ABSTRACT

A NOVEL STORAGE AREA NETWORK (SAN) EXTENSION SOLUTION OVER PASSIVE OPTICAL NETWORKS (PONs)

by
Si Yin

After 9/11 and the 2003 power grid failure in North America, storage area network (SAN) extension has emerged as a critical option to ensure business continuity. However, SAN extension encounters challenges in the access network including the scalability, cost, bandwidth bottleneck, and throughput. In this dissertation, a new solution, SAN extension over passive optical networks (S-PONs), has been proposed to address the above problems. PONs are the mainstream wireline technology for upgrading the megabit-level access solutions (such as xDSL and Cable Modem) into gigabit-level broadband access. To tackle the scalability problem and cost challenge, the S-PON architecture has been designed based on the existing point-to-multiple-point (P2MP) PON infrastructure. To address the bandwidth bottlenecks in SAN extension, three solutions have been proposed to carrying storage signals with gigabit-level transmission. In addition, this dissertation introduces a new device, XtenOLT, which is an upgraded optical line terminal (OLT) with storage provisioning capacity. Two core functional modules, the dynamic resource management (DRM) module and the transmission module, are implemented in XtenOLT, respectively. In DRM, a new buffer management scheme, Tetris, is proposed to manage the buffer pools of XtenOLT, in order to improve SAN extension throughput and utility. Our experimental results show that, in the physical layer, the proposed S-PON transmission technologies successfully deliver SAN traffic to the

long-haul at the rate of 2.5 Gb/s; in the network layer, S-PON with XtenOLT dramatically enhances deliverable throughput and utility over long-distance transmission.

The transmission module further adopts one of the mostly used transmission schemes, time division multiple access (TMDA). In TDMA S-PON, the upstream bandwidth allocation (BA) is one of the critical issues. In the past several years, numerous BA algorithms have been proposed, but most of them are presented in an ad-hoc manner, lacking a generic framework under which these algorithms can be evaluated, compared, and further improved. This dissertation proposes a novel state space model to represent the PON BA algorithms with state variables and input variables under a unified framework. Using this new model, the system level characteristics of diverse BA algorithms have been analyzed. Their performance difference in delay, throughput, and packet loss has also been analyzed from the system point of view. Within the framework of the proposed model, a suitable controller and compensator have been proposed to meet the prescribed objectives such as system robustness, accuracy, and transient performance. Lastly, the established state space model has been extended to the non-linear predictor-based dynamic bandwidth allocation scenario.

A NOVEL STORAGE AREA NETWORK (SAN) EXTENSION SOLUTION OVER PASSIVE OPTICAL NETWORKS (PONs)

by
Si Yin

A Dissertation
Submitted to the Faculty of
New Jersey Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Electrical Engineering

Department of Electrical and Computer Engineering

January 2009

APPROVAL PAGE

A NOVEL STORAGE AREA NETWORK (SAN) EXTENSION SOLUTION OVER PASSIVE OPTICAL NETWORKS (PONs)

Si Yin

11/14/2008

Dr. Nirwan Ansari, Dissertation Advisor — Date
Professor, Department of Electrical and Computer Engineering, NJIT

11/14/2008

Dr. Cristian Borcea, Committee Member — Date
Assistant Professor, Department of Computer Science, NJIT

11/14/2008

Dr. Kevin W. Lu, Committee Member — Date
Chief Scientist, Telcordia Technologies, Inc.

11/14/08

Dr. Roberto Rojas-Cessa, Committee Member — Date
Associate Professor, Department of Electrical and Computer Engineering, NJIT

11/14/2008

Dr. Yanchao Zhang, Committee Member — Date
Assistant Professor, Department of Electrical and Computer Engineering, NJIT

BIOGRAPHICAL SKETCH

Author: Si Yin

Degree: Doctor of Philosophy

Date: January 2009

Undergraduate and Graduate Education:

- Master of Engineering in Electrical Engineering,
  National University of Singapore, Singapore, 2005

- Bachelor of Engineering in Electrical Engineering,
  University of Electronic Science and Technology of China, Chengdu, China 2002

Major: Electrical Engineering

Presentations and Publications:

S. Yin and, N. Ansari,
"Non-Linear Predictor-Based Dynamic Bandwidth Allocation over Point-to-Multi-Point (P2MP) Networks: A Control Theoretical Approach", submitted to *IEEE Transactions on Communications*.

S. Yin, Y. Luo, L. Zong, S. Rago, J. Yu, N. Ansari, and T. Wang,
"Storage Area Network Extension over Passive Optical Networks (S-PONs)", *IEEE Communications Magazine*, pp. 44~52, Jan. 2008.

Y. Luo, S. Yin, N. Ansari, and T. Wang,
"Resource Allocation for Broadband Access over TDM Passive Optical Networks",
IEEE Network, vol. 21, issue 5, pp.20-27, Sep./Oct.2007.

S. Yin, Y. Luo, N. Ansari, and T. Wang,
"Stability of Predictor-Based Dynamic Bandwidth Allocation over EPONs", *IEEE Communications Letters*, vol.11, no.6, pp.549-551, Jun. 2007.

N. Ansari and S. Yin,
"Chapter 149: Storage Area Networks: architectures and protocols",

The Handbook Of Computer Networks Volume III: Distributed Networks, Network Planning, Control, Management, and New Trends and Applications (Hossein Bidgoli, ed.), pp. 217-234, John Wiley & Son, ISBN 978-0-471-78458-9, 2008.

S. Yin, Y. Luo, N. Ansari, and T. Wang,
"Non-Linear Predictor-Based Dynamic Bandwidth Allocation over TDM-PONs: Stability Analysis and Controller Design",
Proceedings of IEEE International Conference on Communications 2008 (ICC'08), Beijing, China, May 2008.

S. Yin, Y. Luo, N. Ansari, and T. Wang,
"Pseudowire System with VCCV Insertion (VI) in Optical Access Networks",
Proceedings of Optical Fiber Communication Conference and Exposition (OFC) and The National Fiber Optic Engineers Conference (NFOEC) (OFC/NFOEC'08), San Diego, CA, Feb. 2008

Y. Luo, S. Yin, J. Yu, L. Zong, and T. Wang,
"Storage Area Network Extension over Passive Optical Networks Using Parallel Signal Detection",
Proceedings of Optical Fiber Communication Conference and Exposition (OFC) and The National Fiber Optic Engineers Conference (NFOEC) (OFC/NFOEC'08), San Diego, CA, Feb. 2008

Y. Luo, S. Yin, N. Ansari, and T. Wang,
"Pseudowire Packet Loss Control in Optical Access",
Proceedings of Optical Fiber Communication Conference and Exposition (OFC) and The National Fiber Optic Engineers Conference (NFOEC) (OFC/NFOEC'08), San Diego, CA, Feb. 2008

Y. Luo, S. Yin, N. Ansari, and T. Wang,
"State Space Model for TDM-PON Bandwidth Allocation",
Proceedings of Optical Fiber Communication Conference and Exposition (OFC) and The National Fiber Optic Engineers Conference (NFOEC) (OFC/NFOEC'08), San Diego, CA, Feb. 2008

S. Yin, Y. Luo, N. Ansari, and T. Wang,
"Controllability of Non-Linear Predictor-Based Dynamic Bandwidth Allocation over EPONs",
Proceedings of IEEE Global Communications Conference (GLOBECOM'07), Washington D.C.

Y. Luo, S. Yin, N. Ansari, and T. Wang,
"QoS-Aware Scheduling over Hybrid Optical Wireless Networks,"
Proceedings of Optical Fiber Communication Conference and Exposition (OFC) and The National Fiber Optic Engineers Conference (NFOEC) (OFC/NFOEC'07), Anaheim Marriott Anaheim, CA, Mar. 2007

Y. Luo, S. Yin, T. Wang, S. Nakamura, and N. Ansari,

"Backup Path Multiplexing over Survivable GMPLS Networks",
Proceedings of Optical Fiber Communication Conference and Exposition (OFC) and The National Fiber Optic Engineers Conference (NFOEC) (OFC/NFOEC'07), Anaheim Marriott Anaheim, CA, Mar. 2007

S. Yin, Y. Luo, N. Ansari, and T. Wang,
"Bandwidth Allocation over EPONs: A Controllability Perspective,"
Proceedings of IEEE Global Communications Conference 2006 (GLOBECOM'06), San Francisco, Nov. 2006

To my wife, for her love, sacrifice, patience and support.
To my parents, for their love, support and encouragement.

ACKNOWLEDGMENT

I am fortunate to have Dr. Nirwan Ansari as my doctoral advisor and mentor, who has guided, inspired, encouraged, and supported me during my whole Ph.D. study. His insightful vision on research and invaluable advice has led me through every step in my research and always give me new inspiration. His invaluable suggestions and guidance were, are and will be the source of strength for my future professional endeavor.

Special thanks are given to my committee members, Dr. Cristian Borcea, Dr. Kevin W. Lu, Dr. Roberto Rojas-Cessa, and Dr. Yanchao Zhang, for their constructive discussions, suggestions, and careful reviews of this research.

I am deeply debt to my wife, Ruiqiong, for her love, sacrifice, and constant support. Her invisible hands of care and support are behind the words, graphs, and equations of this dissertation. My deepest appreciation also goes to my parents, Jingming Yin and Fengyun Chen, for their love and constant encouragement.

Last but not the least, I would like to thank my colleagues in the Advanced Networking Laboratory (ANL) and my friends at NJIT. The time we spent together is the source of good memory in my life.

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

TABLE OF CONTENTS
(Continued)

Chapter **Page**

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The rapid growth of data-intensive applications, including multimedia, e-business, e-learning, and Internet protocol television (IPTV), is driving the demand for higher data-storage capacity. Organizations want their huge amount of data to be stored in such a way that it can be easily accessible and manageable. Furthermore, they also require the critical data to be securely transported, stored, and consolidated at high speeds. The Storage Area Network (SAN) is emerging as the data-storage technology of choice because of its significant performance advantages, such as better scalability and higher availability, over the traditional storage architectures [1]. As shown in Figure 1.1, SAN is a high-speed and special-purpose network that interconnects a set of storage devices with associated servers. SAN architectures win attention from large enterprises such as Google, Yahoo, and Amazon that have tremendous amounts of data to backup, consolidate, as well as replicate among different location. After the 9/11 terrorist attack, SANs have been widely deployed as the major data disaster recovery system infrastructure. However, current SANs were originally designed to operate within a limited distance such as a campus. This is not sufficient to safeguard corporate data.

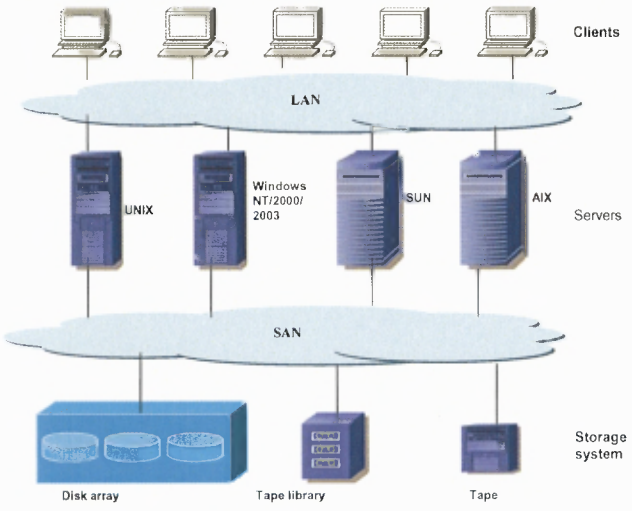


Figure 1.1 Illustration of storage area networks (SANs).

The power-grid failure that occurred in northeastern United States in August 2003 illustrated that a data disaster-recovery system within a small distance cannot ensure business continuity. In this incident, the total business loss was estimated to be 2 billion dollars; the loss is in million dollars per hour during the two days' down time. Hence, the storage sites have to be physically separated up to hundreds or even thousands of miles to avoid severe damage from widespread power outages, earthquakes, fire, and terrorist attacks, such that the damage may be contained within one site only in a disaster [3],[4],[5]. The United States federal regulators, such as the Office of Management and Budget and the General Services Administration, have also adopted a similar disaster recovery strategy into the continuity of operations plan (COOP), which is applicable to all federal agencies, airports, and financial institutes [6]. Recently, one 860-km-long testbed has been set up in European countries to demonstrate the new services over SAN extension [7].

While several solutions have been proposed to extend SANs over long-haul networks, few have addressed the bottleneck in the access network. The real-time and synchronous SAN extension requires gigabit-level data rate among storage devices. This high-bandwidth requirement challenges current telecommunications infrastructures in the access network.

1.1 Motivation

To be a viable solution to support mission-critical storage requirements across long distances, SAN extension has to overcome the following challenges in the access network:

1) Scalability

The conventional extension solution in the access network, such as the fibre channel (FC), uses point-to-point dedicated "dark fiber" to connect SAN into the metro network. This approach lacks scalability and is cost-prohibitive because it requires the available "dark fiber", manpower, and continuously maintained service.

2) Bandwidth bottleneck

The current access technologies, such as DSL, Cable and T1/E1, can only provide up to megabit-level bandwidth, and is far below the gigabit-level bandwidth requirement of SAN [1].

3) Limited buffers

Since SAN is historically designed for local transmission such as campus, only a limited amount of buffers is deployed in SAN switches, thus resulting in poor throughput over the long haul.

The aforementioned three-folded challenges have motivated us to develop a new approach to extend SAN over the access network to improve scalability, to lower cost, to increase the speed of transmission, and to improve throughput.

1.2 Study Scope

In this dissertation, we propose a new solution to address aforementioned problems: SAN extension over passive optical networks (S-PONs). Our scope of study of S-PONs includes,

1) A new architecture to tackle the scalability problem

To tackle the scalability problems and cost challenges, we design the S-PON architecture based on the existing point-to-multiple-point (P2MP) PON infrastructure.

2) A new device in S-PON

We also introduce a new device, XtenOLT, which is an upgraded optical line terminal (OLT) with storage provisioning capacity, to improve SAN extension throughput and utility.

3) New solutions to address the bandwidth bottleneck problem

To address the bandwidth bottlenecks in SAN extension, we propose three solutions for carrying storage signals with gigabit-level transmission, namely, time division multiple access (TDMA), sub-carrier multiple access (SCMA), and wavelength division multiple access (WDMA), respectively.

4) State space model for TDMA schemes

We especially focus on the TDMA schemes of S-PONs and propose a novel state space model as a unified framework to evaluate various TDMA schemes.

5) Non-linear state space model

We also extend the state space model into the non-linear scenario and develop the corresponding non-linear representation model for the system level study.

1.3 Organization

In the rest of this dissertation, Chapter 2 presents the literature review of SAN extension solutions and current TDMA schemes. Chapter 3 introduces a new architecture, S-PON, to tackle the aforementioned challenges of SAN extension over access network. Experiments and simulation results are also discussed. Chapter 4 presents a novel state space model as a unified framework for current TDMA schemes. In Chapter 5, we further discuss the non-linear state space model for TDMA schemes. In Chapter 6, we summarize our contributions and outline the future works.

CHAPTER 2

LITERATURE REVIEW

In this chapter, we firstly review current SAN extension solutions over a long-haul network, including SAN extension over IP, over SONET, and over WDM, respectively. We then survey the existing TDMA schemes for upstream transmission of TDM-PONs, and classify them into three categories.

2.1 SAN Extension Solutions over Long-haul

The existing literature covering SAN extension is mainly about long-haul overlay. Proposed solutions include optical-based extension solutions and IP-based extension solutions. The optical-based extension solutions include extending SAN over synchronous optical network (SONET) and over wavelength division multiplexing (WDM).

SONET-based extension essentially assigns a dedicated SONET channel with fixed bandwidth to each SAN connection [8]. The basic idea of extending SAN over SONET is to map FC frames onto SONET tributaries for transport across metro/regional add-drop (ring) and switching gateways. In SONET, data are passed through dedicated channels at rates from 155 Mbps up to 4 Gbps. Each dedicated channel is given a guaranteed fixed bandwidth for the data.

WDM-based extension divides bandwidth on a fiber into several non-overlapping channels (i.e., wavelengths) and conducts simultaneous message

transmission on different wavelengths in the core network [9]. WDM transmits multiple FC frame exchanges on a single fiber using different wavelengths. WDM is transparent because it will not change the original ways of transmission. Essentially, WDM provides virtual channels for FC transmission with path protection.

Finally, IP-based extension solutions encapsulate data units of SAN traffic into standard IP frames to be transported over core networks [10]. Several protocols, including Internet small computer system interface (iSCSI) [11], fibre channel over TCP/IP (FCIP) [12], and Internet fibre-channel protocol (iFCP) [13] have been introduced to transport the SCSI commands and responses, either by major vendors or the IP Storage Working Group of the Internet Engineering Task Force (IETF).

While several solutions have been proposed to extend SANs over long-haul networks, few have addressed the bottleneck in the access network. The real-time and synchronous SAN extension requires gigabit-level data rate among storage devices. This high-bandwidth requirement challenges current telecommunications infrastructures in the access network.

2.2 Existing Bandwidth Allocation (BA) Algorithms in TDM PONs

As an inexpensive, simple and scalable technology, Passive Optical Networks (PONs) are considered as a promising solution to provide various end users with broadband access [14]. As exemplified in Figure2.1, a PON system is composed of one optical line terminal (OLT) residing in the central office (CO), one passive

optical splitter deployed in the remote node (RN), and multiple optical network units (ONUs) near subscribers' locations. Intermediate powering between the OLT and the ONUs is eliminated by the use of optical fibers and passive optical splitter. Great efforts to expedite the PON standardization process have been made through both the ITU Telecommunication Standardization Sector (ITU-T) and the IEEE Ethernet in the First Mile Task Force (IEEE EFM TF). The ratified PON standard and recommendations are tabulated in Table 2.1. Significant differences among diverse PON "flavors" include the supported line rates and the type of bearer units. The ITU-T G.983.x recommendation series were ratified to specify Broadband PON (BPON) [15], which employs ATM cells to encapsulate the data transmitted between the OLT and ONUs. The IEEE Standard 802.3ah [16] specifies the physical and medium access control (MAC) layer characteristics of Ethernet PON (EPON). EPON carries Ethernet frames with 1 Gb/s symmetric transmission speed. The recently approved IEEE P802.3av Task Force is working on an enhanced version of EPON, 10Gb/s Ethernet Passive Optical Network (10GEPON). 10GEPON upgrades the existing EPON with two solutions, a symmetric solution of 10Gb/s upstream and 10Gb/s downstream transmission, and an asymmetric solution of 10Gb/s downstream and 1Gb/s upstream transmission [17]. Gigabit PON (GPON) [18] is the continuation and evolution of BPON. Besides ATM cells, GPON supports Ethernet frames as well as TDM units by mapping them into GPON Encapsulation

Method (GEM) frames. The maximum transmission speed over GPON reaches 2.448 Gb/s symmetrically.

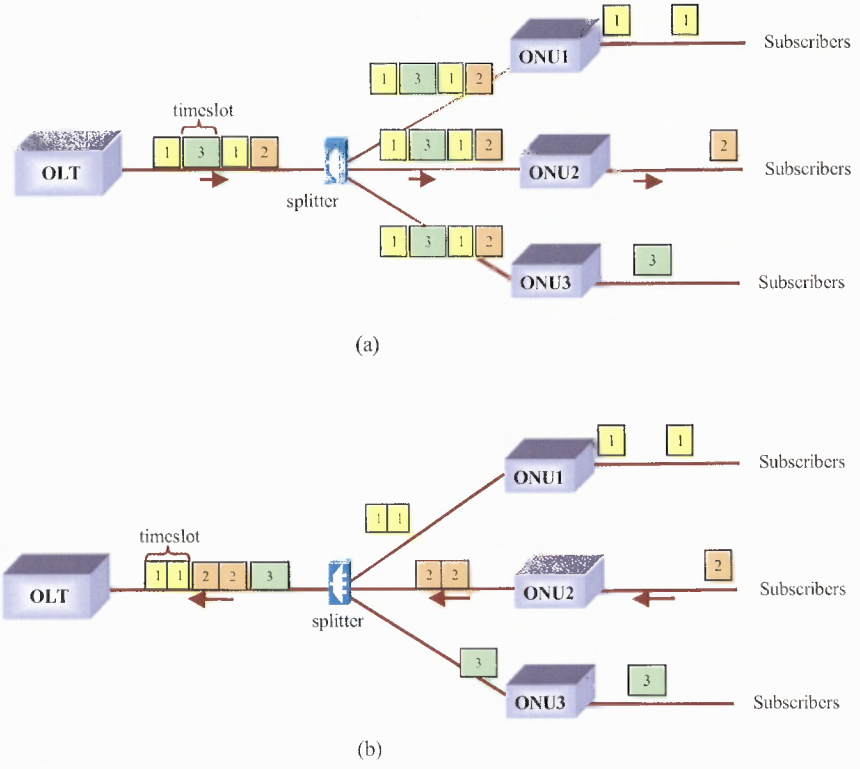


Figure 2.1 Data transmission over EPONs: (a) downstream; (b) upstream.

Table 2. 1 PON Standards and Recommendations

	BPON	EPON	GPON
Standard	ITU-T G.983.x	IEEE 802.3ah	ITU-T G.984.x
Max Speed	1.244 Gb/s downstream, 622.08 Mb/s upstream	1 Gb/s symmetric	2.448 Gb/s symmetric
Data unit	ATM cell	Ethernet frame	GEM frame, ATM cell

BPON, EPON, and GPON are considered as time division multiplexed PONs (TDM-PONs) because their data transmission is divided into timeslots. As shown in

Figure 1 (a), in the downstream from the OLT to the associated ONUs, one wavelength is employed, and time division multiplexing enables data transmission to different ONUs. This is a point-to-multipoint architecture, and data are broadcasted to each ONU through the shared downstream trunk.

In the upstream direction from ONUs to the OLT, another wavelength is employed. As shown in Figure 2.1 (b), each ONU transmits the subscribers' data in dedicated timeslots. This is a multipoint-to-point architecture, which requires a proper mechanism of access control on the shared wavelength. Because of the directional nature of the splitter, each ONU transmits directly to the OLT, but not to other ONUs. Therefore, the ONUs are unable to detect data collision in the upstream direction, and the conventional contention-based mechanism for resource sharing such as carrier sense multiple access with collision avoidance (CSMA/CA) and carrier sense multiple access with collision detection (CSMA/CD) is difficult to implement in TDM-PONs.

Network providers target to build access networks with the lowest cost while achieving the finest granularity. Because neither the EPON standard nor the BPON and GPON recommendations specify any particular resource management mechanism, upstream resource sharing is a critical issue for the TDM-PON performance. Intensive research endeavors have been devoted to this issue recently. In this section, we will provide an overview of the bandwidth management issue over TDM-PONs along with the state-of-art schemes. Although most of the schemes

in the literature address EPON resource management, they can be easily extended to both BPON and GPON scenarios by employing appropriate MAC control cells and fields in the frames.

McGarry et al. [19] summarized upstream bandwidth allocation schemes of EPON. However, many algorithms on bandwidth allocation have been proposed recently that are far beyond the scope reported in Reference [19]. In the following, we provide a new perspective of the state-of-art progress in TDM-PON upstream bandwidth allocation. The major characteristics of the reviewed algorithms are summarized in Table 2.2. Various bandwidth allocation algorithms fall into three major categories: fixed bandwidth allocation (FBA), IPACT-based bandwidth allocation (IBA), and prediction-based bandwidth allocation (PBA). Based on whether they support QoS or not, some categories can be further classified into two sub-categories.

2.2.1 Fixed Bandwidth Allocation (FBA)

FBA grants one ONU a fixed timeslot length for upstream transmission [20]. Without the overhead of resource negotiation as well as bandwidth arbitration, FBA is simple to implement. On the other hand, without considering the instantaneous changes of on-line traffic, FBA uses the upstream wavelength with low efficiency. For example, an ONU will occupy the upstream channel for its assigned timeslot even if it has no data to transmit, while many data could be backlogged in the buffers of other ONUs. Kramer *et al.* [20] evaluated FBA performance, and

concluded that the low efficiency of FBA exacerbates data delay even under medium traffic load, and packet loss is thus deteriorated. Increasing the buffer size could not prevent this phenomenon, mainly because larger buffer only slightly alleviates the congestion, but will continuously increase the burst delay, as more data will be accumulated during the bursts.

2.2.2 IPACT-Based Bandwidth Allocation (IBA)

A. IPACT

Interleaved Polling with Adaptive Cycle Time (IPACT) [21] is the first dynamic bandwidth allocation (DBA) algorithm proposed for TDM-PONs. IPACT adopts a resource negotiation process to facilitate queue report and bandwidth allocation. The OLT polls ONUs and grants timeslots in a round-robin fashion. The granted timeslot is determined by the queue status reported from the ONU. Therefore, the OLT is able to monitor traffic dynamics of each ONU and allocate the upstream bandwidth in accordance with traffic load. IPACT also employs the service level agreement (SLA) parameter to upper-bound the allocated bandwidth to each ONU. This restricts the aggressive competition among ONUs for upstream transmission. As the pioneering bandwidth allocation algorithm, IPACT is regarded as the performance comparison benchmark by most of the later works.

B. IPACT with QoS

Realizing that IPACT solely could not fulfill the multiservice need of subscribers, different IPACT variants have been proposed trying to provision QoS. The state of art includes the following algorithms.

Fair sharing with dual service-level agreements (FSD-SLA) — Banerjee *et al.* [22] proposed to employ dual SLAs in IPACT. The primary SLA specifies the service whose minimum guarantee must be treated with high priority. The secondary SLA describes the service requirement with lower priority. This algorithm fulfills bandwidth allocation by first assigning timeslots to those services with primary SLA, guaranteeing their upstream transmission. After meeting the services with primary SLA, the next round is to meet the secondary SLA services. If bandwidth is not sufficient for the secondary SLA services, the max-min fairness scheme is adopted to distribute the bandwidth among ONUs. In the case that surplus bandwidth is available after arbitration, FSD-SLA distributes the surplus portion first to the primary SLA entities, and then to the secondary SLA entities, both by using the principle of max-min fairness.

Class-of-service oriented packet scheduling (COPS) – Naser and Mouftah [23] proposed the class-of-service oriented packet scheduling (COPS) algorithm to tackle the issue of multiservice. The basic idea is to maintain two groups of leaky-bucket credit pools in the OLT side. One group includes k credit pools, corresponding to k class-of-services (CoSs) in the TDM-PON system. Each pool is used to enforce a long-term average rate of certain CoS traffic transmitted from all ONUs to the OLT.

The other group is composed of m credit pools, corresponding to m ONUs in the TDM-PON system. Each pool is used to control the usage of the upstream channel by an ONU. When processing bandwidth requests, the OLT begins with the highest priority CoS of all ONUs to the lowest priority CoS. As long as the OLT issues any grant, the granted bytes will be subtracted from the corresponding credit pools. New requests will be granted as long as there are enough credits in the pools.

Hybrid granting protocol (HGP)— Shami *et al.* [24] proposed the so-called hybrid granting protocol (HGP) to support diverse QoS requirements. HGP serves the EF traffic in a deterministic manner, while the AF and BE traffic with IPACT. One transmission cycle begins with the EF sub-cycle. The length of the EF sub-cycle is pre-decided. The AF/BE sub-cycle follows the EF sub-cycle, and the AF and BE traffic are transmitted according to the loaded data. In this way, HGP guarantees the service of the delay-sensitive EF traffic, while keeping QoS support for AF and BE services with flexible bandwidth allocation.

2.2.3 Prediction-Based Bandwidth Allocation (PBA)

As shown in Figure 2.2, during upstream transmission, each ONU experiences a waiting time, which ranges from sending the transmission request to sending the buffered data. Since the reported queue status does not consider the data arrived in the waiting time, the OLT defers the transmission of these data (also called waiting-time data). To overcome this drawback, several PBA algorithms

[25],[26],[27] have been proposed. The motivation is to acquire more accurate information of the on-line traffic and to deliver the incoming data as soon as possible.

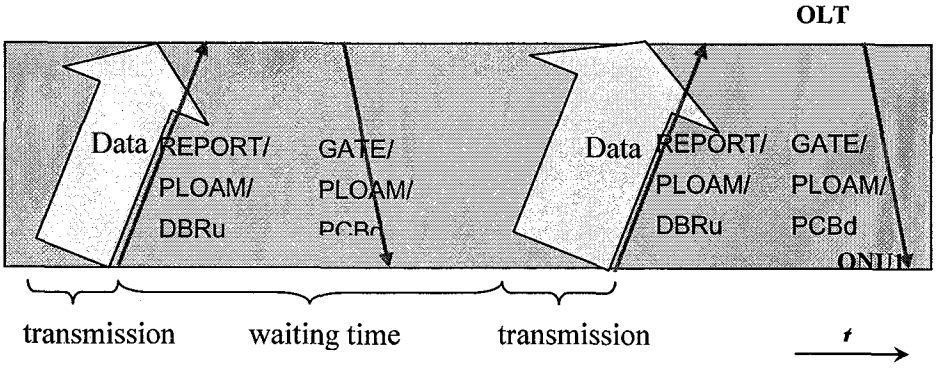


Figure 2. 2 Waiting time in TDM-PON upstream transmission.

A. PBA without QoS

Byun *et al.* [26] addressed the aforementioned issue by estimating the waiting-time data at an ONU and incorporating them into the grant to the ONU. More specifically, a control gain, α, is used to adjust the estimation based on the difference between the departed data and the arrived data in the previous transmission cycle. Simulations with α=0.9 show packet delay reduction as compared to FBA and IBA.

B. PBA with QoS

Dynamic bandwidth allocation 1 (DBA1)— By observing that the delay-sensitive traffic is not able to afford the waiting-time deferral, Assi *et al.* [25] proposed an algorithm (DBA1) to estimate the waiting-time EF data by using the actually arrived EF data in the previous transmission cycle. However, the authors did not

consider the estimation of AF and BE traffic, which dominate the overall access network traffic load and exhibit much severe bursty characteristics.

Limited sharing with traffic prediction (LSTP) — Luo and Ansari [27] proposed to use adaptive filter for traffic prediction. The limited sharing with traffic prediction (LSTP) algorithm estimates each class of waiting-time data based on the actually arrived data of this class in previous transmission cycles. Therefore, the bandwidth requirement is the sum of the estimation and the reported queue length. The OLT arbitrates the upstream bandwidth by using this more accurate information. As a result, LSTP pre-reserves a portion of the upstream bandwidth to deliver the waiting-time data in the earliest transmission cycle, thus mitigating the delay and loss of these data. By using different SLA parameters to restrict different classes of traffic, LSTP facilitates service differentiation.

Table 2. 2 Comparisons of Bandwidth Allocation Algorithms

Algorithm	Category	QoS Support	Delay	Bandwidth Utilization	Brief Description
FBA [8]	FBA	No	High	Low	Both timeslot length and transmission order are fixed; easy to implement; low utilization
IPACT [14]	IBA	No	Medium	Medium	REPORT/GATE mechanism for bandwidth negotiation; bandwidth allocation is upper-bounded by SLA; flexible timeslot length
FSD-SLA [15]	IBA	Yes	Compatible with	Compatible with	Bandwidth requirement is fulfilled by the SLA

			IPACT	IPACT	priority; fair allocation with max-min
COPS [12]	IBA	Yes	Higher than IPACT	Higher than IPACT	Class-of-service oriented manipulation by using two groups of leaky bucket credit pools
HGP [11]	IBA	Yes	50% less than IPACT for EF data	Higher than IPACT	Each transmission cycle is divided into EF sub-cycle (ahead) and AF/BE sub-cycle (afterward) to transmit the differentiated data
Byun et al. [16]	PBA	No	Lower than FBA and IPACT	Higher than FBA and IPACT	A constant control gain is used to estimate the waiting-time data
DBA1 [7]	PBA	Yes	Lower than FBA and IPACT	Higher than FBA and IPACT	The waiting-time EF data are estimated by using the actual data in the previous cycle; AF and BE traffic are treated by IPACT
LSPT [17]	PBA	Yes	Lower than FBA and IPACT	Higher than FBA and IPACT	EF, AF, and BE data in the waiting time are estimated by using adaptive filters

2.3 Chapter Summary

In this chapter, we have reviewed conventional SAN extension solutions over long haul, which covers extension over SONET, over WDM and over IP respectively. To better understand the upstream bandwidth allocation over PON, we classify existing DBA schemes into three categories, namely, fixed bandwidth allocation (FBA), IPACT-based bandwidth allocation (IBA) and predictor-based bandwidth allocation (PBA).

CHAPTER 3

S-PON: TOPOLOGY AND KEY DEVICE AND TRANSMSSION TECHNOLOGIES

As discussed in Chapter 1, SAN extension encounters challenges in the access network, including scalability problems, cost challenges, bandwidth bottlenecks, and low throughput. Recently, we propose a new solution, SAN extension over PON (S-PON), to tackle the aforementioned challenges [28]. This chapter summarizes the topology and key device and transmission technologies of S-PON.

3.1 S-PON Topology

To overcome the scalability problem and cost challenges of dedicated FC, we propose to extend SAN over the passive optical network (PON). The resulting architecture is called S-PON. Instead of using point-to-point "dark fiber" (see Figure 3.1), S-PON employs the point-to-multiple-point (P2MP) architecture of PON, illustrated in Figure 3.2.

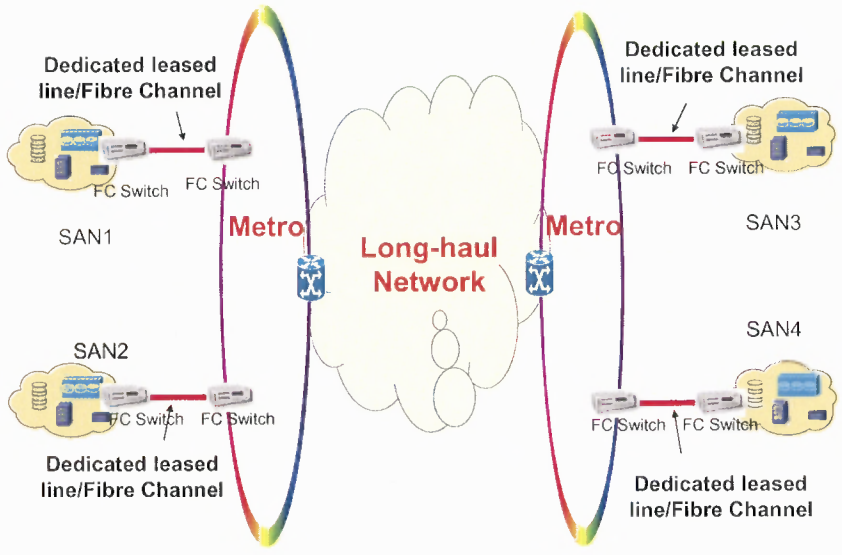


Figure 3.1 Conventional SAN extension topology.

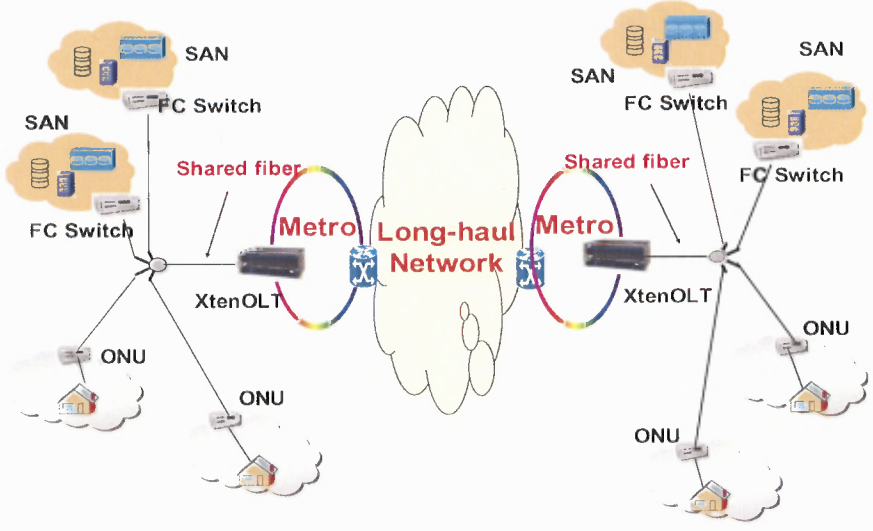


Figure 3. 2 S-PON topology.

The PON infrastructure has been widely deployed in recent years. For example, Verizon's FiOS service, facilitated by PON technologies, has been deployed in 16 different states in the U.S. and is targeted to reach 50% of

households by 2010 [35]. Since PON is leading the trend for next-generation broadband access, S-PON naturally solves the FC scalability problem by building on the growth of PON coverage. Furthermore, the P2MP architecture of S-PON allows SAN to share a single feeder fiber up to 20 km long in the access network with other optical network units (ONUs), thus greatly reducing the cost of SAN extension.

3.2 XtenOLT: Key Device in S-PON

SAN service cannot be directly provided through the PON infrastructure because current OLT in the central office does not support storage service. S-PON essentially requires a new device that can support OLT function and storage provisioning simultaneously. Furthermore, the conventional SAN switch node was designed with few buffers for short-distance transmission. The storage flow control mechanism was implemented with buffers to hold the incoming FC frames before receiving acknowledgements [1]. When transmitting over hundreds of miles, these insufficient buffers lead to low throughput because of the storage flow control sensitivity to the long distance round-trip time.

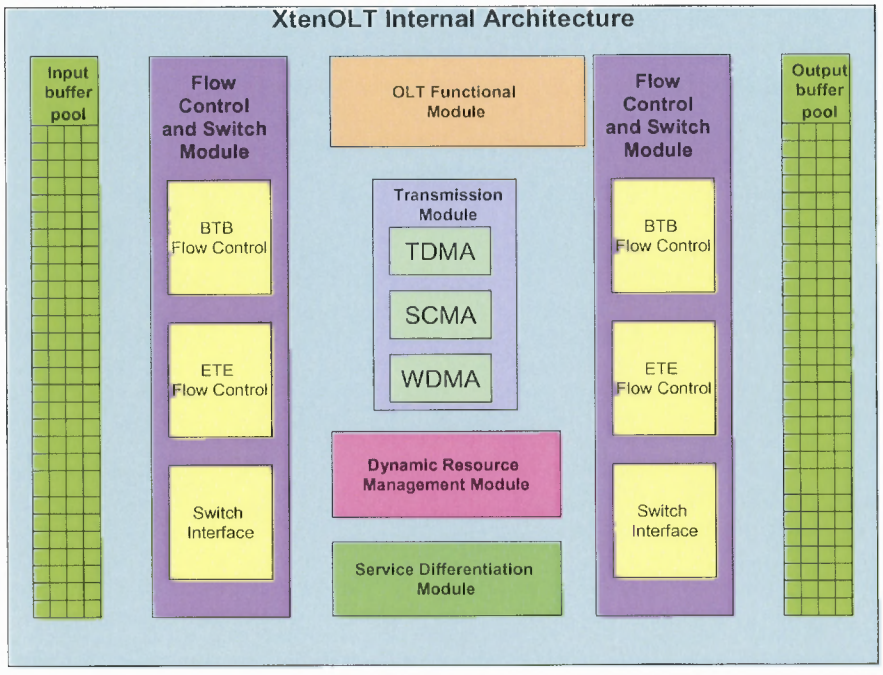


Figure 3.3 Internal architecture of XtenOLT.

To solve this problem, we propose a new device, XtenOLT, in the S-PON architecture. XtenOLT is an enhanced optical line terminal (OLT) in PON with storage service provisioning. The internal architecture of XtenOLT is illustrated in Figure 3.3. Two buffer pools are constructed for buffering the incoming FC frames from local SANs and remote sites, respectively. The flow control and switch module is composed of a buffer-to-buffer (BTB) flow-control sub-module, an end-to-end (ETE) flow-control sub-module, and a switch interface, which are responsible for the BTB and ETE flow control and the switching function [1]. An OLT module is also included, which is responsible for OLT arbitration. The transmission module is responsible for physical layer transmission through the TDMA, SCMA or WDMA sub-modules. The dynamic resource management (DRM) module is responsible for

efficiently managing buffer pools. Lastly, the service differentiation (SD) module is responsible for differentiating services among the SANs.

3.3 Transmission Technologies

To solve the bandwidth bottleneck of current SAN extension techniques, we propose three different transmission technologies for the physical layer: in-band transmission, out-of-band transmission, and out-of-wavelength transmission.

3.3.1 In-band Transmission

With in-band transmission, the SAN shares the upstream channel with other ONUs through time division multiple access (TDMA). In this way, the SAN is regarded as a user in an ONU, sharing the 1 Gb/s bandwidth with other PON users. The TDMA in-band transmission technique is illustrated in Figure 3.4.

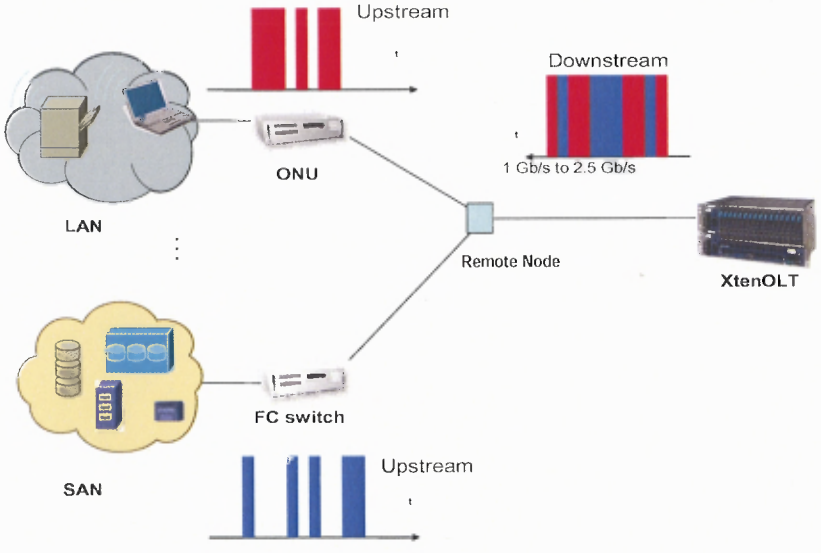


Figure 3.4 In-band transmission.

In the last chapter, we have reviewed numerous TDMA schemes in the literature. To further understand the intrinsic characteristics of these schemes, a unified framework is needed, under which these TDMA schemes can be compared, analyzed, and further improved. To reach this objective, we will extend our discussion in the next chapter, by providing a new state space model as the unified framework for TDMA schemes of S-PON.

3.3.2 Out-of-band Transmission

For more critical SAN applications, S-PON fulfills the bandwidth requirements with out-of-band transmission technology. This is facilitated by sub-carrier multiple access (SCMA), as shown in Figure 3.5. The baseband carrier f_0 is for LAN traffic transmission, while two sub-carriers, f_1 and f_2, are used to transmit the storage data from SAN1 and SAN2, respectively. Either SAN can transmit gigabit-level traffic by using the allocated sub-carrier through the proposed communication infrastructure.

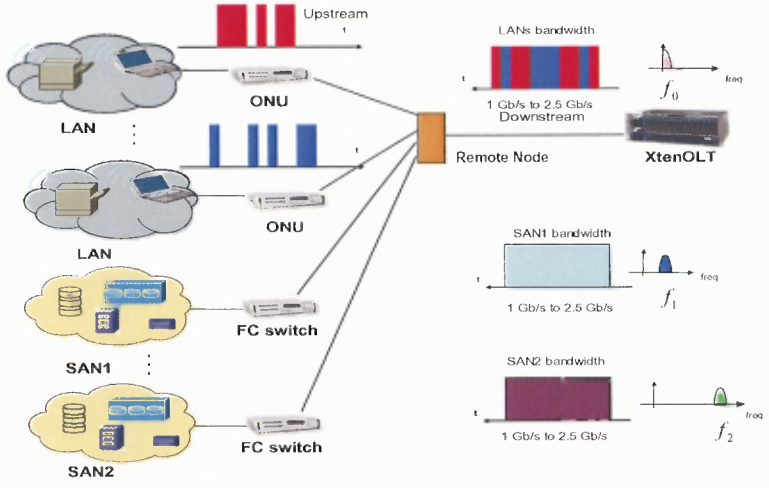


Figure 3. 5 Out-of-band transmission.

3.3.3 Out-of-wavelength Transmission

Out-of-wavelength techniques are employed for the most critical storage data transmission requiring high quality. The practical method takes advantage of wavelength division multiple access (WDMA). As shown in Figure 3.6, LANs are assigned wavelength λ_1 for data transmission, and SAN1 and SAN2 are assigned two other wavelengths, λ_2 and λ_3, respectively, for storage data transmission.

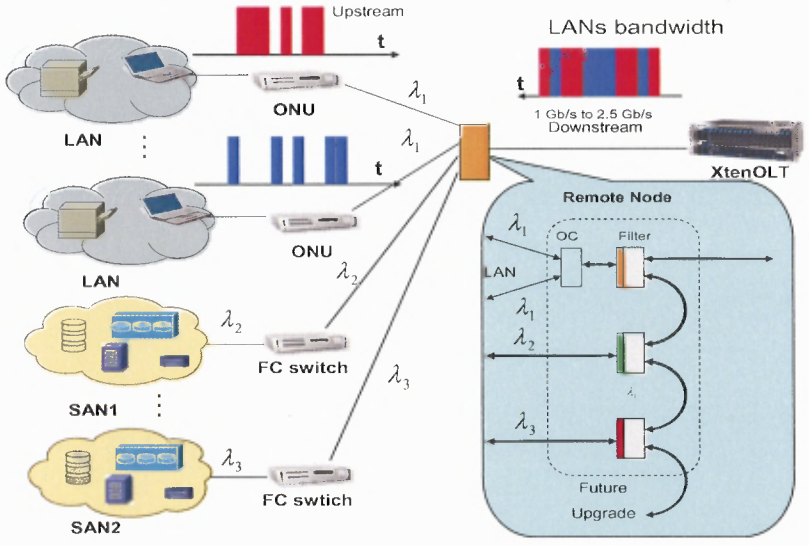


Figure 3. 6 Out-of-wavelength transmission.

The remote node is responsible for multiplexing wavelengths in the upstream direction and demultiplexing in the downstream direction. Regular PON remote nodes, such as optical splitters, can still be used in the TDMA and SCMA scenarios. In the WDMA scenario, however, a modified remote node needs to be implemented, which is shown in the inset of Figure 3.6. The modified remote node separates the downstream LAN data and SAN data with a set of optical filters or an arrayed-waveguide-grating (AWG) to achieve high security and enhanced

transmission rate. For the upstream data, the modified remote node multiplexes the LAN and SAN data and sends them to XtenOLT.

In addition, the optical transmitters and receivers in the FC switches and the OLT need to be upgraded to support sub-carrier and WDM transmission for the SCMA and WDMA scenarios, respectively. In the TDMA scenario, transmitters and receivers for a regular PON can still be used. Table 3.1 summarizes the pros and cons of the three transmission techniques in term of media access, bandwidth, security, and cost.

Table 3. 1 Comparison of Transmission Technologies in S-PON

Transmission techniques	Media access	Bandwidth	Security	Cost
In-band transmission	TDMA	Low	Low	Low
Out-of-band transmission	SCMA	Medium	Medium	Medium
Out-of-Wavelength transmission	WDMA	High	High	High

3.4 Tetris: A New Buffer Management Scheme

Among the various functional modules in XtenOLT, the DRM is the core module for buffer management. Various buffer management schemes have been proposed in the literature. The conventional fixed scheme simply allocates a constant number of buffers to each SAN regardless of the traffic. Under such a scheme, a fixed threshold

is set for each SAN and the arriving packets are discarded if the queue length is beyond the prescribed threshold. It has been shown that the fixed buffering scheme leads to poor performance [29]. Furthermore, the SAN traffic follows a self-similar pattern with bursty characteristics [30],[31],[32]. The fixed buffering scheme also ignores the bursty nature of SAN traffic by preventing the heavily loaded traffic from accessing the free space in the shared buffer pool, thus leading to overall inefficiency.

The linear proportional scheme is another commonly used buffer management technique [33]. Under such a scheme, the number of buffers granted to each SAN is linearly proportional to the request in the previous time interval. Because the linear proportionality scheme favors SANs with large buffer requirements, it causes unfairness and low utility [34].

To overcome the problems of existing buffer management schemes, we propose an algorithm called Tetris, which allocates the buffers to the SANs dynamically. The basic idea of the Tetris algorithm is to grant each SAN the number of buffers equivalent to the minimal request among the SANs. In each time cycle, the Tetris algorithm may take several rounds to complete until all available buffers are successfully granted.

Denote SAN_i as the ith SAN of the system. Let's define the request $R_i(n)$ from SAN_i as the sum of the queued length $Q_i(n)$ of SAN_i at the beginning of

time interval n, and the arrived data length $L_i(n)$ at SAN$_i$ at the end of time interval n. In the linear proportional scheme, the granted buffers to SAN$_i$ is calculated by

$$G_i(n+1) = M \times \frac{Q_i(n) + L_i(n)}{\sum_{i=1}^{k} [Q_i(n) + L_i(n)]}, i = 1, ..., k \tag{3.1}$$

where M is the total amount of buffers available at time interval $n+1$ and k is the number of SANs connected to the switch. The Tetris algorithm can be described as follows: assume there are k SANs requesting buffers at time interval n with their requests $R_i(n)$, queued length $Q_i(n)$, and the arrived data length $L_i(n)$. In round 1, grant $G_i(n+1)$ to each SAN is set to the minimum among requests of k SANs, say $R_1^{\min}$. Without loss of generality, SAN$_i$ is assigned the minimum requirement in round 1, and thus,

$$G_2(n+1) = ... = G_k(n+1) = \min\{Q_i(n) + L_i(n)\} = R_1^{\min}. \tag{3.2}$$

After round 1, there are $k-1$ requests left with the value of $Q_i(n) + L_i(n) - R_1^{\min}, i = 1, 2, ..., k-1$. Assume the minimum value of the leftover request is $R_2^{\min}$, we then have the grant in round 2 as

$$\begin{cases} G_1(n+1) = R_1^{\min} \\ G_2(n+1) = G_3(n+1) = ... = G_k(n+1) = R_1^{\min} + \min\{Q_i(n) + L_i(n) - R_1^{\min}\} = R_1^{\min} + R_2^{\min} \end{cases} \tag{3.3}$$

As long as the available buffer M is larger than $k \times \min\{Q_i(n) + L_i(n), i = 1, 2, ...k\}$, Tetris continues to allocate buffer until the last request is granted, i.e.,

$$\begin{cases} G_1(n+1) = R_1^{\min} \\ G_2(n+1) = R_1^{\min} + R_2^{\min} \\ ... \\ G_k(n+1) = R_1^{\min} + R_2^{\min} + ... + R_{k-1}^{\min} + [Q_k(n) + L_k(n) - R_1^{\min} - R_2^{\min} - ... - R_{k-1}^{\min}] = Q_k(n) + L_k(n) \end{cases} \tag{3.4}$$

A critical condition for deploying the Tetris algorithm is to ensure that $M > k \times \min\{Q_i(n) + L_i(n), i = 1,2,...k\}$ always holds. However, it is possible that the available buffers are not larger than $k \times \min\{Q_i(n) + L_i(n), i = 1,2,...k\}$ after several rounds. In this case, the leftover available buffers will be distributed to each SAN following a certain remainder distribution policy (RDP).

Figure 3.7 shows a simple illustration of this algorithm. Assume there are four SANs requesting buffers in time interval *n* and their requests are represented by four columns. In round 1, the granted buffers to each SAN are equal to the minimal request, which is request 2. Thereafter, the granted buffers (i.e., request 2) are then chopped from each request, as illustrated by the dashed line in round 1. Request 2 is therefore 100% fulfilled in round 1. In round 2, there are only three requests, requests 1, 3 and 4. Similarly, the granted buffers to each SAN are equal to the minimal request, which is request 3. The granted buffers (i.e., request 3) are then chopped from each request. Therefore, request 3 is fulfilled. By following the same process, round 3 fulfills request 1, and round 4 fulfills request 4.

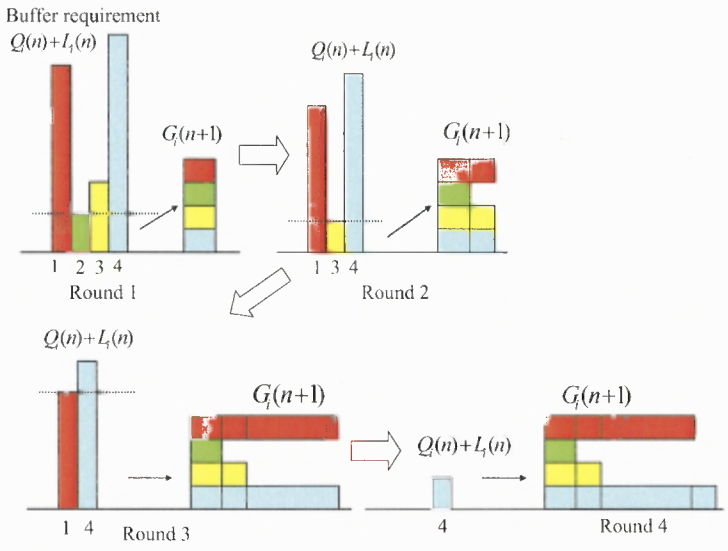


Figure 3. 7 The Tetris algorithm.

3.5 Experiments and Simulation Results

We performed experiments to evaluate the various physical layer transmission techniques. We also simulated several buffer management schemes to evaluate their performance. This section summarizes our results.

3.5.1 Physical Layer Simulation

Our S-PON experiment setup is shown in Figure 3.8. One wavelength is employed to carry the 2.5-Gb/s data signal (to emulate LAN traffic) and the 2.5-Gb/s storage signal (to emulate SAN traffic) in the upstream direction. The storage signal is mixed with a 10-GHz carrier before they are used to drive the modulator to generate the sub-carrier multiplexing signal. One photodetector (PD) is employed after an erbium doped fiber amplifier (EDFA) and a tunable optical filter to receive both data and storage signals. A low-pass filter is used to receive the data signal. To receive

the storage signal, a high-pass filter, a 10-GHz mixer and an electronic amplifier are employed.

Figs. 3.9a–d show the eye diagrams measured for the signals. The experimental results demonstrate that by employing the low-cost electronic filters, the baseband data signal and the modulated storage signal are correctly detected simultaneously at the OLT side, and thus, the extended storage service can be provided by using the widely deployed PON access network architecture.

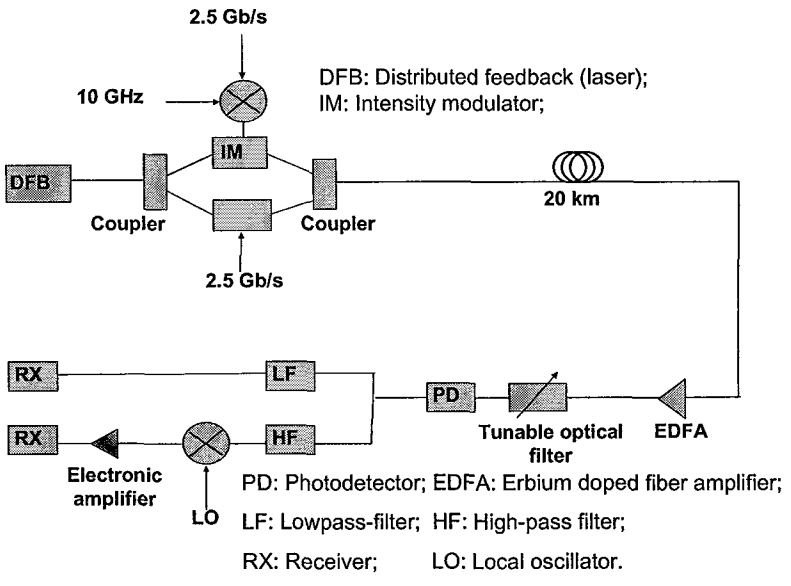


Figure 3.8 Experimental set-up.

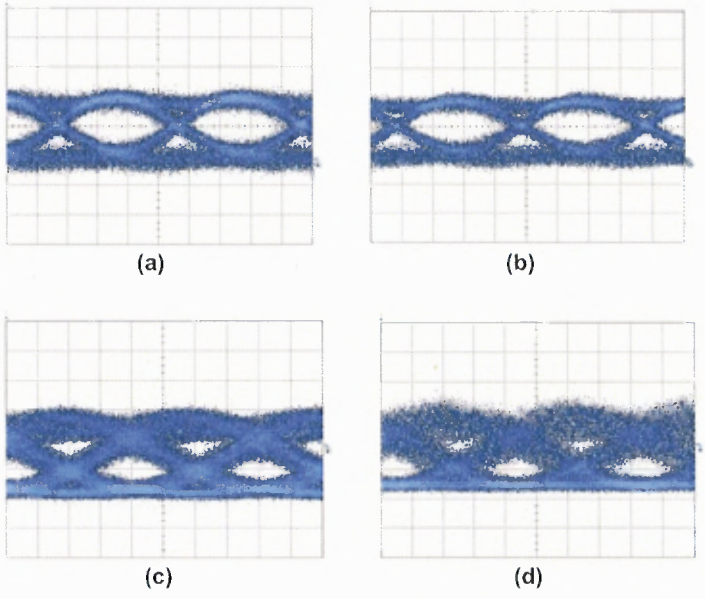
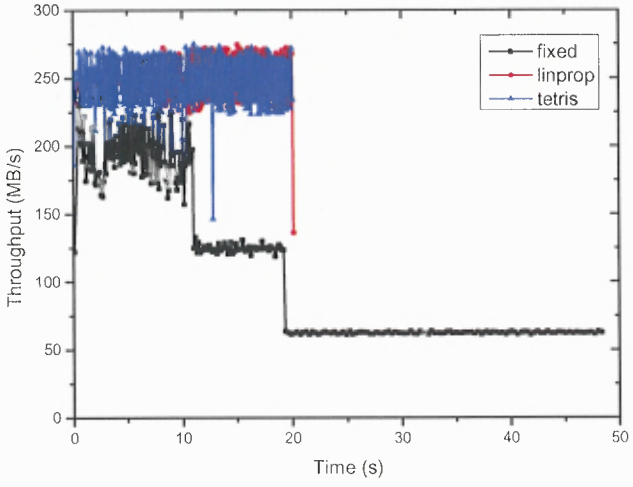


Figure 3. 9 The S-PON experiment (a) Eye diagram for 2.5-Gb/s data signal before transmission, (b) Eye diagram for 2.5-Gb-s data signal after 20-km transmission, (c) Eye diagram for 2.5-Gb/s storage signal before transmission, (d) Eye diagram for 2.5-Gb/s storage signal after 20-km transmission.

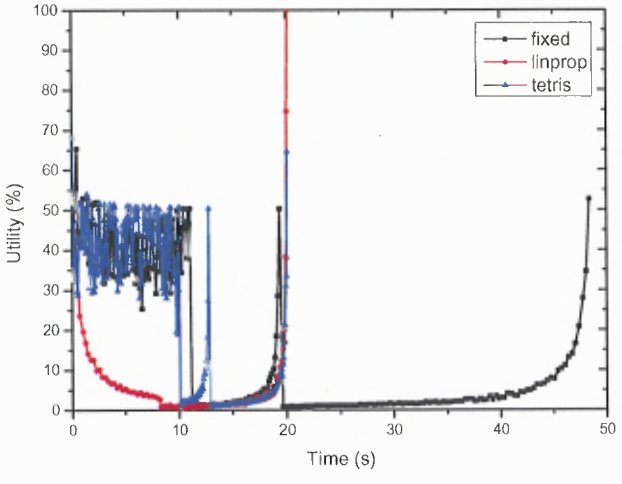
3.5.2 Buffer Management Simulations

We simulated our Tetris buffer management algorithm to evaluate its performance. The experimental S-PON connects two sites about 5000 km apart, one in New York City and one in San Francisco. Each site consists of four SANs, which are connected to the XtenOLT node through the PON architecture. In the simulation, each SAN carries its own local traffic, which are 100 Mb/s, 500 Mb/s, 1 Gb/s and 2.5 Gb/s, respectively. All the traffic patterns are simulated by using a self-similar traffic generator, with the Hurst parameter H set to 0.8. This parameter, with a range of 0.5–1, is a measure of the self-similarity of a time series of traffic. The generated traffic exhibits higher self-similarity when H is closer to the value of 1, and lower self-similarity when H is closer to 0.5 [30]. The long-distance link capacity is set to

be 2.5 Gb/s (i.e., 320 MB/s), and 4800 buffers are configured in XtenOLT. We also compared the performance of the Tetris scheme with other two other buffer-management schemes, namely, the fixed and linear proportional schemes. The simulation results are shown in Figure 3.10 (a) and (b).



(a) Comparison of instantaneous throughput of three algorithms.



(b) Comparison of overall utility of three algorithms.

Figure 3.10 The performance comparison of fixed, proportional and Tetris schemes.

Figure 3.10a shows the instantaneous throughput comparison of the three algorithms. It shows that both the Tetris and the linear proportional scheme achieve around 250 MB/s, which is 78% of the link capacity. Since both schemes make full use of free buffer space, the instantaneous aggregated throughput of the two algorithms overlaps most of the time, and so it is difficult to distinguish the difference between the two in the throughput graph (one color obscures the other). On the other hand, the fixed scheme achieves an average throughput of 100 MB/s, which is 31% of the link capacity. The fixed scheme achieves low throughput because it ignores the bursty nature of SAN traffic and prevents the heavily loaded traffic from accessing the free space in the shared buffer pool. Figure 3.10a also shows that the fixed scheme may cause severe congestion when the queue length reaches a certain threshold level (i.e., the throughput of the fixed scheme in the12th and 19th seconds), which also explains why the fixed scheme takes longer than Tetris and linear proportional schemes to transmit the SAN traffic in Figure 3.10a.

The instantaneous measurements of the overall utility of the three algorithms in the simulation are compared in Figure 3.10b. Here, the overall utility is defined as the request-to-grant ratio in each time cycle, which is a measure of the degree of customer satisfaction. Specifically, the utility of SANi in time interval n is defined as

$$u_i(n) = \frac{G_i(n)}{R_i(n)} \tag{3.5}$$

where $G_i(n)$ is the granted buffer, and $R_i(n)$ is the sum of the queued length $Q_i(n)$ at the beginning of time interval n and the arrived data length $L_i(n)$ at the end of time interval n. $u_i(n+1)$ essentially represents how much of the ratio of the requests are granted in each time interval. Assuming that there are k SANs, the overall utility in the time interval n is defined as

$$u_i^{overall}(n) = \frac{\sum_{i=1}^{k} u_i(n)}{k} \tag{3.6}$$

Figure 3.10b shows that the Tetris, fixed and linear proportional schemes achieve 23%, 20% and 5% average overall utility, respectively. The linear proportional scheme has the lowest overall utility because when heavily-loaded SANs constantly request large numbers of buffers, the linear proportional scheme has no way to prevent the heavy traffic from monopolizing the buffer pool. Consequently, the lightly loaded traffic begins to starve, leading to low utility. The Tetris scheme, on the other hand, always satisfies the SAN with the minimal request, and thus prevents the heavily loaded traffic from monopolizing the buffer pool. In this way, the overall utility is greatly enhanced, as shown in Figure 3.10b.

In the simulation, the fixed scheme provides higher utility than the linear proportional scheme, because the low-traffic SAN requests are always fully satisfied by the buffers allotted to the SAN. On the other hand, the linear proportional scheme provides better throughput than the fixed scheme, because underutilized buffers do not remain idle, and instead are used to satisfy requests from other SANs. The Tetris

scheme exhibits the higher throughput than that of the linear proportional scheme as well as the higher utility than that of the fixed scheme.

3.6 Chapter Summary

In this chapter, we have proposed a new solution, S-PON, to tackle the challenges of extending the SAN into the long-haul network. S-PON adopts the P2MP architecture and leverages the existing PON infrastructure to solve the key issues of scalability and cost. Furthermore, three transmission technologies, TDMA, SCMA and WDMA, were investigated to tackle the legacy transmission bottleneck. We have also proposed a new device to deliver storage service over PON and to solve the low throughput of conventional SAN extension. Our experiments and simulations have shown that, in the physical layer, the proposed S-PON transmission technologies successfully deliver SAN traffic to the long-haul at the rate of 2.5 Gb/s; in the network layer, and XtenOLT with the Tetris buffer-management scheme dramatically enhances the deliverable throughput and overall utility.

CHAPTER 4

STATE SPACE MODEL FOR TDMA TRANSMISSION IN S-PON

In order to provide a general representation of the resource management issue for upstream in S-PON, we will establish a state space model in this chapter. This model describes the S-PON system as a threesome of on-line traffic load, bandwidth arbitration decision, as well as queue status at ONUs. The resource allocation of transmission cycle (n+1) is related to that of transmission cycle n by differential equations. This time-domain approach provides a convenient and compact way to model and analyze the S-PON system with multiple inputs from and outputs to the associated ONUs and SANs.

4.1 System Model

To make the model more generic, we start our discussion with a general point-to-multiple-point (P2MP) network, which comprises a root station (RS) and a number of leaf stations (LSs). Any media with a RS broadcasting packets through a single trunk (such as frequency, wavelength, or wireless channel) to LSs is referred to as downstream, and with LSs unicasting packets through branches and trunk to the RS is referred to as upstream. In addition, the LSs may not communicate with each in a peer-to-peer manner. In the case of S-PON, XtenOLT serves as the root station, and ONUs/SANs serve as the leaf stations. In the downstream, packets are broadcasted through wavelength λ_1 to each LS (i.e., ONU or SAN). While in

the upstream, each LS (i.e., ONU or SAN) unicasts its packets to the RS (i.e., XtenOLT) through a shared wavelength λ_2.

Consider a P2MP system with one RS and y LSs, as illustrated in Figure 4.1. The RS serves each LS once in a service cycle. The following notations are adopted for our analysis. With service cycle n defined as the time interval $[t_n, t_{n+1})$, where t_n stands for the time when cycle n starts, and t_{n+1} cycle n ends, the following notations are adopted for our analysis.

$Q_i(n)$: the reported queued length by the piggybacked REPORT message from LSi ($1 \le i \le y$) at the beginning of service cycle n;

$\lambda_i(n)$: the actually arrived data of LSi at the end of service cycle n;

$\hat{\lambda}_i(n)$: the predicted arrival data at LSi at the beginning of service cycle n;

$d_i(n+1)$: the departed data from LSi at the end of service cycle n;

$R_i(n+1)$: the bandwidth requirement of LSi at the end of service cycle n (it may or may not be the same as $G_i(n+1)$, depending on the particular bandwidth allocation scheme);

$G_i(n+1)$: the allocated timeslot to LS $_i$ at the end of service cycle n;

$G_i^{\max}$: the maximum timeslot length prescribed by the service level agreement (SLA).

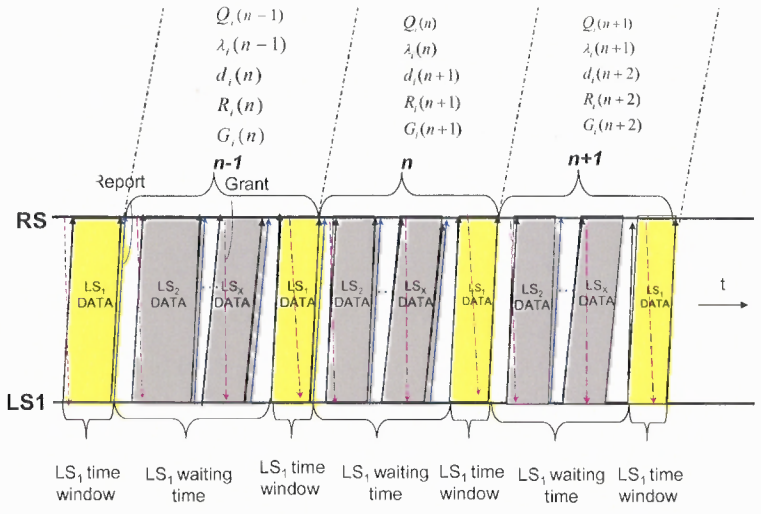


Figure 4.1 Resource management in P2MP networks.

Since no queue status report is conducted in the fixed bandwidth allocation (FBA) scheme, the reported queue length equals zero, and we have

$$Q_i(n+1) = 0 \tag{4.1a}$$

for FBA.

In the IPACT-based bandwidth allocation (IBA) scheme, the reported queue length of transmission cycle n+1 is determined by the difference of the injected data, which include the transmission residual of cycle n (i.e., $Q_i(n)$) as well as the incoming data arrived in the waiting time at ONUi in transmission cycle n (i.e., $\lambda_i(n)$), and the delivered data (i.e., $d_i(n+1)$), i.e.,

$$Q_i(n+1) = Q_i(n) + \lambda_i(n) - d_i(n+1) \tag{4.1b}$$

In the predictor-based bandwidth allocation (PBA) scheme, it is possible that "over-grant" occurs. This "over-grant" is adjusted by reporting the difference between the injected data (i.e., $Q_i(n)+\lambda_i(n)$) and the grant $G_i(n+1)$, i.e.,

$$Q_i(n+1) = Q_i(n) + \lambda_i(n) - G_i(n+1) \tag{4.1c}$$

Eq.(4.1a)~(4.1c) are further formulated as

$$Q_i(n+1) = \begin{cases} 0, & for\ FBA \\ Q_i(n) + \lambda_i(n) - d_i(n+1), & for\ IBA \\ Q_i(n) + \lambda_i(n) - G_i(n+1), & for\ PBA \end{cases} \tag{4.1}$$

On the other hand, the resource request $R_i(n)$ of LS$_i$ for service cycle n is determined by perspective resource allocation schemes. For FBA, the resource request of LSi in service cycle (n+1) (i.e., $R_i(n+1)$) is the fixed value R_{fix}, i.e.,

$$R_i(n+1) = R_{fix} \tag{4.2a}$$

While in IBA, $R_i(n+1)$ is determined by the reported queue length, i.e.,

$$R_i(n+1) = Q_i(n), \tag{4.2b}$$

When traffic predictor is employed, as the case in PBA, $R_i(n+1)$ is determined by the sum of the reported queue length and the predicted arrival data, i.e.,

$$R_i(n+1) = Q_i(n) + \hat{\lambda}_i(n) \tag{4.2c}$$

where $\hat{\lambda}_i(n)$ is the predicted arrival data at LS$_i$ in service cycle n+1. Eq.(2a)~(2c) is further represented by

$$R_i(n+1) = \begin{cases} R_{fix}, & for\ FBA \\ Q_i(n), & for\ IBA \\ Q_i(n) + \hat{\lambda}_i(n), & for\ PBA \end{cases} \tag{4.2}$$

After processing the request, the RS allocates time windows $G_i(n+1)$ to LS$_i$. In FBA, the assigned resource to LSi in transmission cycle n+1 (*i.e.*, $G_i(n+1)$) is the

fixed value R_{fix}. While in both IBA and PBA, $G_i(n+1)$ is the smaller value of the bandwidth request (*i.e.*, $R_i(n+1)$) and the SLA parameter (*i.e.*, $G_i^{\max}$), i.e.,

$$G_i(n+1) = \begin{cases} R_{fix}, & FBA \\ \min[R_i(n+1), G_i^{\max}], & IBA \\ \min[R_i(n+1), G_i^{\max}], & PBA \end{cases} . \quad (4.3)$$

After receiving the bandwidth allocation decision, LS*i* schedules its upstream transmission indicated by $G_i(n+1)$, and the delivered data $d_i(n+1)$ is

$$d_i(n+1) = \min\{G_i(n+1), Q_i(n) + \lambda_i(n)\} \quad (4.4).$$

According to the modern control theory [36],[37], a unified state space model can be constructed for FBA, IBA, and PBA based on Eq.(4.1) and (4.2), as follows

$$X_i(n+1) = AX_i(n) + BU_i(n), \quad (4.5)$$

where $X_i(n) = [R_i(n) \quad Q_i(n)]^T$ is the state vector, indicating the bandwidth requirement and the queue length of LS_i, and $U_i(n)$ is the input vector, representing the arrived data during the waiting time and the SLA parameter. A and B are the matrices for the state vector and input vector, respectively, determining the intrinsic characteristics of each scheme at the system level.

Hence, a unified model for upstream resource allocation schemes over a P2MP system is established through the state space equation Eq. (4.5), with Eq.(4.3) and Eq.(4.4) as the performance constraints. The model essentially exhibits the threesome relationship between the input (*i.e.*, on-line network traffic load), output (*i.e.*, bandwidth allocation decision), and state variables (*i.e.*, queue length and

resource requirement). The state space representation provides a convenient and compact way to model and analyze various resource allocation schemes for the P2MP system from the control theory point of view. In this way, a specific resource allocation scheme essentially defines its particular coefficient matrices A and B to assign the upstream resource in a different way.

Changes of the on-line traffic imply the following four scenarios when the RS arbitrates the upstream resource allocation,

Scenario 1: $R_i(n+1) < G_i^{\max}, G_i(n+1) < Q_i(n) + \lambda_i(n)$

Scenario 2: $R_i(n+1) < G_i^{\max}, G_i(n+1) \geq Q_i(n) + \lambda_i(n)$,

Scenario 3: $R_i(n+1) \geq G_i^{\max}, G_i(n+1) < Q_i(n) + \lambda_i(n)$, and

Scenario 4: $R_i(n+1) \geq G_i^{\max}, G_i(n+1) \geq Q_i(n) + \lambda_i(n)$.

The first condition in the above scenarios determines the granted timeslot length as formulated by Eq. (4.3), and the second condition decides the departed data in a service interval, which is defined by Eq. (4.4). Once $G_i(n+1)$ and $d_i(n+1)$ are settled, so is Eq.(4.5), such that the state space model can be determined.

In the following sections, based on the proposed state space model, we will further discuss the intrinsic characteristics, i.e., controllability and stability, of each scheme.

4.2 System Controllability Analysis

The first critical characteristic of the resource allocation schemes at the system level is their controllability. By "controllability", we mean that the shared upstream resource in a P2MP system can be arbitrated properly among multiple LSs, even when the loaded traffic changes drastically [38],[39]. An individual resource allocation scheme designs its own state vector matrix A and input vector matrix B to arbitrate the upstream resource in a different way. A well-designed resource allocation scheme is expected to properly manage the upstream resource, implying that the scheme can meet the dynamic traffic input from multiple LSs, and steer the resource efficiency over the P2MP system from any initial value to the optimum state within a limited time window. This kind of controllability property is a crucial factor in many other P2MP management issues, such as queue stabilization, delay bounds, and optimal resource control. It is impossible to tune a P2MP system into the optimum state within prescribed time window if the applied resource management scheme is uncontrollable. In this section, we will analyze FBA, IBA and PBA schemes in detail, elaborating their performance differences from the system controllability point of view.

4.2.1 PBA Controllability Analysis

The traffic forecast in PBA can be generalized as

$$\hat{\lambda}_i(n) = \alpha_i \lambda_i(n-1), \tag{4.6}$$

where α_i is the estimation index to extract the correlation from the history traffic. α_i indicates the impact of the input history data on the output prediction, and is a positive real number [40],[41].

A. Scenario 1: $R_i(n+1) < G_i^{\max}, G_i(n+1) < Q_i(n) + \lambda_i(n)$

In this scenario, the resource request is no larger than the SLA specification, and the granted timeslot is no larger than the sum of residual data and arrived data in a service cycle. From Eq.(4.3) and (4.4), we will have $G_i(n+1) = R_i(n+1)$. This is the case when the LSs are well behaved within the SLA specification. The state space is thus

$$R_i(n+1) = Q_i(n) + \hat{\lambda}_i(n) = Q_i(n) + \alpha_i \lambda_i(n-1) \tag{4.7a}$$

$$Q_i(n+1) = Q_i(n) + \lambda_i(n) - R_i(n+1). \tag{4.7b}$$

Substituting Eqs. (4.6a) into (4.6b), we obtain,

$$Q_i(n+1) = \lambda_i(n) - \alpha_i \lambda_i(n-1) \tag{4.7c}$$

Eqs.(4.6a) and (4.6c) are discrete linear system represented by

$$\begin{bmatrix} R_i(n+1) \\ Q_i(n+1) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} R_i(n) \\ Q_i(n) \end{bmatrix} + \begin{bmatrix} 0 & \alpha_i \\ 1 & -\alpha_i \end{bmatrix} \begin{bmatrix} \lambda_i(n) \\ \lambda_i(n-1) \end{bmatrix}. \tag{4.7}$$

Let $A_1 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, B_1 = \begin{bmatrix} 0 & \alpha_i \\ 1 & -\alpha_i \end{bmatrix}$, and $U_i(n) = \begin{bmatrix} \lambda_i(n) \\ \lambda_i(n-1) \end{bmatrix}$; we have

$$X_i(n+1) = A_1 X_i(n) + B_1 U_i(n) \quad . \tag{4.8}$$

Theorem 4.1 The system described by Eq. (4.8) is controllable.

Proof: A system is controllable if and only if the $n \times nr$ controllability matrix $U = \begin{bmatrix} B & AB & \dots & A^{n-1}B \end{bmatrix}$ is full row rank [37]. In the above scenario,

$U=[B_1 \quad A_1B_1]=\begin{bmatrix} 0 & \alpha_i & 1 & -\alpha_i \\ 1 & -\alpha_i & 0 & 0 \end{bmatrix}$. Since α_i is defined as a positive real number, it is obvious to see the controllability matrix U is full row rank. Hence, the system denoted by Eq. (4.8) is controllable.

B. Scenario 2: $R_i(n+1) < G_i^{\max}, G_i(n+1) \geq Q_i(n) + \lambda_i(n)$

In this scenario, the resource request is no larger than the SLA specification, implying $G_i(n+1) = R_i(n+1)$. However, the granted timeslot is larger than the sum of residual data and arrived data. This "over-grant" is adjusted by reporting the difference between the granted timeslot and bandwidth requirement. To facilitate this mechanism, the "negative" queue length is used to measure the "over-grant". Hence, we will have $Q_i(n+1) = Q_i(n) + \lambda_i(n) - G_i(n+1)$, i.e.,

$$Q_i(n+1) = Q_i(n) + \lambda_i(n) - R_i(n+1). \tag{4.9a}$$

Note that the "negative" REPORT implies "over-grant" with empty queue, and thus the resource requirement is only the estimated arrival data, i.e.,

$$R_i(n+1) = \alpha_i \lambda_i(n-1), \tag{4.9b}$$

From Eqs. (4.9a) and (4.9b), we obtain

$$Q_i(n+1) = Q_i(n) + \lambda_i(n) - \alpha_i \lambda_i(n-1). \tag{4.9c}$$

Eqs.(4.9b) and (4.9c) give the discrete linear system

$$X_i(n+1) = A_2 X_i(n) + B_2 U_i(n), \tag{4.9}$$

where $U_i(n)=[\lambda_i(n) \quad \lambda_i(n-1)]^T$, $A_2 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$, and $B_2 = \begin{bmatrix} 0 & \alpha_i \\ 1 & -\alpha_i \end{bmatrix}$. Similar to the previous scenario, the following theorem is established.

Theorem 4.2 The system described by Eq. (4.9) is controllable.

Proof: The system is controllable if and only if the $n \times nr$ controllability matrix $U=[B_2 \quad A_2B_2]=\begin{bmatrix} 0 & \alpha_i & 0 & 1 \\ 1 & -\alpha_i & 0 & -\alpha_i \end{bmatrix}$ is full row rank. Note that with $\alpha_i > 0$, the controllability matrix U is always full row rank. Hence, the system is controllable.

C. Scenario 3: $R_i(n+1) \geq G_i^{max}, G_i(n+1) < Q_i(n) + \lambda_i(n)$

In this scenario, the incoming traffic is heavy, and the RS uses the SLA upper bound G_i^{max} to limit the aggressive resource requirement. The state space turns into

$$R_i(n+1) = Q_i(n) + \alpha_i \lambda_i(n-1), \tag{4.10a}$$

$$Q_i(n+1) = Q_i(n) + \lambda_i(n) - G_i^{max}. \tag{4.10b}$$

The discrete linear system is described by

$$X_i(n+1) = A_3 X_i(n) + B_3 U_i(n), \tag{4.10}$$

where $u_i(n)=\left[\lambda_i(n) - G_i^{max} \quad \lambda_i(n-1)\right]^T$, $A_3 = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}$, and $B_3 = \begin{bmatrix} 0 & \alpha_i \\ 1 & 0 \end{bmatrix}$.

Theorem 4.3 The system described by Eq. (4.10) is controllable.

Proof: The system is controllable if and only if the $n \times nr$ controllability matrix $U=[B_3 \quad A_3B_3]=\begin{bmatrix} 0 & \alpha_i & 1 & 0 \\ 1 & 0 & 1 & 0 \end{bmatrix}$ is full row rank. Note that with $\alpha_i > 0$, the controllability matrix U is always full row rank. Hence, the system is controllable.

D. Scenario 4: $R_i(n+1) \geq G_i^{max}, G_i(n+1) \geq Q_i(n) + \lambda_i(n)$

In this scenario, the request is over the upper bound G_i^{max} and the grant is larger than the sum of residual data and arrived data. The state space turns into

$$R_i(n+1) = \alpha_i \lambda_i(n-1) \tag{4.11a}$$

$$Q_i(n+1) = Q_i(n) + \lambda_i(n) - G_i^{max} \tag{4.11b}$$

The discrete linear system is then described by

$$X_i(n+1) = A_4 X_i(n) + B_4 U_i(n), \tag{4.11}$$

where $u_i(n) = \left[\lambda_i(n) - G_i^{max} \quad \lambda_i(n-1)\right]^T$, $A_4 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$, and $B_4 = \begin{bmatrix} 0 & \alpha_i \\ 1 & 0 \end{bmatrix}$.

Theorem 4.4 The system described by Eq. (4.11) is controllable.

Proof: The system is controllable if and only if the $n \times nr$ controllability matrix $U = [B_3 \quad A_3 B_3] = \begin{bmatrix} 0 & \alpha_i & 0 & 0 \\ 1 & 0 & 1 & 0 \end{bmatrix}$ is full row rank. Note that with $\alpha_i > 0$, the controllability matrix U is always full row rank. Hence, the system is controllable.

4.2.2 IBA Controllability Analysis

A. Scenario 1: $R_i(n+1) < G_i^{max}, G_i(n+1) < Q_i(n) + \lambda_i(n)$

Similar to PBA, this scenario is the case when the LSs are well behaved within the SLA specification, where the resource request is no larger than the SLA specification, and the granted timeslot is no larger than the sum of residual data and arrived data. The state space is thus

$$R_i(n+1) = Q_i(n) \tag{4.12a}$$

$$Q_i(n+1) = Q_i(n) + \lambda_i(n) - R_i(n+1). \tag{4.12b}$$

Substituting Eq. (4.12a) into (4.12b), we obtain,

$$Q_i(n+1) = \lambda_i(n) \tag{4.12c}$$

Eqs.(4.12a) and (4.12c) are the discrete linear system represented by

$$\begin{bmatrix} R_i(n+1) \\ Q_i(n+1) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} R_i(n) \\ Q_i(n) \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \lambda_i(n). \quad (4.12d)$$

Let $A_5 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, B_5 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, and $U_i(n) = \lambda_i(n)$; we have

$$X_i(n+1) = A_5 X_i(n) + B_5 U_i(n) \quad . \quad (4.12)$$

Theorem 4.5 The system described by Eq. (4.12) is controllable.

Proof: It is easy to see the controllability matrix $U = [B_5 \quad A_5 B_5] = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ is full row rank.

Hence, the system is controllable.

B. Scenario 2: $R_i(n+1) < G_i^{\max}, G_i(n+1) \geq Q_i(n) + \lambda_i(n)$

In this scenario, the granted timeslot is larger than the sum of the residual data and arrived data, and the LS*i* sends the data up to the sum of the residual data and the arrived data, and then report zero queue length to the RS at the end of time interval n+1. Hence, we will have

$$Q_i(n+1) = 0 . \quad (4.13a)$$

Since the report queue length is zero and IBA does not estimate the arriving data, we have,

$$R_i(n+1) = 0 . \quad (4.13b)$$

Eqs.(4.13a) and (4.13b) gives the discrete linear system

$$X_i(n+1) = A_6 X_i(n) + B_6 U_i(n), \quad (4.13)$$

where $U_i(n) = \lambda_i(n)$, $A_6 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$, and $B_6 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$.

Theorem 4.6 The system described by Eq. (4.13) is uncontrollable.

Proof: The system is controllable if and only if the controllability matrix U is full row rank. Note $U=[B_6 \quad A_6B_6]=\begin{bmatrix}0 & 0\\ 0 & 0\end{bmatrix}$, and the controllability matrix U is not full row rank. Hence, the system is uncontrollable.

C. Scenario 3: $R_i(n+1) \geq G_i^{\max}, G_i(n+1) < Q_i(n) + \lambda_i(n)$

In this scenario, the incoming traffic is heavy, and the RS uses the SLA upper bound $G_i^{\max}$ to limit the aggressive resource requirement. The state space turns into

$$R_i(n+1) = Q_i(n), \tag{4.14a}$$

$$Q_i(n+1) = Q_i(n) + \lambda_i(n) - G_i^{\max}. \tag{4.14b}$$

The discrete linear system is described by

$$X_i(n+1) = A_7 X_i(n) + B_7 U_i(n), \tag{4.14}$$

where $U_i(n)=\lambda_i(n)-G_i^{\max}$, $A_7=\begin{bmatrix}0 & 1\\ 0 & 1\end{bmatrix}$, and $B_7=\begin{bmatrix}0\\ 1\end{bmatrix}$.

Theorem 4.7 The system described by Eq. (4.14) is controllable.

Proof: The system is controllable if and only if the controllability matrix U is full row rank. Note $U=[B_7 \quad A_7B_7]=\begin{bmatrix}0 & 1\\ 1 & 1\end{bmatrix}$, and the controllability matrix U is full row rank. Hence, the system is controllable.

Scenario 4: $R_i(n+1) \geq G_i^{\max}, G_i(n+1) \geq Q_i(n) + \lambda_i(n)$

Similar to scenario 2, the LSi sends the data up to the sum of the residual data and the arrived data, and reports zero queue length to the RS at the end of time interval n+1. We will then have

$$Q_i(n+1) = 0. \tag{4.15a}$$

Since the report queue length is zero and IBA does not estimate the arriving data, we have,

$$R_i(n+1) = 0, \tag{4.15b}$$

This is the state space

$$X_i(n+1) = A_8 X_i(n) + B_8 U_i(n), \tag{4.15}$$

where $U_i(n) = \lambda_i(n)$, $A_8 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$, and $B_8 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. Obviously, the state vector matrix and input vector matrix under scenario 4 is the same as the one in scenario 2. Since we have proven that the system under scenario 2 is not controllable, the following theorem holds,

Theorem 4.8 The system described by Eq. (4.15) is uncontrollable.

4.2.3 FBA Controllability Analysis

In FBA, the granted timeslot length to LS$_i$ is fixed, and no dynamic bandwidth negotiation is supported between the RS and LSs. Hence, the state space for the system becomes

$$R_i(n+1) = R_{fix}, \tag{4.16a}$$

where R_{fix} is the fixed resource request and RS assigns such request to LS_i accordingly. Moreover, the reported resource requirement is zero in FBA, and we have

$$Q_i(n+1) = 0 \quad . \tag{4.16b}$$

Eq.(4.16a) and (4.16b) are discrete linear system, represented by

$$X_i(n+1) = A_7 X_i(n) + B_7 U_i(n), \tag{4.16}$$

where $A_9 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, B_9 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, and $U_i(n) = R_{fix}$.

Theorem 4.9 The FBA system is uncontrollable.

Proof: The controllability matrix of the SBA system denoted by Eq. (4.16) is $U=[B_9 \quad A_9B_9]=\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$. Since U is not full row rank, this system is uncontrollable. That is, the FBA system is always uncontrollable.

The comparison on system controllability of different schemes is summarized in Table 4.1.

Table 4. 1 Controllability Comparison.

	PBA	IBA	FBA
Scenario 1	controllable	Controllable	uncontrollable
Scenario 2	controllable	Uncontrollable	uncontrollable
Scenario 3	controllable	Controllable	uncontrollable
Scenario 4	controllable	Uncontrollable	uncontrollable

It is not surprised to see that FBA is completely uncontrollable because it blindly allocates the shared resource by ignoring the traffic dynamics. Its

uncontrollability also explains the reason, from the system's point of view, that FBA generates the lowest resource utilization.

The partial controllability implies than IBA is unable to handle some circumstances over a P2MP system. For example, when one LS consistently requests more bandwidth than it actually needs while the other LSs present their real transmission needs, IBA will grant excessive bandwidth to the malicious request. As a result, the shared upstream resource will be wasted by idle timeslots, and under-utilization as well as unfairness occur under the above scenario. Theorems 4.6 and 4.8 reveal, from the system's point of view, the reason IBA leading to the above problems is IBA's uncontrollability under the scenarios 2 and 4, in which the system state variables matrix A and system input matrix B are both zero matrix. Consequently, the system has no way to capture its current state information and input information. As a result, IBA is unable to tune the upstream resource allocation into the optimum state within a prescribed time window. The unfairness and under-utilization problems for the P2MP system with IBA thus occur.

On the other side, PBA is able to manage the shared resource effectively. PBA monitors the actually arrived data by checking $Q_i(n)$ and $\lambda_i(n)$, and the resource arbitration decision is determined by both the queue length and the estimation. In this sense, any malicious LS cannot idly and constantly occupy the upstream channel, and the allocated resource to each LS follows its real traffic load. Theorems 4.1~4.4 reveal the P2MP system with PBA is completely controllable

under all circumstances, implying that PBA is able to tune the upstream resource allocation into the optimum state with high resource utilization and fairness.

4.3 System Stability Analysis

In this section, we further examine the stability characteristics of resource allocation schemes in P2MP architecture and reveal the proper controller that can drive the P2MP system into the stable state. By "stable", we mean that, when the input traffic load changes dramatically, the resource allocation scheme is able to provide the upstream resource fair share among the LSs with optimal bandwidth utilization [42]. For any resource allocation scheme, the stability design is critical because it provides predictability for system behavior and guarantees any generated oscillations to be bounded within a certain range. On the other hand, instability usually leads to unbounded oscillations which lower the overall network efficiency.

The open plant denoted by Eq. (4.5) usually implies an unbounded output. For any resource allocation scheme that is controllable, there always exists a controller,

$$U_i(n) = -K_i X_i(n) + F_i R_i(n), \tag{4.17a}$$

to drive the system into the stable state, which is known as *pole placement* [37]. K_i is a constant matrix, F_i is an pre-defined matrix, and $r_i(n)$ is a reference vector.

Substituting (4.17a) into (4.5) yields,

$$X_i(n+1) = (A_i - B_i K_i) X_i(n) + F r_i(n). \tag{4.17b}$$

Therefore, by implementing the controller of Eq. (4.17a), (A,B) is changed into $(A-BK,F)$. The controllability of a resource allocation scheme is unaltered by state feedback [37]. That is, if (A_i,B_i) is controllable (or uncontrollable), so is $(A_i-B_iK_i,F_i)$ for any K_i. Since F_i and $r_i(n)$ have no impact on the system's stability [36], in this section, we will focus on K_i which dominates the system stability. Assuming the reference the vector $r_i(n)=0$, we have

$$U_i(n)=-K_iX_i(n) \tag{4.18a}$$

and,

$$X_i(n+1)=(A_i-B_iK_i)X_i(n) \tag{4.18b}$$

Hence, after implementing the controller of Eq. (4.18a), the system becomes a close-loop form of Eq. (4.18b).

The stability discussion is meaningful only if the system is controllable. In the previous sections, we have proven that PBA is completely controllable under the four scenarios; IBA is controllable under scenario 1 and 3, and not controllable under scenario 2 and 4; Finally, FBA is not controllable under all four scenarios. As a result, our discussion of the stability for this section is applicable to PBA under all four scenarios and IBA under scenario 1 and 3. In the following, we will further investigate the conditions of K_i to ensure stability of the P2MP system.

4.3.1 PBA Stability Analysis

A. Scenario 1: $R_i(n+1)<G_i^{\max},G_i(n+1)<Q_i(n)+\lambda_i(n)$

In the previous section, we have developed the state space equation for PBA under scenario 1 as

$$X_i(n+1) = A_1 X_i(n) + B_1 U_i(n) \tag{4.18}$$

with $A_1 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, B_1 = \begin{bmatrix} 0 & \alpha_i \\ 1 & -\alpha_i \end{bmatrix}$, and $U_i(n) = \begin{bmatrix} \lambda_i(n) \\ \lambda_i(n-1) \end{bmatrix}$.

Theorem 4.10 In scenario 1, a P2MP system with PBA is stable when implementing the controller

$$U_i(n) = -K_1 X_i(n), \tag{4.19}$$

where $K_1 = \begin{bmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{bmatrix}$ with

$$\begin{cases} \alpha_i k_{12}k_{21} - \alpha_i k_{11}k_{22} - \alpha_i k_{22} + k_{11} + k_{12} + 1 > 0 \\ \alpha_i k_{12}k_{21} - \alpha_i k_{11}k_{22} - 2\alpha_i k_{21} + \alpha_i k_{22} + k_{11} - k_{12} + 1 > 0 \\ -1 < \alpha_i k_{12}k_{21} - \alpha_i k_{11}k_{22} + k_{11} - \alpha_i k_{21} < 1 \end{cases} \tag{4.20}$$

Proof: When implementing controller $U_i(n)$, the system becomes $X_i(n+1) = (A_1 - B_1 K_1) X_i(n)$. This discrete system is stable *iff* eigenvalues of the state matrix $(A_1 - B_1 K_1)$ fall inside the unit circle [43].

Assume $D(z) = \det[zI - (A_1 - B_1 K_1)]$, we will have

$$\begin{aligned} D(z) &= \begin{vmatrix} z + \alpha_i k_{21} & \alpha_i k_{22} - 1 \\ k_{11} - \alpha_i k_{21} & z - \alpha_i k_{22} + k_{12} \end{vmatrix} \\ &= z^2 + (\alpha_i k_{12} + k_{12} - \alpha_i k_{22}) z + (\alpha_i k_{12}k_{21} - \alpha_i k_{11}k_{22} - \alpha_i k_{21} + k_{11}) \end{aligned} \tag{4.20a}$$

Assume $L = \alpha_i k_{12} + k_{12} - \alpha_i k_{22}$, and $M = \alpha_i k_{12}k_{21} - \alpha_i k_{11}k_{22} + k_{11} - \alpha_i k_{21}$, by applying the Jury's criterion [43], this second order system is stable *iff* the following rules are all fulfilled:

$$\text{Rule 1: } D(1) = 1 + L + M > 0, \text{ i.e., } M + L + 1 > 0 \tag{4.20b}$$

$$\text{Rule 2: } (-1)^2 D(-1) = 1 - L + M > 0, \text{ i.e., } M - L + 1 > 0 \tag{4.20c}$$

Rule 3: $|M|<1$, i.e., $-1<M<1$ (4.20d)

From rules 1~ 3, the necessary and sufficient conditions for Eq.(4.13) to be stable are

$$\begin{cases} M+L+1>0 \\ M-L+1>0 \\ -1<M<1 \end{cases}, \text{ i.e.,}$$

$$\begin{cases} \alpha_i k_{12}k_{21}-\alpha_i k_{11}k_{22}-\alpha_i k_{22}+k_{11}+k_{12}+1>0 \\ \alpha_i k_{12}k_{21}-\alpha_i k_{11}k_{22}-2\alpha_i k_{21}+\alpha_i k_{22}+k_{11}-k_{12}+1>0 \\ -1<\alpha_i k_{12}k_{21}-\alpha_i k_{11}k_{22}-\alpha_i k_{21}+k_{11}<1 \end{cases} \quad (4.20)$$

B. Scenario 2: $R_i(n+1)<G_i^{\max}, G_i(n+1)\geq Q_i(n)+\lambda_i(n)$

The state space equation under this scenario has been developed as follows,

$$X_i(n+1)=A_2X_i(n)+B_2U_i(n), \quad (4.9)$$

where $U_i(n)=[\lambda_i(n)\ \ \lambda_i(n-1)]^T$, $A_2=\begin{bmatrix}0 & 0\\ 0 & 1\end{bmatrix}$, and $B_2=\begin{bmatrix}0 & \alpha_i\\ 1 & -\alpha_i\end{bmatrix}$.

Theorem 4.11 In scenario 2, a P2MP system with PBA is stable by implementing the controller

$$U_i(n)=-K_2X_i(n), \quad (4.21)$$

where $K_2=\begin{bmatrix}p_{11} & p_{12}\\ p_{21} & p_{22}\end{bmatrix}$, with

$$\begin{cases} \alpha_i p_{12}p_{21}-\alpha_i p_{11}p_{22}-\alpha_i p_{22}+p_{12}>0 \\ \alpha_i p_{12}p_{21}-\alpha_i p_{11}p_{22}-2\alpha_i p_{21}+\alpha_i p_{22}-p_{12}+2>0 \\ -1<\alpha_i p_{12}p_{21}-\alpha_i p_{11}p_{22}-\alpha_i p_{21}<1 \end{cases}. \quad \textbf{(4.22)}$$

Proof: When implementing controller $U_i(n)$, the system becomes $X_i(n+1)=(A_2-B_2K_2)X_i(n)$. This discrete system is stable *iff* eigenvalues of the state matrix $(A_2-B_2K_2)$ fall inside the unit circle. Assume $D(z)=\det[zI-(A_2-B_2K_2)]$, we will have

$$D(z)=z^2+(\alpha_i p_{21}-\alpha_i p_{22}+p_{12}-1)z+(\alpha_i p_{12}p_{21}-\alpha_i p_{21}-\alpha_i p_{11}p_{22}) \quad (4.22a)$$

Similarly, by applying the Jury's criterion for the second order discrete system, this system is stable *iff* rule 1~ rule 3 specified by Eqs. (4.21b)~(4.21d) are all fulfilled, i.e.,

$$\begin{cases} \alpha_i p_{12} p_{21} - \alpha_i p_{11} p_{22} - \alpha_i p_{22} + p_{12} > 0 \\ \alpha_i p_{12} p_{21} - \alpha_i p_{11} p_{22} - 2\alpha_i p_{21} + \alpha_i p_{22} - p_{12} + 2 > 0 \\ -1 < \alpha_i p_{12} p_{21} - \alpha_i p_{11} p_{22} - \alpha_i p_{21} < 1 \end{cases} \tag{4.22}$$

C. Scenario 3: $R_i(n+1) \geq G_i^{\max}, G_i(n+1) < Q_i(n) + \lambda_i(n)$

Under this scenario, the state space equations for PBA are

$$X_i(n+1) = A_3 X_i(n) + B_3 U_i(n), \tag{4.10}$$

where $u_i(n) = \left[\lambda_i(n) - G_i^{\max} \quad \lambda_i(n-1)\right]^T$, $A_3 = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}$, and $B_3 = \begin{bmatrix} 0 & \alpha_i \\ 1 & 0 \end{bmatrix}$.

Theorem 4.12 In scenario 3, a P2MP system with PBA is stable by implementing the controller

$$U_i(n) = -K_3 X_i(n), \tag{4.23}$$

where $K_3 = \begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix}$, with

$$\begin{cases} \alpha_i q_{12} q_{21} - \alpha_i q_{11} q_{22} + q_{11} + q_{12} > 0 \\ \alpha_i q_{12} q_{21} - \alpha_i q_{11} q_{22} - 2\alpha_i q_{21} + q_{11} - q_{12} + 2 > 0 \\ -1 < \alpha_i q_{12} q_{21} - \alpha_i q_{11} q_{22} - \alpha_i q_{21} + q_{11} < 1 \end{cases} \tag{4.24}$$

Proof: When implementing controller $U_i(n)$, the system becomes $X_i(n+1) = (A_3 - B_3 K_3) X_i(n)$. This discrete system is stable *iff* eigenvalues of the state matrix $(A_3 - B_3 K_3)$ fall inside the unit circle. Assume $D(z) = \det[zI - (A_3 - B_3 K_3)]$, we have

$$D(z) = z^2 + (\alpha_i q_{21} + q_{12} - 1)z + (\alpha_i q_{12} q_{21} - \alpha_i q_{11} q_{22} - \alpha_i q_{21} + q_{11}) \tag{4.24a}$$

Similarly, by applying the Jury's criterion for the second order discrete system, this system is stable *iff* rule 1~ rule 3 specified by Eqs. (4.21b)~(4.21d) are all fulfilled, i.e.,

$$\begin{cases} \alpha_i q_{12}q_{21} - \alpha_i q_{11}q_{22} + q_{11} + q_{12} > 0 \\ \alpha_i q_{12}q_{21} - \alpha_i q_{11}q_{22} - 2\alpha_i q_{21} + q_{11} - q_{12} + 2 > 0 \\ -1 < \alpha_i q_{12}q_{21} - \alpha_i q_{11}q_{22} - \alpha_i q_{21} + q_{11} < 1 \end{cases} \tag{4.24}$$

D. Scenario 4: $R_i(n+1) \geq G_i^{\max}, G_i(n+1) > Q_i(n) + \lambda_i(n)$

Under this scenario, the state space equations for PBA are

$$X_i(n+1) = A_4 X_i(n) + B_4 U_i(n), \tag{4.11}$$

where $u_i(n)=\left[\lambda_i(n)-G_i^{\max} \quad \lambda_i(n-1)\right]^T$, $A_4 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$, and $B_4 = \begin{bmatrix} 0 & \alpha_i \\ 1 & 0 \end{bmatrix}$.

Theorem 4.13 In scenario 4, a P2MP system with PBA is stable by implementing the controller

$$U_i(n) = -K_4 X_i(n), \tag{4.25}$$

where $K_4 = \begin{bmatrix} l_{11} & l_{12} \\ l_{21} & l_{22} \end{bmatrix}$, with

$$\begin{cases} l_{12} + \alpha_i l_{21} - \alpha_i l_{11} l_{22} > 0 \\ l_{12} + \alpha_i l_{21} - \alpha_i l_{11} l_{22} - 2 > 0 \\ -1 < \alpha_i l_{11} l_{22} < 1 \end{cases} \tag{4.26}$$

Proof: When implementing controller $U_i(n)$, the system becomes $X_i(n+1) = (A_4 - B_4 K_4) X_i(n)$. This discrete system is stable *iff* eigenvalues of the state matrix $(A_4 - B_4 K_4)$ fall inside the unit circle. Assume $D(z) = \det[zI - (A_4 - B_4 K_4)]$, we have

$$D(z) = z^2 + (l_{12} + \alpha_i l_{21} - 1)z - \alpha_i l_{11} l_{22} \tag{4.26a}$$

. Similarly, by applying the Jury's criterion for the second order discrete system, this system is stable *iff* rule 1~ rule 3 specified by Eqs. (4.21b)~(4.21d) are all fulfilled, i.e.,

$$\begin{cases} l_{12} + \alpha_i l_{21} - \alpha_i l_{11} l_{22} > 0 \\ l_{12} + \alpha_i l_{21} - \alpha_i l_{11} l_{22} - 2 > 0 \\ -1 < \alpha_i l_{11} l_{22} < 1 \end{cases} \tag{4.26}$$

4.3.2 IBA Stability Analysis

The IBA scheme is controllable only under scenarios 1 and 3, and therefore, the stability analysis is meaningful only under the above two scenarios. In this section, we will continue the stability analysis for IBA.

A. Scenario 1: $R_i(n+1) < G_i^{\max}, G_i(n+1) < Q_i(n) + \lambda_i(n)$

As discussed in the previous section, the state space equations under scenario 1 for IBA is

$$X_i(n+1) = A_5 X_i(n) + B_5 U_i(n), . \tag{4.12}$$

where $A_4 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, B_4 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, and $U_i(n) = \lambda_i(n)$.

Theorem 4.14 In scenario 1, a P2MP system with IBA is stable when implementing the controller

$$U_i(n) = -K_5 X_i(n), \tag{4.27}$$

where $K_5 = [h_1 \;\; h_2]$ with

$$\begin{cases} -1 < h_1 < 1 \\ -h_1 - 1 < h_2 < h_1 + 1 \end{cases} \tag{4.28}$$

Proof: When implementing controller $U_i(n)$, the system becomes $X_i(n+1) = (A_5 - B_5 K_5) X_i(n)$. This discrete system is stable *iff* eigenvalues of the state matrix $(A_5 - B_5 K_5)$ fall inside the unit circle [43].

Assuming $D(z) = \det[zI - (A_5 - B_5 K_5)]$, we have

$$D(z) = \begin{vmatrix} z & -1 \\ h_1 & z + h_2 \end{vmatrix} = z^2 + h_2 z + h_1 \tag{4.28a}$$

According to the Jury's criterion [43], this second order system is stable *iff* the following rules are all fulfilled:

$$\text{Rule 1: } D(1) = 1 + h_2 + h_1 > 0, \tag{4.28b}$$

$$\text{Rule 2: } (-1)^2 D(-1) = 1 - h_2 + h_1 > 0, \tag{4.28c}$$

$$\text{Rule 3: } |h_1| < 1, \text{ i.e., } -1 < h_1 < 1 \tag{4.28d}$$

From rules 1~ 3, the necessary and sufficient conditions for Eq.(4.13) to be stable are

$$\begin{cases} 1 + h_2 + h_1 > 0 > 0 \\ 1 - h_2 + h_1 > 0 \\ -1 < h_1 < 1 \end{cases} \quad \text{i.e.,} \quad \begin{cases} -1 < h_1 < 1 \\ -h_1 - 1 < h_2 < h_1 + 1 \end{cases} \tag{4.28}$$

B. Scenario 3: $R_i(n+1) \geq G_i^{\max}, G_i(n+1) < Q_i(n) + \lambda_i(n)$

Under this scenario, the state space equations for IBA are

$$X_i(n+1) = A_7 X_i(n) + B_7 U_i(n), \tag{4.14}$$

where $U_i(n) = \lambda_i(n) - G_i^{\max}$, $A_7 = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}$, and $B_7 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$.

Theorem 4.15 In scenario 1, a P2MP system with IBA is stable when implementing the controller

$$U_i(n) = -K_7 X_i(n), \tag{4.29}$$

where $K_4 = [j_1 \quad j_2]$ with

$$\begin{cases} -1 < j_1 < 1 \\ -j_1 < j_2 < j_1 + 2 \end{cases} \tag{4.30}$$

Proof: When implementing controller $U_i(n)$, the system becomes $X_i(n+1)=(A_7-B_7K_7)X_i(n)$. This discrete system is stable *iff* eigenvalues of the state matrix $(A_7-B_7K_7)$ fall inside the unit circle.

Assuming $D(z)=\det[zI-(A_7-B_7K_7)]$, we have

$$D(z)=\begin{vmatrix} z & -1 \\ j_1 & z+j_2-1 \end{vmatrix}=z^2+(j_2-1)z+j_1 \tag{4.30a}$$

According to the Jury's criterion, this second order system is stable *iff* the following rules are all fulfilled:

$$\text{Rule 1: } D(1)=1+j_2+j_1>0, \tag{4.30b}$$

$$\text{Rule 2: } (-1)^2D(-1)=1-j_2+1+j_1>0, \tag{4.30c}$$

$$\text{Rule 3: } |j_1|<1, \text{ i.e., } -1<j_1<1 \tag{4.30d}$$

From rules 1~ 3, the necessary and sufficient conditions for Eq.(4.15) to be stable are

$$\begin{cases} j_2+j_1>0 \\ 2-j_2+j_1>0 \\ -1<j_1<1 \end{cases} \quad \text{i.e.,} \quad \begin{cases} -1<j_1<1 \\ -j_1<j_2<j_1+2 \end{cases} \tag{4.30}$$

Consequently, Theorems 4.10~4.15 give the range of the feedback gain K that guarantees the system's stability such that the upstream resource of the applied P2MP system is fairly shared by LSs. After implementing the controller (i.e., Eq. (4.18a)), with the controller gains $K_i\big|_{i=1,2,3,4,5,6}$ in different scenarios, the system representing PBA and IBA becomes the close form, which was illustrated in Figure 4.2.

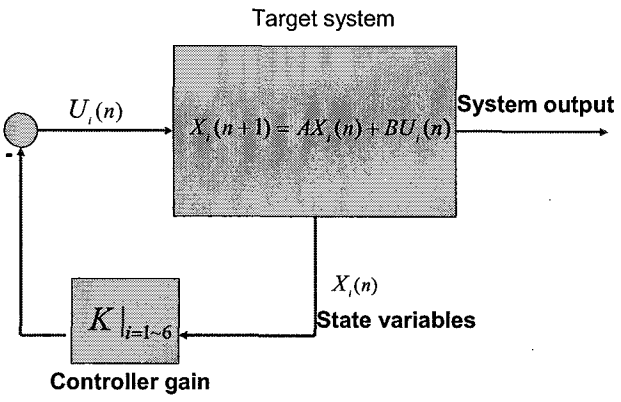


Figure 4.2 The close-loop form of resource allocation schemes of P2MP.

In Figure 4.2, the target system is achieved by feeding back proportional state variables to the control input. The state variables represent the on-line traffic dynamics, which imply changes of the queue length and bandwidth requirement of an LS. The controllers essentially feedbacks the traffic dynamics information, after multiplying the controller gain, to the input of system. By doing so, the eigenvalues of an open plant system, which is usually outside of the unite circle, are driven back into the inside of the unite circle after implementing the controller gains prescribed by Theorem 4.10~4.15. The system is hence driven into the stable state. A particular controller can be facilitated through the proper buffering and intra-LS scheduling schemes at the RS, or the appropriate inter-LS scheduling scheme among LSs. Thus, the RS works as a central controller to tune LSs accordingly, ensuring that the upstream resource of a P2MP system is fairly shared by multiple LSs. The controller gains $K_i|_{i=1,2,3,4,5,6}$ describe the controller characteristics in different scenarios.

For PBA, the RS manipulates the upstream transmission from multiple LSs by using different controllers $U_i(n) = -KX_i(n)$, where the controller $K = K_i|_{i=1\sim4}$. From Theorems 4.10~4.13, it is interesting to notice that the estimation index α_i may affect the system stability when designing a controller for a P2MP system with PBA. Theorems 4.10-4.13 further reveal the relationship between controller gain and estimation index with a set of inequalities groups. On the other hand, for IBA, Theorems 4.14~4.15 describe the range of each vector of the control gain matrix. In both cases, these theorems essentially give the guideline of the controller design to guarantee the system's stability.

4.4 Controller Design

In the previous sections, we discussed a unified system model for different resource allocation schemes (PBA, IBA, and FBA) of an S-PON system (more generically, a P2MP system), followed by the controllability and stability analysis. In this section, we will further discuss the controller design for the P2MP system based on the proposed model.

4.4.1 Design Objectives

The design objective is always the first step for the controller design. In our case, the design objective for resource allocation in an S-PON system (more generically, a P2MP system) is the system robustness, accuracy, and target transient performance.

In general, it is difficult to meet transient performance objectives such as the settling time and the overshoot because of the complexity of mapping these objectives into the corresponding scheduling algorithms and resource management schemes However, the proposed state space model gives a simple and straightforward framework to achieve the objectives by using the state space feedback control techniques.

Consider the measured system output $Y_i(n) = CX_i(n)$, and define matrix $C = [0 \quad 1]$. The system output is essentially the measurement of the report queue length $Q_i(n)$ .The state space system is then described by

$$\begin{aligned} X_i(n+1) &= AX_i(n) + BU_i(n) \\ Y_i(n) &= CX_i(n) \end{aligned} \tag{4.31}$$

Our target is to design a controller

$$U_i(n) = -K_i X_i(n) + F_i r \tag{4.32}$$

to achieve the design objectives of robustness, accuracy and target transient performances.

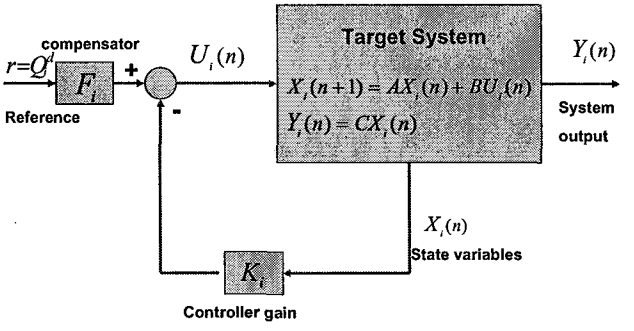


Figure 4.3 The controller design to meet transient performance objectives.

Figure 4.3 illustrates our approach by using the controller of Eq.(4.32) to achieve the prescribed objectives. The reference input r is the desired queue length, and thus $e(n) = Y_i(n) - r$ is the control error. The matrix F_i is a compensator to offset the control error, so that the system output can eventually converge to the input reference (i.e., $e(n) = 0$). Our focus is now on the design of a suitable controller gain K_i and compensator F_i.

4.4.2 System Robustness

P2MP resource allocation schemes need to achieve robustness performance regarding to the system dynamics. It implies that the system should be able to handle different conditions even when the online traffic changes dramatically. The system robustness essentially requires the system to be stable. In Sections 3, Theorems 4.10-4.15 prescribe the range of K_i to guarantee the system's stability. Similarly, these theorems also dictate the controller gain K_i in Eq.(4.32), as the added reference input r and pre-compensator F_i have no impact on the system's stability.

4.4.3 Accuracy

A control system is said to be "accurate" if the measured output converges or become sufficiently close to the reference input [44]. In a P2MP system, the reference input r can be chosen from various Service Level Agreement (SLA)

parameters, or other pre-defined parameters. The measured output $Y_i(n)$ is thus required to converge to r in order to ensure the control objectives are met. In this section, we choose the desired queue length Q_i^d of LS_i as the reference input, i.e., $r = Q_i^d$. The desired queue length Q_i^d is defined as the efficient queue length to achieve high network resource utilization. Theoretically, each LS needs to maintain a desired queue length Q_i^d to avoid overflow or emptiness [45],[46]. If the queue length is too large, data loss and retransmission are inevitable because of the limited available buffers; on the other hand, if the queued length becomes empty, it indicates that the allocated resource for this LS is always more than it actually needs. The network resource is thus wasted with low utilization. Both of the extremes should be avoided by maintaining a desired queue length Q_i^d.

The objective of system accuracy essentially requires that the output $Y_i(n)$, which is the measurement of the report queue length, converges to the system input r, the desired queue length. To reach this objective, a compensator F_i is implemented right after the reference, and is added up to the feedback from the state variable, to form the controller $U_i(n)$, which is illustrated in Figure 4.3. The key issue here is to design a compensator F_i in such way that it can offset the control error, i.e., $e(n) = 0$.

For a particular P2MP system *i*, the compensator F_i is determined by the state matrix A_i, the input matrix B_i, the output matrix C_i, and the controller gain

K_i. According to feedback control theory [44], the compensator F_i that drives the control error $e(n) = Y_i(n) - r$ to zero is given by

$$F_i = [K_i \quad 1] \begin{bmatrix} A_i - I & B_i \\ C_i & 0 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{4.33}$$

where $\begin{bmatrix} A_i - I & B_i \\ C_i & 0 \end{bmatrix}$ is a non-singular matrix. The derivation of F_i is detailed in Appendix A.

Consequently, by implementing the compensator of Eq. (4.34), the controller of Eq.(4.32) is able to force the system output $Y_i(n)$ to track the reference input r, implying that the queue length can be eventually driven into the desired queue length Q_i^d.

4.4.4 Target Transient Performances

In the control system, the settling time T_i and the maximum overshoot M_i are the two main parameters to prescribe the system's target transient performance. The settling time T_i is defined as the time for the P2MP system to reach the steady state. A short settling time is critical to achieve the performance objective, especially when the incoming traffics of LSs have large volatility. In such case, short settling time ensures the system convergence to the stable state before the traffic load changes. On the other side, the maximum overshoot M_i is defined as the difference between the maximum system output $y_{\max}$ and steady-state system output y_{ss} divided by the steady-state system output y_{ss}, i.e.,

$$M_i = \frac{y_{\max} - y_{ss}}{y_{ss}} \tag{4.34}$$

The maximum overshoot gives the upper bound for the output oscillations of a P2MP system.

The settling time T_i and the maximum overshoot M_i are the two critical parameters for the transient performance. For example, the specifications of a S-PON system may require the system to reach stable state within 10 seconds, and the overshoot should be less than 5%. From the control point of view, the settling time and maximum overshoot are determined by the closed loop poles [36]. Recall that controller gain K_i in Eq.(4.32) essentially determines the poles in the closed loop characteristic polynomial $\det[zI-(A_i-B_iK_i)]$, and thus we can achieve the target transient performance T_i and M_i by properly tuning the controller gain K_i.

Without losing generality, let the poles of a second order S-PON system be a pair of complex conjugates $re^{\pm j\theta}$. According to control theory [44], the relationship between the pole parameters r and θ, and the settling time T_i and the maximum overshoot M_i is given by

$$r \approx e^{-4/T_i} \tag{4.35a}$$

$$\theta \approx \pi \frac{\log r}{\log M_i} \tag{4.35b}$$

Considering the eigenvalues of the closed form characteristic polynomial $\det[zI-(A_i-B_iK_i)]$ is $re^{\pm j\theta}$, we have, $\det[zI-(A_i-B_iK_i)]=(z-re^{j\theta})(z-re^{-j\theta})$, i.e.,

$$\det[zI-(A_i-B_iK_i)]=z^2-2r\cos\theta z+r^2 \tag{4.36}$$

From Eq. (4.20a), (4.22a), (4.24a) and (4.26a), we know that the second order close-form characteristic polynomial for PBA is represented by,

$$\det[zI-(A_i-B_iK_i)]=z^2+f_1(k_{11},k_{12},k_{21},k_{22},\alpha_i)z+f_2(k_{11},k_{12},k_{21},k_{22},\alpha_i) \tag{4.37a}$$

where k_{11},k_{12},k_{21} and k_{22} are vectors of K_i, and α_i is the estimate index.

As Eq. (4.36) and Eq.(4.37a) represent the same closed form characteristic polynomial for PBA, they have the same coefficients for each order of z. Thus, we have,

$$\begin{cases} f_1(k_{11},k_{12},k_{21},k_{22},\alpha_i)=-2r\cos\theta \\ f_2(k_{11},k_{12},k_{21},k_{22},\alpha_i)=r^2 \end{cases}. \tag{4.37b}$$

By substituting Eq. (4.35a) and (4.35b) into Eq.(4.37b), we have,

$$\begin{cases} f_1(k_{11},k_{12},k_{21},k_{22},\alpha_i)=-2e^{-4/T_i}\cos(\pi\dfrac{\log r}{\log M_i}) \\ f_1(k_{11},k_{12},k_{21},k_{22},\alpha_i)=e^{-2/T_i} \end{cases} \tag{4.38}$$

Hence, Eq. (4.38) provides the range of each vector of the controller gain K_i to reach the target settling time and maximum overshoot. The solutions of Eq.(4.38) also illustrate the relationships between each vector of the control gain matrix and the estimate index. Although the exact value of each vector and estimate index is not given, Eq. (4.38) essentially provides the guideline to design a suitable controller gain K_i such that the target settling time T_i and maximum overshoot

M_i in PBA can be met. It is also interesting to see that the estimate index α_i has impact on achieving the required transient performance.

On the other hand, from Eq. (4.28a) and (4.30a), we know that the characteristic polynomial for IBA is represented by

$$\det[zI-(A_i-B_iK_i)]=z^2+f_1(k_1,k_2)z+f_2(k_1,k_2) \qquad (4.39a)$$

where k_1 and k_2 are vectors of K_i.

Similarly, since Eq. (4.36) and Eq.(4.39a) represent the same closed form characteristic polynomial for IBA, Eq.(4.36) and (4.39a) have the same coefficients for each order of z. Comparing the coefficients of Eq. (4.36) and (4.39a) yields,

$$\begin{cases} f_1(k_1,k_2)=-2r\cos\theta \\ f_2(k_1,k_2)=r^2 \end{cases} \qquad (4.39b)$$

Similarly, by substituting Eq. (4.35a) and (4.35b) into Eq. (4.39b), we have

$$\begin{cases} f_1(k_1,k_2)=-2e^{-4/T_i}\cos(\pi\dfrac{\log r}{\log M_i}) \\ f_2(k_1,k_2)=e^{-2/T_i} \end{cases} \qquad (4.40)$$

Hence, the solutions of Eq.(4.40) are essentially the guidelines to design a suitable controller gain K_i such that the target settling time T_i and maximum overshoot M_i in IBA can be met.

4.5 Chapter Summary

In this chapter, a unified state space model has been proposed for different bandwidth allocation schemes of S-PON (more generically, a P2MP network), namely, fixed bandwidth allocation (FBA), IPACT-based bandwidth allocation (IBA)

and predictor-based bandwidth allocation (PBA), respectively. By looking into the controllability of each scheme, we conclude that only PBA is completely controllable, and FBA and RBA are either completely uncontrollable or partially controllable. We have also discussed the controller design for PBA and RBA through "pole placement". Finally, we have provided the guideline of constructing the suitable compensator and controller gain to meet the prescribed objectives such as system robustness, accuracy, and system transient performance. The analysis result will be greatly helpful for TDMA scheme design for S-PON.

CHAPTER 5

NON-LINEAR PREDICTOR-BASED STATE SPACE MODEL

In the last chapter, a state space model was established for TDMA transmission of S-PON. Our analysis shows that the predict-based bandwidth allocation (PBA) has the most superior characteristics in terms of the controllability and stability, as compared to other schemes. For more accurate prediction, the non-linear predictor-based dynamic bandwidth allocation (NLPDBA) schemes [26],[41] is usually employed with traffic correlation to predict the incoming data in the next cycle. A non-linear index is employed to extract the time-dependent correlation among traffics in consecutive cycles.

In this chapter, we extend the investigation of NLPDBA from the vantage points of S-PON system characteristics. A well-designed bandwidth allocation scheme is expected to maintain the S-PON performance under dynamic traffic changes, and to guarantee the fair share of the available upstream bandwidth among multiple ONUs/SANs. In particular, we establish a state space model to evaluate the controllability of NLPDBA, illustrating that an S-PON system with NLPDBA is capable of properly steering the upstream bandwidth allocation among multiple ONUs/SANs. We will then discuss the NLPDBA stability over the S-PON system, on which stability analysis is conducted to reveal the requirements of maintaining system performance under dynamic S-PON input traffic. Based on the model, guidelines of the controller design are further established. These guidelines

essentially highlight the framework for designing NLPDBA schemes for S-PONs that ensures stability.

5.1 NLPDBA State Space Model

The upstream bandwidth allocation among multiple ONUs/SANs can be modeled as a point-to-multiple-point (P2MP) system, in which the upstream resource allocation is arbitrated by one master, i.e., the RS, over multiple clients, i.e., LSs. Let's assume one RS serves y LSs, and the RS serves each LS once in a service cycle (Figure 5.1).

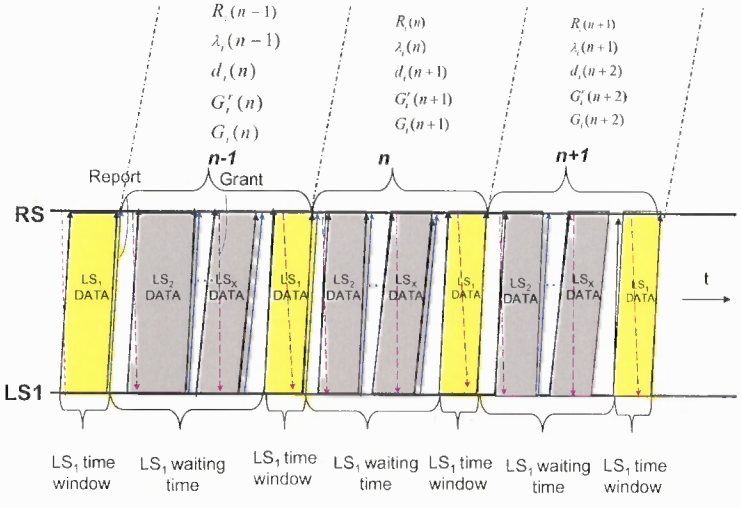


Figure 5.1 NLPDBA in P2MP networks.

With service cycle n defined as the time interval $[t_n, t_{n+1})$, where t_n stands for the time when cycle n starts, and t_{n+1} cycle n ends (or equivalently cycle $n+1$ begins), the following notations are adopted for our analysis.

$R_i(n)$: the reported queued length by the piggybacked REPORT message from LSi ($1 \le i \le y$) at the beginning of service cycle n;

$\lambda_i(n)$: the actually arrived data of LS*i* at the end of service cycle *n*;

$\hat{\lambda}_i(n)$: the predicted arrival data at LS*i* at the beginning of service cycle *n*;

$d_i(n+1)$: the departed data from LS*i* at the end of service cycle *n*;

$G_i(n+1)$: the allocated timeslot to LS*i* at the end of service cycle *n*;

$G_i^r(n+1)$: the bandwidth requirement of LS*i* at the end of service cycle *n* (it may or may not be the same as $R_i(n)$, depending on the particular bandwidth allocation scheme);

G_i^{max}: the maximum timeslot length prescribed by the service level agreement (SLA).

At the beginning of service cycle *n+1*, the queue length of LS*i* is the residual of data transmission, and it is described by

$$R_i(n+1) = R_i(n) + \lambda_i(n) - d_i(n+1). \tag{5.1}$$

A REPORT message is piggybacked at the beginning of service cycle *n*, indicating the awaiting data, which is the current queue length $R_i(n)$. After processing the request, the RS allocates timeslot $G_i(n)$ to LS*i*, and the departed data at the end of service cycle *n* is

$$d_i(n+1) = \min\{G_i(n+1), R_i(n) + \lambda_i(n)\}. \tag{5.2}$$

The granted timeslot is thus represented by the smaller value of the bandwidth requirement and the SLA upper bound, i.e.,

$$G_i(n+1) = \min\{G_i^r(n+1), G_i^{max}\}. \tag{5.3}$$

When traffic predictor is employed, $G_i^r(n+1)$ is determined by

$$G_i^r(n+1) = R_i(n) + \hat{\lambda}_i(n), \tag{5.4}$$

where $\hat{\lambda}_i(n)$ is the predicted arrival data at LS*i* in service cycle *n*.

NLPDRA works as follows: when a queue length report $R_i(n)$ is received, the RS updates the bandwidth requirement from LS*i*, i.e., $G_i^r(n)$, according to Eq. (5.4), and arbitrates the allocated timeslot according to Eq. (5.3). The traffic forecast by NLPDRA is made according to

$$\hat{\lambda}_i(n+1) = \alpha_i(n)\lambda_i(n), \tag{5.5}$$

where $\alpha_i(n)$ is the estimation credit, indicating the impact of the input history data on the output prediction, and $\alpha_i(n)$ is adjusted by least-mean squares (LMS) algorithm [41][47] as

$$\alpha_i(n+1) = \alpha_i(n) + \tau \times \frac{e_i(n)}{\lambda_i(n)}, \tag{5.6}$$

where τ is the step size and is defined as a positive real number, $e_i(n)$ is the prediction error, and

$$e_i(n) = \lambda_i(n) - \hat{\lambda}_i(n) \tag{5.7}$$

Assume $\alpha_i'(n) = \alpha_i(n-1)$ and $\lambda_i'(n) = \lambda_i(n-1)$. From Eqs. (5)-(7), we get

$$\alpha_i(n+1) = \alpha_i(n) + \tau - \tau \times \frac{\alpha_i'(n) \times \lambda_i'(n)}{\lambda_i(n)} \tag{5.8a}$$

Obviously, we know from $\alpha_i'(n) = \alpha_i(n-1)$ that

$$\alpha_i'(n+1) = \alpha_i(n) \tag{5.8b}$$

The NLPDBA scheme is thus represented by

$$x_i(n+1) = Ax_i(n) + Bu_i(n), \tag{5.8}$$

where $x_i(n) = [G_i^r(n) \quad R_i(n) \quad \alpha_i(n) \quad \alpha_i'(n)]^T$ is the S-PON system state vector, indicating the bandwidth requirement, the queue length of LS_i, and the prediction index. The input vector, $u_i(n) = [\lambda_i(n) \quad \lambda_i'(n)]$, represents the arrived data during the waiting time. The state space model represented by Eq. (5.8) provides a convenient and compact way to model and analyze the upstream resource allocation over SPON (more generally, P2MP) networks. In the next section, the controllability of the P2MP system with NLPDRA will be evaluated based on the model represented by Eq. (5.8).

5.2 Controllability of NLPDBA

By "controllability", we mean that the shared upstream bandwidth in a P2MP system can be arbitrated properly among multiple LSs, even when the loaded traffic changes drastically [38]. We expect that the employed bandwidth allocation scheme is capable of adjusting the bandwidth allocated to each LS in accordance with the traffic dynamics, and the arbitration decision is expected to be fair and efficient.

As formulated in Section 5.1, Eq. (5.8) describes the upstream bandwidth allocation over a P2MP system. One element in the state vector, $R_i(n)$, represents the queue length. From Eq. (1) we know that $R_i(n)$ is determined by $d_i(n)$. Combining Eq. (5.2) and (5.3), we get

$$d_i(n+1) = \min\{G_i^r(n+1), R_i(n) + \lambda_i(n), G_i^{\max}\}. \tag{5.9}$$

Depending on the loaded traffic from end users, the P2MP system falls in one of the following three scenarios,

A) $d_i(n+1) = G_i^r(n+1)$,

B) $d_i(n+1) = R_i(n) + \lambda_i(n)$, and

C) $d_i(n+1) = G_i^{\max}$.

A). Scenario 1: $d_i(n) = G_i^r(n+1)$

In this scenario, the bandwidth requirement is no larger than the SLA specification, and the granted timeslot is no larger than the total data arrived at LS_i during a service cycle. This is the case when the end users are well behaved under the guidance of the SLA specification. We have $G_i(n+1) = G_i^r(n+1)$, and $d_i(n) = G_i^r(n)$. The state space is thus

$$G_i^r(n+1) = R_i(n) + \hat{\lambda}_i(n) = R_i(n) + \alpha_i'(n)\lambda_i'(n) \tag{5.10a}$$

$$R_i(n+1) = R_i(n) + \lambda_i(n) - G_i^r(n+1). \tag{5.10b}$$

Substituting Eq. (5.10a) into (5.10b), we obtain,

$$R_i(n+1) = \lambda_i(n) - \alpha_i'(n)\lambda_i'(n) \tag{5.10c}$$

Note that Eqs. (5.10a), (5.10c), (5.8a), and (5.8b) describe a non-linear discrete system, and linearization is necessary to analyze the controllability [36],[37].

Assume the equilibrium point is $(G_{i0}^r, R_{i0}, \alpha_{i0}, \alpha_{i0}', \lambda_{i0}, \lambda_{i0}')$, all of which are positive real numbers; linearizing Eqs. (5.10a), (5.10c), (5.8a) and (5.8b) about the

equilibrium point (see Appendix B for details), we obtain the following linearized system

$$\delta G_i^r(n+1) = \delta R_i(n) + \lambda'_{i0}\delta\alpha'_i(n) + \alpha'_{i0}\delta\lambda'_i(n) \tag{5.11a}$$

$$\delta R_i(n+1) = \delta\lambda_i(n) - \lambda'_{i0}\delta\alpha'_i(n) - \alpha'_{i0}\delta\lambda'_i(n) \tag{5.11b}$$

$$\delta\alpha_i(n+1) = \delta\alpha_i(n) - \frac{\tau\lambda'_{i0}}{\lambda_{i0}}\delta\alpha'_i(n) + \frac{\tau\alpha'_{i0}\lambda'_{i0}}{\lambda_{i0}^2}\delta\lambda_i(n) - \frac{\tau\alpha'_{i0}}{\lambda_{i0}}\delta\lambda'_i(n) \tag{5.11c}$$

$$\delta\alpha'_i(n+1) = \delta\alpha_i(n) \quad , \tag{5.11d}$$

Eqs. (11a)~(11d) can be further represented by

$$\delta x_i(n+1) = A_1\delta x_i(n) + B_1\delta u_i(n) \quad , \tag{5.11}$$

where

$$A_1 = \begin{bmatrix} 0 & 1 & 0 & \lambda'_{i0} \\ 0 & 0 & 0 & -\lambda'_{i0} \\ 0 & 0 & 1 & -\dfrac{\tau\lambda'_{i0}}{\lambda_{i0}} \\ 0 & 0 & 1 & 0 \end{bmatrix}, B_1 = \begin{bmatrix} 0 & \alpha'_{i0} \\ 1 & -\alpha'_{i0} \\ \dfrac{\tau\alpha'_{i0}\lambda'_{i0}}{\lambda_{i0}^2} & -\dfrac{\tau\alpha'_{i0}}{\lambda_{i0}} \\ 0 & 0 \end{bmatrix}.$$

Theorem 5.1 A P2MP system with NLPDBA is controllable when $d_i(n+1) = G_i^r(n+1)$.

Proof: A system described by Eq. (5.11) is controllable *iff* the $n\times nr$ controllability matrix $U = \begin{bmatrix} B & AB & \ldots & A^{n-1}B \end{bmatrix}$ is full row rank.

In the above scenario, $U = \begin{bmatrix} B_1 & A_1B_1 \end{bmatrix} = \begin{bmatrix} 0 & \alpha'_{i0} & 1 & -\alpha'_{i0} \\ 1 & -\alpha'_{i0} & 0 & 0 \\ \dfrac{\tau\alpha'_{i0}\lambda'_{i0}}{\lambda_{i0}^2} & -\dfrac{\tau\alpha'_{i0}}{\lambda_{i0}} & \dfrac{\tau\alpha'_{i0}\lambda'_{i0}}{\lambda_{i0}^2} & -\dfrac{\tau\alpha'_{i0}}{\lambda_{i0}} \\ 0 & 0 & \dfrac{\tau\alpha'_{i0}\lambda'_{i0}}{\lambda_{i0}^2} & -\dfrac{\tau\alpha'_{i0}}{\lambda_{i0}} \end{bmatrix}$. Notice that U is a square matrix, it is full rank when $|U| \neq 0$.

Since $|U| = \begin{vmatrix} 0 & \alpha'_{i0} & 1 & -\alpha'_{i0} \\ 1 & -\alpha'_{i0} & 0 & 0 \\ \frac{\tau\alpha'_{i0}\lambda'_{i0}}{\lambda^2_{i0}} & -\frac{\tau\alpha'_{i0}}{\lambda_{i0}} & \frac{\tau\alpha'_{i0}\lambda'_{i0}}{\lambda^2_{i0}} & -\frac{\tau\alpha'_{i0}}{\lambda_{i0}} \\ 0 & 0 & \frac{\tau\alpha'_{i0}\lambda'_{i0}}{\lambda^2_{i0}} & -\frac{\tau\alpha'_{i0}}{\lambda_{i0}} \end{vmatrix} = \frac{\tau^2\alpha'^2_{i0}(\lambda_{i0} - \alpha'_{i0}\lambda'_{i0})^2}{\lambda^4_{i0}}$, and τ, α'_{i0}, λ'_{i0} and λ_{i0} are defined as positive real numbers, $|U| \neq 0$ *iff* $\alpha'_{i0}\lambda'_{i0} \neq \lambda_{i0}$ holds. Therefore, when $d_i(n+1) = G^r_i(n+1)$, the P2MP system represented by Eq. (5.11) is controllable.

B) Scenario 2: $d_i(n+1) = R_i(n) + \lambda_i(n)$

In this scenario, the granted timeslot is larger than the bandwidth requirement, i.e., $G_i(n+1) > R_i(n) + \lambda_i(n)$. This "over-grant" is adjusted by reporting the difference between the granted timeslot and bandwidth requirement. To facilitate this mechanism, we use "negative" queue length to measure the "over-grant". Hence, we have

$$R_i(n+1) = R_i(n) + \lambda_i(n) - G^r_i(n+1). \tag{5.12a}$$

Note that the "negative" REPORT indicates that the OLT over-grants timeslots to LS_i. The pre-reserved network resource for LS_i is able to deliver all incoming data, and the queue of LS_i is empty after the current service cycle. In this scenario, the bandwidth requirement only contains the estimated arrival data, i.e.,

$$G^r_i(n+1) = \alpha'_i(n)\lambda'_i(n). \tag{5.12b}$$

From Eqs. (12a) and (12b), we obtain

$$R_i(n+1) = R_i(n) + \lambda_i(n) - \alpha'_i(n)\lambda'_i(n). \tag{5.12c}$$

Following the similar linearization procedure in Appendix C, the state space can be linearized to

$$\delta G_i^r(n+1) = \lambda_{i0}'\delta\alpha_i'(n) + \alpha_{i0}'\delta\lambda_i'(n), \tag{5.13a}$$

$$\delta R_i(n+1) = \delta R_i(n) + \delta\lambda_i(n) - \lambda_{i0}'\delta\alpha_i'(n) - \alpha_{i0}'\delta\lambda_i'(n). \tag{5.13b}$$

Eqs. (5.13a), (5.13b), (5.11c), and (5.11d) are the state space which can be represented by

$$\delta x_i(n+1) = A_2\delta x_i(n) + B_2\delta u_i(n), \tag{5.13}$$

where

$$A_2 = \begin{bmatrix} 0 & 0 & 0 & \lambda_{i0}' \\ 0 & 1 & 0 & -\lambda_{i0}' \\ 0 & 0 & 1 & -\dfrac{\tau\lambda_{i0}'}{\lambda_{i0}} \\ 0 & 0 & 1 & 0 \end{bmatrix}, B_2 = \begin{bmatrix} 0 & \alpha_{i0}' \\ 1 & -\alpha_{i0}' \\ \dfrac{\tau\alpha_{i0}'\lambda_{i0}'}{\lambda_{i0}^2} & \dfrac{-\tau\alpha_{i0}'}{\lambda_{i0}} \\ 0 & 0 \end{bmatrix}.$$

Theorem 5.2 A P2MP system with NLPDBA is controllable when $d_i(n+1) = R_i(n) + \lambda_i(n)$.

Proof: Similarly, we analyze the controllability by evaluating matrix U in this scenario, where

$$U = \begin{bmatrix} B_2 & A_2B_2 \end{bmatrix} = \begin{bmatrix} 0 & \alpha_{i0}' & 0 & 0 \\ 1 & -\alpha_{i0}' & 1 & -\alpha_{i0}' \\ \dfrac{\tau\alpha_{i0}'\lambda_{i0}'}{\lambda_{i0}^2} & -\dfrac{\tau\alpha_{i0}'}{\lambda_{i0}} & \dfrac{\tau\alpha_{i0}'\lambda_{i0}'}{\lambda_{i0}^2} & -\dfrac{\tau\alpha_{i0}'}{\lambda_{i0}} \\ 0 & 0 & \dfrac{\tau\alpha_{i0}'\lambda_{i0}'}{\lambda^2{}_{i0}} & -\dfrac{\tau\alpha_{i0}'}{\lambda_{i0}} \end{bmatrix}.$$

Furthermore, we find that

$$|U| = \begin{vmatrix} 0 & \alpha_{i0}' & 0 & 0 \\ 1 & -\alpha_{i0}' & 1 & -\alpha_{i0}' \\ \dfrac{\tau\alpha_{i0}'\lambda_{i0}'}{\lambda_{i0}^2} & -\dfrac{\tau\alpha_{i0}'}{\lambda_{i0}} & \dfrac{\tau\alpha_{i0}'\lambda_{i0}'}{\lambda_{i0}^2} & -\dfrac{\tau\alpha_{i0}'}{\lambda_{i0}} \\ 0 & 0 & \dfrac{\tau\alpha_{i0}'\lambda_{i0}'}{\lambda^2{}_{i0}} & -\dfrac{\tau\alpha_{i0}'}{\lambda_{i0}} \end{vmatrix} = \frac{\tau^2\alpha_{i0}'^3\lambda_{i0}'(\alpha_{i0}'\lambda_{i0}' - \lambda_{i0})}{\lambda_{i0}^4}.$$

Since $\tau, \alpha'_{i0}, \lambda'_{i0}$ and λ_{i0} are defined as positive real numbers, $|U| \neq 0$ *iff* $\alpha'_{i0}\lambda'_{i0} \neq \lambda_{i0}$ holds. Therefore, when $d_i(n) = R_i(n-1) + \lambda_i(n-1)$, the P2MP system represented by Eq. (5.13) is controllable.

Scenario 3: $d_i(n+1) = G_i^{\max}$

In this scenario, the incoming traffic is heavy, and the OLT uses the SLA upper bound $G_i^{\max}$ to limit the aggressive bandwidth requirement. The state space of this scenario turns into

$$G_i^r(n+1) = R_i(n) + \alpha'_i(n)\lambda'_i(n), \tag{5.14a}$$

$$R_i(n+1) = R_i(n) - G_i^{\max} + \lambda_i(n). \tag{5.14b}$$

Following the similar linearization procedure in Appendix D, the above equations can be linearized to

$$\delta G_i^r(n+1) = \delta R_i(n) + \lambda'_{i0}\delta\alpha'_i(n) + \alpha'_{i0}\delta\lambda'_i(n), \tag{5.15a}$$

$$\delta R_i(n+1) = \delta R_i(n) + \delta\lambda_i(n). \tag{5.15b}$$

Eqs. (5.15a), (5.15b), (5.11c), and (5.11d) are essentially the state space represented as

$$\delta x_i(n+1) = A_3\delta x_i(n) + B_3\delta u_i(n) \tag{5.15}$$

where

$$A_3 = \begin{bmatrix} 0 & 1 & 0 & \lambda'_{i0} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -\dfrac{\tau\lambda'_{i0}}{\lambda_{i0}} \\ 0 & 0 & 1 & 0 \end{bmatrix}, B_3 = \begin{bmatrix} 0 & \alpha'_{i0} \\ 1 & 0 \\ \dfrac{\tau\alpha'_{i0}\lambda'_{i0}}{\lambda_{i0}^2} & -\dfrac{\tau\alpha'_{i0}}{\lambda_{i0}} \\ 0 & 0 \end{bmatrix}.$$

Similar to the previous two scenarios, the following theorem is established.

Theorem 5.3 A P2MP system with NLPDBA is controllable when $d_i(n+1)= G_i^{\max}$.

Proof: Similarly, we check the controllability matrix U, which is

$$U=\begin{bmatrix} B_3 & A_3B_3 \end{bmatrix}=\begin{bmatrix} 0 & \alpha'_{i0} & 1 & 0 \\ 1 & 0 & 1 & 0 \\ \frac{\tau\alpha'_{i0}\lambda'_{i0}}{\lambda_{i0}^2} & -\frac{\tau\alpha'_{i0}}{\lambda_{i0}} & \frac{\tau\alpha'_{i0}\lambda'_{i0}}{\lambda_{i0}^2} & -\frac{\tau\alpha'_{i0}}{\lambda_{i0}} \\ 0 & 0 & \frac{\tau\alpha'_{i0}\lambda'_{i0}}{\lambda_{i0}^2} & -\frac{\tau\alpha'_{i0}}{\lambda_{i0}} \end{bmatrix}.$$

Furthermore, the determinant of U is

$$|U|=\begin{vmatrix} 0 & \alpha_{i0} & 1 & 0 \\ 1 & 0 & 1 & 0 \\ \frac{\tau\alpha'_{i0}\lambda'_{i0}}{\lambda_{i0}^2} & -\frac{\tau\alpha'_{i0}}{\lambda_{i0}} & \frac{\tau\alpha'_{i0}\lambda'_{i0}}{\lambda_{i0}^2} & -\frac{\tau\alpha'_{i0}}{\lambda_{i0}} \\ 0 & 0 & \frac{\tau\alpha'_{i0}\lambda'_{i0}}{\lambda_{i0}^2} & -\frac{\tau\alpha'_{i0}}{\lambda_{i0}} \end{vmatrix}=\frac{\tau^2\alpha'^2_{i0}}{\lambda_{i0}^2}+\frac{\tau^2\alpha'^3_{i0}\lambda'_{i0}}{\lambda_{i0}^3}.$$

Since $\tau, \alpha'_{i0}, \lambda'_{i0}$ and λ_{i0} are defined as positive real numbers, $|U|\neq 0$ always holds. Hence, when $d_i(n)= G_i^{\max}$, the P2MP system represented by Eq. (5.15) is controllable.

The above three scenarios summarize all of the possible combinations of loaded traffic and granted transmission in a P2MP system. Theorems 5.1-5.3 testify that the generic P2MP system with NLPDBA is completely controllable.

When the predictor underestimates the traffic, there will be residual data queued up at the LS buffer after one service cycle. This is the so-called "unsatisfied" case and it falls into Scenario 1. Theorem 5.1 shows that the P2MP system with

NLPDBA can self-tune to reach the proper state of bandwidth sharing even when prediction inaccuracy occurs.

When the predictor overestimates the traffic, the total data arrived at an LS (i.e., the ONU/SAN) can be delivered to the RS (i.e., the OLT) within the current service cycle, with a small portion of the timeslot being "idle". This falls into Scenario 2. Theorem 5.2 indicates that the RS with NLPDBA is capable of eliminating the over-reserved bandwidth by taking "over-grant" into consideration.

When the users aggressively request the upstream bandwidth, the RS employs the SLA specification to upper bound their transmission, and this falls into Scenario 3. Theorem 5.3 verifies that the P2MP system with NLPDBA is capable of limiting the aggressive bandwidth competition among users, and the upstream bandwidth is thus arbitrated fairly.

The above conclusions are all applicable to the S-PON system, which is a special case of a P2MP system.

5.3 Stability Analysis and Controller Design of NLPDBA

By "stable", we mean that, when the input traffic load changes dramatically, the resource allocation scheme is able to provide the upstream resource fair share among the LSs with optimal bandwidth utilization [39],[42],[49]. For any resource allocation scheme, the stability design is critical because it provides predictability for system behavior and guarantees any generated oscillations to be bounded within

a certain range. On the other hand, the instability usually leads to unbounded oscillations which lower the overall network efficiency.

The open plant denoted by Eq. (5.8) usually implies an unbounded output. For any resource allocation scheme that is controllable, there always exists a controller,

$$u_i(n) = -K_i x_i(n) + F_i r_i(n), \tag{5.16a}$$

which drives the system into the stable state; this is known as *pole placement* [37]. K_i is a constant matrix, F_i is a pre-defined matrix, and $r_i(n)$ is a reference vector.

Substituting (5.16a) into (5.8) yields,

$$x_i(n+1) = (A_i - B_i K_i) x_i(n) + B_i F_i r_i(n). \tag{5.16b}$$

Therefore, by implementing the controller of Eq. (5.16a), (A_i, B_i) is transformed into $(A_i - B_i K_i, B_i F_i)$. The controllability of a resource allocation scheme is unaltered by state feedback [37]. That is, if (A_i, B_i) is controllable (or uncontrollable), so is $(A_i - B_i K_i, B_i F_i)$ for any K_i and F_i. Since F_i and $r_i(n)$ have no impact on the system's stability [36], we will focus on K_i which dominates the system stability. Assuming the reference vector $r_i(n) = 0$, we will have

$$u_i(n) = -K_i x_i(n) \quad , \tag{5.17a}$$

and

$$x_i(n+1) = (A_i - B_i K_i) x_i(n) \quad . \tag{5.17b}$$

Hence, after implementing the controller of Eq. (5.17a), the system becomes a close-loop form expressed in Eq. (5.17b).

Similarly, Eq. (5.9) gives three different traffic scenarios for the P2MP system, depending on the loaded traffic from end users:

A) $d_i(n) = G_i^r(n)$,

B) $d_i(n) = R_i(n-1) + \lambda_i(n-1)$, and

C) $d_i(n) = G_i^{\max}$.

A). Scenario 1: $d_i(n+1) = G_i^r(n+1)$

In the previous section, we have developed the linearized state space equation for NLPDBA under scenario 1 as

$$\delta x_i(n+1) = A_1 \delta x_i(n) + B_1 \delta u_i(n) \quad , \tag{5.11}$$

where

$$A_1 = \begin{bmatrix} 0 & 1 & 0 & \lambda'_{i0} \\ 0 & 0 & 0 & -\lambda'_{i0} \\ 0 & 0 & 1 & -\dfrac{\tau\lambda'_{i0}}{\lambda_{i0}} \\ 0 & 0 & 1 & 0 \end{bmatrix}, B_1 = \begin{bmatrix} 0 & \alpha'_{i0} \\ 1 & -\alpha'_{i0} \\ \dfrac{\tau\alpha'_{i0}\lambda'_{i0}}{\lambda_{i0}^2} & -\dfrac{\tau\alpha'_{i0}}{\lambda_{i0}} \\ 0 & 0 \end{bmatrix}.$$

Theorem 5.4 In scenario 1, a P2MP system with NLPDBA is stable when $0 < \tau < \dfrac{\lambda_0}{\lambda'_0}$.

Proof: The discrete system represented by Eq. (5.11) is stable *iff* eigenvalues of the state matrix A_1 fall inside the unit circle [43]. Let $|zI - A_1| = 0$; we have,

$$\det(zI - A_1) = D(z) = z^4 - z^3 + \frac{\tau\lambda'_0}{\lambda_0} z^2. \tag{5.18a}$$

According to the Jury's criterion [43], a fourth order system $D(z) = \sum_{i=0}^{4} a_i z^i$ is stable *iff* the following rules are all fulfilled:

Rule 1: $a_0^2 - a_4^2 - a_0 a_3 + a_1 a_4 < 0$ (5.18b)

Rule 2: $a_0^2 - a_4^2 + a_0 a_3 - a_1 a_4 < 0$ (5.18c)

Rule 3: $a_0^3 + 2a_0 a_2 a_4 + a_1 a_3 a_4 - a_0 a_4^2 - a_2 a_4^2 - a_0 a_3^2 - a_0^2 a_4 - a_0^2 a_2 - a_1^2 a_4 + a_4^3 + a_0 a_1 a_3 > 0$ (5.18d)

Rule 4: $D(1) > 0$ (5.18e)

Rule 5: $D(-1) > 0$ (5.18f)

Eq. (5.18a) implies $a_0 = a_1 = 0$, $a_2 = \frac{\tau \lambda_0'}{\lambda_0}$, $a_3 = -1$ and $a_4 = 1$. By applying Rules 1-5, and considering τ a positive real number, the system is stable when $0 < \tau < \frac{\lambda_0}{\lambda_0'}$.

B). Scenario 2: $d_i(n+1) = R_i(n) + \lambda_i(n)$

Similarly, we have developed the state space equation for NLPDBA in the previous section under scenario 2 as

$$\delta x_i(n+1) = A_2 \delta x_i(n) + B_2 \delta u_i(n), \tag{5.13}$$

where

$$A_2 = \begin{bmatrix} 0 & 0 & 0 & \lambda_{i0}' \\ 0 & 1 & 0 & -\lambda_{i0}' \\ 0 & 0 & 1 & -\frac{\tau \lambda_{i0}'}{\lambda_{i0}} \\ 0 & 0 & 1 & 0 \end{bmatrix}, B_2 = \begin{bmatrix} 0 & \alpha_{i0}' \\ 1 & -\alpha_{i0}' \\ \frac{\tau \alpha_{i0}' \lambda_{i0}'}{\lambda_{i0}^2} & \frac{-\tau \alpha_{i0}'}{\lambda_{i0}} \\ 0 & 0 \end{bmatrix}.$$

Theorem 5.5 In scenario 2, a P2MP system with NLPDBA is stable when implementing the controller

$$u_i(n) = -K_1 x_i(n), \tag{5.19}$$

where $K_1 = \begin{bmatrix} k_{11} & k_{12} & k_{13} & k_{14} \\ k_{21} & k_{22} & k_{23} & k_{24} \end{bmatrix}$. The range of vectors of K_1 is given by

$$\begin{cases} L_0^2 - L_0 L_3 + L_1 - 1 < 0 \\ L_0^2 + L_0 L_3 - L_1 - 1 < 0 \\ L_0^3 + 2L_0 L_2 + L_1 L_3 - L_0 - L_2 - L_0 L_3^2 - L_0^2 - L_0^2 L_2 - L_1^2 + L_0 L_1 L_3 + 1 > 0 \\ 1 + L_3 + L_2 + L_1 + L_0 > 0 \\ 1 - L_3 + L_2 - L_1 + L_0 > 0 \end{cases} \tag{5.20}$$

where $L_i|_{i=0\sim4}$ are given by Eq. (5.20c)~(5.20g).

Proof: Let $|zI - A_2| = 0$; we have

$$\det(zI - A_2) = D(z) = z^4 - 2z^3 + (1 + \frac{\tau\lambda_0'}{\lambda_0})z^2 - \frac{\tau\lambda_0'}{\lambda_0}z. \tag{5.20a}$$

It is easy to check whether the coefficients in Eq. (5.20a) violate Jury criteria Rules. However, since NLPDBA is completely controllable [48], there always exists a controller $u_i(n) = -K_1 x_i(n)$ which can drive the system into the stable state [37]. By "controllable", we mean that NLPDBA can schedule the dynamic traffic input from multiple ONUs efficiently over the P2MP system from any initial value to the optimum state within a limited time window. After implementing such a controller, the system becomes $x_i(n+1) = (A_2 - B_2 K_1) x_i(n)$. This discrete system is stable *iff* eigenvalues of the state matrix $(A_2 - B_2 K_1)$ fall inside the unit circle [43]. Let $|zI - (A_2 - B_2 K_1)| = 0$; by solving this 4×4 matrix, we have

$$\det[(zI - (A_2 - B_2 K_1)] = D(z) = z^4 + L_3 z^3 + L_2 z^2 + L_1 z + L_0. \tag{5.20b}$$

The coefficients $L_i|_{i=0\sim4}$ are given by

$$\begin{aligned} L_0 = {} & (\frac{\tau\alpha_0'\lambda_0'}{\lambda_0^2}k_{12} - \frac{\tau\alpha_0'}{\lambda_0}k_{22})[(\lambda_0' - \alpha_0' k_{24} + k_{14})\alpha_0' k_{21} + (k_{11} - \alpha_0' k_{21})(\alpha_0' k_{24} - \lambda_0')] \\ & - \alpha_0' k_{22}(k_{11} - \alpha_0' k_{21})(\frac{\tau\lambda_0'}{\lambda_0} - \frac{\tau\lambda_0'}{\lambda_0}k_{24} + \frac{\tau\alpha_0'\lambda_0'}{\lambda_0^2}k_{14}) \\ & + (\frac{\tau\alpha_0'\lambda_0'}{\lambda_0^2}k_{11} - \frac{\tau\alpha_0'}{\lambda_0}k_{21})[\alpha_0' k_{22}(\lambda_0' - \alpha_0' k_{24} + k_{14}) - (\alpha_0' k_{22} - 1 - k_{12})(\alpha_0' k_{24} - \lambda_0')] \end{aligned} \tag{5.20c}$$

$$L_1 = (\frac{\tau\alpha_0'\lambda_0'}{\lambda_0^2}k_{11} - \frac{\tau\alpha_0'}{\lambda_0}k_{21})[\alpha_0'k_{22}(k_{13} - \alpha_0'k_{23}) - \alpha_0'k_{23}(\alpha_0'k_{22} - k_{12} - 1)] - (\frac{\tau\alpha_0'\lambda_0'}{\lambda_0^2}k_{12} - \frac{\tau\alpha_0'}{\lambda_0}k_{22})$$
$$\times[\alpha_0'k_{21}(k_{13} - \alpha_0'k_{23}) + \alpha_0'k_{23}(k_{11} - \alpha_0'k_{21})] + (\frac{\tau\alpha_0'\lambda_0'}{\lambda_0^2}k_{13} - \frac{\tau\alpha_0'}{\lambda_0}k_{23} - 1)[\alpha_0'k_{21}(\alpha_0'k_{22} - k_{12} - 1)$$
$$-\alpha_0'k_{22}(k_{11} - \alpha_0'k_{21})] + (\frac{\tau\lambda_0'}{\lambda_0} - \frac{\tau\alpha_0'}{\lambda_0}k_{22} + \frac{\tau\alpha_0'\lambda_0'}{\lambda_0^2}k_{14})(2\alpha_0'k_{22} - k_{12}) - (\lambda_0' - \alpha_0'k_{24} + k_{14})$$
$$\times(\frac{\tau\alpha_0'\lambda_0'}{\lambda_0^2}k_{12} - \frac{\tau\alpha_0'}{\lambda_0}k_{22}) - (\frac{\tau\alpha_0'\lambda_0'}{\lambda_0^2}k_{11} - \frac{\tau\alpha_0'}{\lambda_0}k_{21})(\alpha_0'k_{24} - \lambda_0'), \quad (5.20d)$$

$$L_2 = \alpha_0'k_{21}(\alpha_0'k_{22} - k_{12} - 1) - \alpha_0'k_{22}(k_{11} - \alpha_0'k_{21}) + (\frac{\tau\alpha_0'\lambda_0'}{\lambda_0^2}k_{13} - \frac{\tau\alpha_0'}{\lambda_0}k_{23} - 1)\times(\alpha_0'k_{21} + \alpha_0'k_{22} - k_{12} - 1)$$
$$-\alpha_0'k_{23}(\frac{\tau\alpha_0'\lambda_0'}{\lambda_0^2}k_{11} - \frac{\tau\alpha_0'}{\lambda_0}k_{21}) - (k_{13} - \alpha_0'k_{23})(\frac{\tau\alpha_0'\lambda_0'}{\lambda_0^2}k_{12} - \frac{\tau\alpha_0'}{\lambda_0}k_{22}) + \frac{\tau\lambda_0'}{\lambda_0} - \frac{\tau\alpha_0'}{\lambda_0}k_{24} + \frac{\tau\alpha_0'\lambda_0'}{\lambda_0^2}k_{14} \quad (5.20e)$$

$$L_3 = \alpha_0'k_{21} + \alpha_0'k_{22} - k_{12} + \frac{\tau\alpha_0'\lambda_0'}{\lambda_0^2}k_{13} - \frac{\tau\alpha_0'}{\lambda_0}k_{23} - 2 \quad (5.20f)$$

and,

$$L_4 = 1 \quad (5.20g)$$

By applying Jury's criteria, this fourth order system is stable *iff* Rules 1-5 (i.e., Eqs.(5.18b)~(5.18f)) are all fulfilled, i.e.,

$$\begin{cases} L_0^2 - L_0L_3 + L_1 - 1 < 0 \\ L_0^2 + L_0L_3 - L_1 - 1 < 0 \\ L_0^3 + 2L_0L_2 + L_1L_3 - L_0 - L_2 - L_0L_3^2 - L_0^2 - L_0^2L_2 - L_1^2 + L_0L_1L_3 + 1 > 0 \\ 1 + L_3 + L_2 + L_1 + L_0 > 0 \\ 1 - L_3 + L_2 - L_1 + L_0 > 0 \end{cases} \quad (5.20).$$

C. Scenario 3: $d_i(n) = G_i^{\max}$

Similarly, we have developed the state space equation for NLPDBA in the previous section under scenario 2 as

$$\delta x_i(n+1) = A_3\delta x_i(n) + B_3\delta u_i(n) \quad (5.21)$$

where

$$A_3=\begin{bmatrix}0 & 1 & 0 & \lambda'_{i0}\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & -\dfrac{\tau\lambda'_{i0}}{\lambda_{i0}}\\ 0 & 0 & 1 & 0\end{bmatrix}, B_3=\begin{bmatrix}0 & \alpha'_{i0}\\ 1 & 0\\ \dfrac{\tau\alpha'_{i0}\lambda'_{i0}}{\lambda_{i0}^2} & -\dfrac{\tau\alpha'_{i0}}{\lambda_{i0}}\\ 0 & 0\end{bmatrix}.$$

Similar to the previous two scenarios, the following theorem is established.

Theorem 5.6 In scenario 3, a P2MP system with NLPDBA is stable when implementing the controller

$$u_i(n)=-K_2x_i(n), \tag{5.22}$$

where $K_2=\begin{bmatrix}p_{11} & p_{12} & p_{13} & p_{14}\\ p_{21} & p_{22} & p_{23} & p_{24}\end{bmatrix}$. The range of vectors of K_2 is given by

$$\begin{cases}M_0^2-M_0M_3+M_1-1<0\\ M_0^2+M_0M_3-M_1-1<0\\ M_0^3+2M_0M_2+M_1M_3-M_0-M_2-M_0M_3^2-M_0^2-M_0^2M_2-M_1^2+M_0M_1M_3+1>0\\ 1+M_3+M_2+M_1+M_0>0\\ 1-M_3+M_2-M_1+M_0>0\end{cases} \tag{5.23}$$

where $M|_{i=0\sim4}$ are given by

$$\begin{aligned}M_0=&(\frac{\tau\alpha'_0\lambda'_0}{\lambda_0^2}p_{11}-\frac{\tau\alpha'_0}{\lambda_0}p_{21})[p_{14}(\alpha'_0p_{22}-1)-(\alpha'_0p_{24}-\lambda'_0)(p_{12}-1)]+(\frac{\tau\alpha'_0\lambda'_0}{\lambda_0^2}p_{12}-\frac{\tau\alpha'_0}{\lambda_0}p_{22})\\ &\times[\alpha'_0p_{14}p_{21}+p_{11}(\alpha'_0p_{24}-\lambda'_0)]+(\frac{\tau\alpha'_0\lambda'_0}{\lambda_0^2}p_{14}-\frac{\tau\alpha'_0}{\lambda_0}p_{24}+\frac{\tau\lambda'_0}{\lambda_0})[\alpha'_0p_{21}(p_{12}-1)-p_{11}(\alpha'_0p_{22}-1)]\end{aligned} \tag{5.23a}$$

$$\begin{aligned}M_1=&p_{13}(\alpha'_0p_{22}-1)(\alpha'_0p_{24}-\lambda'_0)-(\frac{\tau\alpha'_0\lambda'_0}{\lambda_0^2}p_{11}-\frac{\tau\alpha'_0}{\lambda_0}p_{21})\times[\alpha'_0p_{23}(p_{12}-1)+\alpha'_0p_{24}-\lambda'_0]\\ &+(\frac{\tau\alpha'_0\lambda'_0}{\lambda_0^2}p_{12}-\frac{\tau\alpha'_0}{\lambda_0}p_{22})(\alpha'_0p_{11}p_{23}-\alpha'_0p_{13}p_{21}-p_{14})\end{aligned} \tag{5.23b}$$

$$\begin{aligned}M_2=&(\frac{\tau\alpha'_0\lambda'_0}{\lambda_0^2}p_{13}-\frac{\tau\alpha'_0}{\lambda_0}p_{23}-1)(\alpha'_0p_{21}+p_{12}-1)-p_{11}(\alpha'_0p_{22}-1)-\alpha'_0p_{23}(\frac{\tau\alpha'_0\lambda'_0}{\lambda_0^2}p_{11}-\frac{\tau\alpha'_0}{\lambda_0}p_{21})\\ &-p_{13}(\frac{\tau\alpha'_0\lambda'_0}{\lambda_0^2}p_{12}-\frac{\tau\alpha'_0}{\lambda_0}p_{22})+(\frac{\tau\alpha'_0\lambda'_0}{\lambda_0^2}p_{14}-\frac{\tau\alpha'_0}{\lambda_0}p_{24}+\frac{\tau\lambda'_0}{\lambda_0})\end{aligned} \tag{5.23c}$$

$$M_3=\alpha'_0p_{21}+p_{12}+\frac{\tau\alpha'_0\lambda'_0}{\lambda_0^2}p_{13}-\frac{\tau\alpha'_0}{\lambda_0}p_{23}-2 \tag{5.23d}$$

and,

$$M_4 = 1 \quad . \tag{5.23e}$$

The above three scenarios summarize all of the possible combinations of loaded traffic and granted transmission in a P2MP system.

When the predictor underestimates the traffic, there will be residual data queued up at the LS buffer after one service cycle. This is the so-called "unsatisfied" case and it falls into scenario 1. Theorem 5.4 shows that the P2MP system with NLPDBA can self-tune to reach the stable state of bandwidth sharing even with prediction inaccuracy.

When the predictor overestimates the traffic, the total data arrived at an LS can be delivered to the RS within the current service cycle, with a small portion of the timeslot being "idle". This falls into scenario 2. Theorem 5.5 indicates that, by implementing the suitable controller Eq. (5.19), the OLT works as a central controller to tune LSs accordingly, ensuring that the upstream bandwidth of a P2MP system is fairly shared by multiple LSs.

When the users aggressively request the upstream bandwidth, the RS employs the SLA specification to upper bound their transmission, and this falls into scenario 3. Theorem 5.6 shows that the P2MP system with NLPDBA is capable of guaranteeing the system's stability by implementing the controller Eq. (5.22). In the last two scenarios, K_1 and K_2 essentially describe the controller characteristics in different scenarios, and their relationship to the estimation index has been revealed in Theorems 5.5 and 5.6.

5.4 Chapter Summary

In this chapter, we have analyzed and verified through the state space model that the implementation of NLPDBA in a P2MP system maintains the system controllability. The traffic predictor is robust to dynamic traffic load. A P2MP system with NLPDBA is able to reach the optimum state of upstream bandwidth allocation, no matter how dynamic the input traffic is. Furthermore, we have also analyzed and verified through the state space model that the implementation of NLPDBA with suitable controllers maintains the P2MP system stability. The employed traffic predictor is robust to dynamic traffic load, and the bandwidth utilization can be improved by adaptive control at the RS (i.e. OLT) side. Although this chapter focuses on generic P2MP system, all results are applicable to the S-PON system, as the S-PON system is a special case of the P2MP system.

CHAPTER 6

CONCLUSIONS AND FUTURE WORK

6.1 Contributions

In this dissertation, we have proposed a new solution, SAN extension over passive optical networks (S-PONs), to address the challenges of SAN extension over the access network. To tackle the scalability problems and cost challenges, we have designed the S-PON architecture based on the existing point-to-multiple-point (P2MP) PON infrastructure. To address the bandwidth bottlenecks in SAN extension, we have also proposed three solutions for carrying storage signals with gigabit-level transmission. We have also introduced a new device, XtenOLT, to improve SAN extension throughput and utility.

Upstream bandwidth allocation through TDMA is critical to the access network performance. In S-PONs, it implies dynamic information exchange between the OLT and ONUs/SANs, upstream transmission scheduling, as well as upstream bandwidth arbitration. In this dissertation, following the introduction of challenges related to resource management over TDM-PONs, we have provided an overview of the state-of-art dissertation in the literature. The state space representation has been introduced as a general model to evaluate various proposed algorithms. Our discussion explains the performance difference among the major upstream bandwidth allocation schemes from the perspective of system control. Original contributions of this dissertation include the following

1) A novel S-PON architecture to tackle the scalability problem and high cost based on the proposed point-to-multiple-point (P2MP) PON infrastructure.

2) Three transmission technologies, namely, TDMA, SCMA, and WDMA, for S-PON that provide up to 2.5Gbps transmission capacity to address the current bandwidth bottleneck.

3) A new device XtenOLT with buffer pools managed by a new proposed Tetris buffer management algorithm to tackle the throughput and utility problem.

4) The creation of a new generic unified model for resource allocation in S-PON (more generically, any P2MP networks) that dramatically changes the way of tackling resource allocation at network edges.

5) The establishment of a novel state space model for studying the controllability characteristics of various resource allocation schemes in P2MP networks.

6) The facilitation of the state space model for analyzing the stability of various resource allocation schemes and for providing guidelines in designing a stable P2MP system.

7) The invention of a suitable controller and compensator to meet the prescribed objectives such as system robustness, accuracy, and transient performance.

8) The extension of the established state space model to the non-linear predictor-based dynamic bandwidth allocation scenario.

6.2 Future Work

In addition to the above contributions, this dissertation has also created the following future research opportunities:

1) To enhance XtenOLT by including additional modules such as wireless module and service differentiation module. One possible wireless module could leverage on 802.16 WiMAX air interface to support the P2MP mode. In this way, S-PON enhances system flexibility by ensuring critical data transmission continuity in case of a fiber failure, and at the same time reduces system cost by decreasing the amount of cabling between the switch and storage devices. Furthermore, regarding to diverse requirements of various enterprises, SAN service differentiation administration and provisioning is a critical issue to consider. Upgrading XtenOLT with a new service differentiation (SD) function module could combine the pros of WiMAX and PON to be implemented in the XtenOLT, and takes advantage of the optical wireless integration (OWI) flexibility of broadband access.

2) To extend the proposed state space model to the WDM-PON and wireless scenario. It is quite a challenge to adapt the state space model for the WDM-PON and wireless systems. The future access networks are most likely evolved into the hybrid architecture of WDM and TDM by utilizing the WDM technology to reach the curb and neighborhood, while employing the TDM technology to reach the end users. Apparently, the hybrid PON system requires not only bandwidth allocation, but also wavelength allocation. This may require the extension of the current

two-dimensional state vector model to the three-dimensional state vector model by including wavelength allocation. The future research could address this issue by imposing suitable resource management migration to making the best use of the low-cost TDM resource while reducing the system cost of the expensive WDM part.

APPENDIX A

THE DERIVATION OF THE COMPENSATOR F_i

According to feedback control theory [44], when the system output converges to the input reference (i.e. $e(n) = Y_i(n) - r = 0$), the state variable $X_i(n)$ reaches its steady state X_i^{ss}. Assume the associated steady state input is U_i^{ss}, the controller represented by Eq. (4.32) can thus be re-written as

$$U_i(n) = -K_i(X_i(n) - X_i^{ss}) + U_i^{ss} \tag{A.1a}$$

Obviously, from Eq.(A.1a), it is easy to see that the system input $U_i(n)$ reaches its steady state U_i^{ss} when the state variables $X_i(n)$ reaches X_i^{ss}. Eq.(A.1a) can be further re-written as $U_i(n) = -K_i X_i(n) + K_i X_i^{ss} + U_i^{ss}$, i.e.,

$$U_i(n) = -K_i X_i(n) + \begin{bmatrix} K_i & 1 \end{bmatrix} \begin{bmatrix} X_i^{ss} \\ U_i^{ss} \end{bmatrix} \tag{A.1b}$$

When the system reaches the steady state, the following equations hold,

$$\begin{aligned} X_i^{ss} &= AX_i^{ss} + BU_i^{ss} \\ Y_i^{ss} &= CX_i^{ss} \\ Y_i^{ss} &= r \end{aligned} \tag{A.2a}$$

Eq. (A.2a) further yields,

$$\begin{aligned} &(A-I)X_i^{ss} + BU_i^{ss} = 0 \\ &CX_i^{ss} = r \end{aligned} \quad \text{, i.e, } \begin{bmatrix} A-I & B \\ C & 0 \end{bmatrix} \begin{bmatrix} X_i^{ss} \\ U_i^{ss} \end{bmatrix} = \begin{bmatrix} 0 \\ r \end{bmatrix} \tag{A.2b}$$

From Eq. (A.2b), we get

$$\begin{bmatrix} X_i^{ss} \\ U_i^{ss} \end{bmatrix} = \begin{bmatrix} A-I & B \\ C & 0 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ r \end{bmatrix} \tag{A.2c}$$

provided that $\begin{bmatrix} A-I & B \\ C & 0 \end{bmatrix}$ is a non-singular matrix.

From Eq.(A.1b) and (A.2c), we have the controller in the following expression,

$$U_i(n) = -K_i X_i(n) + \begin{bmatrix} K_i & 1 \end{bmatrix} \begin{bmatrix} A-I & B \\ C & 0 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix} r \quad \text{(A.3)}$$

By comparing Eqs. (A.1b) and (A.3), the compensator F_i to offset the control error is

$$F_i = \begin{bmatrix} K_i & 1 \end{bmatrix} \begin{bmatrix} A-I & B \\ C & 0 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad \text{(A.4)}$$

APPENDIX B
LINEARIZATION OF THE SYSTEM IN SCENARIO 1

We first define the right-hand sides of Eqs. (5.10a), (5.10c), (5.8a), and (5.8b) as

$$f_1(R_i, \alpha_i', \lambda_i') \doteq R_i + \alpha_i' \lambda_i' \quad , \tag{B.1a}$$

$$f_2(\lambda_i, \alpha_i', \lambda_i') \doteq \lambda_i - \alpha_i' \lambda_i' \quad , \tag{B.1b}$$

$$f_3(\alpha_i, \alpha_i', \lambda_i, \lambda_i') \doteq \alpha_i + \tau - \tau \times \frac{\alpha_i' \times \lambda_i'}{\lambda_i} \quad , \tag{B.1c}$$

$$f_4(\alpha_i) \doteq \alpha_i \quad . \tag{B.1d}$$

Taking partial derivatives at the equilibrium point $(G_{i0}^r, R_{i0}, \alpha_{i0}, \alpha_{i0}', \lambda_{i0}, \lambda_{i0}')$ yields

$$\frac{\partial f_1}{\partial R_i} = 1, \frac{\partial f_1}{\partial \alpha_i'} = \lambda_{i0}', \frac{\partial f_1}{\partial \lambda_i'} = \alpha_{i0}', \tag{B.1e}$$

$$\frac{\partial f_2}{\partial \lambda_i} = 1, \frac{\partial f_2}{\partial \alpha_i'} = -\lambda_{i0}', \frac{\partial f_2}{\partial \lambda_i'} = -\alpha_{i0}', \tag{B.1f}$$

$$\frac{\partial f_3}{\partial \alpha_i} = 1, \frac{\partial f_3}{\partial \alpha_i'} = \frac{\tau \lambda_{i0}'}{\lambda_{i0}}, \frac{\partial f_3}{\partial \lambda_i} = -\frac{\tau \alpha_{i0}' \lambda_{i0}'}{\lambda_{i0}^2}, \frac{\partial f_3}{\partial \lambda_i'} = \frac{\tau \alpha_{i0}'}{\lambda_{i0}}, \tag{B.1g}$$

$$\frac{\partial f_4}{\partial \alpha_i} = 1 . \tag{B.1h}$$

Therefore, the system denoted by Eqs. (5.10a), (5.10c), (5.8a), and (5.8b) can be linearized at the equilibrium point $(G_{i0}^r, R_{i0}, \alpha_{i0}, \alpha_{i0}', \lambda_{i0}, \lambda_{i0}')$ as Eqs. (5.11a)-(5.11d).

APPENDIX C
LINEARIZATION OF THE SYSTEM IN SCENARIO 2

In this scenario, the linearization of Eqs. (5.8a) and (5.8b) is unchanged because it follows the same LMS algorithm to update the estimation index $\alpha_i(n+1)$. We then focus on the linearization of loaded traffic and queue length in Scenario 2. We define the right-hand sides of Eqs. (5.12b) and (5.12c) by

$$f_5(\alpha_i', \lambda_i') \doteq \alpha_i' \lambda_i' \tag{C.1a}$$

$$f_6(R_i, \alpha_i', \lambda_i, \lambda_i') \doteq R_i + \lambda_i - \alpha_i' \lambda_i' \tag{C.1b}$$

Taking partial derivatives at the equilibrium point $(G_{i0}^r, R_{i0}, \alpha_{i0}, \alpha_{i0}', \lambda_{i0}, \lambda_{i0}')$ yields

$$\frac{\partial f_5}{\partial \alpha_i} = \lambda_{i0}', \frac{\partial f_5}{\partial \lambda_i} = \alpha_{i0}', \tag{C.1c}$$

$$\frac{\partial f_6}{\partial R_i} = 1, \frac{\partial f_6}{\partial \alpha_i'} = -\lambda_{i0}', \frac{\partial f_6}{\partial \lambda_i} = 1, \frac{\partial f_6}{\partial \lambda_i'} = -\alpha_{i0}'. \tag{C.1d}$$

Therefore, the system denoted by Eqs. (5.12b), (5.12c), (5.8a), and (5.8b) can be linearized at the equilibrium point $(G_{i0}^r, R_{i0}, \alpha_{i0}, \alpha_{i0}', \lambda_{i0}, \lambda_{i0}')$ as Eqs. (5.13a), (5.13b), (5.11c), and (5.11d).

APPENDIX D
LINEARIZATION OF THE SYSTEM IN SCENARIO 3

Similarly, we define the right-hand sides of Eqs. (5.14a) and (5.14b) by

$$f_7(R_i, \alpha_i', \lambda_i') = R_i + \alpha_i' \lambda_i', \tag{D.1a}$$

$$f_8(R_i, \lambda_i) = R_i - G_i^{\max} + \lambda_i. \tag{D.1b}$$

Taking partial derivatives at the equilibrium point $(G_{i0}^r, R_{i0}, \alpha_{i0}, \alpha_{i0}', \lambda_{i0}, \lambda_{i0}')$ yields

$$\frac{\partial f_7}{\partial R_i} = 1, \frac{\partial f_7}{\partial \alpha_i'} = \lambda_{i0}', \frac{\partial f_7}{\partial \lambda_i'} = \alpha_{i0}', \tag{D.1c}$$

$$\frac{\partial f_8}{\partial R_i} = 1, \frac{\partial f_8}{\partial \lambda_i} = 1. \tag{D.1d}$$

The linearization of Eqs. (5.8a) and (5.8b) follows the same process as in Scenario 1, resulting in Eqs. (5.11c) and (5.11d). The system denoted by Eqs. (5.14a), (5.14b), (5.8a), and (5.8b) can be linearized at the equilibrium point $(G_{i0}^r, R_{i0}, \alpha_{i0}, \alpha_{i0}', \lambda_{i0}, \lambda_{i0}')$ as Eqs. (5.15a), (5.15b), (5.11c), and (5.11d).

REFERENCES

[1] N. Ansari and S. Yin, "Chapter 149: Storage Area Networks: architectures and protocols", *The Handbook of Computer Networks Volume III: Distributed Networks, Network Planning, Control, Management, and New Trends and Applications* (Hossein Bidgoli, ed.), pp. 217-234, John Wiley & Son, ISBN 978-0-471-78458-9, 2008.

[2] T. Clark, *Designing Storage Area Networks: A Practical Reference for Implementing Fibre Channel and IP SANs*, Reading, MA: Addison-Wesley, 2003.

[3] Telikepalli, R., T. Drwiega and J. Yan. "Storage area network extension solutions and their performance assessment," *IEEE Communications Magazine*, pp.56–63, Apr. 2004.

[4] -- "Improving business resilience with data replication and storage area network extension technology," April 2007, Hitachi whit paper. http://www.hds.com/assets/pdf/wp-improving-business-resilience-with-data-replication-and-storage-area-network-extension-technology.pdf

[5] X. Qiu, R. Telikepalli, T. Drwiega, and J. Yan, "Reliability and availability assessment of storage area network extension solutions," *IEEE Communications Magazine*, pp. 80~85, Mar. 2005.

[6] --, "Thinking storage network extension," Ciena Inc. whitepaper, http://www.fcw.com/vendorsolutions/THINK_Ciena-2-final2.pdf

[7] T. Kessler, A. Ehrhardt, A. Weber, "SAN extension testbed within the global seamless networks demonstrator," *Proc. of 2nd International Conference on Testbeds and Research Infrastructures for the Development of Networks and Communities*, pp. 209 – 215, Mar. 2006.

[8] A. Helland, "The economics of large-scale data transfer: the benefits of fibre channel and SONET for high performance, cost-effective data transport through WAN storage networking," *Computer Technology Review*, Dec. 2002.

[9] --, "EMC and Cisco metro optical storage networking solutions," Cisco Systems Inc. White paper 2001. http://whitepapers.zdnet.co.uk/0,39025945,60031308p-39000566q,00.htm

[10]T. Clark, *IP SANs: A Guide to iSCSI, iFCP, and FCIP Protocols for Storage Area Networks*, Upper Saddle River, NJ : Pearson Education, 2002.

[11] M.Bakke, J. Muchow , M. Krueger , and T. McSweeney, iSCSI, IETF.draft-ietf-ips-iscsi-mib-11.txt.2005. www.ietf.org/internet-drafts/draft-ietf-ips-iscsi-mib-11.txt

[12] R.Natarajan, A. Rijhsinghani, Fibre Channel over TCP/IP (FCIP). IETF, draft-ietf-ips-fcip-mib-09.txt, 2005. http://www.ietf.org/internet-drafts/draft-ietf-ips-fcip-mib-09.txt

[13] K. Gibbons, K.,Monia, J.Tseng, F. Travostino, iFCP — A Protocol for Internet Fibre Channel Storage Networking, IETF, draft-ietf-ips-ifcp-mib-07.txt, 2005. http://www.ietf.org/internet-drafts/draft-ietf-ips-ifcp-mib-07.txt

[14] A. Girard, *FTTx PON Technology and Testing*, EXFO Electro-Engineering Inc., Quebec City, Canada, 2005.

[15] ITU-T G.983.1 Recommendation, "Broadband optical access systems based on passive optical networks (PON)," Oct. 1998.

[16] IEEE Standard 802.3ah-2004.

[17] IEEE P802.3av Task Force, "Five Criteria Responses," http://www.ieee802.org/3/av/tf_docs/10gepon_5criteria_0506.pdf.

[18] ITU-T G.984.1 Recommendation, "Gigabit-capable passive optical networks (GPON): general characteristics," Mar. 2003.

[19] M. McGarry, M. Maier, and M. Reisslein, "Ethernet PONs: a survey of dynamic bandwidth allocation (DBA) algorithms," *IEEE (Optical) Communications Magazine*, vol.42, no.8, pp.s8-s15, Aug.2004.

[20] G. Kramer. B. Mukherjee, and G. Perawnto, "Ethernet PON (ePON): design and analysis of an optical access network," *Photonic Network Communications*, vol. 3. no. 3. pp. 307-319, Jul. 2001.

[21] G. Kramer, B. Mukherjee, and G. Pesavento, "IPACT: a dynamic protocol for an Ethernet PON (EPON)," *IEEE Communications Magazine,* vol. 40, no. 2, pp. 74-80, Feb. 2002.

[22] A. Banerjee, G. Kramer, and B. Mukherjee, "Fair sharing using dual service-level agreements to achieve open access in a passive optical network," *IEEE Journal on Selected Areas in Communications*, Vol. 24, no. 8, pp. 32-44, Aug. 2006.

[23] H. Naser and H. Mouftah, "A joint-ONU interval-based dynamic scheduling algorithm for Ethernet passive optical networks," IEEE/ACM Transactions on Networking, vol. 14, no. 4, pp. 889-899, Aug. 2006.

[24] A. Shami, X. Bai, D. Assi, and N. Ghani, "Jitter performance in Ethernet passive optical networks," Journal of Lightwave Technology, vol. 23, no. 4, pp. 1745-1753, April 2005.

[25] C. Assi, Y. Ye, S. Dixi, and M. Ali, "Dynamic bandwidth allocation for quality-of-service over Ethernet PON", IEEE Journal on Selected Areas in Communications, vol. 12, no. 9, pp. 1467- 1477, Nov. 2003.

[26] H. Byun, J. Nho and J. Lim, "Dynamic bandwidth allocation algorithm in Ethernet passive optical networks," IEE Electronic letters, vol. 39, no. 13, pp. 1001-1002, June 2003.

[27] Y. Luo and N. Ansari, "Limited sharing with traffic prediction for dynamic bandwidth allocation and QoS provisioning over EPONs," Journal of Optical Networking, vol. 4, no.9, pp. 561-572, Sep. 2005.

[28] S. Yin, Y. Luo, L. Zong, S. Rago, J. Yu, N. Ansari, and T. Wang, "Storage Area Network Extension over Passive Optical Networks (S-PONs)", to appear in IEEE Applications & Practice Magazine, the supplement to IEEE communication magazine, Jan. 2008;

[29] A.K. Choudhury and E.L. Hahne, "Dynamic Queue Length Threshold for Shared-Memory Packet Switches," IEEE/ACM Trans. Network, Vol. 6, no. 2, Apr. 1998, pp. 130-140.

[30] X. Molero, F. Silla, V. Santonja and J. Duato, "Modeling and Simulation of Storage Area Networks," *Proc. 8th International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems*, pp. 307–314, Aug. 29–Sept. 1, 2000,

[31] M.E. Gomez and V. Santonja, "Analysis of Self-Similarity in I/O Workload Using Structural Modeling," *Proc. 7th International Symposium on IEEE*

Analysis and Simulation of Computer and Telecommunication Systems, Oct.24–28, 1999.

[32] M.E. Gomez and V. Santonja, "Self-Similarity in I/O Workload: Analysis and Modeling," *Proc. IEEE Workload Characterization: Methodology and Case Studies*, pp. 97–104, Nov. 1998.

[33] Y. Afek, M. Cohen, E. Haalman and Y. Mansour, "Dynamic Bandwidth Allocation Policies," *Proc. IEEE INFOCOM'96*, San Francisco.

[34] S. Biswas and R. Izmailov, "Design of a Fair Bandwidth Allocation Policy for VBR Traffic in ATM Networks," *IEEE/ACM Trans. on Networking*, Vol. 8, pp. 212–223, Issue 2, Apr. 2000.

[35] V. Ruesterholz, "Verizon FiOS: executing strategy and revenue opportunities", http://investor.verizon.com/news/20060927/20060927.pdf

[36] Z. Bubnicki, Modern control theory, Spring Berlin Heidelberg New York 2005.

[37] C.T. Chen, *Linear System theory and design, 3rd ed.*, Oxford University Press, 1999.

[38] S. Yin, Y. Luo, N. Ansari, and T. Wang, "Bandwidth Allocation over EPONs: A Controllability Perspective", *Proc. IEEE GLOBCOM 2006*, Nov.-Dec. 2006, San Francisco, CA.

[39] Y. Luo, S. Yin, N. Ansari, and T. Wang, "Resource Allocation for Broadband Access over TDM Passive Optical Networks", *IEEE network*, vol. 21, issue 5, pp.20-27, Sept./Oct.2007;

[40] W. Willinger, M. S. Taqqu, R. Sherman, and D. V. Wilson, "Self-similarity through high-variability: statistical analysis of Ethernet LAN traffic at the source level," *IEEE/ACM Transactions on Networking*, vol. 5, no. 1, pp. 71–86., Feb. 1997.

[41] Y. Luo and N. Ansari, "Bandwidth allocation for multiservice access on EPONs," *IEEE Communications Magazine*, vol.43, no.2, pp. S16-S21, Feb. 2005.

[42] S. Yin, Y. Luo, N. Ansari, and T. Wang, "Stability of predictor-based dynamic bandwidth allocation over EPONs," *IEEE Communications Letters*, vol. 11, no. 6, pp. 549-551, Jun. 2007.

[43] Jury, E. I., *Inners and Stability of Dynamic Systems, 2nd ed.*, Krieger, Malabar, FA, 1982.

[44] J. Hellerstein, Y. Diao, S. Parekh, and D. Tilbury, *Feedback Control of Computing Systems*, Wiley-Interscience, Hoboken, NJ, 2004

[45] C.V.Hollot, V. Misra, D. Towsley and W. Gong, "A control theoretic analysis of RED", *Proc. of IEEE INFOCOM 2001,* Vol. 3, pp: 1510 – 1519, Apr. 2001, Anchorage, Alaska.

[46] K.Y.Kim, B.S.Kim, Y.B.Choi, S.T.Ko, and K.S.Kim, "Optimal rate based flow control for ABR services in ATM networks", *Proc. of IEEE TENCON'99*, Vol. 1, pp. 773 - 776. Sep. 1999, Korea.

[47] S.Haykin, *Adaptive Filter Theory, 3rd ed.* (Prentice-Hall, 1996)

[48] S. Yin, Y. Luo, N. Ansari, and T. Wang, "Controllability of Non-Linear Predictor-Based Dynamic Bandwidth Allocation over EPONs", *Proc. Of IEEE Global Communications Conference (GLOBECOM'07)*, Nov. 2006, Washington D.C.

[49] S. Yin, Y. Luo, N. Ansari, and T. Wang, "Non-Linear Predictor-Based Dynamic Bandwidth Allocation over TDM-PONs: Stability Analysis and Controller Design", *Proceeding of IEEE International Conference on Communications 2008 (ICC'08)*, Beijing, China, May 2008.