the granted transmission window.

In BGP, the GRANT message contains the MAC address of the destined ONU, granted transmission window size, and the start time permitting the ONU to transmit. Accordingly, the REPLY message consists of the ONU’s own MAC address, the transmission window size that the OLT grants to the ONU through the previous GRANT message, and the buffer occupancy status in the ONU.

4.2.3 Operations of the OLT and ONUs

By the BGP scheme, all entries in the Entry Table are allocated to fixed BG ONUs or dynamically assigned to non-BG ONUs, so the Entry Table together with the non-BG ONU List will determine the scheduling order of ONUs, based on which the OLT polls ONUs one after another cyclically. When the OLT decides to poll the next entry in the Table, it will first locate which ONU is allocated this entry and then grant a transmission window to this ONU by sending a GRANT message. Upon receiving the message, the ONU sends a REPLY message to the OLT reporting its buffer occupancy status and transmits the granted amount of data packets. The interactive operations of the OLT and ONUs can be described as follows.

1. Initially, the OLT initializes the Entry Table and the non-BG ONU List according to the SLA contracts. The OLT also initializes the system parameters like the maximum transmission window size W_{max} .

2. The OLT starts to poll BG ONUs in the order as determined in the Entry Table, or poll non-BG ONUs dynamically according to the sequence of List, by sending the GRANT message with appropriate window size G through the downstream channel.

3. On receiving a GRANT message destined to it, the ONU (including BG and non-BG ONUs) will:

- (1) Get the granted window size G (in byte) from the message;
- (2) Send a REPLY message to the OLT, reporting the granted window size G and the buffer occupancy L_o (in byte) before transmission;
- (3) Transmit data packets in the buffer according to the limited service:
 - ① if L_o is less than G , the ONU will send all data packets in the buffer;
 - ② if L_o is larger than G , the ONU will only send data packets up to G ;

Chapter 4. A Systematic Multiple Access Scheme

4. The OLT keeps on receiving data packets transmitted from ONUs. On receiving a REPLY message from an ONU, the OLT will:

- (1) First get the granted window size G from the message and compare it with W_{max} ;
- (2) If G equals to W_{max} , then this REPLY message is generated by an ONU assigned the whole entry, either a BG ONU or a non-BG ONU. Get the buffer length L_o :
 - ① if L_o equals to zero, poll next entry in the Entry Table immediately;
 - ② if L_o is between zero and the threshold T , the superfluous transmission window in this entry can be used by a non-BG ONU, poll next non-BG ONU in the List with the granted window size $G=(W_{max}-L_o)$ immediately;
 - ③ if L_o is between the threshold T and W_{max} , give up the superfluous transmission window in this entry, poll next entry in the Entry Table after the time of this entry passes by;
 - ④ if L_o is larger than W_{max} , poll next entry in the Entry Table immediately;
- (3) If G doesn't equal to W_{max} , this Reply message is generated by a non-BG ONU that has been assigned the superfluous transmission window. Get the buffer length L_o :
 - ① if L_o is less than G , give up the superfluous transmission window in this entry, poll next entry in the Entry Table after the time of this entry passes by;
 - ② if L_o is larger than G , poll next entry in the Entry Table immediately.

Figure 4.2 and Figure 4.3 show the operation flowcharts of the ONUs and OLT in the BGP scheme respectively.

Chapter 4. A Systematic Multiple Access Scheme

Figure 4.2 Flowchart of the ONU Operations in BGP

Figure 4.3 Flowchart of the OLT Operations in BGP

4.2.4 Implementation of the BGP Scheme

To implement the BGP scheduling scheme, two pointers $p1$ and $p2$ are required, with $p1$ pointing to the entries in Entry Table and $p2$ pointing to the non-BG ONUs in the List. The OLT cyclically polls ONUs one by one in the order of the entry sequence in the Entry Table, by pointer $p1$ indicating the current entry. If an entry in the Entry Table is not allocated to a BG ONU, it can be used to poll a non-BG ONU indicated by pointer $p2$ as current ONU being served in the non-BG ONUs' List. If an entry allocated to either a BG or a non-BG ONU has the superfluous transmission window left, it can be used to poll another non-BG ONU. The detailed implementation of the pointers is concluded as follows.

1. During the initialization, $p1$ is set to point to the first entry in Entry Table, and $p2$ is set to point to the first non-BG ONU in the List. Whenever the OLT receives a REPLY message from an ONU, it decides the next action according to the BGP scheduling scheme.
2. If the OLT decides to poll next entry in the Entry Table, $p1$ will point to the next entry in the Entry Table ($p1++$) and get the value:
 - (1) If the value isn't 0, then this entry is occupied by a BG ONU whose ID can be get from the entry's contents. The OLT will send a GRANT message with granted window size $G = W_{max}$ to this BG ONU.
 - (2) If the value equals to 0, then this entry will be assigned to a non-BG ONU. The OLT knows the non-BG ONU from the List, whose ID can be obtained from the contents of the item pointed by $p2$. Then the OLT sends a GRANT message with granted window size $G = W_{max}$ to this non-BG ONU, and lets $p2$ point to the next non-BG ONU in the List ($p2++$).
3. If the OLT decides to poll the next non-BG ONU in the List, it processes as 2 (2).

4. If $p1$ points to the last entry in the Entry Table, next time it will point to the first one in the Table. Similarly, if $p2$ points to the last non-BG ONU in the List, next time it will point to the first one in the List.

4.2.5 Even Distribution Algorithm (EDA)

We also propose an Even Distribution Algorithm (EDA) to evenly distribute entries to the same BG ONU in the Entry Table, in order to facilitate the implementation of the BGP scheduling scheme. It will be run at the OLT during the initialization period to prepare the Entry Table and the non-BG ONU List. As mentioned before, a BG ONU may occupy more than one entry in the Entry Table. Then it can be polled more than one time in a round of polling. EDA evenly distributes these entries (of the same ONU) among all entries in the Table. The entry that is occupied by a BG ONU will be filled with the ID of this ONU in the Table. For example, in a 100-entry Table (Figure 4.4), both ONU1 and ONU2 are allocated 4 entries, and ONU3 is allocated 10 entries, then ONU1 can occupy entry 15, 40, 65, 90; ONU2 can occupy entry 2, 27, 52, 77; ONU3 occupies entry 9, 19, 29, 39, 49, 59, 69, 79, 89 and 99.

1	2	3	4	5	6	7	8	9	10
	2							3	
11	12	13	14	15	16	17	18	19	20
				1				3	
21	22	23	24	25	26	27	28	29	30
						2		3	
31	32	33	34	35	36	37	38	39	40
								3	1
41	42	43	44	45	46	47	48	49	50
								3	
51	52	53	54	55	56	57	58	59	60
	2							3	
61	62	63	64	65	66	67	68	69	70
				1				3	
71	72	73	74	75	76	77	78	79	80
						2		3	
81	82	83	84	85	86	87	88	89	90
								3	1
91	92	93	94	95	96	97	98	99	100
								3	

Figure 4.4 An Example of Entry Table with 100 Entries

Chapter 4. A Systematic Multiple Access Scheme

According to the EDA algorithm, all entries of the same BG ONU are separated from the same distances in the Entry Table, so that bandwidth of this ONU can be evenly distributed over the upstream channel when they are polled based on the Entry Table. It will benefit to the BG ONUs with high traffic load, since the packets in the buffer can be allowed to transmit after the same intervals according to BGP. Thus there will be fewer bursts introduced to the data transmissions, and the buffer of ONUs will also be released from over-burdened. If EDA is not deployed, an ONU occupying multiple entries will be polled in the consecutive order. Thus, it will be provided more bandwidth than it needs at the polling time. On the other hand, other ONUs have to wait for longer time for transmission due to the consecutive polling of this ONU. Moreover, it will be more liable to bring burst to the data transmission and the buffer will also have more burden than the case where EDA is adopted.

The operation of the EDA algorithm is described as follows, and Figure 4.5 shows the implementation flowchart, where $T[N]$ is the array of the Entry Table.

1. Fill the Entry Table with N entries, starting from the BG ONU with the maximum amount of entries to the one with the minimum amount of entries.
2. For each BG ONU i with N_i entries, determines:
 - (1) First entry number $E_i[1]$: let entry i to be the first entry of ONU i , $E_i[1]=i$. If entry i is occupied by other ONU, check its neighbor entries $E_i[1]=i \pm n$, let $n=1,2,3,\dots$ until the nearest entry that is not occupied by other BG ONUs, and fill this entry with the ONU ID i .
 - (2) Other entry number $E_i[j]$ ($j=2,3,\dots,N_i$): the j^{th} entry number of this BG ONU should be $E_i[j]=\text{mod}((E_i[1]+\text{int}((j-1)*N/N_i)), N)$, where N is the total entry number of the Entry Table. If the j^{th} entry is occupied by other ONU, check its neighbor entries $E_i[j]=\text{mod}((E_i[1]+\text{int}((j-1)*N/N_i) \pm n), N)$, let $n=1,2,3,\dots$ until the nearest entry that is not occupied by other BG ONUs, and fill this entry with the ONU ID i .

3. All entries that are not occupied by BG ONUs will be filled with 0 and assigned to non-BG ONUs dynamically.
4. The List for non-BG ONUs is sequenced in ascending order based on non-BG ONUs' IDs. The total number of the elements in the List is not fixed.

Figure 4.5 Flowchart of the EDA Algorithm

4.3 Issues of the Bandwidth Utilization

The main purpose of our BGP scheme is to provide ONUs with the corresponding service according to SLA contracts in EPON networks. It makes use of the control messages to avoid the collisions in the upstream data transmissions, and carries out the EDA algorithm to separate the heavy load of ONUs in order to control the traffic burst. Besides, the BGP scheme allows one transmission window carry data packets from up to two different ONUs

by using the superfluous window. All these mechanisms help BGP to achieve the efficient bandwidth allocation for multiple ONUs. Nevertheless, we should also pay attention to the effect of these mechanisms to the bandwidth utilization in EPON networks deploying the BGP scheme.

The bandwidth utilization is the ratio of the maximum throughput to the network capacity, and throughput is numerically the amount of information being transmitted by the network in a unit of time. The utilization can be easily measured by finding out the overhead components related to the network. In this section, we will discuss the overhead of the proposed BGP scheme over the EPON networks. It mainly includes the control message overhead, guard time overhead, and remainder window overhead.

4.3.1. Control Message Overhead

In the BGP scheme, the overhead of the control message transmission represents the bandwidth loss due to the transmission of GRANT and REPLY messages. For the GRANT message transmission, BGP follows the same way of interleaved polling in IPACT [27] to reduce the overhead of the downstream messages. In the BGP scheduling scheme, the propagation and transmission time of the GRANT message to the next ONU can be interleaved with the current ONU's data transmission except the number of packets in current transmission is or almost is 0. The interleaving can be implemented if the OLT can send out the GRANT message at a proper time so that the message can arrive at the destination ONU to activate its data transmission within the time period of current ONU's data transmission. By this interleaving, the total average delay experienced by a packet can be reduced and BGP can achieve the efficient bandwidth utilization. So the overhead of downstream GRANT message transmissions can be ignored in the BGP scheme.

Chapter 4. A Systematic Multiple Access Scheme

As to the overhead of upstream REPLY message transmissions O_{up} , it cannot be eliminated completely in the BGP scheme, and it exists in all other polling-based bandwidth allocation schemes including IPACT. In the BGP scheme, since the REPLY messages are transmitted together with the data packets, there is no extra propagation time for the upstream messages transmission unless the ONU being polled has no data in the buffer. Then we need to consider the transmission time for the REPLY message overhead. Recall that, in BGP, an entry can be used to poll at least one ONU in each polling cycle. If the traffic load is light enough, the OLT can poll two different ONUs for an entry. Then the overhead of upstream messages in one polling cycle can be expressed as:

$$O_{up} = \frac{t_{msg} \cdot N'}{t_c} \quad (4-1)$$

where $N \leq N' \leq 2N$ and N' is the polling times in a cycle, N is the number of entries in the Entry Table, t_{msg} is the transmission time of a control message, and t_c is the cycle time. For IPACT, the overhead of the upstream message transmission is $\frac{t_{msg} \cdot M}{t_c}$, where M is the number of ONUs in the EPON system and also the polling times in one cycle.

At the heavy load, almost all entries cannot poll two different ONUs by the BGP scheme, thus the overhead is close to: $\frac{t_{msg} \cdot N}{t_c}$ because N' approximates to N . In this case, the overhead of the upstream message transmission is limited to a low value, and the ratio between BGP's overhead and IPACT's overhead regarding to control message transmission is: $\frac{O_{BGP}}{O_{IPACT}} = \frac{N}{M}$. When the number of entries N is larger than the number of ONUs M in the EPON system, the control message overhead of BGP will be worse than that of IPACT. This

is a trade off between the bandwidth utilization and the efficiency of BGP scheme, which will be fully discussed in Section 4.3.4.

When the load is light, more entries are liable to poll a second ONU, leading to N' larger than N . Accordingly, the overhead may exceed that of IPACT at the light load. However, the difference between the overheads is insignificant due to the small value of t_{msg} comparing to t_c . Moreover, since it is unnecessary to use the full channel capacity to transmit data at the light load, the unconsumed capacity can handle the overhead effortlessly. Hence, the increase of the upstream message overhead at the light load will not impact the data transmissions much.

4.3.2. Guard Time Overhead

When the transmission from one ONU finishes and the OLT prepares to receive data packets from another ONU, the receiver in the OLT should be adjusted to the new power level and be synchronized on the new bit/byte boundary. Hence a small interval (guard time) is required between transmissions from different ONUs. The guard time overhead is not particular to the BGP scheme. Instead, it is common to all DBA schemes for EPON networks. Generally, the guard time between any pair of ONUs is set to an identical value, t_g . In the BGP scheme, the overhead of guard time in one cycle can be calculated as:

$$O_{guard} = \frac{t_g \cdot N'}{t_c} \quad (4-2)$$

Similar to the overhead of upstream control messages, the guard time overhead will be limited to near: $\frac{t_g \cdot N}{t_c}$ when the load is heavy, and the ratio to that of IPACT is also:

$$\frac{O_{BGP}}{O_{IPACT}} = \frac{N}{M}. \text{ When } N \text{ is larger than } M \text{ in the EPON system, the guard time overhead of BGP}$$

will be worse than that of IPACT. The increase of overhead at the light load will not influence the data transmissions much as well.

4.3.3. Remainder Window Overhead

In the BGP scheme, the usage of the superfluous transmission window leads to another kind of overhead. Recall that when an entry occupied by an ONU cannot be used up for a whole transmission window, the superfluous window can be used to poll another ONU if it has enough size. If the superfluous window has a size less than the predefined value, it will be wasted leading to the remainder window overhead. The overhead also results from the fact when the superfluous transmission window is used by another ONU while it cannot consume the window completely.

Let W_{max} be the maximum transmission window size, and T the predefined threshold that determines whether the superfluous transmission window can be used to poll a second ONU or not. Then according to the BGP scheme, the OLT will grant the ONU k with the granted transmission window size:

$$G_k = \begin{cases} W_{max} - L_{k-1} , & \text{if } G_{k-1} = W_{max} \text{ and } 0 < L_{k-1} \leq T \\ W_{max} , & \text{otherwise} \end{cases} \quad (4-3)$$

where L_{k-1} is the buffer occupancy of the previous ONU ($k-1$) and G_{k-1} is the granted window size to ONU ($k-1$).

Let $R1$ to be the average remainder size of the entry polling only one ONU, and $R2$ the average remainder size of the entry used to poll two ONUs. Based on the BGP scheduling, $R1$ and $R2$ can be calculated as:

$$R_1 = \sum_{m=T}^{W_{max}} f(m) \cdot (W_{max} - m)$$

Chapter 4. A Systematic Multiple Access Scheme

$$R_2 = \sum_{m=0}^T f(m) \cdot \sum_{n=0}^{W_{\max}-m} f(n) \cdot (W_{\max} - m - n)$$

where $f(x)$ is the probability distribution function of the ONU's transmission amount in one polling. Then the remainder window overhead of the BGP scheme is:

$$O_{\text{remainder}} = \frac{R_1 \cdot (N - P) + R_2 \cdot P}{W_{\max} \cdot N}$$

where P is the number of entries that can be used to poll a second ONU. The value of P is mainly determined by the threshold T and the traffic load in the EPON network.

4.3.4. Discussion

From the expressions (4-1) and (4-2), the polling times in a cycle, N' , is a key parameter to decide the control message overhead and guard time overhead. When the entry number of the Table N is fixed, N' should be as little as possible to reduce the overhead from the expressions. However, with the fixed N , a too small N' indicates that most entries in the Table will be used to poll only one ONU. Many remainder transmission windows will be wasted accordingly if the traffic load is light, which may result in the much more bandwidth waste than a big N' can bring. There should be a balanced decision for the value of N' .

Another consideration is to reduce the number of entries in the Table N , thus N' can also be a small value to control the overhead. However, in the BGP scheme, the entry number of the Table determines the distribution of the ONU's traffic over the upstream transmission link. Recall that an ONU with heavy traffic load will request multiple entries for transmission. And its traffic will be evenly distributed among the multiple entries according to the EDA algorithm in order to avoid burst. The larger entry number of the Table, the more entries an ONU will be allocated. Consequently, the traffic of the ONU will

be distributed among more entries for transmission, leading to the fewer chance of traffic burst from this ONU. Hence, a large entry number N of the Table will benefit to the real-time traffic with better performance by reducing burst in the transmission. It is a trade off between the bandwidth utilization and the performance of the desirable service to the real-time traffic in EPON networks.

For the remainder window overhead, the threshold T is very important to decide this kind of overhead. Assuming P is the number of entries that can be used to poll a second ONU. A small T will lead to the increased P , making more entries being used to poll a second ONU, so the polling times in each cycle N' will also grow. However, a too large N' may result in the high overhead of control message and guard time. In contrast, if T is too large making P too small, large number of superfluous transmission windows will be wasted, leading to the high remainder window overhead. The threshold T should be carefully set according to the traffic load, arrival rate and other system parameters, in order to achieve the balance between different kinds of overhead for the high bandwidth utilization.

4.4 Performance Estimation

In this section, we will estimate the performance of the proposed BPG scheme mathematically. Since the EPON network is actually a polling system when deploying BGP, we will analyze it as a roll-call polling system [61][62] for the estimation. The major performance of this analysis is the average delay experienced by a packet from the time of its arriving at an ONU to the time of its arriving at the OLT.

The point-to-point average delay obviously consists of three parts: the access delay, transmission delay and propagation delay. The access delay is the time from the packet arrives at an ONU to the time it starts to be transmitted over the upstream channel. The access delay A experienced by a packet can be divided into two parts: $A = W_1 + W_2$, where W_1

is the waiting delay while other ONUs are being polled, the ONU, which contains this packet, is idle and waiting for its turn to be polled; W_2 is the waiting delay while the ONU is being polled and transmitting earlier packets in the buffer. The transmission delay of a packet is the time period for the ONU to send out the whole data frame. The propagation delay is the time period for the packet traveling from the ONU to the OLT in the upstream channel. Based on the above considerations, we can get the expression of the average packet delay as:

$$E(D) = E(A) + \bar{t} + \bar{\tau} \quad (4-4)$$

where $\bar{t}$ is the average transmission time for a packet, $\bar{\tau}$ is the average propagation delay for each ONU to the OLT, and:

$$E(A) = E(W_1) + E(W_2) \quad (4-5)$$

4.4.1 Average Cycle Time

In order to evaluate the average waiting time in a polling system, we need to consider the cycle time t_c , which is the time needed to complete one cycle of polling to all users. The cycle time t_c is equal to the total data transmission time t plus the walk time WT , which is the total time used to transmit the control messages in a polling cycle.

In our BGP scheme, an entire polling will include transmissions from both BG and non-BG ONUs presented in all the entries of the Entry Table. The data transmission time for each entry equals to the average transmission time for a packet times the maximum transmission window size. To poll all the entries in the Entry Table, it will take time t for transmitting all the packets:

$$t = NW_{\max} \bar{t} \quad (4-6)$$

Chapter 4. A Systematic Multiple Access Scheme

where N is the total number of entries in the Entry Table, W_{max} is the maximum transmission window size, and $\bar{t}$ is the average transmission time for a packet.

The walk time in a polling cycle equals to the sum of the control message propagation time, control message transmission time and synchronization time. We consider that the synchronization time and control message transmission time can be ignored because the size of the control messages are very small compared to the data packets and their transmission can be interleaved with the data transmissions according to BGP. Then the walk time can be divided into:

$$WT = WT_{up} + WT_{down} \quad (4-7)$$

where WT_{up} is the upstream walk time, the propagation time experienced for all REPLY messages to be transmitted from ONUs to the OLT, WT_{down} is the downstream walk time, the propagation time for all GRANT messages traveling from the OLT to ONUs. As one entry may be used to poll one or two ONUs in terms of different traffic load to the whole network, ρ and the threshold of the superfluous transmission window T , WT_{up} will range from $N\bar{\tau}$ to $2N\bar{\tau}$, where $\bar{\tau}$ is the average propagation delay from an ONU to the OLT. For WT_{down} , the downstream message propagation time can be ignored when the previous ONU sends packets with full transmission window, because the message transmission can be interleaved with the upstream data transmission in the BGP scheduling scheme. Only those GRANT messages pertaining to the entries used to poll two ONUs are taken into account in the evaluation. Let P to be the number of entries that can be used to poll a second ONU, which is a random variable that varies in terms of different ρ . Then: $WT_{up} = N\bar{\tau} + P\bar{\tau}$, $WT_{down} = 2P\bar{\tau}$. The average walk time will be:

$$WT = N\bar{\tau} + 3P\bar{\tau} \quad (4-8)$$

Combining Formula (4-6) and (4-8), the average cycle time can be got as:

Chapter 4. A Systematic Multiple Access Scheme

$$\bar{t}_c = WT + t = N\bar{\tau} + 3P\bar{\tau} + NW_{\max}\bar{t} \quad (4-9)$$

To determine the value of P , the concept of effective traffic load ρ_e is introduced as follows. As the full channel capacity μ is divided into N bandwidth units in our BGP model and each ONU is assigned different bandwidth units according to the number of entries it occupies, the effective service rate of an ONU can be expressed as: $n\mu/N$, where n is the number of entries distributed to this ONU. In an EPON system with M ONUs, the average traffic arrival rate of each ONU is $\lambda_i = \rho\mu/M$, so a specific ONU's effective traffic load can be defined as:

$$\rho_e = \frac{\lambda_i}{n\mu/N} = \frac{\rho\mu/M}{n\mu/N} = \frac{N}{nM} \times \rho \quad (4-10)$$

We can also express the number of entries being used to poll a second ONU P as:

$$P = \sum_{n=1}^N nm_n \times p_n \quad (4-11)$$

where m_n is the number of BG ONUs that have n entries, p_n is the probability that one entry can be used to poll two ONUs, defined as:

$$p_n = \begin{cases} \gamma, & \text{if } \rho_e < \rho_s \\ 0, & \text{if } \rho_e \geq \rho_s \end{cases} \quad (4-12)$$

ρ_s is the saturate traffic load defined as the point of effective traffic load when a specific ONU starts to lose packets due to its finite buffer, determined by the total traffic load ρ and threshold of the transmission window T . Let's assume that p_n is uniformly distributed with value γ , $0 \leq \gamma \leq 1$, then the average value of P is:

$$\bar{P} = \sum_{n=1}^N nm_n \times \bar{p}_n \quad (4-13)$$

where the mean value of p_n is 0.5 when $\rho_e < \rho_s$.

4.4.2 Average Waiting Delay $E(W_1)$

With the average cycle time $\bar{t}_c$, it is possible for us to derive the average waiting delay $E(W_1)$. Recall that W_1 is the waiting period while other ONUs are being polled, and the ONU containing the observed packet is idle and waiting for its turn to be polled. We derive $E(W_1)$ for BG and non-BG ONUs respectively.

4.4.2.1 $E(W_1)$ for BG ONUs

During the scan time $\bar{t}_c$, the number of packets arrived at each ONU is: $\lambda_i \bar{t}_c$. To a specific BG ONU with n entries, if $\lambda_i \bar{t}_c < nW_{\max}$, we regard that it could send out all its waiting packets when it is polled. Because this ONU is polled every $(\bar{t}_c/n)$ time, the packets arrived during this time period is: $\lambda_i \bar{t}_c/n$. The transmission time for all these packets is: $\bar{t} \times \lambda_i \bar{t}_c/n = \rho_i \bar{t}_c/n$. Then the remaining part of the ONU's polling cycle $(\bar{t}_c/n)$, during which this ONU is idle, is: $(1-\rho_i) \bar{t}_c/n$. It is also the waiting time for this ONU to be polled. Hence, at the steady state, the average waiting delay that a packet experienced is:

$$E(W_1) = \frac{1}{2n} (1 - \rho_i) \bar{t}_c \quad (4-14)$$

If $\lambda_i \bar{t}_c > nW_{\max}$, then this ONU can only send out packets up to the $W_{\max}$ when it is polled. The transmission time for these packets is: $W_{\max} \bar{t}$, and the time this ONU is waiting for a poll is: $\bar{t}_c/n - W_{\max} \bar{t}$. At the steady state, the average waiting delay that a packet experienced is:

$$E(W_1) = \frac{1}{2} (\bar{t}_c/n - W_{\max} \bar{t}) \quad (4-15)$$

4.4.2.2 $E(W_1)$ for non-BG ONUs

Let m_n to be the number of BG ONUs occupying n entries and m' the number of non-BG ONUs. Then the remaining entries that are not allocated to BG ONUs are: $(N - \sum_{n=1}^N nm_n)$, which will be dynamically assigned to the m' non-BG ONUs. In addition, the superfluous transmission window in one entry that is not fully consumed by an ONU can also be assigned to a non-BG ONU dynamically. As mentioned before, there are $\bar{P}$ on average superfluous windows for non-BG ONUs. So the average number of entries that a non-BG ONU is assigned can be approximate to:

$$\bar{n} = \frac{1}{m'} \left[(N - \sum_{n=1}^N nm_n) + \bar{P} \right] \quad (4-16)$$

The transmission window size for an entire entry is $W_{\max}$, while the superfluous transmission windows have random sizes (less than $W_{\max}$) due to the first ONU's consumption. If the packets arrived at a BG ONU with n entries during $(\bar{t}_c/n)$ time is less than the threshold T , then the superfluous window $(W_{\max} - \lambda_i \bar{t}_c/n)$ can be used to poll a non-BG ONU. As the first ONU can take window size up to threshold T , the average window size for the first ONU is $T/2$, and the average superfluous window size for a non-BG ONU is $(W_{\max} - T/2)$. So all window size for non-BG ONUs is:

$$W_{\text{non}} = (N - \sum_{n=1}^N nm_n) W_{\max} + \bar{P} (W_{\max} - T/2) \quad (4-17)$$

and the average window size for a non-BG ONU is:

$$\bar{W}_{\text{non}} = \frac{W_{\text{non}}}{m'} = \frac{1}{m'} \left[(N - \sum_{n=1}^N nm_n) W_{\max} + \bar{P} (W_{\max} - T/2) \right] \quad (4-18)$$

For a specific non-BG ONU, if $\lambda_i \bar{t}_c < \bar{W}_{\text{non}}$, then we can regard this non-BG ONU will send out all its waiting packets when it is polled. Because this ONU is polled every $(\bar{t}_c/\bar{n})$

Chapter 4. A Systematic Multiple Access Scheme

time, the packets arrived during this period is: $\lambda_i \bar{t}_c / \bar{n}$. The transmission time for all these packets is: $\bar{t} \times \lambda_i \bar{t}_c / \bar{n} = \rho_i \bar{t}_c / \bar{n}$. Then the remaining part of the polling cycle ($\bar{t}_c / \bar{n}$), during which this ONU is idle, is $(1 - \rho_i) \bar{t}_c / \bar{n}$. It is also the time the ONU is waiting for a poll. At the steady state, the average waiting delay that a packet experienced is:

$$E(W_1) = \frac{1}{2n} (1 - \rho_i) \bar{t}_c \quad (4-19)$$

If $\lambda_i \bar{t}_c > \bar{W}_{non}$, the non-BG ONU can only send out packets up to the $\bar{W}_{non} / \bar{n}$ when it is polled. The transmission time for these packets is: $\bar{W}_{non} \times \bar{t} / \bar{n}$, and the time for this ONU waiting for a poll is: $(\bar{t}_c - \bar{W}_{non} \times \bar{t}) / \bar{n}$. At the steady state, the average waiting delay that a packet experienced is:

$$E(W_1) = \frac{1}{2n} (\bar{t}_c - \bar{W}_{non} \times \bar{t}) \quad (4-20)$$

4.4.3 Average Waiting Delay $E(W_2)$

W_2 is the delay that the observed packet is waiting for its turn to be transmitted when the ONU is being polled and transmitting other packets. Since each ONU will experience three stages of traffic: the light, moderate and heavy traffic load, each of which is referred to the effective transmission capacity of the ONU, we can evaluate $E(W_2)$ for all kinds of ONUs according to different traffic loads.

4.4.3.1 $E(W_2)$ for the Light Load

When the traffic load ρ is low, that means $\lambda_i \bar{t}_c < nW_{max}$ for the BG ONUs or $\lambda_i \bar{t}_c < \bar{W}_{non}$ for the non-BG ONUs. The whole system can be approximated to multiple M/G/1 queues, each of which corresponds to one kind of ONUs. We define: $\rho' = \frac{m_n}{M} \rho$ as the traffic load for

each kind of ONUs. Then the average waiting delay for an ONU (no matter BG or non-BG ONU) can be expressed based on the assumption that all new arrival packets can be transmitted during the same polling cycle when they arrive:

$$E(W_2) = \frac{\lambda \cdot E(\tau^2)}{2(1-\rho)} = \frac{\rho' \bar{t}}{2(1-\rho')} \quad (4-21)$$

4.4.3.2 $E(W_2)$ for the Moderate Load

With the increase of the load ρ , $\lambda_i \bar{t}_c$ of BG ONUs approaches to nW_{max} gradually. BG ONUs experience a moderate traffic load. Now we can assume $\lambda_i \bar{t}_c \geq wnW_{max}$ where w (around 95%) is the weight of effective transmission capacity, and there is no packet loss for the BG ONUs. Since the new arrival packets can't be sent out during the same polling cycle when they arrive, we assume they will be transmitted after another k cycles of polling, where k is the random variable and $\bar{k}$ is the average value. $E(W_2)$ for the BG ONU can be expressed as:

$$E(W_2) = \frac{\bar{k}}{n} \bar{t}_c + \frac{\rho' \bar{t}}{2(1-\rho')} \quad (4-22)$$

If we regard $\bar{n}$ as the number of entries that a non-BG ONU occupies, then the analysis will be similar to that of the BG ONUs. For non-BG ONUs, when $\lambda_i \bar{t}_c \geq w\bar{W}_{non}$, they experience the moderate traffic load. And $E(W_2)$ for non-BG ONU can be expressed as:

$$E(W_2) = \frac{\bar{k}}{n} \bar{t}_c + \frac{\rho' \bar{t}}{2(1-\rho')} \quad (4-23)$$

4.4.3.3 $E(W_2)$ for the Heavy Load

When the load ρ is heavy, that means $\lambda_i \bar{t}_c \geq nW_{max}$. For the BG ONUs, there will be data loss from the ONUs due to the finite buffer filled with waiting data packets. Let L_B to be the

ONU's buffer size, the new arrival data should be sent out after waiting for enough polling cycles to transmit the previous (L_B-1) data. So the average waiting delay for BG ONUs can be expressed as:

$$E(W_2) = \left\lfloor \frac{L_B}{nW_{max}} \right\rfloor \cdot \bar{t}_c + \frac{\rho' \bar{t}}{2(1-\rho')} \quad (4-24)$$

For non-BG ONUs with heavy traffic load, $\lambda_i \bar{t}_c \geq \overline{W_{non}}$, there will also be packet loss from the ONUs. Then the waiting delay for non-BG ONUs is expressed as:

$$E(W_2) = \left\lfloor \frac{L_B}{\overline{W_{non}}} \right\rfloor \cdot \bar{t}_c + \frac{\rho' \bar{t}}{2(1-\rho')} \quad (4-25)$$

4.4.4 Summary

According to the above analysis, we can get the point-to-point delay $E(D)$ at last with the network parameters. First the average cycle time $\bar{t}_c$ is derived from formula (4-9) with the estimation of the number P of entries that can be used to poll a second ONU. With $\bar{t}_c$, the average waiting delay $E(W_1)$ can be obtained for different groups of ONUs, which represents the waiting period while the ONU containing the observed packet is idle and waiting for its turn to be polled. Then we get the average waiting delay $E(W_2)$ that the observed packet is waiting when the ONU is being polled and transmitting the earlier packets. Recall that $E(W_2)$ is derived according to the ONU's traffic load. Due to different bandwidth units/entries allocation, different groups of ONUs have variant stages of light, moderate and heavy traffic load accordingly.

In conclusion, formulas (4-26), (4-27) and (4-28) express the average waiting delays for both BG and non-BG ONUs with different traffic load in the BGP scheme.

$$E(D) = E(W_1) + E(W_2) + \bar{t} + \bar{\tau} \quad (4-26)$$

$$E(W_1) = \begin{cases} \frac{1}{2n}(1-\rho_i)\bar{t}_c & BG\ ONU, \lambda_i\bar{t}_c < n W_{\max} \\ \frac{1}{2}(\bar{t}_c/n - W_{\max}\bar{t}) & BG\ ONU, \lambda_i\bar{t}_c \geq n W_{\max} \\ \frac{1}{2n}(1-\rho_i)\bar{t}_c & non-BG\ ONU, \lambda_i\bar{t}_c < \overline{W_{non}} \\ \frac{1}{2n}(\bar{t}_c - \overline{W_{non}}\bar{t}) & non-BG\ ONU, \lambda_i\bar{t}_c \geq \overline{W_{non}} \end{cases} \quad (4-27)$$

$$E(W_2) = \begin{cases} \frac{\rho' \bar{t}}{2(1-\rho')} & low\ load, BG\ and\ non-BG\ ONU \\ \frac{\bar{k}}{n}\bar{t}_c + \frac{\rho' \bar{t}}{2(1-\rho')} & moderate\ load, BG\ ONU \\ \frac{\bar{k}}{n}\bar{t}_c + \frac{\rho' \bar{t}}{2(1-\rho')} & moderate\ load, non-BG\ ONU \\ \left\lfloor \frac{L_B}{nW_{\max}} \right\rfloor \cdot \bar{t}_c + \frac{\rho' \bar{t}}{2(1-\rho')} & high\ load, BG\ ONU \\ \left\lfloor \frac{L_B}{\overline{W_{non}}} \right\rfloor \cdot \bar{t}_c + \frac{\rho' \bar{t}}{2(1-\rho')} & high\ load, non-BG\ ONU \end{cases} \quad (4-28)$$

4.5 A Continuous-Time Markov Chains (CTMC) Model

In the literature, many investigations have addressed the multi-queue systems with a single server. Various polling mechanisms and several service disciplines, such as exhaustive, gated and limited service, are considered in [63-68]. All these studies are dedicated to analyze the infinite buffer multi-queue systems, which are not practical for a real system. Few models have been proposed to analyze the multi-queue systems with finite buffers [69][70] due to the computing complication, among which [69] establishes the analysis method using the technique of the imbedded Markov chain to investigate such systems.

In this section, the numerical analytical mode developed in [69] will be applied to analyze the EPON network with the BGP scheme. The analysis makes use of the Continuous-Time Markov Chains (CTMC) and a two-moment approximation for the cycle

times, in order to obtain the desirable system characteristics such as the blocking probability and the mean waiting delay.

4.5.1 Model Description

The basic model of the single server multi-queue polling system with finite buffers is shown in Figure 4.6, including a server and g queues. The central server polls the multiple queues in a cyclic order with the limited service. The server will only transmit one message if there is at least one message waiting in the queue's buffer. If the queue is empty, the server will move to the next queue with the inter-polling period containing just the switchover time. For a queue j , it can have the independent distribution functions for the message service time T_{hj} and switchover time T_{uj} different from other queues. The arrival process of queue j is assumed to be Poisson distributed with the individual arrival rate λ_j , and its finite buffer size is S_j . The system can be modeled to be either symmetrically or asymmetrically loaded, which is applicable to the EPON network employing the BGP scheme.

Figure 4.6 Basic Model of the Multi-Queue Polling System with Finite Buffers

In order to emulate the above multi-queue polling system, let us recall the principle of the BGP scheme. In BGP, the OLT can be regarded as the central server to cyclically poll multiple ONUs for upstream transmissions with the order decided by the Entry Table and

Chapter 4. A Systematic Multiple Access Scheme

List. Different ONUs are allocated different number of entries according to the SLA contracts, leading to the asymmetrical polling chances for ONUs in the network. A BG ONU occupying multiple entries will be polled with multiple times by the OLT in one cycle of polling, while a non-BG ONU may not be polled once for several cycles. The chance of polling in BGP can be mapped to the individual queue in the multi-queue polling model. Recall that, if an ONU occupying an entry cannot consume the whole transmission window, the superfluous window can be used to poll a non-BG ONU if the remainder is large enough. Then an entry actually represents two chances of polling, each for one ONU's transmission. The entry that is not used to poll a second ONU can be considered to have two polling chances for the same ONU. In an EPON network with N entries in the Table, the OLT will have $2N$ polling chances in one cycle. Hence, the EPON network deploying the BGP scheme can be emulated to the multi-queue polling model that contains $2N$ queues in a whole ($g=2N$), with each entry representing two queues.

Assume an EPON network with M ONUs, let ρ be the system traffic load and μ the upstream channel capacity. Each ONU has the identical packet arrival rate ($\lambda = \rho\mu / M$) of Poisson distribution in the BGP scheme. For a specific BG ONU occupying n entries, its traffic will be evenly distributed into the n entries according to the EDA algorithm. If the traffic load is light, the entry allocated to the BG ONU can be used to poll a second ONU. Then the two queues represented by the entry will belong to two different ONUs. The first queue is for the specific BG ONU with arrival rate $\lambda_j = \lambda / n$, and the second queue is for a dynamic non-BG ONU. If the traffic load is high enough and the entry can only be used to poll the specific BG ONU itself, then the two queues represented by the entry will both belong to this ONU with the same arrival rate $\lambda_j = \lambda / 2n$ (assume the ONU separates the traffic into these two queues averagely). For the non-BG ONUs, since they are averagely

Chapter 4. A Systematic Multiple Access Scheme

allocated the remained entries that are not occupied by BG ONUs in the Entry Table and the superfluous transmission windows, the number of entry occupied by a non-BG ONU can be approximate to $\bar{n}$ according to formula (4-16). Then the arrival rate for the queues represented by the whole entries remained in the Table will be $\lambda_j = \lambda / \bar{n}$ at the light load and $\lambda_j = \lambda / 2\bar{n}$ at the high load. For the queues represented by the superfluous transmission windows, the arrival rate is always $\lambda_j = \lambda / \bar{n}$ in spite of the traffic load.

Now let us consider the message service time. Since every time when the OLT polls an entry, it will consume the time period of transmitting the whole transmission window unless the buffer of the ONU occupying the entry is empty. Now an entry represents two queues in emulating the multi-queue polling model. If there are data packets waiting in the specific ONU's buffer, the OLT will spend the time period of half transmission window for the queue belonging to this ONU. So the half window transmission time in BGP can be regarded as the message service time in the multi-queue polling model accordingly. Then the service time for all queues is identical to be: $T_{hj} = \frac{W_{\max}}{2\mu}$, where $W_{\max}$ is the maximum window size in byte.

The switchover time contains mainly the upstream and downstream propagation delay, which is also identical to all queues as $T_{uj} = 2\bar{\tau}$, where $\bar{\tau}$ is the average propagation time from the ONU to the OLT. The capacities for all queues are same as $S_j = \left\lfloor \frac{2L_B}{W_{\max}} \right\rfloor$, where L_B is the buffer size in byte for each ONU. In conclusion, the EPON network with the BGP scheme can be emulated to the finite buffer multi-queue polling system with the following parameters:

$$\text{queues of BG ONUs: } \lambda_j = \begin{cases} \frac{\rho\mu}{nM} & \text{at the light load} \\ \frac{\rho\mu}{2nM} & \text{at the high load} \end{cases} \quad (4-29)$$

$$\text{queues of non-BG ONUs: } \lambda_j = \begin{cases} \frac{\rho\mu}{nM} & \text{at the light load} \\ \frac{\rho\mu}{2nM} & \text{at the high load} \end{cases} \quad (4-30)$$

$$T_{hj} = \frac{W_{\max}}{2\mu}, \quad T_{uj} = 2\bar{\tau}, \quad S_j = \left\lfloor \frac{2L_B}{W_{\max}} \right\rfloor, \quad j=1,2,\dots,g=2N \quad (4-31)$$

4.5.2 Markov Chain State Probabilities

A particular queue j is observed in the model. Let t_n be the time of the n^{th} polling epoch and $X^{(n)}(0^-)$ the number of messages in the queue at time t_n^- , just prior to the n^{th} polling epoch.

The Markov chain state probabilities and steady state probabilities are defined as:

$$P_{k,j}^{(n)} = pr\{X^{(n)}(0^-) = k\}, \quad k=0,1,\dots,S_j \quad (4-32)$$

$$P_{k,j} = pr\{X(0^-) = k\}, \quad k=0,1,\dots,S_j \quad (4-33)$$

For ease of reading, the subscript j will be suppressed, and the steady state probabilities will use the notation P_k instead of $P_{k,j}$. In order to calculate the transition probabilities of the Markov chain:

$$P_{jk} = pr\{X^{(n+1)}(0^-) = k \mid X^{(n)}(0^-) = j\} \quad (4-34)$$

We observe the system state $X^{(n)}(t)$ of the queue at time $(t_n + t)$. Considering the pure birth process in the queue between two consecutive polling epochs, the state probabilities at time $(t_n + t)$ denoted by:

$$P_k^{(n)}(t) = P_{k,j}^{(n)}(t) = pr\{X^{(n)}(t) = k\}, \quad k=0,1,\dots,S_j \quad (4-35)$$

can be obtained as follows:

Chapter 4. A Systematic Multiple Access Scheme

$$\begin{cases} P_k^{(n)}(t) = P_0^{(n)} \cdot a_k(t) + \sum_{i=0}^k P_{i+1}^{(n)} \cdot a_{k-i}(t), & k = 0, 1, \dots, S_{j-1} \\ P_{S_j}^{(n)}(t) = P_0^{(n)} \cdot \sum_{i=S_j}^{\infty} a_i(t) + \sum_{i=0}^{S_j-1} \left(P_{i+1}^{(n)} \cdot \sum_{m=S_j-i}^{\infty} a_m(t) \right) \end{cases} \quad (4-36)$$

where $a_m(t) = \frac{(\lambda_j t)^m \cdot e^{-\lambda_j t}}{m!}$ is the arrival process with Poisson distribution of queue j .

With the consideration of the conditional cycle time [67], the following random variables are defined for the analysis model.

$T_{c,j}$: the cycle time with respect to the observed queue j ;

$T'_{c,j}$: conditional cycle time of queue j when it is empty in the previous scanning period
(without service of queue j during the polling cycle);

$T''_{c,j}$: conditional cycle time of queue j when it is not empty in the previous scanning period
(with service of queue j during the polling cycle).

The state probability equations that implicitly contain the transition probabilities can be:

$$\begin{cases} P_k^{(n+1)} = P_0^{(n)} \cdot \int_0^{\infty} a_k(t) f_{c'}(t) dt + \sum_{i=1}^{k+1} P_i^{(n)} \cdot \int_0^{\infty} a_{k-i+1}(t) f_{c''}(t) dt, & k = 0, 1, \dots, S_{j-1} \\ P_{S_j}^{(n+1)} = P_0^{(n)} \cdot \sum_{i=S_j}^{\infty} \int_0^{\infty} a_i(t) f_{c'}(t) dt + \sum_{i=1}^{S_j} \left(P_i^{(n)} \cdot \sum_{m=S_j-i+1}^{\infty} \int_0^{\infty} a_m(t) f_{c''}(t) dt \right) \end{cases} \quad (4-37)$$

Defining the probabilities for m arrivals during a conditional cycle of type T'_c or T''_c as the arrival probabilities:

$$b'_m = \int_0^{\infty} a_m(t) f_{c'}(t) dt \quad (4-38)$$

$$b''_m = \int_0^{\infty} a_m(t) f_{c''}(t) dt \quad (4-39)$$

Then the Markov chain state equations can be obtained as follows:

$$\begin{cases} P_k^{(n+1)} = P_0^{(n)} \cdot b_k' + \sum_{i=1}^{k+1} P_i^{(n)} \cdot b_{k-i+1}'' , & k = 0, 1, \dots, S_{j-1} \\ P_{S_j}^{(n+1)} = P_0^{(n)} \cdot \sum_{i=S_j}^{\infty} b_i' + \sum_{i=1}^{S_j} \left(P_i^{(n)} \cdot \sum_{m=S_j-i+1}^{\infty} b_m'' \right) \end{cases} \quad (4-40)$$

These equations will be used for the numerical calculation of the Markov chain steady state probabilities. The arrival probabilities $b_{m,j}'$ and $b_{m,j}''$ are decided by the following conditional cycle time approximation.

4.5.3 Conditional Cycle Time Approximation

Let $T_{E,j}$ to be the time interval between the polling epochs of queue j and queue $(j+1)$, then its first two moments are:

$$\bar{T}_{E,j} = \bar{T}_{uj} + (1 - P_{0,j}) \cdot \bar{T}_{hj} \quad (4-41)$$

$$Var[T_{E,j}] = Var[T_{uj}] + (1 - P_{0,j}) \cdot Var[T_{hj}] + P_{0,j} \cdot (1 - P_{0,j}) \cdot \bar{T}_{hj}^2 \quad (4-42)$$

By means of the Laplace Stieltjes Transform (LST), the first two moments of the conditional cycle time can be yielded as:

$$\bar{T}_{c,j}' = \bar{T}_{uj} + \sum_{\substack{k=1 \\ k \neq j}}^g \bar{T}_{E,k} , \quad Var[T_{c,j}'] = Var[T_{uj}] + \sum_{\substack{k=1 \\ k \neq j}}^g Var[T_{E,k}] \quad (4-43)$$

$$\bar{T}_{c,j}'' = \bar{T}_{uj} + \bar{T}_{hj} + \sum_{\substack{k=1 \\ k \neq j}}^g \bar{T}_{E,k} , \quad Var[T_{c,j}''] = Var[T_{uj}] + Var[T_{hj}] + \sum_{\substack{k=1 \\ k \neq j}}^g Var[T_{E,k}] \quad (4-44)$$

The arrival probabilities $b_{m,j}'$ and $b_{m,j}''$ can be obtained using the two-moment approximation method presented in [71]. According to the method, a given random variable T , which is characterized by the mean $\bar{T}$ and the coefficient of variance c , is approximated by means of the substitute Probability Distribution Function (PDF):

Chapter 4. A Systematic Multiple Access Scheme

$$0 \leq c \leq 1: F(t) = \begin{cases} 0 & 0 \leq t \leq t_1 \\ 1 - e^{-(t-t_1)/t_2} & t \geq t_1 \end{cases}, \text{ where } t_1 = \bar{T} \cdot (1-c) \text{ and } t_2 = \bar{T} \cdot c \quad (4-45)$$

$$c \geq 1: F(t) = 1 - pe^{-t/t_1} - (1-p)e^{-t/t_2}, \text{ where } t_{1,2} = \bar{T} \cdot \left(1 \pm \sqrt{\frac{c^2-1}{c^2+1}}\right)^{-1} \text{ and } p = \bar{T} / 2t_1 \quad (4-46)$$

Using the above substitute PDF, the conditional cycle times T_c' and T_c'' with the means and variances (4-43) (4-44) are approximately described. Then the arrival probabilities $b_{m,j}'$ and $b_{m,j}''$ can be calculated using formula (4-38) (4-39) accordingly.

4.5.4 Recursive Algorithm for the CTMC Model

A numerical algorithm is developed using the expressions derived for the Markov chain state probabilities and the conditional cycle times. The recursive algorithm alternates iteration of the Markov chain state probabilities and the cycle times. During each iteration cycle, the state probabilities of all queues are determined in a cyclic manner, depending on which the conditional cycle times are updated. These values will be used in the next iteration cycle. The termination of the iteration is decided by the convergence criteria. The detailed algorithm is described as follows.

Step 1: Initialize the Markov chain state probabilities for all queues, and the mean and coefficient of variation for the time interval between the polling epochs. Considering all queues are empty during the initialization period, then for a particular queue j :

$$P_{i,j}^{(0)} = \begin{cases} 1 & i = 0 \\ 0 & i = 1, 2, \dots, S_j \end{cases}$$

$$\bar{T}_{E,j} = \bar{T}_{uj} + (1 - P_{0,j}^{(0)}) \cdot \bar{T}_{hj}$$

$$Var[T_{E,j}] = Var[T_{uj}] + (1 - P_{0,j}^{(0)}) \cdot Var[T_{hj}] + P_{0,j}^{(0)} \cdot (1 - P_{0,j}^{(0)}) \cdot \bar{T}_{hj}^2$$

Step 2: Repeat (n^{th} iteration cycle, $n > 0$)

Chapter 4. A Systematic Multiple Access Scheme

For queue j do

Begin:

1. Calculate the mean and variation for the conditional cycle time $T'_{c,j}$ and $T''_{c,j}$.

$$\bar{T}'_{c,j} = \bar{T}_{uj} + \sum_{\substack{k=1 \\ k \neq j}}^g \bar{T}_{E,k}, \quad \bar{T}''_{c,j} = \bar{T}_{uj} + \bar{T}_{hj} + \sum_{\substack{k=1 \\ k \neq j}}^g \bar{T}_{E,k}$$

$$Var[T'_{c,j}] = Var[T_{uj}] + \sum_{\substack{k=1 \\ k \neq j}}^g Var[T_{E,k}], \quad Var[T''_{c,j}] = Var[T_{uj}] + Var[T_{hj}] + \sum_{\substack{k=1 \\ k \neq j}}^g Var[T_{E,k}]$$

$$c'_{c,j} = \sqrt{Var[T'_{c,j}]} / \bar{T}'_{c,j}, \quad c''_{c,j} = \sqrt{Var[T''_{c,j}]} / \bar{T}''_{c,j}$$

2. Calculate the arrival probabilities $b'_{m,j}$ and $b''_{m,j}$ by means of the two-moment approximation method.

$$\text{when } c'_{c,j} = 0: b'_{m,j} = \frac{(\lambda_j t'_{1,j})^m \cdot e^{-\lambda_j t'_{1,j}}}{m!}, \text{ where } t'_{1,j} = \bar{T}'_{c,j}$$

$$\text{when } 0 < c'_{c,j} < 1: b'_{m,j} = \frac{(\lambda_j t'_{2,j})^m \cdot e^{-\lambda_j t'_{2,j}}}{(1 + \lambda_j t'_{2,j})^{m+1}} \cdot \sum_{k=0}^m \left(\frac{t'_{1,j}}{t'_{2,j}} \right)^k \cdot \frac{(1 + \lambda_j t'_{2,j})^k}{k!}$$

$$\text{where } t'_{1,j} = \bar{T}'_{c,j} \cdot (1 - c'_{c,j}) \text{ and } t'_{2,j} = \bar{T}'_{c,j} \cdot c'_{c,j}$$

$$\text{when } c'_{c,j} \geq 1: b'_{m,j} = p'_j \cdot \frac{(\lambda_j t'_{1,j})^m}{(1 + \lambda_j t'_{1,j})^{m+1}} + (1 - p'_j) \cdot \frac{(\lambda_j t'_{2,j})^m}{(1 + \lambda_j t'_{2,j})^{m+1}}$$

$$\text{where } t'_{1,j} = \bar{T}'_{c,j} / \left(1 + \sqrt{\frac{c'^2_{c,j} - 1}{c'^2_{c,j} + 1}} \right), t'_{2,j} = \bar{T}'_{c,j} / \left(1 - \sqrt{\frac{c'^2_{c,j} - 1}{c'^2_{c,j} + 1}} \right), \text{ and } p'_j = \bar{T}'_{c,j} / 2t'_{1,j}$$

$$\text{when } c''_{c,j} = 0: b''_{m,j} = \frac{(\lambda_j t''_{1,j})^m \cdot e^{-\lambda_j t''_{1,j}}}{m!}, \text{ where } t''_{1,j} = \bar{T}''_{c,j}$$

$$\text{when } 0 < c''_{c,j} < 1: b''_{m,j} = \frac{(\lambda_j t''_{2,j})^m \cdot e^{-\lambda_j t''_{2,j}}}{(1 + \lambda_j t''_{2,j})^{m+1}} \cdot \sum_{k=0}^m \left(\frac{t''_{1,j}}{t''_{2,j}} \right)^k \cdot \frac{(1 + \lambda_j t''_{2,j})^k}{k!}$$

$$\text{where } t''_{1,j} = \bar{T}''_{c,j} \times (1 - c''_{c,j}) \text{ and } t''_{2,j} = \bar{T}''_{c,j} \cdot c''_{c,j}$$

Chapter 4. A Systematic Multiple Access Scheme

$$\text{when } c_{c,j}'' \geq 1: b_{m,j}'' = p_j'' \cdot \frac{(\lambda_j t_{1,j}'')^m}{(1 + \lambda_j t_{1,j}'')^{m+1}} + (1 - p_j'') \cdot \frac{(\lambda_j t_{2,j}'')^m}{(1 + \lambda_j t_{2,j}'')^{m+1}}$$

$$\text{where } t_{1,j}'' = \bar{T}_{c,j}'' / \left(1 + \sqrt{\frac{c_{c,j}''^2 - 1}{c_{c,j}''^2 + 1}} \right), t_{2,j}'' = \bar{T}_{c,j}'' / \left(1 - \sqrt{\frac{c_{c,j}''^2 - 1}{c_{c,j}''^2 + 1}} \right), \text{ and } p_j'' = \bar{T}_{c,j}'' / 2t_{1,j}''$$

3. Calculate the Markov-chain state probabilities for the current iteration cycle.

$$P_{k,j}^{(n+1)} = P_{0,j}^{(n)} \cdot b_{k,j}' + \sum_{i=1}^{k+1} P_{i,j}^{(n)} \cdot b_{k+1-i,j}'' \quad k = 0, 1, \dots, S_j - 1$$

$$P_{S_j,j}^{(n+1)} = P_{0,j}^{(n)} \cdot \left(\sum_{i=S_j}^{\infty} b_{i,j}' \right) + \sum_{i=1}^{S_j} P_{i,j}^{(n)} \cdot \left(\sum_{m=S_j+1-i}^{\infty} b_{m,j}'' \right)$$

4. Update the mean and coefficient of variation for the time interval between the scanning epochs.

$$\bar{T}_{E,j} = \bar{T}_{uj} + (1 - P_{0,j}^{(n+1)}) \cdot \bar{T}_{hj}$$

$$Var[T_{E,j}] = Var[T_{uj}] + (1 - P_{0,j}^{(n+1)}) \cdot Var[T_{hj}] + P_{0,j}^{(n+1)} \cdot (1 - P_{0,j}^{(n+1)}) \cdot \bar{T}_{hj}^2$$

5. Calculate the current convergence indicator e according to:

$$e = \sum_{j=1}^g \Delta_j, \text{ with } \Delta_j = \sum_{k=1}^{S_j} k \cdot P_{k,j}^{(n+1)} - \sum_{k=1}^{S_j} k \cdot P_{k,j}^{(n)}$$

End

Until: $e < 10^{-6}$

Step 3: Calculate the system characteristics through the arbitrary time state probabilities and the arrival probabilities during the backward recurrence time [69].

1. Calculate the arrival probabilities during the backward recurrence time $b_{m,j}^{**}$ and $b_{m,j}^*$.

$$b_{m,j}^{**} = \begin{cases} \frac{1}{\lambda_j \cdot (t_{1,j}' + t_{2,j}')} \cdot \left(1 - \sum_{k=0}^m \frac{(\lambda_j t_{1,j}')^k}{k!} \cdot e^{-\lambda_j t_{1,j}'} \right) + \frac{t_{2,j}'}{t_{1,j}' + t_{2,j}'} \cdot b_{m,j}' & 0 \leq c_{c,j}' < 1 \\ \frac{(\lambda_j t_{1,j}')^m}{2(1 + \lambda_j t_{1,j}')^{m+1}} + \frac{(\lambda_j t_{2,j}')^m}{2(1 + \lambda_j t_{2,j}')^{m+1}} & c_{c,j}' \geq 1 \end{cases}$$

$$b_{m,j}^{**} = \begin{cases} \frac{1}{\lambda_j \cdot (t_{1,j}^{**} + t_{2,j}^{**})} \cdot \left(1 - \sum_{k=0}^m \frac{(\lambda_j t_{1,j}^{**})^k}{k!} \cdot e^{-\lambda_j t_{1,j}^{**}} \right) + \frac{t_{2,j}^{**}}{t_{1,j}^{**} + t_{2,j}^{**}} \cdot b_{m,j}^{**} & 0 \leq c_{c,j}^{**} < 1 \\ \frac{(\lambda_j t_{1,j}^{**})^m}{2(1 + \lambda_j t_{1,j}^{**})^{m+1}} + \frac{(\lambda_j t_{2,j}^{**})^m}{2(1 + \lambda_j t_{2,j}^{**})^{m+1}} & c_{c,j}^{**} \geq 1 \end{cases}$$

2. Calculate the arbitrary time state probabilities according to:

$$P_{k,j}^{**} = \frac{\bar{T}_{c,j}'}{\bar{T}_{c,j}} \cdot P_{0,j} \cdot b_{k,j}^{**} + \frac{\bar{T}_{c,j}''}{\bar{T}_{c,j}} \cdot \sum_{i=1}^{k+1} P_{i,j} \cdot b_{k+1-i,j}^{**} \quad k = 0, 1, \dots, S_j - 1$$

$$P_{S_j,j}^{**} = \frac{\bar{T}_{c,j}'}{\bar{T}_{c,j}} \cdot P_{0,j} \cdot \left(\sum_{i=S_j}^{\infty} b_{i,j}^{**} \right) + \frac{\bar{T}_{c,j}''}{\bar{T}_{c,j}} \cdot \sum_{i=1}^{S_j} P_{i,j} \cdot \left(\sum_{m=S_j+1-i}^{\infty} b_{m,j}^{**} \right)$$

$$\text{where } \bar{T}_{c,j} = P_{0,j} \cdot \bar{T}_{c,j}' + (1 - P_{0,j}) \cdot \bar{T}_{c,j}''$$

3. Calculate the system characteristics using the arbitrary time state probabilities.

$$\text{Block probability in queue } j: b_j = P_{S_j,j}^{**}$$

$$\text{Mean waiting delay in queue } j: \bar{T}_{w,j} = \frac{L_j}{\lambda_j \cdot (1 - b_j)}, \text{ where } L_j = \sum_{k=1}^{S_j} k \cdot P_{k,j}^{**}$$

4.6 BGP-Based MAC Protocol

In this section, we will first propose a parameter-based Call Admission Control (CAC) mechanism based on the BGP scheme. It will classify ONUs in consideration of bandwidth availability and QoS requirements expressed in SLA parameters from different end users. Then the CAC mechanism will be combined with the bandwidth allocation scheme to compose a systematic MAC protocol for EPON networks based on the BGP scheduling scheme. The novel BGP-based MAC protocol will contain three parts that operate autonomously but are closely related to each other: the parameter-based CAC mechanism, the EDA algorithm, and the BGP scheme. The kernel of the protocol is the BGP scheduling scheme.

In the BGP-based MAC protocol, the parameter-based CAC mechanism will divide total ONUs into BG ONU and non-BG ONU groups according to their requirements. In addition, CAC will decide how many entries/bandwidth units should be assigned to each BG ONU. With the CAC mechanism, some of the QoS requirements from certain traffic flows can be ensured while some could not be achieved because of unavailability of enough bandwidth. The traffic flows without bandwidth guarantee will be considered as the traffic flows from non-BG ONUs with best-effort service provided.

Based on the partition of ONUs by the CAC scheme, the EDA algorithm will generate the Entry Table and the non-BG ONU List to set the polling sequences for both BG ONUs and non-BG ONUs. The objective of the EDA algorithm is to prevent or reduce the bursts of the traffic, which are caused by the scheduling procedure during the data transmission. Then by the BGP scheduling, the OLT will poll ONUs one after another based on the results of EDA, inviting ONUs transmit data over the channel. The entire protocol will be implemented at the OLT.

Among the three parts of the MAC protocol, the BGP scheme and EDA algorithm have been detailed developed in the previous sections and will not be iterated in length here. The newly proposed parameter-based CAC mechanism will be described in detail as follows.

4.6.1 Parameter-based CAC Mechanism

In general, the CAC mechanism decides which transmission requests can be admitted for transmission. Its decision is based on whether the QoS requirements of each request can be met under the limited network resources. CAC mechanisms can be classified as non-statistical allocation (also known as peak rate allocation) and statistical allocation [72]. According to [73], there are two basic approaches for admission control: the parameter-based and measurement-based approaches.

Chapter 4. A Systematic Multiple Access Scheme

In our proposal, we employ a parameter-based CAC mechanism that uses the peak rate for decision. The beauty of our proposed CAC mechanism is that it will not reject any transmission request from ONUs, rather it will accept all requests with certain QoS requirement at all circumstance. If the requirements of the transmission cannot be satisfied through analysis, the CAC mechanism will suggest to provide it with the downgraded service. At the system initialization, CAC decides the members of the BG ONU and non-BG ONU groups according to the SLA parameters and determines the number of entries each BG ONU can be allocated. During the operation of the system, CAC will decide whether or not to admit a new ONU as a BG ONU based on the utilization of network resources and QoS requirements of the new and existing BG ONUs. If a new ONU cannot be considered as a BG ONU, it could also be served as a non-BG ONU with best-effort service.

We assume that the SLA contract includes bandwidth and delay requirements, expressed as the parameter pair $\langle B, D \rangle$. Parameter B specifies the minimum peak bandwidth an ONU requests, while D indicates the maximum waiting delay the ONU can tolerate. Each ONU has an individual parameter pair according to the requirements of their end users.

During the initialization period, based on the bandwidth requirement B , the OLT will pre-calculate the worst case average delay $E(D)$ for each ONU at system load $\rho=1$. The formulas used to do the pre-calculation are given in (4-26) (4-27) (4-28).

After getting the worst case delay $E(D)$, the OLT compares $E(D)$ with the delay requirement D for each ONU. If the delay requirement is violated for an ONU, the OLT will re-calculate the bandwidth B^* required to guaranteed D for this ONU. Then the OLT checks the utilization of network resources. If the network cannot provide bandwidth B^* , the ONU will be grouped as a non-BG ONU. Otherwise, the OLT will negotiate B^* with the ONU. If this ONU refuses to pay for the bandwidth, it will be grouped as a non-BG ONU; otherwise,

Chapter 4. A Systematic Multiple Access Scheme

the bandwidth units will be allocated to this ONU, which is grouped as a BG ONU. Figure 4.7 is the flowchart of the CAC mechanism for initialization procedure.

Figure 4.7 Flowchart of the CAC Operations During the Initialization

While not all ONUs have data packets to transmit during the initialization period, some ONUs may be inactive at first and request for network transmission service during the operation period. In this case, the OLT will make the admission decision for the new ONUs. The procedure is similar to that in the initialization period by the CAC mechanism.

According to those formulas, the OLT gets the worst case delay $E(D)$ for the new ONU k and other existing ONUs. If $E(D_k) > D_k$ for the new ONU, the OLT will re-calculate bandwidth B_k^* required to guarantee D_k . Note that the re-calculation may change other ONUs' worst-case delays. If other ONUs' delay requirements are violated due to the re-calculation, the new ONU will be grouped as a non-BG ONU. Otherwise, the OLT checks the utilization of the network resources and negotiates B_k^* with the new ONU, then makes the final admission decision.

4.6.2 Implementation of the BGP-Based MAC Protocol

Composed by the admission control and bandwidth allocation mechanisms, the BGP-based MAC protocol controls the upstream transmission in the EPON networks, to permit the multiple ONUs delivering data packets over the channel without collision. During the initialization period of EPON, the parameter-based CAC mechanism is first invoked to process the ONUs' requirements expressed in SLA parameters, decide the members of the BG and non-BG ONU groups and determine the number of entries each BG ONU can be allocated.

After implementing the CAC mechanism, the OLT employs the EDA algorithm to prepare the Entry Table and the non-BG ONU List that decide the polling sequence for transmission. Since a BG ONU may need more than one bandwidth unit, it could be polled more than one time in each polling cycle. EDA evenly distributes the multiple entries of the same ONU among all entries in the Entry Table to space out the traffic from this ONU. So the traffic of each ONU can be evenly distributed over the upstream channel when it is polled based on the Entry Table. It will benefit to the BG ONUs with high traffic load since the packets in the buffer can be allowed to transmit after the same intervals. Thus there will be reduced bursts introduced to the data transmission in the access network.

Chapter 4. A Systematic Multiple Access Scheme

Finally the BGP scheme is invoked to control the upstream data transmission from multiple ONUs to the OLT after the Entry Table and non-BG ONU List have been ready. Since all entries in the Entry Table are allocated to fixed BG ONUs or dynamically assigned to non-BG ONUs, the Entry Table together with the non-BG ONU List will determine the scheduling order of ONUs, based on which the OLT will poll ONUs one after another. When the OLT decides to poll the next entry in the Table, it will first locate which ONU is allocated this entry and then grants a transmission window to this ONU by sending a GRANT message. Upon receiving the message, the ONU reports its buffer occupancy status by sending a REPLY message to the OLT and transmits data packets up to the duration of the transmission window. The detailed operation of the EDA algorithm and BGP scheme can be found in the previous sections.

During the operation of EPON systems, inactive ONUs may request transmission thus leading to the re-initialization of the BGP-based MAC protocol. The protocol will also be re-initialized in the case that non-BG ONUs request to convert to BG ONUs for transmitting QoS traffic, and the case that traffic for the active ONU is changed leading to the variation of ONU's transmission requirements. We suggest the re-initialization of MAC protocol to be processed in the back-end and the system's operation is not necessary to be interrupted.

When the MAC protocol is to be re-initialized due to the inactive ONUs request for transmission or any of the other two cases, the backup processor of OLT will re-invoke CAC and EDA to work out the new Entry Table and non-BG List. At the same time, the working processor will continue the current operation to schedule ONUs using the old Entry Table and non-BG List by BGP. Only after the new Entry Table and non-BG List are ready, and a whole polling cycle is finished, then the OLT will switch to use the backup processor to arbitrate the upstream data transmission by the new Entry Table and non-BG List from next polling cycle. And the previous working processor will become the backup processor. By

this method, the BGP-based MAC protocol can be re-invoked without interrupting the current operation, thus the impact of re-initialization to the system can be very limited.

4.7 Simulation Experiments

In order to evaluate the efficiency of the proposed BGP scheme, we build a tree-based experimental model of EPON using the simulation tool — OPNET Modeler. First the extensive discrete event simulation experiments have been conducted to show the performances of the BGP scheme with comparison to the performances obtained by the famous IPACT scheme under the same experimental design. Then the simulation results are compared with the results of the mathematical estimation model regarding to the average waiting delay experienced by packets in the BGP scheme. Finally we present the performances of the novel BGP-based MAC protocol obtained from the simulation experiments.

Microsoft Windows 2000 is the platform used by OPNET Modeler. To prove the correctness of simulations, we have first simulated the EPON system with IPACT scheme and compared the results with previous papers related. After getting the similar results, we then conducted simulations for BGP scheme accordingly. The correctness of simulations has been further proved by comparing with the BGP mathematical model in Section 4.7.1.3. For each data point in the experiments, we have collected results from multiple simulation trials and average to get the mean value. It is verified that the mean value of results falls within the 95% confidence interval.

All simulation experiments have been conducted with system parameters close to the real world. We will present the experiment design as well as the experiment results in the following subsections. The mathematical estimation model take the same network parameters and scenario as those used in the simulation experiments.

4.7.1 Experiments on the BGP Scheme

4.7.1.1 Simulation Designs

In the experiments, we design our simulation model with reference of the real EPON architecture and system parameters. We built up the system model as described in Figure 4.1. The total number of the ONUs in the EPON is 64. The 64 ONUs are divided into two groups: 20 BG ONUs and 44 non-BG ONUs. The IDs of BG and non-BG ONUs are assumed to be continuous respectively. ONU₁~ONU₂₀ are BG ONUs and ONU₂₁~ONU₆₄ are non-BG ONUs. This assumption makes non-BG ONUs' List simple to implement. It composes of ONUs with continuous IDs that start from ONU₂₁ and end at ONU₆₄. Among those BG ONUs, ONU₅ has 20 entries; ONU₈, ONU₁₂, ONU₁₇ have 10 entries; ONU₁, ONU₃, ONU₆, ONU₁₀, ONU₁₅, ONU₁₈ have 4 entries; all of the 10 remaining ONUs have only 1 entry. As the BG ONUs occupy 84 entries totally, the remaining 16 entries could be assigned to non-BG ONUs.

The Entry Table is filled by EDA algorithm. According to EDA, we first allocate entries to ONU₅ that has the maximum number of entries. Then ONU₈, ONU₁₂ and ONU₁₇ are allocated. Following them, ONU₁, ONU₃, ONU₆, ONU₁₀, ONU₁₅ and ONU₁₈ will be allocated. Finally, the BG ONUs requiring only one entry will be assigned entries. Let's see an example of ONU₈. The first entry of ONU₈ is entry 8 (number of the ONU ID) based on EDA. Since ONU₈ occupies 10 entries in all 100 entries, the distance between two neighboring entries of ONU₈ should be $100/10 = 10$. Other entries of ONU₈ should be entry 18, 28, 38, 48, 58, 68, 78, 88, and 98. The final result of entry assignment is shown in Figure 4.8.

Chapter 4. A Systematic Multiple Access Scheme

1	1	2	12	3	3	4	2	5	5	6	6	7	17	8	8	9	4	10	5
11	10	12	12	13	7	14	9	15	5	16	15	17	17	18	8	19	18	20	5
21	11	22	12	23	13	24	14	25	5	26	1	27	17	28	8	29	3	30	5
31	6	32	12	33	16	34	19	35	5	36	10	37	17	38	8	39	20	40	5
41	15	42	12	43		44	18	45	5	46		47	17	48	8	49		50	5
51	1	52	12	53	3	54		55	5	56	6	57	17	58	8	59		60	5
61	10	62	12	63		64		65	5	66	15	67	17	68	8	69	18	70	5
71		72	12	73		74		75	5	76	1	77	17	78	8	79	3	80	5
81	6	82	12	83		84		85	5	86	10	87	17	88	8	89		90	5
91	15	92	12	93		94	18	95	5	96		97	17	98	8	99		100	5

Figure 4.8 Entry Table with 100 Entries for the Simulation Experiments

According to IEEE 802.3, the Ethernet frame can be variable from 64 bytes to 1518 bytes. In the model, we design packet with fixed size as 500 bytes or 4000 bits. Each ONU has its own source generating the fixed size packets, with exponential distributed inter-arrival time, ranging from 0.233ms to 2.56ms making total traffic load of the entire network changing from 0.1 to 1.1 accordingly.

In the simulation model, the OLT generates GRANT messages to poll ONUs with the granted window size. ONUs produce REPLY messages to inform the OLT about the number of packets in transmission. The GRANT message is formatted to contain the ONU ID being polled and the granted transmission window size. The REPLY message contains its own ID, the number of packets in transmission and the transmission window size being granted by the OLT. Every ONU has a round-trip propagation delay from the OLT to the ONU. The propagation delay for each ONU is assumed to be randomly (uniformly) over the interval $[50 \mu s, 100 \mu s]$. These values correspond to distances between the OLT and ONUs ranging from 10 to 20 km.

In this study, we consider the transmission speed of the EPON network is 1Gb/s, the upstream link rate is divided into 100 bandwidth units, each with 10Mb/s. It makes the Entry Table have 100 entries. The maximum transmission window size W_{max} is set to 15000 bytes

or 30 packets. The threshold T is set to $2/3W_{max}$, equal to 20 packets. Every ONU has a finite memory buffer of size, which is set to 10 Mbytes.

4.7.1.2 Simulation Results of BGP Comparing with IPACT

In this section, we will present and explain a series of simulation results. In addition to the performance of different kinds of ONUs in the BGP scheme, we also present the performance of the ONUs in the IPACT scheme for comparison. The simulation design for the IPACT model is the same as that for the BGP scheme, with the exception that all ONUs in IPACT are treated in the same way [27], no difference on bandwidth requirements from the ONUs.

Figure 4.9 shows the relationship between the average packet delay for different kinds of ONUs' and the entire network load. It shows that, in the BGP scheme, the more entries a BG ONU occupies, the less average delay it suffers. The average delays for the BG ONUs with 20 and 10 entries are kept at very low values because these ONUs are allocated enough bandwidth units/entries for transmission. They are almost unchanged despite of the increase of the traffic load. The curve of average delay for the BG ONUs with 4 entries is kept smooth and low when the load is less than 0.9, after which it increases rapidly. The BG ONUs with only one entry experience much more average delay than other BG ONUs. The curve increases in the very beginning with the network load of 0.1. It quickly becomes smooth and at quite high value after the network load reaches 0.3. The delay for non-BG ONUs is larger than those of BG ONUs with multiple entries, while a little less than that of BG ONUs with only one entry. The reason is that many ONUs are allocated much more entries than they can consume in this scenario. Hence these superfluous transmission windows are dynamically assigned to the non-BG ONUs, leading to the average window of a non-BG ONU exceeds the transmission window of one entry. So the performances of non-

Chapter 4. A Systematic Multiple Access Scheme

BG ONUs are slightly better than those of the BG ONUs with only one entry. As for the average delay of IPACT ONUs, it is larger than the delays of BG ONUs with multiple entries and less than the delay of non-BG ONUs in the BGP scheme. The curve increases slowly before the load of 0.6, after which it promptly climbs up. The average delay reaches a balanced point after the load of 0.7.

Figure 4.9 Average Delay of Different ONUs in BGP and IPACT

Figure 4.10 shows the relationship between the throughput of different groups of ONUs and the entire network load. In the BGP scheme, as the number of non-BG ONUs is more than that of BG ONUs, the throughput of the former group preponderates over that of the latter one when the traffic load is light. With the increase of the network load, the throughput of the BG ONUs grows gradually. While the throughput of the non-BG ONUs decreases after the load 0.4, because fewer and fewer superfluous transmission windows can be used to poll the non-BG ONUs due to the higher traffic load of BG ONUs. At last, the throughput of the BG ONUs exceeds that of the non-BG ONUs at the load of 1.0. The throughput of all ONUs in the IPACT scheme is very similar to that of all ONUs in the BGP scheme when the load is very low. After the load 0.3, it increases more rapidly than the throughput of BGP

ONUs. The reason is that, with BGP scheme, the traffic load of the same ONU may be separated for transmission in one polling cycle. Thus there will be bandwidth loss due to more control message transmission and more guard time between different ONUs data transmissions. This is a trade off between the throughput and the BGP's efficiency to transmit the real-time traffic with reduced burst, which have been discussed in the previous section.

Figure 4.10 Throughput of Different ONUs in BGP and IPACT

Figure 4.11 shows the relationship between the buffer occupancy of different kinds of ONUs and the entire network load. The curve shapes in this figure are very similar to those of average delay vs. network load in Figure 4.9. In the BGP scheme, the more entries a BG ONU occupies, the shorter queue length existing in its buffer. The queue lengths of the BG ONU with 20 and 10 entries are kept at the lowest values, far less than the finite buffer size of 10 Mbytes. They increase slowly as the network load increases. The curve of the queue length for BG ONUs with 4 entries goes up slowly when the network traffic load is less than 0.9. There is a sharp increase after the load 1.0. Finally it reaches the bound of buffer size when the load reaches 1.1. When the traffic load is light, non-BG ONUs have the smaller

Chapter 4. A Systematic Multiple Access Scheme

queue length than BG ONUs with only one entry. After the network load 0.4, the catch that of the latter ones and queues of both kinds of ONUs come to the limited length of buffer size. Among all kinds of ONUs, BG ONUs with one entry are the earliest ones to reach the upper bound buffer size, when load is only 0.3. That is because they are allocated only one entry each cycle in the scenario, whose transmission window is even less than the average window of non-BG ONUs with many superfluous windows assigned. The curve of ONUs queue length in the IPACT is located between the curves of BG ONUs with multiple entries and the curve of non-BG ONUs in the BGP scheme. It increases slowly before the load 0.6. After that it promptly climbs to the limited buffer size of 10 Mbytes at the load 0.7.

Figure 4.11 Average Buffer Occupancy of Different ONUs in BGP and IPACT

Figure 4.12 shows the relationship between the average loss rate of different kinds of ONUs and the entire network load. In the BGP scheme, the BG ONUs with 20 and 10 entries have no data loss despite the increase of the network load since they have been allocated enough bandwidth units for transmission. The BG ONUs with 4 entries experience data loss after the network load exceeds 1.0. The BG ONUs with only one entry begin to drop packets after the load 0.2, earlier than the non-BG ONUs, whose loss starts to occur at the load 0.3.

Chapter 4. A Systematic Multiple Access Scheme

The average loss rate of the non-BG ONUs is less than that of the BG ONUs with one entry, because they have more average transmission window than that of one entry due to the many superfluous windows assigned. In the IPACT scheme, the ONUs have no data loss before the network load of 0.6. After that, they begin to drop data. The loss rate increases continuously and locates between the curves of the BG ONUs with multiple entries and the non-BG ONUs in the BGP scheme.

Figure 4.12 Average Loss Rate of Different ONUs in BGP and IPACT

We can see from this special scenario that, it is possible for the non-BG ONUs to get better performances than 1-entry BG ONUs, if there are too many superfluous transmission windows in the system causing the non-BG ONUs being assigned more than one entry in average. But non-BG ONUs cannot always perform better than 1-entry BG ONUs. In the system with more non-BG ONUs than BG ONUs and only a few superfluous transmission windows being assigned to non-BG ONUs, the performances of 1-entry ONUs will definitely be better than those of non-BG ONUs. It is also possible for 1-entry BG ONUs in BGP scheme have worse performances than ONUs in IPACT. That is because the ONUs are only getting average performances in IPACT, but different ONUs in BGP can obtain

different performances according to the SLA contracts. In BGP, although 1-entry BG ONUs and non-BG ONUs may perform worse than ONUs in IPACT, ONUs with multiple ONUs can get much better performances than ONUs in IPACT.

Through the experimental simulations for the BGP scheme and comparison with the IPACT scheme, we have the following conclusions. In the BGP scheme, the more entries a BG ONU occupies, the better performance it can behave in terms of less average delay, higher throughput, shorter buffer occupancy and lower loss rate. A BG ONU with multiple entries has better performance than an ONU in IPACT scheme. As a sacrifice, the performance of a non-BG ONU is worse than that of an IPACT ONU. The experiments show that, although the low-demand users are served on the best-effort basis, the proposed BGP scheme can guarantee bandwidth for high-demand users, and provide good performance to them. Incorporating SLA contracts into the DBA scheme design, the BGP scheme is proved to be a powerful bandwidth allocation scheme for the upstream transmission in EPON networks.

4.7.1.3 Comparison with the Results of Mathematical Estimation Model

According to the mathematical estimation model developed in the previous section, we can get a group of outcomes on the point-to-point average delay $E(D)$ by formulas (4-26), (4-27) and (4-28) for different ONUs in the BGP scheme. To compare with the simulation results, the estimation adopts the same network parameters as used in the simulation experiment. The average waiting delay $E(W_1)$ can be estimated for BG ONUs and non-BG ONUs separately from formula (4-27) directly. When calculating the average delay $E(W_2)$, there are different expressions for different stages of low, moderate and heavy traffic load. Note that different groups of ONUs have different stages of traffic load. $E(W_2)$ increases rapidly

Chapter 4. A Systematic Multiple Access Scheme

when the traffic rises from low load to heavy load, and the moderate load between is a very short stage comparatively.

We calculate the average delay $E(W_2)$ for different kinds of ONUs with entire network traffic load changing from 0.1 to 1.1. For the BG ONUs with 20 and 10 entries, the load is always low because they are allocated enough bandwidth units and their traffic load can be separated over the multiple units for transmission. So the first expression of formula (4-28) can be simply used to calculate $E(W_2)$ for these ONUs.

For the BG ONUs with 4 entries, the first expression is adopted for the light load of $E(W_2)$ when the load is from 0.1 to 0.9. The load becomes moderate when the system load is between 0.9 and 1.05. The delay for BG ONUs in this stage should not exceed 50ms based on experience. Since the scan time of a polling cycle in our scenario ranges from 30ms to 40ms, the new arrival data should wait for another k (maximum value of k is 1 or 2) polling cycles for transmission in the moderate load stage. As k is a random variable, the average value $\bar{k}$ should be $0.5 \leq \bar{k} \leq 1$ for the moderate load of $E(W_2)$ from formula (4-28).

For the BG ONUs with 1 entry, the stage of moderate load ranges from the system load of 0.15 to that of 0.25. The average value $\bar{k}$ is also $0.5 \leq \bar{k} \leq 1$ for the moderate load $E(W_2)$. After the load 0.25, the load is heavy for ONUs with only 1 entry. Then the fourth expression of formula (4-28) will be used to calculate $E(W_2)$ for these ONUs.

For non-BG ONUs, the load is moderate when the system load is between 0.3 and 0.4. The delay in this stage for non-BG ONUs should not exceed 100ms based on experience. Since the scan time of a polling cycle in our model ranges 30~40ms, the new arrival data should wait for another k (maximum value of k is 1~4) polling cycles for transmission. Accordingly, the average value $\bar{k}$ should be $0.5 \leq \bar{k} \leq 2$ for the moderate load of $E(W_2)$.

As to the light and heavy traffic load, the first and last expression of formula (4-28) will be adopted for non-BG ONUs respectively.

Figure 4.13 compares the mathematical estimation outcomes with the simulation results of different kinds of ONUs in the BGP scheme. It shows that the estimation results match well with the simulation results, especially for the BG ONUs with multiple entries. There are some discrepancies existing between the estimation and simulation results for the non-BG ONUs, because some approximations have been made in the estimation to get the average number of entries $\bar{P}$ used to poll a second ONU, the average number of entries $\bar{n}$ assigned to a non-BG ONU, and the average transmission window size $\bar{W}_{non}$ for all non-BG ONUs, etc. Our mathematical estimation results are reasonably well to prove the efficiency of the proposed BGP scheme and testify the rationality of the simulation results. The matching with the simulation results also verifies the accuracy of the mathematical estimation model itself.

Figure 4.13 Comparison of Average Delay by Estimation and Simulation for BGP

4.7.2 Experiments on the BGP-Based MAC Protocol

We have conducted another group of simulation experiments to evaluate the performance of the entire BGP-based MAC protocol including the CAC mechanism, EDA algorithm and BGP scheme. A general scenario is considered under which the bandwidth and delay requirements of each ONU are set to random values.

4.7.2.1 Simulation Designs

The simulation model has been designed similar to the previous experiment for the BGP scheme. We take a general scenario for the simulation where each ONU's bandwidth and delay requirements are set based on the different random SLA requirements. According to the BGP-based MAC protocol, the OLT will process these requirements to decide the members of BG ONU and non-BG ONU groups by the parameter-based CAC mechanism, produce the Entry Table and the List by the EDA algorithm, and schedule multiple ONUs by the BGP scheme respectively.

Table 4.1 Original ONUs Parameters for the BGP-Based MAC Protocol

ONU ID	B_i (M bps)	D_i (s)	U_i (M bps)	ONU ID	B_i (M bps)	D_i (s)	U_i (M bps)
1	35.05	0.01	50	33	40	0.005	40
2	0	0	0	34	0	0	0
3	0	0.05	20	35	0	0	0
4	9.54	1	50	36	100	0.005	110
5	0	1	40	37	0	0	0
6	33	0.015	50	38	22	0.2	50
7	10.78	0	50	39	0	0	0
8	94	0.001	120	40	31.5	0.03	50
9	0	0	0	41	0	0	0
10	38.2	0.005	50	42	0	0.001	50
11	10	25	50	43	0	0	0
12	0	0	0	44	0	0	0
13	0	0	0	45	3.8	25	50
14	8	0	50	46	0	0	0
15	36	0.02	50	47	0	0	0
16	0	0	0	48	50	0.005	100
17	99	0.002	120	49	0	0	0
18	0	0	0	50	0	0	0
19	10	20	50	51	26	0	40
20	6.3	0	50	52	0	0	0
21	0	0	0	53	0	5	50
22	0	0	0	54	0	0	0

Chapter 4. A Systematic Multiple Access Scheme

23	18	25	50	55	0	0	0
24	0	0	0	56	0	0	0
25	0	0.01	50	57	12	0.05	50
26	10	10	20	58	0	0	0
27	0	0	0	59	0	0	0
28	0	0	0	60	0	0	0
29	196	0.001	250	61	0	0.03	30
30	5	0.01	50	62	0	0	0
31	0	0	0	63	20	0	20
32	0	0	0	64	0	0	0

There are 64 ONUs requesting data transmission service in the initialization period of the EPON network. The ONUs' original parameters are listed in Table 4.1. Besides the minimum bandwidth requirement B_i and maximum tolerable delay D_i for each ONU, Table 4.1 also includes an item U_i that indicates the maximum bandwidth each ONU affords to pay. The value of 0 in the table indicates that an ONU has no specific requirement for the corresponding item.

4.7.2.2 Simulation Results

In the BGP-based MAC protocol, the CAC mechanism is first invoked to process the ONUs original parameters. According to the original bandwidth requirement B , the OLT can pre-assign the ONUs to BG and non-BG ONU groups for the CAC implementation. It pre-allocates entries to BG ONUs as follows: ONU 29 is assigned 20 entries; ONU 8, 17, 36 have 10 entries; ONU 48 has 5 entries; ONU 1, 6, 10, 15, 33, 40 have 4 entries; ONU 38 and ONU 51 have 3 entries; ONU 7, 23, 57, 63 have 2 entries; ONU 4, 11, 14, 19, 20, 26, 30, 45 have only 1 entry. All remaining ONUs are pre-assigned as non-BG ONUs whose values are 0 as shown in Table 4.2.

Then the OLT pre-calculates the worst case delay $E(D)$ for each ONU based on the pre-allocated entry, and compares $E(D)$ with the delay requirement D for each ONU. In this scenario, there are totally 18 ONUs whose delay requirements D are violated by $E(D)$: ONU 1, 3, 4, 5, 8, 10, 19, 25, 26, 30, 33, 38, 42, 53, 57, 61. According to the CAC mechanism, the

Chapter 4. A Systematic Multiple Access Scheme

OLT will re-calculate the bandwidth B^* required to guaranteed D for each of the 18 ONUs one after another. After each re-calculation, the OLT checks up the network resources and negotiates with the corresponding ONU. The ONU is grouped as the BG ONU only when the network has enough bandwidth B^* available and the ONU can afford to pay for B^* . Otherwise the ONU will be grouped as the non-BG ONU.

Table 4.2 Results of the CAC Mechanism

ONU ID	Pre-allocated Entry Number	Entries Allocated After CAC	ONU ID	Pre-allocated Entry Number	Entries Allocated After CAC
1	4	5	33	4	0
2	0	0	34	0	0
3	0	0	35	0	0
4	1	4	36	10	10
5	0	0	37	0	0
6	4	4	38	3	4
7	2	2	39	0	0
8	10	0	40	4	4
9	0	0	41	0	0
10	4	5	42	0	0
11	1	1	43	0	0
12	0	0	44	0	0
13	0	0	45	1	1
14	1	1	46	0	0
15	4	4	47	0	0
16	0	0	48	5	5
17	10	10	49	0	0
18	0	0	50	0	0
19	1	2	51	3	3
20	1	1	52	0	0
21	0	0	53	0	0
22	0	0	54	0	0
23	2	2	55	0	0
24	0	0	56	0	0
25	0	0	57	2	4
26	1	0	58	0	0
27	0	0	59	0	0
28	0	0	60	0	0
29	20	20	61	0	0
30	1	5	62	0	0
31	0	0	63	2	2
32	0	0	64	0	0

Let's take ONU 30 as an example, since its bandwidth requirement is 5 Mbps originally, it is pre-allocate 1 entry by the OLT because each entry represents one bandwidth unit with 10 Mbps in this scenario. The result of pre-calculation $E(D_{30}) > D_{30}$, so the OLT should re-calculate bandwidth B_{30}^* required to guaranteed D_{30} , that is $B_{30}^* = 50$ Mbps. Since the

Chapter 4. A Systematic Multiple Access Scheme

network can provide the bandwidth, the OLT negotiates with ONU 30. ONU 30 agrees to bandwidth B_{30}^* because the maximum bandwidth it affords to pay is 50 Mbps . So ONU 30 will be allocate 5 entries after the CAC implementation instead of 1 entry as it requires originally. While in other examples of ONU 8 and ONU 53, they also need to be re-calculated because $E(D_8) > D_8$ and $E(D_{53}) > D_{53}$ after the pre-calculation, and the bandwidths required to guarantee the delay requirements are $B_8^* = 180 \text{ Mbps}$ and $B_{53}^* = 40 \text{ Mbps}$ respectively. However, the maximum bandwidth ONU 8 affords to pay is only 120 Mbps from Table 4.1. For ONU 53, the network cannot provide the required bandwidth B_{53}^* . So both ONUs are grouped to non-BG ONUs after the CAC mechanism.

After implementing the CAC mechanism, the OLT finalizes the members of both ONU groups (Table 4.2) and the entries allocated to BG ONUs: 20 entries to ONU 29; 10 entries to ONU 17 and ONU 36; 5 entries to ONU 1, 10, 30, 48; 4 entries to ONU 4, 6, 15, 38, 40, 57; 3 entries to ONU 51; 2 entries to ONU 7, 19, 23, 63; and 1 entry to ONU 11, 14, 20, 45. All other ONUs are non-BG ONUs. The effectiveness of the CAC mechanism can be shown as follows. There are 7 ONUs allocated more entries than required originally (ONU 1, 4, 10, 19, 30, 38, 57), and 3 ONUs assigned fewer entries (ONU 8, 26, 33). The number of non-BG ONUs has been changed from 39 to 42.

Based on the above results, the Entry Table and non-BG ONU List can be obtained by the EDA algorithm. The Entry Table is shown in Figure 4.14 . Since the BG ONUs occupy 99 entries in total, only 1 entry is remained to assign to non-BG ONUs. The List for non-BG ONUs is sequenced according to the non-BG ONUs' ID in ascending order, starting from ONU 2 and ending at ONU 64.

Chapter 4. A Systematic Multiple Access Scheme

1	1	2	7	3	6	4	29	5	4	6	36	7	17	8	48	9	29	10	10
11	30	12	57	13	38	14	29	15	15	16	36	17	17	18	40	19	29	20	51
21	1	22	19	23	23	24	29	25	6	26	36	27	17	28	48	29	29	30	10
31	30	32	4	33	57	34	29	35	11	36	36	37	17	38	38	39	29	40	15
41	1	42	40	43	14	44	29	45	7	46	36	47	17	48	48	49	29	50	10
51	30	52	51	53	6	54	29	55	4	56	36	57	17	58	57	59	29	60	20
61	1	62	63	63	38	64	29	65	15	66	36	67	17	68	48	69	29	70	10
71	30	72	40	73	19	74	29	75	23	76	36	77	17	78	6	79	29	80	4
81	1	82	51	83	57	84	29	85	38	86	36	87	17	88	48	89	29	90	10
91	30	92	15	93	40	94	29	95	45	96	36	97	17	98		99	29	00	63

Figure 4.14 Results of the EDA Algorithm

Then the OLT implements the BGP scheme based on the Entry Table and List, to poll ONUs and let them transmit data packets over the shared upstream channel. Figure 4.15 shows the relationship between average waiting delay and the entire network load for different kinds of ONUs in the scenario. As expected in the BGP scheme, the more entries a BG ONU occupies, the less average delay it experiences. The waiting delays for BG ONUs with 20, 10, 5 and 4 entries are almost unchanged despite of the increase of the network load, because they have been allocated enough bandwidth units for transmission. The average waiting delays for BG ONUs with 3 and 2 entries are low when the network load is light; they increase rapidly after the load reaches 0.6 and 0.4, respectively. BG ONUs with only one entry experience a much higher average waiting delay than other BG ONUs. The delay increases rapidly from the network load 0.1, and quickly levels off when the load reaches 0.3. The average delay for non-BG ONUs is more than those of BG ONUs with multiple entries consistently, and exceeds that of BG ONUs with one entry after the system load 0.7. The average delay of BG ONUs with one entry is even larger than that of the non-BG ONUs when the load is less than 0.7. The reason is that, when the traffic load is light, many BG ONUs with multiple entries cannot consume the transmission windows and the remainders will be used to poll the non-BG ONUs. Thus the average transmission window of

non-BG ONUs is possible to exceed that of one entry, leading to the better performances than the BG ONUs with only one entry at the light load.

Figure 4.16 shows the relationship between the average loss rate and the network load due to the limited buffer size for different kinds of ONUs. BG ONUs with 20, 10, 5 and 4 entries experience no data loss even when the network load reaches to 1. BG ONUs with 3 and 2 entries suffer data loss after the network load reaches 0.6 and 0.4, respectively. BG ONUs with only one entry begin to drop data frames after the load of 0.2, earlier than non-BG ONUs after the load of 0.3. The loss rate of non-BG ONUs exceeds that of BG ONUs with one entry when the load exceeds 0.7. Again, the enough superfluous transmission windows being assigned to the non-BG ONUs lead to the lower loss rate than BG ONUs with only one entry when the traffic load is low.

Figure 4.15 Average Delay of the BGP-Based MAC Protocol

Figure 4.16 Average Loss Rate of the BGP-Based MAC Protocol

4.8 Summary

The major contribution of this chapter is to propose a novel DBA scheme for the EPON upstream transmission within the MPCP framework. The proposed BGP scheme can provide the absolute QoS services to different kinds of ONUs with diverse bandwidth requirements by incorporating the SLA contracts into the bandwidth allocation scheme design. Another contribution is that a mathematical estimation model and a CTMC model have been developed to evaluate the performances of the BGP scheme. The CTMC model makes use of the Markov Chain state probabilities and the conditional cycle times to derive the related system characteristics.

Based on the BGP scheduling scheme, we have further developed a parameter-based CAC mechanism and combined it with the bandwidth allocation scheme to compose a systematic MAC protocol for the EPON networks. The BGP-based MAC protocol consists of three components: the CAC mechanism, the EDA algorithm and the BGP scheme, which can explicitly incorporate the SLA requirements into the system design to achieve bandwidth

Chapter 4. A Systematic Multiple Access Scheme

guarantee for the end users that subscribe a premium bandwidth-guaranteed service while providing best-effort service to the users that are more price-sensitive and prefer a lower-cost best-effort service.

The extensive simulation experiments have been conducted to evaluate the performances of the BGP scheme and the systematic BGP-based MAC protocol. The simulation results show that, although the low bandwidth-demand users are served on the best-effort basis, the proposed BGP scheme and MAC protocol can guarantee bandwidth availability for high bandwidth-demand users and can provide absolute QoS service to them. The BG ONUs with multiple bandwidth units have better performance than ONUs in the IPACT scheme, which are not explicitly differentiated on bandwidth requirements. The BGP scheme as well as the BGP-based MAC protocol is demonstrated to be efficient by the simulation results comparing with the IPACT scheme. In addition, the simulation results of BGP have been compared with the outcomes of the mathematical estimation model with regard to the packet waiting delay. The matching of the results testifies the rationality of the BGP scheme and verifies the efficiency of the model at the same time.

Chapter 5 . Proposed Solutions to Support DiffServ

Supporting Differentiated Services (DiffServ) with various requirements is highly regarded as a substantial requirement of the EPON access networks. An EPON must be capable of classifying the traffic into classes and serving each class differently to meet the service requirements. Different types of incoming traffic, such as voice communications, standard and high-definition video, video conferencing, real time transactions, and data traffic, can be mapped into the Quality of Service (QoS) traffic or best-effort traffic according to their requirements. The IEEE 802.1D together with its extensions P802.1P and P802.1Q [74], which are regarded as the QoS protocols at the MAC level, has provided the related standard for EPONs. The IEEE 802.1D supports all scheduling algorithms so long as they meet the transmission ordering requirements specified in clause 7.7.3 [74].

Our aim in this chapter is to work out solutions that can support DiffServ in EPONs by means of MAC protocols within the framework of IEEE 802.3ah MPCP protocol [4] and IEEE 802.1D. These solutions will extend the BGP-based MAC protocol proposed in chapter 4 to achieve this goal. A hierarchical scheduling architecture is suggested at the MAC layer to include the inter-ONU and intra-ONU scheduling. Adopted as the inter-ONU

scheduling scheme, the BGP-based MAC protocol can effectively allocate the upstream transmission capacity among multiple ONUs. Three intra-ONU scheduling schemes, the Strict Priority Queuing (SPQ) scheme, the Two-Stage Queuing (TSQ) scheme, and the Urgency Fair Queuing (UFQ) scheme, are proposed to combine with the inter-ONU scheduling scheme separately to support the differentiated services. They can provide different levels of fairness to different types of traffic in EPONs.

5.1 Hierarchical Scheduling

In EPONs, multiple ONUs should share the upstream channel capacity to transmit packets over the link. At the same time, different traffic streams in the same ONU will also share the bandwidth allocated to the ONU for transmission. The two levels of bandwidth allocation require the hierarchical MAC layer scheduling in EPONs in order to effectively support DiffServ as well as share the upstream transmission capacity.

Figure 5.1 Hierarchical Scheduling Architecture in EPONs

We deploy the hierarchical scheduling architecture shown in Figure 5.1 for EPON upstream transmissions, which consists of the inter-ONU scheduling and intra-ONU scheduling. The hierarchical scheduling will efficiently assign the upstream bandwidth for different ONUs and provide DiffServ to different traffic streams. The inter-ONU scheduling exploits the multiple access control schemes for bandwidth allocation among multiple ONUs within the IEEE 802.3ah MPCP framework. And the intra-ONU scheduling adopts the mechanisms supported by the IEEE 802.1D to provide DiffServ to different traffic streams in the same ONU.

5.1.1 Inter-ONU Scheduling

The scheduling among the multiple ONUs is the inter-ONU scheduling, which is usually performed by the OLT. As the central controller, the OLT cyclically polls multiple ONUs according to the adaptive sequence to control the upstream channel sharing among different ONUs. Contentions between the ONUs can be eliminated by employing efficient MAC protocols in the inter-ONU scheduling.

The BGP-based MAC protocol that has been proposed in chapter 4 is the foundation to support the inter-ONU scheduling in our solutions. The adopted MAC protocol consists of three parts operating autonomously but closely related to each other: the parameter-based CAC mechanism, the EDA algorithm, and the BGP scheduling scheme. Since the EDA algorithm and the BGP scheme have been presented in detail in chapter 4, we will not iterate in length here. As to the CAC mechanism, it will be modified to consider the individual traffic's requirements for DiffServ support based on the mechanism proposed in chapter 4. In the next section, the modified CAC mechanism will be described in detail.

5.1.1.1 Parameter-Based CAC Mechanism

Taking the specific QoS requirements of the individual traffic into account, the parameter-based CAC mechanism decides which QoS traffic can be admitted for transmission and what kind of service will be provided to the traffic. Its decision is based on whether the QoS requirements of all traffic streams can be met under the limited network resources. Our CAC scheme will never reject any traffic transmission request from the end user; instead it will accept all requests with certain QoS requirements at all circumstance. If the requirements of the traffic stream cannot be satisfied through analysis, the CAC scheme will suggest providing it with the downgraded service.

In the EPON networks, an ONU is connected with multiple end users and a user may contain multiple traffic streams with different QoS requirements for transmission. Hence, it will be a heavy burden for the OLT alone to decide the admission control for each individual traffic stream. In our modified CAC mechanism, the ONUs participate in the admission control in order to relieve the OLT's pressure, which is different from the mechanism presented in chapter 4 that is only performed by the OLT. During the system initialization, each ONU collects the QoS requirements from all traffic streams and groups them into up to eight traffic classes complying with the MPCP protocol, then reports the requirements of each traffic class to the OLT. Based on the information from all ONUs, the OLT determines which classes of QoS traffic can be supported and which can only be treated as the best-effort traffic for each ONU according to the CAC mechanism, and informs ONUs about the decisions. Then each ONU will further decide the services provided to different traffic streams after negotiating with the end users. After the CAC mechanism, the BG and non-BG ONU groups will be determined.

The requirement of each traffic stream is expressed as the parameter pair $\langle B, D, U \rangle$, where B specifies the minimum peak bandwidth required to transmit the traffic, D indicates

the maximum waiting delay the traffic can tolerate, and U is the maximum bandwidth the related user can afford to pay for the traffic transmission with QoS requirements. Each traffic stream has an individual parameter pair associated with the transmission requirements. The initialization of EPONs with regard to the CAC mechanism is processed as follows.

1. Each ONU classifies all its traffic streams into up to eight classes and reports the QoS requirements of each class $\langle B, D, U \rangle$ to the OLT.
2. For each ONU, the OLT selects the strictest bandwidth requirement (the largest value of B) from its traffic classes to set as the ONU's bandwidth requirement B .
3. The OLT pre-calculates the worst case average delay $E(D)$ for each ONU based on all ONUs bandwidth requirement B .
4. The OLT compares $E(D)$ with the delay requirement D for all traffic classes in each ONU. The following processes are repeated for all ONUs.
 - (1) If the delay requirement D is violated for one traffic class, the OLT recalculates the bandwidth B^* for the ONU that is required to guarantee D , and/or
 - (2) The OLT checks the network resources,
 - ① if the network cannot provide bandwidth B^* , the related traffic class cannot be supported in the system and will be provided with the best-effort service.
 - ② if bandwidth B^* is available in the network, the OLT will check whether B^* is affordable by the related traffic,
 - if all classes of traffic agree to the payment, the ONU will be grouped as an BG ONU and be allocated with B^* .
 - if there are some classes of traffic disagreeing with the payment, the disagreeable classes cannot be supported in the system and will be provided with the best-effort service.

5. After the OLT makes the admission control decisions for all ONUs, it informs each ONU about which classes of traffic can be supported with related QoS service and which classes will be provided with the best-effort service in the network.
6. Acquiring the total bandwidth allocation from the OLT, the ONU further makes the admission control decision for its traffic streams, similar to the OLT's operation.

The calculation of the worst-case delay is based on the system load of $\rho=1$ using formulas (4-26), (4-27) and (4-28) developed in chapter 4. In each ONU, the payment of end users should be proportional to their QoS requirements. The operations of the CAC mechanism regarding to the initialization period of EPONs are shown in Figure 5.2 .

During the operation of the system, there may be a new traffic stream joining the existing ONU. The CAC mechanism determines the service provided to the new stream by the joint operations of the ONU and OLT. Firstly, the new traffic stream is grouped by the ONU into related traffic class according to its QoS requirements. The OLT decides the service provided to this ONU and all other ONUs based on the utilization of network resources. Then the ONU further decides what kind of service will be provided to this new traffic stream. If this stream can be supported with the required QoS service without violating the services acquiring by the existing traffic streams, then the new traffic stream can be admitted as the QoS traffic. Otherwise it will be provided with the best-effort service. Another case is that some ONUs may be inactive at the initialization and request for network transmission service during the operation period. CAC will decide the group of the new ONU based on its traffic requirements and the bandwidth utilization. Both procedures are similar to that in the initialization period by the CAC mechanism.

Chapter 5. Proposed Solutions to Support DiffServ

Figure 5.2 Flowchart of the CAC Operations with Consideration of Individual Traffic

5.1.1.2 Implementation of the Inter-ONU Scheduling

The modified BGP-based MAC protocol arbitrates the upstream transmissions for the inter-ONU scheduling in EPONs, to permit the multiple ONUs delivering data packets over the shared channel without collisions. During the initialization of the EPON network, the parameter-based CAC mechanism processes the customers' requirements expressed in SLA parameters, decides which group each ONU belongs to (BG or non-BG group), and determines how many entries each BG ONU is allocated. Based on the results of CAC, the EDA algorithm generates the Entry Table and the non-BG ONU List to set the polling sequence for transmissions. Then during the subsequent operation of the EPON network, the BGP scheme controls the upstream bandwidth allocation and data transmissions from multiple ONUs to the OLT, according to the adaptive order decided by the Entry Table and non-BG ONU List.

If there are new traffic streams joining the existing ONUs or inactive ONUs requesting for network transmission service during the operation period, the operating BGP scheme will be terminated. In this case, the CAC mechanism is re-invoked to regroup the ONUs and reassign entries to each BG ONU. Then the EDA algorithm produces the new Entry Table and non-BG ONU List, based on which the BGP scheme will arbitrate the upstream transmissions from different ONUs.

5.1.2 Intra-ONU Scheduling

In the hierarchical scheduling architecture, the intra-ONU scheduling takes charge of packet transmissions within an ONU. Each ONU classifies traffic streams into classes according to their QoS requirements. It stores and transmits packets from different traffic streams in accordance with the specific scheme supported by the IEEE 802.1D standard, to provide DiffServ to different types of traffic.

By means of setting priority information in the packet header, the IEEE 802.1D defines up to 7 classes of traffic: network control, voice, video, controlled load, excellent effort, best effort, and background traffic. If less than 7 queues are supported, some of the traffic classes can be grouped together [74]. Figure 5.3 shows the grouping of the traffic classes recommended by the standard.

Number of queues	Traffic types queue assignments						
1	Network Control	Voice	Video	Controlled Load	Excellent Effort	Best Effort	Background
2	Network Control	Voice	Video	Controlled Load	Excellent Effort	Best Effort	Background
3	Network Control	Voice	Video	Controlled Load	Excellent Effort	Best Effort	Background
4	Network Control	Voice	Video	Controlled Load	Excellent Effort	Best Effort	Background
5	Network Control	Voice	Video	Controlled Load	Excellent Effort	Best Effort	Background
6	Network Control	Voice	Video	Controlled Load	Excellent Effort	Best Effort	Background
7	Network Control	Voice	Video	Controlled Load	Excellent Effort	Best Effort	Background

Figure 5.3 Mapping of the Traffic Classes into Priority Queues

There is an easy mapping between the IEEE 802.1D traffic classes and the DiffServ's classes of Expedited Forwarding (EF), Assured Forwarding (AF), and Best Effort (BE). According to the IETF DiffServ Working Group (WG) [75], Internet traffic is classified and forwarded in the Per-Hop Behavior (PHB) according to their service requirements. The EF PHB is designed to provide the "low loss, low delay, low jitter, assured bandwidth, end-to-end" service. Generally, the EF PHB simulates a virtual leased line to support highly reliable voice or video applications and to emulate dedicated circuit services. The AF PHB is a group of PHBs designed to ensure that packets are forwarded with a high probability of delivery, as long as the aggregate traffic in a forwarding class does not exceed the subscribed information rate. The AF PHB group includes four traffic classes. The packets within each AF class can be further marked with one of three possible drop-precedence values. The

classification allows service providers to offer differentiated levels of forwarding assurances for IP packets. The default BE PHB is specified as the conventional best-effort forwarding behavior. When no other agreements are in place, all packets are assumed to belong to this traffic aggregate.

In the remaining of this chapter, three priority-based schemes, SPQ, TSQ, and UFQ, are separately deployed for the intra-ONU scheduling. Supported by the IEEE 802.1D standard, each of these three schemes will be combined with the BGP-based MAC protocol adopted in the inter-ONU scheduling, in order to constitute the efficient solutions to provide DiffServ to different types of traffic with different levels of fairness.

5.2 Solution with the SPQ Scheme

The Strict Priority Queuing (SPQ) Scheme is specified by the IEEE 802.1D standard as the default PHB of bridges corresponding to the intra-ONU scheduling in the hierarchical scheduling architecture. In [41], SPQ is combined with the IPACT scheme to support DiffServ in EPONs. While it causes the light-load penalty where delays of lower-priority traffic increase with the decrease of the load. In this section, we present our solution to deploy the SPQ scheme for the intra-ONU scheduling and the BGP-based MAC protocol for the inter-ONU scheduling in a hierarchical scheduling architecture. The proposed solution can not only provide DiffServ to different types of traffic in EPONs, but also greatly mitigate the light-load penalty that is an obvious phenomenon in the solution combining IPACT with SPQ [41].

5.2.1 SPQ Algorithm

In the SPQ scheme adopted in the intra-ONU scheduling, each ONU receives data packets from end users, and buffers them until its turn to transmit. Figure 5.4 illustrates the structure

of the ONU in SPQ. Data packets sending from users will go through the classifier in the ONU, where they are classified to n classes of traffic and then are sequenced into the n priority queues ($Q_0, Q_1, \dots, Q_n$) according to the packets characteristics. Data in Q_0 have the highest priority P_0 , and those in Q_n have the lowest priority P_n .

Figure 5.4 The ONU Structure in the SPQ Scheme

When arriving at the ONU, a packet will be marked with the appropriate priority P_i and sorted into related queue Q_i according to the packet's parameters. If a packet arriving with priority P_i finds the finite buffer is full, it can preempt other packets with lower priority P_k ($k > i$) in queue Q_k . In this case, packets with P_n are first preempted until Q_n is empty; then packets with $P_{n-1}, P_{n-2}, \dots, P_{i+1}$ are preempted until the new packet can be accommodated in its corresponding queue Q_i .

When the ONU is polled by the OLT to transmit data, it serves the n queues based on the strict priority scheduling, i.e., the ONU will transmit all the packets in Q_0 first, then send packets in Q_1 , and so on until Q_n . The operation of the ONU pertaining to the SPQ scheduling algorithm is as follows.

1. On receiving a GRANT message that matches its ID, an ONU will get the granted transmission window size G .
2. The ONU sends a REPLY message to the OLT, reporting the size l_i (in bytes) of packets in each priority queue.
3. The ONU transmits packets in the buffer from the highest priority queue Q_0 to the lowest priority queue Q_n :

- (1) If $l_0 > G$, transmits only packets in Q_0 up to G ;
- (2) Otherwise, transmits all packets in Q_0 , and:
 - If $(l_0 + l_1) > G$, continues transmitting packets in Q_1 up to $(G - l_0)$;
 - Otherwise, continues transmitting all packets in Q_1 , and:
 -
 - If $(l_0 + l_1 + \dots + l_n) > G$, continues sending packets in Q_n up to $(G - l_0 - l_1 - \dots - l_{n-1})$;
 - Otherwise, continues sending all packets in Q_n .

Figure 5.5 shows the flowchart of the ONU's operation of the SPQ algorithm when it is polled by the OLT to transmit data packets.

Figure 5.5 Flowchart of the ONU Operations in the SPQ Scheme

5.2.2 Benefits to the Light Traffic Load

When the EPON network deploys the SPQ scheme for the intra-ONU scheduling and IPACT for the inter-ONU scheduling, it leads to the unexpected network performance of light-load penalty, where the delays for some types of traffic increase with the decrease of the network load [41]. Such phenomenon occurs only at the light traffic load. Whereas, the light-load penalty can be greatly mitigated in our solution that adopts SPQ as the intra-ONU scheduling scheme and the BGP-based MAC protocol as the inter-ONU scheduling scheme. That is the considerable virtue of the BGP scheduling scheme.

The cause of the light-load penalty is the bandwidth allocation mechanism. In the IPACT scheme, the OLT grants the transmission window to each ONU based on its request. Assume ONU k sends to the OLT a REPORT message informing the entire buffer occupancy L_k . The OLT then grants to ONU k for the next polling cycle with the granted window size:

$$G_k = \begin{cases} L_k, & \text{if } L_k \leq W_{\max} \\ W_{\max}, & \text{if } L_k > W_{\max} \end{cases} \quad (5-1)$$

At the light load, the grant size G_k equals to L_k because L_k is always less than $W_{\max}$. However, during the time lag between $t1$ (when ONU k sends REPORT) and $t2$ (when ONU k receives GATE and the assigned window), more packets arrive at ONU k and the buffer length reaches L_k' ($L_k' > L_k$) (Figure 5.6). According to SPQ, newly arrived packets with higher-priority will be transmitted before the old lower-priority packets already stored in the buffer. Since these new packets have not been reported in previous REPORT, the granted window cannot carry all packets. Thus some lower-priority packets with length $(L_k' - L_k)$ have to be left in the buffer. This situation may repeat many times causing these packets to be delayed for multiple cycle times. Such light-load penalty only happens to the lower-priority traffic. As the load increases, more lower-priority packets accumulate in the buffer than higher-priority packets. Hence, besides carrying the old and newly arrived higher-priority

packets, the granted window can take more lower-priority packets than at the light load. Although some lower-priority packets are still postponed for cycles, the delayed time becomes much shorter than at the light load. So the light-load penalty decreases with the increase of the load.

Figure 5.6 Illustration of the Light-Load Penalty

While the light-load penalty is reduced greatly in the BGP scheduling scheme because it deploys a bandwidth allocation mechanism different from IPACT. In the BGP scheme, the OLT decides the transmission window size granted to the next ONU according to the REPLY message from the previous ONU and the maximum transmission window size W_{max} . Assume the previous ONU ($k-1$) sends a REPLY message to the OLT, reporting the granted transmission window size G_{k-1} obtaining from the OLT and its entire buffer occupancy L_{k-1} . By analyzing whether the previous ONU ($k-1$) can consume up the whole assigned window or not, the OLT grants the next ONU k through a GRANT message with the granted transmission window size:

$$G_k = \begin{cases} W_{max} & \text{if } G_{k-1} < W_{max} \\ W_{max} - L_{k-1} & \text{if } G_{k-1} = W_{max} \text{ and } 0 < L_{k-1} \leq T \\ W_{max} & \text{otherwise} \end{cases} \quad (5-2)$$

where T is the predefined threshold that determines whether the superfluous transmission window can be used to poll a non-BG ONU or not. From the formula, next ONU k is polled

with the superfluous transmission window ($W_{max}-L_{k-1}$) when the previous ONU ($k-1$) is allocated the whole transmission window with W_{max} while can only consumes part of it. The remained window is large enough to poll a non-BG ONU. While in other cases, ONU k is allocated the whole transmission window with size W_{max} .

In BGP, since the transmission window size granted to each ONU is not based on its own request, G_k equals to either W_{max} or ($W_{max}-L_{k-1}$) that is always much larger than the ONU's residual buffer occupancy L_k when the load is light. Although more packets arrive at the ONU between the time $t1$ and $t2$ making the buffer length reach to L_k' (Figure 5.6), and the new higher-priority packets will be transmitted before the old lower-priority packets according to SPQ, the granted transmission window is still large enough to accommodate all the packets stored in the buffer. So the lower-priority packets will not be delayed for transmission at the light load as happen in the IPACT scheme. The light-load penalty is almost overcome in the BGP scheduling scheme.

Consequently, the hierarchical scheduling solutions that adopt the BGP-based MAC protocol for the inter-ONU scheduling will avoid the light-load penalty phenomenon, no matter which scheme is used for the intra-ONU scheduling.

5.2.3 Simulation Experiments

Extensive simulation experiments have been conducted to evaluate the performance of the hierarchical scheduling solution containing the inter-ONU and intra-ONU scheduling schemes. The OLT implements the BGP-based MAC protocol to poll ONUs and let them transmit data over the shared upstream channel. Being polled by the OLT, each ONU will further schedule packets of different traffic types based on the SPQ scheme in the intra-ONU scheduling. OPNET Modeler is used for experimental simulations, which is running on Microsoft Windows 2000 platform. For each data point in the experiments, we have

collected results from multiple simulation trials and average to get the mean value. It is verified that the mean value of results falls within the 95% confidence interval.

5.2.3.1 Experiment Design

We define $n=3$ classes of traffics in the experiments, which have priority P_0 , P_1 and P_2 corresponding to the DiffServ's Expedited Forwarding (EF), Assured Forwarding (AF), and Best Effort (BE) PHB [75]. Modeled as high reliable EF PHB, packets of traffic P_0 have the fixed size of 70 bytes, and the constant inter-arrival time of $125 \mu s$. So traffic P_0 has the constant data rate at 4.48Mb/s. Traffic P_1 and P_2 with lower priorities are modeled to be AF and BE PHBs with variable rates. They comply with the exponentially distributed inter-arrival time with sizes ranging from 64 bytes to 1518 bytes. End users generate data packets of traffic P_0 , P_1 and P_2 independently to each BG ONU. While only traffic P_1 and P_2 are produced in non-BG ONUs because such ONUs are provided with best-effort services according to the BGP scheme, which is not suitable to support the high reliable applications like traffic P_0 . Besides the constant load of traffic P_0 in BG ONUs, each of traffic P_1 and P_2 is supposed to occupy 50% of the remaining network load.

The simulation system is designed according to the real EPON environment. The tree-based topology is deployed where the system consists of one OLT and M ONUs that are connected through a passive splitter. In this study, we consider a system with ONU number $M = 16$ and upstream bandwidth $\mu = 1\text{Gb/s}$. The upstream channel capacity is divided into 20 bandwidth units, each with 50Mb/s (thus there are $N = 20$ entries in the Entry Table). Each ONU has a propagation delay from the ONU to the OLT. It is assigned uniformly over the interval $[50 \mu s, 100 \mu s]$ corresponding to the distances between the OLT and ONUs. The maximum transmission window size W_{max} is set to 15000 bytes. The threshold T is

$2/3 W_{max}$. Each ONU has a finite memory buffer (1 Mbytes) in the model, which is shared by the three priority queues.

In the simulation scenario, 16 ONUs are connected to the OLT with 6 BG ONUs and 10 non-BG ONUs. The entries are fixed allocated to the BG ONUs as: 5 entries to ONU1; 3 entries to ONU2~ONU4; 2 entry to ONU5; 1 entry to ONU6. Since the BG ONUs occupy 17 entries in total, the remaining 3 entries are dynamically assigned to the non-BG ONUs. The Entry Table is shown in Figure 5.7. The List for non-BG ONUs is sequenced according to the non-BG ONUs' ID in the ascending order, starting from ONU7 and ending at ONU16.

1	1	2	2	3	3	4	4	5	1
6	5	7	6	8	2	9	1	10	3
11	4	12		13	1	14	5	15	2
16	3	17	1	18	4	19		20	

Figure 5.7 Entry Table with 20 Entries for the Simulation Experiments

5.2.3.2 Experiment Results

We collect the average and maximum waiting delays for the three types of traffic at each kind of BG ONUs. As we have analyzed, there is no light-load penalty for both average delays and maximum delays in our solution.

The average delays and maximum delays of BG ONUs with 5 and 3 entries are shown in Figure 5.8 and Figure 5.9. For BG ONUs with 5 entries, the delay variances between three types of traffic are very small. All delays are in low values and almost unchanged. For BG ONUs with 3 entries, delay of traffic P_2 increases with the growth of the traffic load, especially after the load 0.8. Delays of traffic P_0 and P_1 are almost unchanged.

Figure 5.8 Average Delay of ONUs with 5 and 3 Entries in the SPQ Solution

Figure 5.9 Maximum Delay of ONUs with 5 and 3 Entries in the SPQ Solution

Figure 5.10 and Figure 5.11 show the average delays and maximum delays of BG ONUs with 2 and 1 entry. Delays of traffic P_0 and P_1 for BG ONUs with 2 entries are almost unchanged, so are the delays of traffic P_0 for BG ONUs with only one entry. For traffic P_2 of ONUs with 1 entry, there is no sense to do statistics on the delays after the network load 0.4, since the data are almost 100% preempted by the higher-priority traffic.

Figure 5.10 Average Delay of ONUs with 2 and 1 Entry in the SPQ Solution

Figure 5.11 Maximum Delay of ONUs with 2 and 1 Entry in the SPQ Solution

The figures show that, with the effects of the BGP scheme, ONUs with more entries/ bandwidth units always have better performances than those with fewer entries/ bandwidth units. As the results of the SPQ scheme, the higher-priority traffic will get smaller delays than the lower-priority traffic in the same ONU. The proposed solution combining the BGP-

based MAC protocol and the SPQ scheme can provide DiffServ to different types of traffic and greatly reduce the light-load penalty effectively.

5.3 Solution with the TSQ Scheme

Although being defined as the default mechanism by the IEEE 802.1D, the SPQ scheme polarizes the performances between different traffic streams in the same ONU. In order to overcome the traffic's performance polarization introduced by the solution with SPQ, we have generated a new solution that deploys another mechanism supported by the IEEE 802.1D, the Two-Stage Queuing (TSQ) scheme, for the intra-ONU scheduling and the same BGP-based MAC protocol for the inter-ONU scheduling. This solution with the TSQ scheme can support the differentiated services and achieve the fairness among different types of traffic at the same time.

Fairness is one of the desirable objectives pursuing by a scheduling scheme. In a system with ill-behaved users, the scheduling scheme should provide protection to other users sharing the same resource, so that the ill-behaved users can only have a limited negative impact on well-behaved users. Fair distribution of bandwidth ensures that the ill-behaved users cannot get more than their fair share, automatically guaranteeing the well-behaved users performances. According to [76] and [77], a fair scheduling scheme allocates the bandwidth to each user in proportional to its reserved weight. The excess bandwidth should also be distributed in a fair manner among the users whose demands are not satisfied. The widely accepted fair policy is the max-min fairness that has three conditions: first, no user receives more bandwidth than its request; second, no other scheme satisfying the first condition has a higher minimum bandwidth allocation; finally, the first two conditions remain recursively true when the minimal user is removed from the system and the total

resource is reduced accordingly. The max-min fairness can maximize the minimum bandwidth sharing among those users whose requests cannot be fully satisfied [76].

All in all, a fair scheduling scheme should at least ensure that all backlogged users get bandwidth allocation at all circumstances. While SPQ cannot ensure that all traffic streams receive transmission services at all traffic loads. In SPQ, packets are transmitted strictly according to their priorities, and the lower-priority traffic will only be served after higher-priority traffic being transmitted. Besides, a newly arriving higher-priority packet can preempt lower-priority packets when the ONU's buffer is not enough to accommodate it. All measures ensure the good performances for the higher-priority traffic, while the lower-priority traffic will be sacrificed with very poor performances. Especially at the heavy load, the lower-priority traffic can hardly get any service because almost all the lower-priority packets are preempted and dropped by the higher-priority packets. Such a mechanism will result in the performance polarization between different traffic streams in the same ONU, where the higher-priority traffic gets better-than-required services while the lower-priority traffic starves with the high traffic load. The SPQ scheme is not capable to provide the fair scheduling for different traffic streams at the same ONU in EPONs.

5.3.1 TSQ Algorithm

To support DiffServ as well as the fairness among different types of traffic, we suggest employing the TSQ scheme for the intra-ONU scheduling within the ONU. In the structure of each ONU in TSQ (Figure 5.12), two stages of queues are included in the buffer: the n priority queues in stage 1 ($Q_0, Q_1, \dots, Q_n$) and a First-In-First-Out (FIFO) queue in stage 2. The classifier processes packets received from the connected users first, where they will be classified to n traffic classes and then are buffered into the n priority queues in the first stage. Packets in Q_0 have the highest priority P_0 and packets in Q_n have the lowest priority P_n .

Figure 5.12 The ONU Structure in the TSQ Scheme

When arriving at an ONU, a packet is classified to priority P_i and is sequenced in the related priority queue Q_i of the first stage according to its QoS requirements. Higher-priority packets cannot preempt lower-priority packets. If the buffer is not enough to accommodate it, the newly arriving packet will be dropped regardless of its priority.

Figure 5.13 Flowchart of the ONU Operations in the TSQ Scheme

When the ONU is polled by the OLT, data packets in the priority queues will first be advanced to the second stage of FIFO queue strictly based on their priorities. Then packets are transmitted from the FIFO queue directly to the transmission link. The FIFO queue is a dynamic queue with the variable size in different polling cycles, which equals to the granted transmission window size G obtained by the ONU. Since packets are popped up to the FIFO queue with the strict order of $P_0, P_1, \dots, P_n$, packets in each transmission window are also with the strict priority sequence. The operation of the ONU that consists of data transmission and two-stage queuing scheduling is shown in Figure 5.13, where l_i indicates the length of the priority queue Q_i .

5.3.2 Benefits to the Lower-Priority Traffic

The TSQ scheme can provide the better performance to the lower-priority traffic and achieve the better fairness among different types of traffic than the SPQ scheme. When the ONU is polled for transmission in TSQ, it pops up packets from the priority queues to the FIFO queue before delivering them to the upstream link. This transmission method through the second stage of FIFO queue can prevent the undesirable behaviors of some higher-priority packets that preempt lower-priority packets for transmission. Such behaviors result from the principle of the priority-based scheduling mechanism. When the ONU is polled by the OLT and transmitting packets, new packets keep on arriving including some higher-priority packets. If there is only one stage of priority queues in the ONU, these new higher-priority packets will be transmitted earlier than the existing lower-priority packets according to the priority scheduling. While in the TSQ scheme, packets are advanced to the FIFO queue before being transmitted over the channel, and newly arriving packets cannot replace the old packets already sequenced in the FIFO queue. Hence, the new higher-priority packets will not be served earlier than the existing lower-priority packets. The lower-priority traffic can

be provided with better services than in SPQ. The transmission from the FIFO queue can benefit the lower-priority traffic and help to reduce the performance polarization between different traffic streams in the same ONU.

Additionally, newly arriving higher-priority packets in TSQ cannot preempt existing lower-priority packets in the first stage of priority queues. Thus lower-priority packets will experience less data drop than in the SPQ scheme. With the protection of the second stage of FIFO queue and without the preemption by higher-priority packets, the lower-priority traffic can always get services no matter how heavy the load is. The solution with the TSQ scheme can provide better services to the lower-priority traffic to achieve the better fairness than the solution with SPQ.

5.3.3 Simulation Experiments

Using OPNET Modeler on the Microsoft Windows 2000 platform, we have conducted wide simulation experiments to evaluate the performance of the new proposed solution for a tree-based EPON network. The solution deploys the TSQ scheme for the intra-ONU scheduling, which is combined with the BGP-based MAC protocol in the inter-ONU scheduling, to support DiffServ and provide the fairness among different types of traffic. For each data point in the experiments, we have collected results from multiple simulation trials and average to get the mean value. It is verified that the mean value of results falls within the 95% confidence interval.

In order to compare the performances between the solution with TSQ and solution with SPQ, we have conducted another group of simulation experiments that deploy the TSQ and SPQ schemes for the intra-ONU scheduling respectively, with the same BGP-based MAC protocol for the inter-ONU scheduling. The traffic types, simulation parameters and scenario designs are same in both simulation experiments, similar to the experiment design in section

5.2.3.1. Note that in TSQ, the second stage of FIFO queue in the ONU is a temporary cache whose size is not larger than the maximum transmission window size W_{max} , which is not part of the buffer. The Entry Table used in the simulation is shown in Figure 5.7 .

5.3.3.1 Performance of the Solution with TSQ

In the first group of simulations, we collect the average message delays for different types of traffic at ONUs for the proposed solution with TSQ.

Figure 5.14 and Figure 5.15 show different types of traffic's average message delays and maximum delays for BG ONUs with 5 and 3 entries respectively. For BG ONUs with 5 entries, the delays of all traffic are very low with the slight changes when the network load increases. For ONUs with 3 entries, the delays of traffic P_2 grow faster than traffic P_0 and P_1 , especially after the system load 0.9.

Figure 5.14 Average Delay of ONUs with 5 and 3 Entries in the TSQ Solution

Figure 5.15 Maximum Delay of ONUs with 5 and 3 Entries in the TSQ Solution

Average message delays and maximum delays of different traffic streams for BG ONUs with 2 and 1 entries are shown in Figure 5.16 and Figure 5.17, respectively. For both kinds of BG ONUs, traffic P_2 has much larger delay values than those of traffic P_0 and P_1 . For ONUs with 2 entries, the delays for all types of traffic experience rapid growths at the load 0.5. The growth point advances to 0.2 for ONUs with only one entry.

Figure 5.16 Average Delay of ONUs with 2 and 1 Entry in the TSQ Solution

Figure 5.17 Maximum Delay of ONUs with 2 and 1 Entry in the TSQ Solution

Due to the limited buffer size at each ONU, we have also considered the loss rates for different traffic streams at each kind of ONUs. Since BG ONUs with 5 entries are allocated enough bandwidth for data transmissions, their data packets with any priority will never be dropped and the loss rate is always 0 despite of the network load.

Figure 5.18 Loss Rate of BG ONUs in the TSQ Solution

Figure 5.18 shows the relationship between the loss rates of different traffic streams and the network load for BG ONUs with 3, 2 and 1 entries, respectively. For the ONUs with 3 entries, all traffic streams experience no data loss before the load being 0.9. The traffic at BG ONUs with 2 entries starts to drop packets after load of 0.5 with the loss rates for traffic P_l and P_2 being almost same and higher than that of traffic P_0 . For BG ONUs with only one entry, all types of traffic begin to drop data when traffic load is 0.2.

We can get from the results that, with the effect of the BGP scheme in the inter-ONU scheduling, ONUs with more entries obtain better performances (smaller delays and smaller loss rates) than ONUs with fewer entries. Besides, for the same kind of ONUs, different types of traffic are provided with differentiated services as the result of the TSQ scheme in the intra-ONU scheduling, with the higher-priority traffic having better performance than the lower-priority traffic.

5.3.3.2 Performance Comparison between Solutions with TSQ and SPQ

By the TSQ scheme, although lower-priority traffic streams get worse performances than higher-priority traffic streams, they can always get transmission services without starvation. In contrast, the SPQ scheme can hardly provide any service to the lower-priority traffic at the heavy load, especially in the ONUs with few bandwidth units/entries. The solution with TSQ can achieve better fairness than the solution with SPQ by providing better services to lower-priority traffic streams. In the second group of simulation experiments, we collect and compare the average message delays and loss rates for each kind of ONUs between the two solutions with the TSQ and SPQ schemes respectively.

Figure 5.19 and Figure 5.20 show and compare different traffic streams average message delays and loss rates for BG ONUs with 3 entries in the solutions with TSQ and SPQ, respectively. Performances of traffic P_0 and P_l with TSQ are a little worse than those

of the corresponding traffic streams with SPQ, while traffic P_2 with TSQ has better performance than traffic P_2 with SPQ.

Figure 5.19 Comparison of Average Delay for 3-Entry ONUs in TSQ and SPQ Solutions

Figure 5.20 Comparison of Loss Rates for 3-Entry ONUs in TSQ and SPQ Solutions

Average message delays and loss rates of BG ONUs with 2 entries in the solutions with TSQ and SPQ are compared in Figure 5.21 and Figure 5.22 respectively. Similarly, traffic

P_0 and P_1 with TSQ have worse performances than the corresponding traffic with SPQ while performances of traffic P_2 with TSQ is much better than those of traffic P_2 with SPQ.

Figure 5.21 Comparison of Average Delay for 2-Entry ONUs in TSQ and SPQ Solutions

Figure 5.22 Comparison of Loss Rates for 2-Entry ONUs in TSQ and SPQ Solutions

Figure 5.23 and Figure 5.24 compare different types of traffic's performances for BG ONUs with only one entry in the solutions with TSQ and SPQ respectively. Except that traffic P_0 with TSQ gets worse performance than traffic P_0 with SPQ, traffic P_1 and P_2 with

TSQ have much better delay performance than the corresponding traffic with SPQ. For traffic P_2 with SPQ, data packets are almost 100% preempted by the higher-priority traffic after the network load 0.4, so there is no sense to do statistics on the delays after that load. The lower-priority traffic can hardly get any transmission services in the solution with the SPQ scheme, and even starve from the medium load.

Figure 5.23 Comparison of Average Delay for 1-Entry ONUs in TSQ and SPQ Solutions

Figure 5.24 Comparison of Loss Rates for 1-Entry ONUs in TSQ and SPQ Solutions

We can conclude from the comparison that, although the higher-priority traffic streams by the SPQ have quite good performances, the lower-priority traffic streams perform much worse than those by the TSQ, exacerbating the unfairness among different types of traffic in the same ONU. Some lower-priority traffic streams even starve without any transmission service at the heavy load. Contrarily, lower-priority traffic streams can always be served for transmission without starvation in the solution with TSQ. The proposed solution with the TSQ scheme can reduce the performance polarization to achieve the fairness among different traffic streams in EPONs.

5.4 Solution with the UFQ Scheme

In a network with both QoS and best-effort traffic, fair scheduling is important for different types of traffic to share the bandwidth resource effectively. However, it is not always desirable that two types of traffic share the bandwidth with the absolute fairness. The bandwidth that is not strictly needed to enforce the guarantees for the QoS flows can be allocated to the best-effort flows to improve their performances [78]. Such a bandwidth allocation scheme will maximize the bandwidth instantly given to the best-effort traffic while still ensure the service requirements of different types of QoS traffic, which can achieve the fairness between the QoS traffic and best-effort traffic effectively.

Although the TSQ scheme can achieve some kinds of fairness and improve the performance of the lower-priority traffic than the SPQ scheme, it has not taken into account the specific requirements of various traffic streams. So the solution with the TSQ scheme can hardly obtain the objective of providing QoS service with fair scheduling. In order to achieve the fairness between the QoS traffic and best-effort traffic in EPONs, we propose a novel Urgency Fair Queuing (UFQ) scheme for the intra-ONU scheduling under the IEEE 802.1D standard. Incorporating the individual traffic's requirements into the scheduling

design, UFQ can provide the most superfluous bandwidth to the best-effort traffic and guarantee the requirements of different types of QoS traffic at the same time. The solution containing the newly devised UFQ scheme in the intra-ONU scheduling and the BGP-based MAC protocol in the inter-ONU scheduling will support the differentiated services and fair scheduling between the QoS traffic and the best-effort traffic in EPONs.

5.4.1 UFQ Algorithm

The UFQ scheme considers the service requirements, especially the delay bound requirement for scheduling. In the EPON network, we assume to have several kinds of QoS traffic with different delay bound requirements and a best-effort traffic with no delay bound requirement. By taking into account the traffic's QoS requirements, the UFQ scheme newly proposed for the intra-ONU scheduling can provide better service to the best-effort traffic than other schemes without violating the service requirements of the QoS traffic streams. In UFQ, the ONU's buffer is divided into multiple queues to accommodate different types of traffic, including multiple QoS queues ($Q1, Q2, \dots, Qn$) and a best-effort queue BE (Figure 5.25). According to their QoS requirements and characteristics, packets from different end users are classified to different classes by the classifier and are sequenced into different QoS queues before transmission. Packets without any QoS requirement are simply arrayed in the BE queue. There is no packet preemption in the UFQ scheme. If the ONU's buffer is full, the newly arriving packet will be dropped no matter which class it belongs to.

Figure 5.25 The ONU Structure in the UFQ Scheme

5.4.1.1 Principle of the UFQ Algorithm

The key point of the UFQ scheme is to assign as much as possible bandwidth to the best-effort traffic as long as the requirements of different QoS traffic streams can be guaranteed. To achieve this point, delay bound requirements of the QoS packets are considered in UFQ. We introduce the Urgency Parameter $U_k^j(t)$ to indicate the urgency of packet k belonging to the QoS traffic j to be transmitted in the ONU. It takes into account the packet's delay bound, transmission time and lingering time in the ONU. The Urgency Parameter is used as the sole criterion for QoS packet transmissions in the UFQ scheme. Packets with smaller Urgency Parameter will be served earlier than packets with larger one. $U_k^j(t)$ is expressed as:

$$U_k^j(t) = d_k - v_k(t) - s_k / C \quad (5-3)$$

where d_k is the delay bound requirement of packet k , $v_k(t)$ is the packet lingering time in the ONU, s_k is the size of packet k , and C is the link rate. The lingering time $v_k(t)$ is actually the outcome of the system current time $S(t)$ minus the packet arrival time a_k : $v_k(t) = S(t) - a_k$. Since $v_k(t)$ increases as the time elapses, the value of $U_k^j(t)$ is decreasing with the packet waiting in the buffer.

In the UFQ scheme, packets in different QoS traffic streams are transmitted based on their Urgency Parameters, which can be regarded as the transmission priorities. Since the packet's delay bound, transmission time and lingering time have been mapped into the Urgency Parameter, it can represent the urgency/priority of the QoS packets to be transmitted. The smaller of the $U_k^j(t)$ value, the more urgent this packet requires to be sent out. A packet with $U_k^j(t) < 0$ will be dropped because its delay bound is already infringed.

Meanwhile, a packet with large $U_k^j(t)$ can be delayed for transmission, giving way to the best-effort packets to go through the link.

The best-effort packets have no related Urgency Parameters in the UFQ scheme. They are transmitted when all packets in the QoS traffic are deferred for transmission due to the large values of Urgency Parameters. By using some bandwidth that the QoS traffic cannot consume, the best-effort traffic will get better performances than in other mechanisms. Hence, the UFQ scheme can effectively provide the fairness between the QoS traffic and best-effort traffic in the same ONU, by offering better performances to the best-effort traffic while still ensuring the service requirements of different types of QoS traffic.

5.4.1.2 Packet Arrival Procedure

When arriving at the ONU, a new packet k with QoS requirements will be assigned the original value of the Urgency Parameter $U_k(0)$. After being stamped with $U_k(0)$ and the arrival time a_k , packet k is inserted into the relevant QoS queue Qj according to its original Urgency Parameter value $U_k^j(0)$. Since the arrival time a_k equals to the system current time $S(0)$ when the packet newly arrives at the ONU, the packet lingering time $v_k(0)=0$. So the original Urgency Parameter value of packet k in queue Qj is:

$$U_k^j(0) = d_k - s_k / C \quad \forall Qj \in (Q1, Q2, ..., Qn) \quad (5-4)$$

After comparing with the Urgency Parameter values of the existing packets in queue Qj , packet k will be inserted into queue Qj after packets with less $U_m^j(t) \leq U_k^j(0)$ and before packets with larger $U_m^j(t) > U_k^j(0)$. The urgency value of an existing packet m in queue Qj can be calculated as follows, whose $U_m^j(0)$ and a_m have been stamped when it arrived at the ONU:

$$U_m^j(t) = d_m - v_m(t) - s_m / C = U_m^j(0) + a_m - S(t) \quad \forall Qj \in (Q1, Q2, ..., Qn) \quad (5-5)$$

Newly arriving packets without QoS requirement are simply inserted in the tail of the *BE* queue in the buffer. So packets in the *BE* queue are sequenced in a FIFO manner. The operations of ONU receiving a packet in UFQ can be concluded as follows.

The ONU receives packet k

if (the buffer is not full)

if (packet k has no QoS requirements)

*insert packet k at the tail of the *BE* queue;*

else

find the corresponding QoS queue Q_j ;

stamp packet k with the arrival time a_k and the original urgency value: $U_k^j(0)$

$= d_k - s_k / C$;

if (queue Q_j is not empty)

get the head packet $m=0$;

while ($U_m^j(t) < U_k^j(0)$)

$m=m+1$;

insert packet k before packet m ;

else

insert packet k at the tail of QoS queue Q_j ;

else

drop packet k ;

5.4.1.3 Packet Transmission Procedure

When the ONU is polled by the OLT, it will transmit QoS packets in the increasing order of their Urgency Parameter values $U_k^j(t)$, and serve the best-effort packets during the interval

when all QoS packets are delayed for transmission due to the large values of the Urgency Parameter.

Before transmission, the ONU selects the packet with the minimum Urgency Parameter from all QoS packets. Since packets in each QoS queue are sequenced in the increasing order of $U_k(t)$, the ONU only needs to compare $U_k(t)$ of the head packets ($k=0$) in different QoS queues ($Q1, Q2, \dots, Qn$), instead of all QoS packets in the buffer. Assume that the head packet in QoS queue Qj is selected with the minimum $U_0^j(t)$ value. This packet is also the one most urgent for transmission among all QoS packets. The Urgency Parameter value of this head packet is:

$$U_0^j(t) = U_0^j(0) + a_0 - S(t) \quad (5-6)$$

If the $U_0^j(t)$ value of the selected packet is less than 0, then this packet will be dropped because its delay bound will be violated even though it is transmitted immediately. If this $U_0^j(t)$ value is between 0 and the cycle time plus a guard time ($TC_i + G$), then the selected packet will be sent out by the ONU immediately. If $U_0^j(t)$ is larger than ($TC_i + G$), this packet and all other packets in QoS queues will be deferred for transmission. The reason is that Urgency Parameters of all QoS packets are large enough, which means their delay bounds are far from being violated. So all QoS packets are not urgent for transmission and can wait in the buffer for longer time. In this case, the head packet in the best-effort queue will be transmitted.

The guard time G_T is set to ensure the QoS packets be transmitted before their delay bound violation. It is different from the guard time between the adjacent time slots in the inter-ONU scheduling scheme. G_T is a system parameter that is related to the system traffic load and traffic requirements. A large value of G_T can give good protection for the QoS traffic and provide worse performance to the best-effort traffic, and vice versa. The cycle

time TC_i is the time period between the two consecutive times that ONU i is polled. According to the mathematical estimation of the polling-based scheme in chapter 4, for an EPON network with average cycle time of $\bar{t}_c$, the cycle time of the ONU with n bandwidth units/entries can be expressed as follows:

$$TC_i = \bar{t}_c / n = (N * \bar{\tau} + 3P * \bar{\tau} + N * W_{\max}) / n \quad (5-7)$$

where N is the total number of entries in the system, $\bar{\tau}$ is the average propagation delay from the ONU to OLT, $W_{\max}$ is the maximum transmission window, and P is the number of entries that can be used to poll a second ONU. The operations of ONU transmitting packets in UFQ are as follows:

ONU i transmits the packets

while (the granted transmission size is not used up)

calculate the urgency values of all head packet ($k=0$) in different QoS queues: $U_0(t)$

$= U_0(0) + a_0 - S(t);$

for (all QoS queues)

find queue Q_j whose head packet has the minimum $U_0^j(0);$

if ($U_0^j(0) < 0$)

drop the head packet in queue $Q_j;$

if ($0 \leq U_0^j(0) < (TC_i + G_T)$)

transmit the head packet in queue Q_j immediately;

else

transmit the head packet in the best-effort queue;

In the UFQ scheme, packets from different classes of traffic will be interleaved when being transmitted. Thus the QoS traffic can give some extra bandwidth to the best-effort traffic and will still be provided with the necessary QoS service. At the same time, the best-

effort traffic can get the most superfluous bandwidth and obtain better performances than in other schemes. By combining the traffic classification with fair scheduling, UFQ can achieve the fairness between the QoS traffic and best-effort traffic in the same ONU.

5.4.2 Simulation Experiments

We have conducted extensive simulation experiments to model a tree-based EPON network using OPNET Modular, to evaluate the performance of the combined inter-ONU and intra-ONU scheduling solution. The newly proposed UFQ scheme is deployed for the intra-ONU scheduling in the solution. It is combined with the BGP-based inter-ONU scheduling MAC protocol to support DiffServ and fair scheduling between the QoS traffic and best-effort traffic. In order to compare the performances of the solution with UFQ and the solution with TSQ, we have included two schemes in the simulation experiments that deployed the same BGP-based MAC protocol for the inter-ONU scheduling while UFQ and TSQ for the intra-ONU scheduling, respectively. The traffic types, simulation parameters, scenario designs and delay bound requirements of different QoS traffic types are same in both simulation experiments. For each data point in the experiments, we have collected results from multiple simulation trials and average to get the mean value. It is verified that the mean value of results falls within the 95% confidence interval.

5.4.2.1 Experiment Design

In the experimental simulations, we define three types of traffic: QoS traffic $Q1$, $Q2$ and best-effort traffic BE , corresponding to the IETF specification [75]. Since most types of traffic are featured by the self-similarity and Long-Range Dependencies (LRD) in communication networks, we use a Pareto model to generate the self-similar traffic for $Q2$ and BE traffic classes. The traffic is produced by an ON/OFF source model, and the ON time of the traffic follows the Pareto distribution. Packet sizes are uniformly distributed between

64 and 1518 bytes. For QoS traffic $Q2$, the delay bound requirements are not identical for all packets; instead they are exponentially set with mean value of 400ms. For the BE traffic, there is no delay bound requirement. The ONU produces variable rate for these two types of traffic with the increase of the system load. QoS traffic $Q1$ has the highest QoS requirements. It is modeled as the interactive application with high reliability. The traffic's delay bound requirement is strict and identical for all packets (set to 10ms). Each BG ONU generates this type of traffic with fixed size of 70 bytes and constant inter-arrival time of 125 μs . So traffic $Q1$ has the constant data rate of 4.48Mb/s.

Each BG ONU generates the above three types of traffic independently, and transmits different types of traffic using the UFQ scheme for the intra-ONU scheduling. While non-BG ONUs only produce BE traffic. Besides the fixed load of traffic $Q1$ generated by BG ONUs, non-BG ONUs produce BE traffic occupying 60% of the remainder system traffic load. The other 40% load is averagely distributed between traffic $Q2$ and BE in BG ONUs. The experiment system is designed as same as the one in section 5.2.3.1. The Entry Table used for the simulation is shown in Figure 5.7.

5.4.2.1 Experiment Results

For each kind of BG ONUs, we collect the average message delays for different traffic types, with the system load ranging from 0.1 to 1.0. We can see from the figures that the delay bound requirements for all types of QoS traffic are satisfied well, while the best-effort traffic is provided with good performances. The solution with the UFQ scheme can provide better services to the best-effort traffic than the solution with the TSQ scheme, and achieve better fairness since it takes into account the QoS requirements of different users.

For BG ONUs with 5 and 3 entries, because they are allocated with enough bandwidth to transmit both the QoS and best-effort traffic, corresponding traffic's delays are almost same

for solutions with UFQ and TSQ. All types of traffic in the same group of ONUs experience the similar waiting delays, and delays are very small with values below the delay bound requirements of both QoS traffic types.

Figure 5.26 compares different traffic's average delays for BG ONUs with 2 entries in solutions with UFQ and TSQ respectively. In the solution with UFQ, delays for all types of traffic are very similar to each other and almost unchanged all the time, with the values far below the delay bounds of the two QoS traffic streams. However, in the solution with TSQ, although traffic *Q1* can obtain the delay performance similar to that with UFQ, delays of traffic *Q2* and *BE* are much larger than the corresponding traffic with UFQ. With the effect of the UFQ scheme, the best-effort traffic can get much more bandwidth as well as better performance than with TSQ. At the same time, the two types of QoS traffic with UFQ are provided with the guaranteed services that can satisfy their transmission requirements.

Figure 5.26 Comparison of Average Delay for 2-Entry ONUs in UFQ and TSQ Solutions

Figure 5.27 shows the average delays of BG ONUs versus network load with 1 entry in solutions with UFQ and TSQ. In the solution with UFQ, although the delay values for traffic *Q1* and *Q2* are higher than those with TSQ, they have not violated the bounds of the delay

requirements. More important, with the extra bandwidth dedicated by the QoS traffic, the best-effort traffic with UFQ can experience lower delay than that with TSQ. With the UFQ scheme, the proposed solution can allocate as much as possible bandwidth to the best-effort traffic and ensure the requirements of the QoS traffic simultaneously.

Figure 5.27 Comparison of Average Delay for 1-Entry ONUs in UFQ and TSQ Solutions

The simulation results suggest that, the solution with the UFQ scheme can obviously improve the delay performance of the best-effort traffic, without violating the delay bound requirements of different types of QoS traffic.

In order to prove that the better delay performances in the solution with UFQ have not weakened the loss rate performances by dropping more packets, we collect the loss rates for different ONUs' traffic in solutions with UFQ and TSQ respectively for comparison. The packet with QoS requirements will be dropped if its packet delay bound is violated or the finite buffer is full. So the *integrated loss rate* of the QoS traffic also consists of two parts: the loss rate due to the violation of the delay bound (we call it *loss rate 1* hereafter) and the loss rate due to the fullness of the buffer (we call it *loss rate 2* hereafter).

Chapter 5. Proposed Solutions to Support DiffServ

Figure 5.28 Comparison of Loss Rate 1 for 5-Entry ONUs in UFQ and TSQ Solutions

Figure 5.29 Comparison of Loss Rate 1 for 3-Entry ONUs in UFQ and TSQ Solutions

For BG ONUs with 5 and 3 entries, *loss rates 2* are zero with both UFQ and TSQ schemes because these ONUs are allocated enough bandwidth. So *loss rates 1* are actually the *integrated loss rates* of ONUs as shown in Figure 5.28 and Figure 5.29. The corresponding curves of the two solutions are almost overlapped with each other with the same drop rates. For these two kinds of BG ONUs, the drop rates for traffic *Q1* are zero all

the while. At some loads, loss rates for traffic $Q2$ with UFQ are even lower than those with TSQ. The loss rate performances have not been impaired in the solution with UFQ.

Figure 5.30 and Figure 5.31 compare the *loss rates 1* and *integrated loss rates* for BG ONUs with 2 entries in solutions with UFQ and TSQ respectively. Considering the data drop due to the buffer saturation, loss rates in Figure 5.31 are much higher than those in Figure 5.30. From Figure 5.30, loss rates of traffic $Q1$ with both UFQ and TSQ schemes are always zero. Traffic $Q2$ with UFQ experiences very low loss rate at all times, similar to that with the TSQ scheme. In Figure 5.31, all the three types of traffic experience very low drop rate before the load 0.7. Then the curves climb up gradually with traffic $Q1$ slower than traffic $Q2$ and BE . The corresponding loss rates with UFQ are almost same with those with TSQ all the while. It is proved again that, the improvements of delay performance in the solution with the UFQ have not aggravated the loss rates of different traffic types.

Figure 5.30 Comparison of Loss Rate 1 for 2-Entry ONUs in UFQ and TSQ Solutions

Figure 5.31 Comparison of Integrated Loss Rate for 2-Entry ONUs in UFQ and TSQ Solutions

Figure 5.32 and Figure 5.33 are different traffic's *loss rates* and *integrated loss rates* for BG ONUs with only one entry in solutions with UFQ versus TSQ respectively. From Figure 5.32, traffic *Q1* with both the UFQ and TSQ schemes almost experience no data loss due to the delay violation at all times. After the load 0.4, *loss rates* of traffic *Q2* with both UFQ and TSQ are decreasing with the increase of the traffic load. The reason is that, more packets are discarded due to the fullness of the buffer with the increase of the traffic load, leading to fewer packets being dropped due to the violation of the delay bounds. So *loss rates* of the QoS traffic *Q2* in both solutions decrease after the saturation load 0.4. For the *integrated loss rates* shown in Figure 5.33, all the three types of traffic drop very small amount of packets before the load 0.3. Then the curves climb up gradually with traffic *Q1* slower than the other two types of traffic. The corresponding curves in the two solutions are almost overlapped with each other with the same drop rates. We can see from the figures that, the loss rates in the solution with UFQ are not impaired as the price to improve the delay performances.

Chapter 5. Proposed Solutions to Support DiffServ

Figure 5.32 Comparison of Loss Rate 1 for 1-Entry ONUs in UFQ and TSQ Solutions

Figure 5.33 Comparison of Integrated Loss Rate for 1-Entry ONUs in UFQ and TSQ Solutions

The simulation results demonstrate that, in the solution with the UFQ scheme, both types of the QoS traffic in the BG ONUs are satisfied with the bounded delays, and the best-effort traffic is provided with fairly good performance. The QoS requirements are well considered to support DiffServ and fair scheduling in the novel UFQ scheme. Whereas, in the solution

with the TSQ scheme, bandwidth is not allocated as fairly as in the solution with UFQ. The QoS traffic streams are provided with much more bandwidth than they need, while the best-effort traffic can only get poor services. The solution with UFQ can achieve better fairness between the QoS traffic and best-effort traffic than the solution with TSQ, while can guarantee the performances of QoS traffic as well. Moreover, the improvements of delay performance in the solution with UFQ have not weakened the performances of different traffic's loss rates.

5.5 Summary

In this chapter, three intra-ONU scheduling schemes, SPQ, TSQ, and UFQ, are presented to constitute three solutions by combining with the BGP-based MAC protocol in the inter-ONU scheduling. All the proposed solutions can effectively provide DiffServ to different types of traffic and also arbitrate the upstream transmission by means of the hierarchical MAC scheduling in the EPON environment. The simulation results show that, among the three classes of priority traffic at each ONU, traffic with higher priorities can always outperform the traffic with lower priorities. By the effect of the BGP scheduling algorithm in the inter-ONU scheduling, all the presented solutions can greatly mitigate the light-load penalty phenomenon that is common and obvious in other proposals.

Comparing to the solution with SPQ that cannot support the fair scheduling, the solution with TSQ scheme can provide some kinds of fairness to traffic streams with different priorities in EPONs. It diminishes the performance polarization between the higher-priority and lower-priority traffic at the same ONU. Lower-priority traffic in the solution with TSQ drops fewer data packets than in the solution with SPQ and can always get transmission services with fairness. In order to achieve the fairness between QoS and best-effort traffic streams, we have designed the novel UFQ scheme for the intra-ONU scheduling, where

Chapter 5. Proposed Solutions to Support DiffServ

packets in an ONU are scheduled based on the urgency regarding to the delay bound requirements. QoS packets can give way to the best-effort packets for transmission if they are not so urgent for transmission. The solution with the proposed UFQ scheme is effective to improve the performance of the best-effort traffic and also efficient to meet the necessary requirements of different QoS traffic streams. It can achieve better fairness than the solution with TSQ in bandwidth sharing between the QoS traffic and the best-effort traffic.

Chapter 6 . Conclusions

6.1 Summary

EPON is widely regarded as the best candidature for next generation access networks for it merges the virtues of inexpensive Ethernet equipments and high-bandwidth fiber transmission media. In the EPON upstream transmission, multiple ONUs should share the single fiber channel to transmit data packets to the OLT. This thesis mainly focuses on the design and analysis of Dynamic Bandwidth Allocation (DBA) schemes for EPONs to address the multiple access issue in the upstream transmission.

We have proposed an enhancement to the famous IPACT scheme, IPACT-GE, to improve the performance of IPACT in chapter 3. In IPACT, the OLT grants the ONU based on its request size in the previous polling cycle. Since packets keep arriving after the request time, there is the inevitable discrepancy between the grant size and instant buffer occupancy when the ONU is polled for transmission. When the load is light, the average waiting delays experienced by the packets are unnecessarily prolonged due to such discrepancy. To overcome the weakness of IPACT, in the IPACT-GE scheme, the amount of new packets arriving between pollings is estimated by the ONU and reported to the OLT. Then the OLT decides the granted transmission window size for the ONU according to the request size and

the estimated amount of new packets. The estimation-based IPACT-GE scheme can diminish the discrepancy between the grant size and the instant buffer length, resulting in the shorter waiting delay and smaller buffer occupancy than IPACT at the light load. The experimental simulations have demonstrated that, besides the improvement to the IPACT performance, the IPACT-GE scheme can achieve better performance than another enhancement to IPACT, the EB-DBA scheme. Furthermore, when combined with the SPQ mechanism for the DiffServ support in EPONs, IPACT-GE can greatly mitigate the light-load penalty that is an obvious phenomenon in the solution of IPACT combined with SPQ. The IPACT-GE scheme is proved to be an efficient enhancement to the pioneer IPACT scheme.

Based on the comprehensive review of existing bandwidth allocation schemes, a novel DBA scheme, BGP, has been proposed in chapter 4 for EPON upstream transmissions within the IEEE 802.3ah MPCP framework. The roll-call polling-based BGP scheme approximates to a TDM system with dynamic scheduling ability. It incorporates SLA contracts into the DBA scheme design, in order to provide absolute QoS assurances to different ONUs with diverse bandwidth requirements. By dividing ONUs into different groups according to the SLA, the BGP scheme can guarantee bandwidth for high-demand users, while provide low-demand users with the best-effort service. Since the ONUs with high bandwidth requirements will be polled more than one time in a cycle, we also develop an EDA algorithm to distribute the multiple bandwidth units to such ONUs, thus to facilitate the implementation of BGP. By evenly allocating bandwidth units to BG ONUs, EDA can keep the burst from the high-load ONUs at a low level.

We have further developed a parameter-based CAC mechanism based on the BGP scheme, to admit the transmission requests from ONUs. Combining the proposed CAC mechanism, EDA algorithm, and BGP scheme, a novel systematic MAC protocol has been

presented for EPONs. In this BGP-based MAC protocol, the CAC mechanism is responsible for processing the users requirements expressed in SLA contracts, to decide the group of each ONU and the bandwidth units allocated. Then the EDA algorithm prepares the Entry Table and non-BG ONU List to determine the polling sequence for ONUs. Finally the BGP scheme, which is the kernel of the MAC protocol, performs the scheduling function to allow ONUs transmit data packets over the shared upstream channel without collisions.

To numerically analyze the proposed BGP scheme, we have presented a mathematical estimation model and a CTMC model separately. Some approximations have been taken in the estimation model to get the average bandwidth units for the non-BG ONU. The EPON system with BGP is emulated to a multi-queue polling system when applying the CTMC model. The outcomes of the mathematical estimation model are matching with the simulation results of BGP quite well, which testifies the rationality of the BGP scheme. Other experimental simulations have proved the efficiency of the proposed BGP scheme and the BGP-based MAC protocol by comparing with the IPACT scheme. BG ONUs with multiple bandwidth units in BGP are demonstrated to achieve the guaranteed bandwidth with better performance than ONUs in IPACT. However, the non-BG ONUs are only provided with the best-effort service, which is the main effect of the proposed BGP scheme. The newly designed BGP scheme is verified to be a powerful DBA scheme for the upstream transmission in EPON networks.

In chapter 5, we have proposed several solutions based on the BGP scheduling scheme to support DiffServ in EPONs under the framework of MPCP and IEEE 802.1D. In order to support the bandwidth allocation in both the OLT and ONU levels, a hierarchical scheduling architecture has been introduced for EPONs including the inter-ONU and intra-ONU scheduling respectively. The BGP-based MAC protocol has been modified and adopted for the inter-ONU scheduling, to effectively allocate the upstream channel capacity among

multiple ONUs. The SPQ scheme, which is specified as the default mechanism by IEEE 802.1D, has first been applied as the intra-ONU scheduling scheme to conduct different classes of traffic transmissions within the ONU. Shown by the simulation results, the solution with the SPQ scheme can provide DiffServ to different traffic flows effectively, with the better performance of higher-priority traffic than that of the lower-priority traffic. Moreover, due to the efficiency of the BGP scheduling scheme, the solution with SPQ can greatly reduce the light-load penalty that is apparent in the combined IPACT and SPQ solution.

Then we have proposed to combine the TSQ scheme with the BGP-based MAC protocol to provide the Diffserv support in EPONs. The second stage of FIFO queue in the TSQ scheme allows the lower-priority traffic always getting service even under the heavy load, thus the performance polarization is diminished between the higher-priority and lower-priority traffic in the same ONU. The simulation results have shown that, the solution with the TSQ scheme can provide some kinds of fairness to different types of traffic, where the lower-priority traffic will achieve better performance than in the solution with SPQ.

At last, the solution combining the BGP-based MAC protocol with the new proposed UFQ scheme has been presented to support DiffServ in EPONs and provide the fair scheduling at the same time. The UFQ scheme is newly devised for the intra-ONU scheduling in order to fairly distribute bandwidth between the QoS traffic and best-effort traffic in the same ONU. Packets in an ONU are scheduled based on the urgency regarding to the delay bound requirements. QoS packets will give way to the best-effort packets for transmission if they are not so urgent for transmission. Since the requirements of individual traffic are taken into account, the best-effort traffic can obtain the most superfluous bandwidth that is not required by the QoS traffic. The experimental simulation has demonstrated that, the solution with the UFQ scheme can provide better performance to the

best-effort traffic than by TSQ and SPQ, and still ensures the service requirements of different types of QoS traffic. It can achieve the better fairness between the QoS traffic and best-effort traffic than the solution with TSQ. The UFQ scheme is proved to be efficient for intra-ONU scheduling in EPONs with the fairness provision.

6.2 Recommendation to the Future Research

6.2.1 New DBA Schemes for EPON Networks

One direction of the future research in EPONs is to develop new DBA schemes for the upstream transmission from multiple ONUs. In an EPON network with various traffic flows, an ONU should be equipped with multiple queues serving different classes of traffic (i.e., voice, video, and data) with different QoS requirements. To satisfy the different requirements of traffic flows, a DBA scheme should meet several objectives. The DBA scheme must be able to guarantee a minimum bandwidth and a maximum delay to each queue, with different parameters for different traffic flows. Fairness of bandwidth distribution is another objective that should be satisfied by the DBA scheme. Then a DBA scheme should be able to effectively isolate and limit the misbehaving user, to protect other users with other services. Finally, the DBA scheme should be robust enough to prevent the system collapse.

In the future research, we will consider to develop new DBA schemes that can provide fair scheduling to all individual queues from multiple ONUs. In the schemes, any queue cannot consume bandwidth exceeding its request. Excess bandwidth left by the queues with low traffic load should be allocated to the backlogged queues in all ONUs in a fair and predictable manner. So the fairness of bandwidth distribution will be achieved for all queues from different ONUs. To fulfill such fair scheduling schemes, the OLT may require to send multiple GATE messages to each ONU, each for an individual queue. It will receive a

separate REPORT message from each independent queue as well. By this mean, the OLT can easily limit the bandwidth to one queue and give more excess bandwidth to other queues. The ONU under the schemes will be very simple because there is no need for lower-level scheduling within the ONU. Several issues should be considered for the new DBA schemes, such as the potential increasing overhead, the different cycle time for different queues, and so on.

6.2.2 Network Management for EPON Networks

Since EPON is aimed to evolve to a converged platform for integrated data, voice and video services, a comprehensive access network management solution is needed to provide the enhanced Fault, Configuration, Accounting, Performance, and Security (FCAPS) functions for service providers to control and manage the access networks. This is an important direction for the future research in the EPON area.

Network management includes all the technological and administrative processes that are necessary and sufficient to achieve a specified set of management objectives. The main objective of network management systems is to provide all the information concerning the resource and the associated operations, and to configure, maintain, and disconnect the connections to the users, and to bill the users for the services provided. Referring to the integrated management of various types of networks, FCAPS covers the Open Systems Interconnection (OSI) five functional areas [79] as follows:

1. Fault management: Network problems or faults should be found and fixed by the fault management facilities, which include mechanisms to detect, isolate, correct, test and record errors or failures occurring in the network. The fault management provides the procedures to report the occurrence of faults, analyze the logs of reported events, execute diagnostic tests, and initiate the corrective actions.

Chapter 6. Conclusions

2. Configuration management: The configuration management is divided into service management and resource management, under which the network is monitored and controlled effectively. It includes mechanisms to change the configuration of the network, which is necessary to alleviate congestion, isolate faults, or meet new user demands. Additionally, it provides procedures to collect and disseminate data concerning the current state of the network, and to modify the operating parameters in network elements.
3. Accounting management: The accounting management deals with the generation and processing functions of subscriber usage information. This includes measuring the subscribers and managing the connection detail information. After collecting the billing data, the access usage record can be obtained and be transferred to the billing system to render invoices to the subscribers. Fraud detection and subscriber-related profile information is also a function of accounting management.
4. Performance management: It can be divided into real-time (or short-term) and long-term management. Performance management capabilities include mechanisms to monitor and analyze the performance of the network and the service. Results of the performance analysis may trigger diagnostic testing procedures and initiate configuration changes in order to maintain the prescribed levels of performance. The performance management provides procedures to collect data, set thresholds, make and test changes, and maintain and analyze the performance logs.
5. Security management: Security management includes mechanisms to protect the access to the communication resources. By the network management, only those who really need to access the specific network resources are allowed to use them. The security management provides support for the management of authorization facilities, access control, encryption and key management, authentication, and the maintenance and manipulation of security logs.

To support the above FCAPS functions, the future research will aim to investigate and develop new network control and management technologies to manage the high-speed EPON networks with multi-vendor products, in order to support DiffServ and provide QoS guarantees to the subscribers. To achieve this objective, the network management solution will be divided into three subsystems for EPONs:

- The Element Management System (EMS) is responsible for configuring, monitoring and managing the access network products from multi-vendors. Interoperability will be provided to various products from different vendors.
- The Network Management System (NMS) provides the network configuration, dynamic resource allocation and multicasting support, which is relates to the utilization of the network resources.
- The Service Management System (SMS) enables the efficient multiple services/applications and supports the flexible subscriber access. It will provide the policy management and subscriber management capabilities.

The above subsystems will be designed and developed respectively to work out the integrated management framework for EPONs.

6.2.3 Adapting EPON to Wireless Access Networks

Although EPON has been developed for optical access networks with fixed lines, it can also be adapted to the wireless access environment to provide multimedia services to users. This is another prospective direction for the future research on EPONs.

The characteristics of EPON make it a promising candidate for optical fiber distribution to the base stations in wireless communication networks [80]. In an EPON, the fiber connected to the OLT is split by a passive optical splitter and shared between multiple ONUs. Correspondingly, a large amount of base stations can be regarded as ONUs and be

Chapter 6. Conclusions

supported by a single fiber with high transmission rate in wireless access networks. The fiber length in an EPON can extend up to 20 km, which allows a cluster of base stations to be connected to the splitter spreading over a wide area of 40 km in diameter. Finally, the inherited broadcast functionality of the EPON transmission can be exploited in a wireless access network. Altering the downlink to a broadcast link supports more flexible support for unbalanced traffic volumes in the different radio cells, and at the same time open for distribution of IP packets to several cells, without decreasing the throughput on the fiber.

In the future research, the proposed DBA schemes in this dissertation can be applied to wireless access networks. It may require modifications to those schemes due to the properties of wireless communications.

Author's Publications

1. Papers (Accepted):

Journal Papers:

1. Maode Ma, Yongqing Zhu, Tee Hiang Cheng, "A Systematic Scheme for Multiple Access in Ethernet Passive Optical Access Networks", *IEEE/OSA Journal of Lightwave Technology*, Vol. 23, No.11, November 2005, pp. 3671-3682.
2. Yongqing Zhu, Maode Ma, Tee Hiang Cheng, "Hierarchical Scheduling to Support Differentiated Services in Ethernet Passive Optical Networks", *Computer Networks*, Vol. 50, Issue 3, February 2006, pp. 350-366.
3. Yongqing Zhu and Maode Ma, "Supporting Differentiated Services with Fairness By an Urgency Fair Queuing Scheduling Scheme in EPONs", *Photonic Network Communications*, Accepted for publication.
4. Yongqing Zhu, Maode Ma, and T. H. Cheng, "An Efficient Solution for Mitigating Light Load Penalty in EPONs," *Journal of Computers and Electrical Engineering*, Accepted for publication.

Conference Papers:

5. Maode Ma, Yongqing Zhu, Tee Hiang Cheng, "A Bandwidth Guaranteed Polling MAC Protocol for Ethernet Passive Optical Networks", *Proceedings of IEEE INFOCOM'2003*, Vol. 1, Mar-Apr. 2003, pp. 22-31.
6. Yongqing Zhu, Maode Ma, Tee Hiang Cheng, "A Novel Multiple Access Scheme for Ethernet Passive Optical Networks", *Proceedings of IEEE GLOBECOM'2003*, Vol. 5, Dec. 2003, pp. 2649-2653.

7. Yongqing Zhu, Maode Ma, Tee Hiang Cheng, "Differentiated services supported by bandwidth guarantee polling (BGP) scheme in Ethernet passive optical networks", *Proceedings of The Ninth IEEE International Conference on Communications Systems (ICCS'2004)*, Sep. 2004, pp. 371–375.
8. Yongqing Zhu, Maode Ma, Tee Hiang Cheng, "An efficient multiple access control scheme to support differentiated services in EPONs", *Proceedings of The IEEE International Conference on Networks (ICON'2004)*, Nov. 2004, pp. 147–151.
9. Yongqing Zhu and Maode Ma, "An Urgency Fair Queuing Scheduling to Support Differentiated Services in EPONs", *Proceedings of IEEE GLOBECOM'2005*, Vol. 4, Dec. 2005, pp.1925–1929.

2. Papers (Under Review)

Conference Papers:

10. Maode Ma and Yongqing Zhu, "A Prediction-based Interleaved Polling MAC Scheme for Ethernet Passive Optical Networks", Submitted to *GLOBECOM'2006*.

Bibliography

- [1] G. Kramer, B. Mukherjee, and A. Maislos, “Multiprotocol over DWDM: Building the Next Generation Optical Internet”, *John Wiley & Sons*, Mar. 2003.
- [2] Glen Kramer, Gerry Pesavento, “Ethernet Passive Optical Network (EPON): Building a Next-Generation Optical Access Network”, *IEEE Communications Magazine*, Vol. 40, Issue 2, Feb. 2002, pp. 66–73.
- [3] Howard Frazier, Gerry Pesavento, “Ethernet Takes on the First Mile”, *IT Professional*, Vol. 3, Issue 4, July-Aug. 2001, pp. 17–22.
- [4] IEEE 802.3ah Ethernet in the First Mile Task Force: <http://www.ieee802.org/3/efm/>.
- [5] H. Ueda, K. Okada, B. Ford, G. Mahony, S. Hornung, D. Faulkner, J. Abiven, S. Durel, R. Ballard, J. Ericson, “Deployment Status and Common Technical Specifications for a B-PON System”, *IEEE Communications Magazine*, Vol. 39, Issue 12, Dec. 2001, pp. 134–141.
- [6] ITU-T Recommendation G.983.1, “Broadband Optical Access Systems Based on Passive Optical Networks (PON),” Oct. 1998.
- [7] J.D. Angelopoulos, H-C. Leligou, T. Argyriou, S. Zontos, “Efficient Transport of Packets with QoS in an FSAN-Aligned GPON”, *IEEE Communications Magazine*, Vol. 42, Issue 2, Feb. 2004, pp. 92–98.
- [8] ITU-T Recommendation G.984.3, “Gigabit-capable Passive Optical Networks (GPON): Transmission Convergence Layer Specification”, Oct. 2003.
- [9] Jongwook Jang, E.K. Park, “Dynamic resource allocation for quality of service on a PON with home networks”, *IEEE Communications Magazine*, Vol. 38 Issue 6, June 2000, pp. 184–190.

Bibliography

- [10] Ringoot, E.; Janssens, N.; Tassent, A.; Angeloupoulos, J.; Blondia, C.; Vetter, P.;
“Demonstration of dynamic medium access control for APON and SuperPON”,
Proceedings of IEEE GLOBECOM'2001. Vol. 3, Nov. 2001, pp. 1570–1574.
- [11] John D. Angelopoulos, Iakovos S. Venieris, George I. Stassinopoulos, “A TDMA
Based Access Control Scheme for APONs”, *IEEE/OSA Journal of Lightwave
Technology*, Vol. 11, Issue 5, May/June 1993, pp. 1095–1103.
- [12] John D. Angelopoulos, Iakovos S. Venieris, E.N. Protonotarios, “A transparent MAC
method for bandwidth sharing and CDV control at the ATM layer of passive optical
networks”, *IEEE/OSA Journal of Lightwave Technology*, Vol. 14, Issue 12, Dec. 1996,
pp. 2625–2635.
- [13] Frans J.M. Panken, “A TDMA based access control scheme for an ATM passive optical
tree network”, *Proceedings of the International IFIP-IEEE Conference, Global
Infrastructure for the Information Age Broadband Communications*, Apr. 1996, pp.
321–332.
- [14] John D. Angelopoulos, Nikos I. Lepidas, E. K. Fragouloupoulos, Iakovos S. Venieris,
“TDMA multiplexing of ATM cells in a residential access SuperPON”, *IEEE Journal
on Selected Areas in Communications*, Vol. 16, Issue 7, Sep. 1998, pp.1123–1133.
- [15] Hakyong Kim, Haewook Lee, Kiseon Kim, “DSP: a MAC scheme with double-side
scheduling for APON”, *Processings of IEEE Pacific Rim Conference (PACRIM'2001)*,
Communications, Computers and signal, Vol. 2, Aug. 2001, pp. 530–533.
- [16] Youngjin MOON, Changhwan OH, Kiseon KIM, “A MAC Scheme Based on the Cell
Arrival Timing Information for Multimedia Services over ATM-Based PON”, *IEICE
Transactions on Communications*, Vol. E82-B, No.10, Oct. 1999, pp. 1702–1705.
- [17] Kae Hsiang Kwong, David Harle, and Ivan Andonovic, “Dynamic Bandwidth
Allocation Algorithm For Differentiated Services over WDM EPONs”, *Proceedings of*

Bibliography

- The Ninth IEEE International Conference on Communications Systems (ICCS'2004)*, Sep. 2004, pp. 116–120.
- [18] G. Thomas, “Improved Performance in Ethernet Passive Optical Networks”, *Proceedings of the First International Workshop on Community Networks and FTTH/P/x (CNFT)*, Oct. 2003.
- [19] Chang-Joon Chae; Wong, E.; Tucker, R.S.; “Optical CSMA/CD Media Access Scheme for Ethernet Over Passive Optical Network”, *IEEE Photonics Technology Letters*, Vol. 14, No. 5, May 2002, pp. 711–713.
- [20] E. Wong and C.-J. Chae, “CSMA/CD-based Ethernet over Passive Optical Network for Delivery of Voice-over-IP Traffic”, *Proceedings of The 16th Annual Meeting of the IEEE Lasar & Electro-Optics Society*, Oct. 2003, pp. 447–448.
- [21] Wong, E.; Chang-Joon Chae; “CSMA/CD-Based Ethernet Passive Optical Network with Optical Internetworking Capability Among Users”, *IEEE Photonics Technology Letters*, Vol. 16, No. 9, Sep. 2004, pp. 2195–2197.
- [22] Yang Wei, Cheng Tee Hiang, “A Hybrid-Slot Decentralized Control Protocol With Quality of Service for Passive Optical Networks”, *Proceedings of The Ninth IEEE International Conference on Communications Systems (ICCS'2004)*, Sep. 2004, pp. 505–507.
- [23] Foh, C.; Andrew, L.; Zukerman, M.; Wong, E.; “FULL-RCMA: A High Utilization EPON”, *Proceedings of Optical Fiber Communications Conference (OFC'2003)*, Vol. 1, Mar. 2003, pp. 282–284.
- [24] Foh, C.H.; Andrew, L.; Wong, E.; Zukerman, M.; “FULL-RCMA: A High Utilization EPON”, *IEEE Journal on Selected Areas in Communications*, Vol. 22, No. 8, Oct. 2004, pp. 1514–524.

Bibliography

- [25] A. Hadjiantonis, S. Sherif, A. Khalil, T. Rahman, G. Ellinas, M. F. Arend, and M. A. Ali, “A Novel Decentralized Ethernet-based Passive Optical Network Architecture”, *Proceedings of IEEE International Conference on Communications (ICC’2004)*, Vol. 3, June 2004, pp. 1781–1785.
- [26] Micheal P. McGarry, Martin Maier, and Martin Reisslein, “Ethernet PONs: A Survey of Dynamic Bandwidth Allocation (DBA) Algorithms”, *IEEE Communications Magazine*, Vol. 42, Issue 8, Aug. 2004, pp. S8–S15.
- [27] Glen Kramer, Biswanath Mukherjee, and Gerry Pesavento, “IPACT: A Dynamic Protocol for an Ethernet PON (EPON)”, *IEEE Communications Magazine*, Vol. 40, Issue 2, Feb. 2002, pp. 74–80.
- [28] G. Kramer, B. Mukherjee, and G. Pesavento, “Interleaved Polling with Adaptive Cycle Time (IPACT): A Dynamic Bandwidth Distribution Scheme in an Optical Access Network”, *Photonic Network Communications*, Vol. 4, No.1, Jan. 2002, pp.89–107.
- [29] Hee-Jung Byun, Ji-Myung Nho, Jong-Tae Lim: “Dynamic bandwidth allocation algorithm in Ethernet passive optical networks”, *Electronics Letters*, Vol. 39, Issue 13, June 2003, pp. 1001–1002.
- [30] Kyuho Son, Hyungkeun Ryu, Song Chong, Taewhan Yoo, “Dynamic bandwidth allocation schemes to improve utilization under non-uniform traffic in ethernet passive optical networks”, *Proceedings of IEEE International Conference on Communications (ICC’2004)*, Vol. 3, June 2004, pp. 1766–1770.
- [31] Tang Shan, Ji Yang, Cheng Sheng, “EPON upstream multiple access scheme”, *Proceedings of International Conferences on Info-tech and Info-net (ICII’2001)*, Vol. 2, Oct. 2001, pp. 273–278.
- [32] Jihyoung Cho, Taesik Lim, Ikpyo Hong; Sunkyoung Kwon, Batchuluun, “Modified of polling mechanism with threshold and precondition for market activation in Ethernet

Bibliography

- PON”, *Proceedings of International Conference on Communication Technology (ICCT'2003)*, Vol. 1, Apr. 2003, pp. 738–742.
- [33] Gumaste, A.; Chlamtac, I.; “A protocol to implement Ethernet over PON”, *Proceedings of IEEE International Conference on Communications (ICC'2003)*, Vol. 2, May 2003, pp. 1345–1349.
- [34] Maode Ma, Yongqing Zhu, Tee Hiang Cheng, “A Bandwidth Guaranteed Polling MAC Protocol for Ethernet Passive Optical Networks”, *Proceedings of IEEE INFOCOM'2003*, Vol. 1, Mar-Apr. 2003, pp. 22–31.
- [35] Yongqing Zhu, Maode Ma, Tee Hiang Cheng, “A Novel Multiple Access Scheme for Ethernet Passive Optical Networks”, *Proceedings of IEEE GLOBECOM'2003*, Vol. 5, Dec. 2003, pp. 2649–2653.
- [36] Yongqing Zhu, Maode Ma, Tee Hiang Cheng, “Differentiated services supported by bandwidth guarantee polling (BGP) scheme in Ethernet passive optical networks”, *Proceedings of The Ninth IEEE International Conference on Communications Systems (ICCS'2004)*, Sep. 2004, pp. 371–375.
- [37] Yongqing Zhu, Maode Ma, Tee Hiang Cheng, “An efficient multiple access control scheme to support differentiated services in EPONs”, *Proceedings of The IEEE International Conference on Networks (ICON'2004)*, Nov. 2004, pp. 147–151.
- [38] Lin Zhang, Eung-Suk An, Chan-Hyun Youn, Hwan-Geun Yeo, and Sunhee Yang, “Dual DEB-GPS Scheduler for Delay-Constraint Applications in Ethernet Passive Optical Networks”, *IEICE Transactions on Communications*, Vol. E86-B, No.5, May 2003, pp. 1575–1584.
- [39] Lin Zhang and Gee-Swee Poo, “Delay Constraint Dynamic Bandwidth Allocation in Ethernet Passive Optical Networks”, *Proceedings of The Ninth IEEE International Conference on Communications Systems (ICCS'2004)*, Sep. 2004, pp. 126–130.

Bibliography

- [40] Xue Chen, Meihong Yu, Yang Zhang, Yu Deng, "A novel upstream dynamic bandwidth assignment scheme for Ethernet PONs", *Proceedings of International Conference on Communication Technology (ICCT'2003)*, Vol. 1, Apr. 2003, pp. 748–750.
- [41] Glen Kramer, Biswanath Mukherjee, and Ryan Hirth, "Supporting Differentiated Classes of Service in Ethernet Passive Optical Networks", *Journal of Optical Networking*, Vol. 1, No. 8&9, 2002, pp. 280–298.
- [42] Maode Ma, Lishan Liu, and Tee Hiang Cheng, "Adaptive scheduling for differentiated services in ethernet passive optical networks", *Proceedings of The Ninth IEEE International Conference on Communications Systems (ICCS'2004)*, Sep. 2004, pp. 102–106
- [43] Ghani, N.; Shami, A.; Assi, C.; Raja, M.Y.A.; "Quality of service in Ethernet passive optical networks" *Proceedings of IEEE/Sarnoff Symposium on Advances in Wired and Wireless Communication*, Apr. 2004, pp. 161–165.
- [44] D. Nikolova, B. Van Houdt, and C. Blondia, "Dynamic bandwidth allocation algorithms in EPON: a simulation study", *Proceedings of Opticomm'2003*, pp. 369–380.
- [45] D. Nikolova, B. Van Houdt, and C. Blondia, "QoS issues in EPON", *Proceedings of the First International Workshop on Community Networks and FTTH/P/x*, Oct. 2003.
- [46] Kye-Hyun Ahn, Kyeong-Eun Han, and Young-Chon Kim, "Hierarchical Dynamic Bandwidth Allocation Algorithm for Multimedia Services over EPONs", *ETRI Journal*, Vol. 26, No. 4, Aug. 2004, pp. 321–330.
- [47] G. Kramer, A. Banerjee, N. Singhal, B. Mukherjee, S. Dixit, Y. Ye, "Fair Queuing with Service Envelopes (FQSE): A cousin-fair hierarchical scheduler for Ethernet PONs",

Bibliography

- Proceedings of Optical Fiber Communications Conference (OFC'2004)*, Vol. 1, Feb. 2004. pp. 833–835.
- [48] Kramer, G.; Banerjee, A.; Singhal, N.; Mukherjee, B.; Dixit, S.; Ye, Y.; “Fair Queueing With Service Envelopes (FQSE): A Cousin-Fair Hierarchical Scheduler for Subscriber Access Networks”, *IEEE Journal on Selected Areas in Communications*, Vol. 22, Issue 8, Oct. 2004, pp. 1497–1513.
- [49] Chadi M. Assi, Yinghua Ye, Sudhir Dixit, and Mohamed A. Ali, “Dynamic Bandwidth Allocation for Quality-of-Service Over Ethernet PONs”, *IEEE Journal on Selected Areas in Communications*, Vol. 21, No. 9, Nov. 2003, pp. 1467–1476.
- [50] Assi, C.; Yinghua Ye; Sudhir Dixit; “Support of QoS in IP-based Ethernet-PON”, *Proceedings of IEEE GLOBECOM'2003*, Vol. 7, Dec. 2003, pp. 3737–3741.
- [51] Su-il Choi, “Cyclic Polling-Based Dynamic Bandwidth Allocation for Differentiated Classes of Service in Ethernet Passive Optical Networks”, *Photonic Network Communications*, Vol. 7, No. 1, Jan. 2004, pp. 87–96.
- [52] Su-il Choi and Jae-doo Huh, “Dynamic Bandwidth Allocation Algorithm for Multimedia Services over Ethernet PONs,” *ETRI Journal*, Vol. 24, No. 6, Dec. 2002, pp. 465–468.
- [53] An, F.; Bae, H.; Hsueh, Y.; Rogge, M.; Kazovsky, L.; Kim, K.; “A new media access control protocol guaranteeing fairness among users in Ethernet-based passive optical networks”, *Proceedings of Optical Fiber Communications Conference (OFC'2003)*, Vol. 1, Mar. 2003, pp. 134–135.
- [54] Jing Xie, Shengming Jiang, Yuming Jiang, “A dynamic bandwidth allocation scheme for differentiated services in EPONs”, *IEEE Communications Magazine*, Vol. 42, Issue 8, Aug. 2004, pp. S32–S39.

Bibliography

- [55] Jing Xie, Shengming Jiang, Yuming Jiang, “A Class-based Dynamic Bandwidth Allocation Scheme for EPONs”, *Proceedings of The Ninth IEEE International Conference on Communications Systems (ICCS'2004)*, Sep. 2004, pp. 357–360.
- [56] Yeon-Mo Yang, Ji-Myong Nho, and Byung-Ha Ahn, “An Enhanced Burst-Polling based Delta Dynamic Bandwidth Allocation Scheme for QoS over E-PONs”, *Proceedings of the ACM Workshop on Next-generation Residential Broadband Challenges (NRBC'2004)*, 2004, pp. 31–36.
- [57] Ho-Sook Lee, Tae-Whan Yoo, Ji-Hyun Moon, and Hyeong-Ho Lee, “A Two-Step Scheduling Algorithm to Support Dual Bandwidth Allocation Policies in an Ethernet Passive Optical Network”, *ETRI Journal*, Vol. 26, No. 2, Apr. 2004, pp. 185–187.
- [58] Miyoshi, H.; Inoue, T.; Yamashita, K.; “QoS-aware dynamic bandwidth allocation scheme in gigabit-ethernet passive optical networks”, *Proceedings of IEEE International Conference on Communications (ICC'2004)* , Vol. 1, June 2004, pp. 90–94.
- [59] Jong-Tae Park, Jong-Wook Baek, J. Won-Ki Hong, “Management of service level agreements for multimedia Internet service using a utility model”, *IEEE Communications Magazine*, Vol. 39, Issue 5, May 2001, pp.100–106.
- [60] Eric Bouillet, Debasis Mitra, K.G. Ramakrishnan, “The structure and management of service level agreements in networks”, *IEEE Journal on Selected Areas in Communications*, Vol. 20, Issue 4, May 2002, pp.691–699.
- [61] Mischa Schwartz: “Telecommunication Networks: Protocols, Modeling and Analysis”, *Addison-Wesley*, 1986
- [62] Joseph L. Hammond, and Peter P. O'Reilly: “Performance Analysis of Local Computer Networks”, *Addison-Wesley*, 1986

Bibliography

- [63] R. B. Cooper and G. Murray, "Queues served in cyclic order", *The Bell System Technical Journal*, Vol. 48, 1969, pp. 675–689.
- [64] R. B. Cooper, "Queues served in cyclic order: waiting times", *The Bell System Technical Journal*, Vol. 49, 1970, pp. 399–413.
- [65] M. Eisenberg, "Queues with periodic service and changeover times", *Operations Research*, Vol. 20, 1972, pp. 440–451.
- [66] O. Hashida, "Analysis of multiqueue", *Rev. El. Commun. Lab.*, Vol. 20, 1972, pp. 189–199.
- [67] P. J. Kuehn, "Multiqueue systems with nonexhaustive cyclic service", *The Bell System Technical Journal*, Vol. 58, 1979, pp. 671–699.
- [68] H. Takagi, L. Kleinrock, "Analysis of Polling Systems", *Japan Science Institute Research Report*, 1985.
- [69] Phuoc Tran-Gia, Thomas Raith, "Multiqueue systems with finite capacity and nonexhaustive cyclic services", *Computer Networking and Performance Evaluation*, 1986, pp. 213–225.
- [70] Phuoc Tran-Gia, "Analysis of polling systems with general input process and finite capacity", *IEEE Transactions on Communications*, Vol. 40, No. 2, 1992, pp. 337–344.
- [71] P. J. Kuehn, "Approximate analysis of general queuing networks by decomposition", *IEEE Transactions on Communications*, Vol. 27, 1979, pp. 113–126.
- [72] Harry G. Perros, Khaled M. Elsayed, "Call admission control schemes: a review", *IEEE Communications Magazine*, Vol. 34, Issue 11, Nov. 1996, pp. 82–91.
- [73] Sugih Jamin, Scott J. Shenker, Peter B. Danzig, "Comparison of Measurement-based Admission Control Algorithms for Controlled-Load Service", *Proceedings of IEEE INFOCOM'1997*, Vol. 3, Apr. 1997, pp. 973–980.

Bibliography

- [74] ANSI/IEEE Standard 802.1D, “IEEE standard for information technology—Telecommunications and information exchange between systems—Local and metropolitan area networks—Common specifications. Part 3: media access control (MAC) bridges”, *IEEE*, 1998, <http://standards.ieee.org/getieee802/download/802.1D-1998.pdf>.
- [75] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, “An Architecture for Differentiated Services”, *IETF RFC 2475*, Dec. 1998.
- [76] Alan Demers, Srinivasan Keshav, and Scott Shenker, “Analysis and simulation of a fair queueing algorithm”, *ACM SIGCOMM Computer Communication Review*, Vol. 19, Issue 4, Sep. 1989, pp. 1–12.
- [77] Chaskar, H. M.; and Madhow, U.; “Fair scheduling with tunable latency: a round-robin approach”, *IEEE/ACM Transactions on Networking*, Vol. 11, Issue 4, Aug. 2003, pp. 592–601.
- [78] Lenzini, L.; Mingozzi, E.; Stea, G.; “A unifying service discipline for providing rate-based guaranteed and fair queueing services based on the Timed Token protocol”, *IEEE Transactions on Computers*, Vol. 51 , Issue 9 , Sep. 2002, pp. 1011–1025.
- [79] Rayes, A.; Sage, K.; “Integrated management architecture for IP-based networks”, *IEEE Communications Magazine*, Vol. 38, Issue 4, Apr. 2000, pp. 48–53.
- [80] Lambertsen, G.; Yamada, T.; “All IP future mobile wireless access network; passive optical network, dynamic logical macro-cell, and IP multicast”, *Proceedings of IEEE GLOBECOM'2002*, Vol. 1, Nov. 2002, pp. 625–629.