# 國立交通大學資訊學院
# 資 訊 工 程 學 系
# 博士論文

## 全光分波多工封包交換都會環狀網路之
## 光標頭處理及存取控制系統的設計與實現

# Design and Experimentation of an Optical-Header Processing and Access Control System for a Packet-Switched WDM Metro Ring Network

研　　究　　生：王雅纖

指　導　教　授：楊啟瑞　博士

中 華 民 國 九 十 九 年 二 月

# 全光分波多工封包交換都會環狀網路之
# 光標頭處理及存取控制系統的設計與實現

# Design and Experimentation of an
# Optical-Header Processing and Access Control
# System for a Packet-Switched WDM Metro
# Ring Network

研 究 生：王雅纖　　　　**Student: Ya-Shian Wang**

指導教授：楊啟瑞　博士　　**Advisor: Dr. Maria C. Yuang**

國立交通大學　資訊學院

資訊工程學系

博士論文

中 華 民 國 九 十 九 年 二 月

# 全光分波多工封包交換都會環狀網路之光標頭處理及存取控制系統的設計與實現

研究生：王雅纖　　　　　　指導教授：楊啟瑞　博士

國立交通大學　資訊工程學系

## Abstract in Chinese

　　下一世代全光都會型網路(metropolitan area networks; MANs)旨在支援各類型要求高頻寬之網路應用程式以及訊務特性趨於動態變化之網路應用程式，全光封包交換技術(optical packet switching; OPS)，能夠滿足此類網路之需求，被視為是未來全光都會網路的一個典範。此篇論文提出之全光標頭交換與存取控制系統(optical-header processing and access control system; OPACS) 設計與實現，是應用在全光分波多工(wavelength division multiplexing; WDM)封包交換都會環型槽狀網路。OPACS 的設計具有兩項獨一無二的特色。首先， OPACS 設計之內頻控制(in-band)分時多工(time division multiplexing; TDM)全光標頭訊號技術，每個訊槽包含控制標頭以及資料負載，藉由波長與時間之轉換，OPACS 使所有全光平行控制標頭可在成本效益考量下，進行標頭的接收、修改、以及重送。再者，OPACS 系統提出之多用途媒介存取控制(medium access control ; MAC) 設計，稱為分散式多重粒度與視窗預訂(distributed multi-granularity and multi-window reservation; DMGWR)，DMGWR 的動態頻寬配置設計，特別適合應用在訊務量很高且訊務特性趨於動態變化之網路。基本上， DMGWR 為了確保每個網路節點能夠公平地存取網路頻寬，要求網路節點傳送資料前必須先提出預約需求，並藉由全域分散式佇列(global distributed queue)來達到網路頻寬公平性之配置。DMGWR 的多重粒度設計，讓節點可以一次預約多個訊槽。DMGWR 的多重視窗設計，當節點還有資料須要傳送(如大量突發訊務)，即便節點原預約資料還未

傳送完畢前，當預約次數還在系統視窗範圍(window size)內時，都能再提出新的預約需求。透過實驗模擬結果得知，相對於現存的兩種主要動態頻寬配置 HORNET DQBR 以及 WDMA 網路，OPACS 可以達到更為優異的系統輸出、頻寬效率、接取延遲、公平性以及大量突發訊務適應性表現。實驗結果也顯示，全光標頭交換能夠在一完全同步的方式進行標頭的刪除與整合，證明 OPACS 系統的可行性。

# Design and Experimentation of an Optical-Header Processing and Access Control System for a Packet-Switched WDM Metro Ring Network

*Student*: Ya-Shian Wang          *Advisor*: Dr. Maria C. Yuang

Department of Computer Science
National Chiao Tung University, Taiwan

## Abstract

Optical packet switching (OPS) has been considered to be a promising paradigm to support a wide range of applications with different time-varying and high bandwidth demands for future optical metropolitan area networks (MANs). This thesis presents the design of an experimental optical-header processing and access control system (OPACS) for an OPS WDM metro slotted-ring network. OPMACS is endowed with two distinctive features. First, OPMACS has been designed for a dual unidirectional slotted ring network using in-band signaling control. Each control header is in-band time-division-multiplexed with its corresponding payload within a slot. OPACS enables the optical headers across all parallel wavelengths to be efficiently received, modified, and re-transmitted by means of a wavelength-time conversion technique. Moreover, OPACS embodies a versatile medium access control (MAC) scheme, referred to as the distributed multi-granularity and multi-window reservation (DMGWR) mechanism, which is particularly advantageous for traffic of high and varying loads and burstiness. Basically, DMGWR requires each node to make reservation requests prior to transmissions while maintaining a distributed queue for ensuring fair access of bandwidth. By "multi-granularity", each node can

make a reservation of multiple slots at a time. By "multi-window", each node is allowed to have multiple outstanding reservations within the window size. Simulation pit the OPACS network against two other existing networks, simulation results show that the OPACS network outperforms these networks with respect to throughput, access delay, and fairness under various traffic patterns. Experimental results demonstrate that all optical headers are removed and combined with the data in a fully synchronous manner, justifying the viability of the system.

# **CONTENTS**

# List of Figures

# List of Tables

# Symbols

$N$        Number of nodes

$W$        Number of wavelengths

$D$        Inter-nodal distance

$L$        Offer load

$B$        Burstiness

$\alpha$        Probability of changing from State H to Ľ in a slot

$\beta$        Probability of changing from State Ľ to H in a slot

$\lambda_H$        Probability of changing from State H to Ľ in a slot

$\lambda_{Ľ}$        Probability of changing from State Ľ to H in a slot

$\lambda$        Mean arrival rate

# Acronyms

| | |
|---|---|
| ADM | Add/Drop Multiplexer |
| ASK | Amplitude Shift Keying |
| ATMR | Asynchronous Transfer Mode Ring |
| AWG | Arrayed Waveguide Gratings |
| CD | Countdown Counter |
| CSMA/CA | Carrier Sense Multiple Access with Collision Avoidance |
| DAVID | Data And Voice Integration over DWDM |
| DMGWR | Distributed Multi-Granularity and Multi-Window Reservation |
| DSWR | Distributed Single-Window Reservation |
| DPSK | Differential Phase Shift Keying |
| DQDB | Distributed Queue Dual Bus |
| DQBR | Distributed Queue Bidirectional Ring |
| EAM | Electro-Absorption Modulator |
| FBG | Fiber Bragg Gratings |
| FCFS | First Come First Serve |
| FIFO | First In First Out |
| FSK | Frequency Shift Keying |
| HOL | Head-Of-Line |
| HORNET | Hybrid Optoelectronic Ring Network |
| IP | Internet Protocol |
| MAN | Metropolitan Area Network |
| M-ATMR | Multiple Asynchronous Transfer Mode Ring |
| MMPP | Markov Modulated Poisson Process |
| MMR | Muliple MetaRing |
| MTIT | Multitoken Interarrival Time |
| NRS | The total Number of Reserved Slots |
| NRZ | NonReturn to Zero |
| OCS | Optical Circuit Switching |
| OCSS | Optical Carrier Suppression and Separation |
| OPS | Optical Packet Switching |
| OPACS | Optical-Header Processing and Access Control System |
| PQOC | Probabilistic Quota plus Credit |
| PRQ | Pending Request Queues |
| PSK | Phase-Shift Keying |
| RAM | Random Access Memory |

RingO       The Italian Ring Optical Network

RQ          Request Counter

RZ          Return to Zero

SCM         SubCarrier Multiplexed

SDH         Synchronous Digital Hierarchy

SOA         Semiconductor Optical Amplifier

SONET       Synchronous Optical Network

SRR         Synchronous Round Robin

TDM         Time Division Multiplexing

VOQ         Virtual Output Queue

WDM         Wavelength Division Multiplexing

WDMA        WDM Access Protocol

WS          Window Size

# Chapter 1. Introduction

## *1.1 Optical Networks: An Overview*

Over the last decade, advances in Internet technology brought about the proliferation of Internet-based multimedia applications, such as IPTV, remote terminal services, and on-line gaming. These applications virtually require the satisfaction of different time-varying and high bandwidth demand and stringent delay-throughput performance. Optical wavelength division multiplexing (WDM) [1-5] has been shown successful in providing virtually unlimited bandwidth to support a large amount of steady traffic based on the optical circuit switching (OCS) paradigm for long-haul backbone networks. Future optical networks, especially metropolitan area networks (MANs) [4-5] and local area networks (LANs) networks, are expected to flexibly and cost-effectively satisfy a wide range of applications having time-varying and high bandwidth demands and stringent delay requirements.

Regarding MANs, some slightly different optimization parameters are required [5,6]: (1) *Flexible upgrade*: The pace of bandwidth demand in metro calls for new solutions, much more flexible and scalable than traditional synchronous optical network/synchronous digital hierarchy (SONET/SDH) rings. Scalability is not just reaching huge capacities, but more being able to upgrade smoothly the system during operation with limited initial investment cost. WDM is obviously entering this market, and is expected to contribute to the network scalability. (2) *Optimized resource utilization*: Data traffic burstiness is obviously higher than in the backbone due to less efficient statistical multiplexing in a network much closer to the access and usually with simpler topologies. Next-generation metropolitan solutions will need to propose more sophisticated bandwidth and resource allocation management schemes, to

propose bandwidth-flexible services at affordable costs. (3) *Cost per transferred bit*: Cost is clearly of major importance in metropolitan area networks. Capacity and flexibility obviously have to be traded off with the added cost, although higher utilization of available resources will drive it down. (4) *Transparency*: Since a much greater variety of protocols coexist in the metropolitan market, compared to the core, a high level of transparency with respect to these protocols is expected to preserve the past investment of network operators.

Such facts bring about the need of exploiting the optical packet-switching (OPS) [2-5] paradigm that takes advantage of statistical multiplexing (i.e., fine-grained on-demand channel allocation) to efficiently share wavelength channels among multiple users and connections. OPS [2-5] has thus been considered to be a preeminent paradigm capable of supporting such applications over future optical WDM MANs. It is worth noticing that the OPS technique studied here excludes the use of optical random access memory (RAM) [4] and precise optical packet time synchronous technology, which is a significant technological limitations OPS faces.

In general, a WDM OPS switch consists of four parts: the input interface, switching fabric, output interface, and control unit [2-5]. The input interface is mainly used for packet delineation and alignment, packet header information extraction and packet header removal. The switch fabric is the core of the switch and is used to switch packets optically. The output interface is used to regenerate the optical signals and insert the packet header. The control unit controls the switch using the information in the packet headers. Because of synchronization requirements, optical packet switches are typically designed for fixed-size packets.

When a packet arrives at a WDM optical packet switch, it is first processed by the input interface. The header and payload of the packet are separated, and the header is converted into the electrical domain and processed by the control unit electronically.

The payload remains an optical signal throughout the switch. After the payload passes through the switching fabric, it is recombined with the header, which is converted back into the optical domain at the output interface.

## *1.2 Motivation and Objectives*

Numerous topologies and architectures [7-32] for OPS WDM MANs have been proposed in recent years. Of these proposals, the structure of slotted rings [7-28] receives the most attention. While most of the work [13,14,16-26] is simulation driven, only a handful [7-11,25-32] undertakes experimental prototypes. Two key challenges pertaining to OPS-based WDM networks are the header control and medium access controls. The header control can be in-band [25-32], where both header and payload are modulated and transported via the same wavelength, or out-of-band [7-24], where control headers are carried via a dedicated control wavelength. While both control methods have their merits, from a carrier's perspective, an out-of-band control system appears impractical due to the additional cost of a fixed transceiver on each node.

In addition, in-band control has several advantages over out-of-band control. First, in a mesh network, a layer-two wavelength switch can switch the header and payload together to an output port without examining the header. With out-of-band control, complicated control signal processing and routing are needed (e.g., optical burst switching) because the header and payload are carried on different wavelengths. Second, in-band headers can be used as a performance monitoring signal. Physical impairments on the payload can be monitored by detecting header signal's quality. Third, because all wavelengths' control information has to fit within a time slot, out-of-band control requires a costly high-data-rate control wavelength in order to support more channels, resulting in a scalability problem. Thus, this work focuses on

the in-band header control and processing.

There are three basic in-band header control techniques: subcarrier multiplexed (SCM) [25-28,33-50], orthogonal modulation [51-64], and time-domain-multiplexing (TDM) [26-28,65-67]. With the SCM technique, the header information can be carried on a subcarrier frequency that is separated from the baseband payload frequency. SCM requires stringent wavelength accuracy and stability if a fixed optical notch filer (e.g., a fiber Bragg grating) is used to remove the header at each node. Most traditional SCM methods cannot potentially scale up well with the payload data rate because an expanding baseband may eventually overlap with the subcarrier frequency. The optical carrier suppression and separation (OCSS) technique [33], however, was shown to be able to generate the header at very high subcarrier frequency and with high bit rate and extinction ratio. Nevertheless, SCM still requires stringent wavelength accuracy and stability while using a fixed optical filter to remove the header at each node.

The orthogonal modulation technique, which includes amplitude shift keying (ASK) [38], frequency shift keying (FSK) [54], ASK/differential phase shift keying (DPSK) [55,56], and DPSK/FSK [57], exhibits severe transmission system penalty due to the inherently low extinction ratio of a high-speed payload signal. Finally, with the TDM technique, the header and payload are serially connected in the time domain, interspaced with an optical guard time to facilitate header extraction and modification. The bit rates of the header and payload can either be the same [65], or different [66,67]. Generally, traditional TDM-based approaches require an extremely precise control timing and alignment to perform header erasing and rewriting operations. The first goal of this work is to propose a simple and highly efficient TDM-based optical header processing scheme. As will be demonstrated, in our system optical headers can also be easily modified by taking advantage of the particularly notable MAC design.

Another key performance-enhancing feature pertaining to OPS-based networks is the design of the medium access control (MAC) mechanism. The MAC scheme should be designed to offer fair and versatile bandwidth allocation, achieving satisfied throughput and delay performance under a wide range of traffic loads and burstiness. Moreover, the MAC protocol should take into account the scalability problem with respect to the number of wavelengths. While numerous MAC protocols for OPS-based slotted-ring networks have been proposed in the literature [10-14], our second goal is to explore a variant of a reservation-based mechanism, IEEE 802.6 Distributed Queue Dual Bus (DQDB) [72], for the multi-channel WDM metro networks. In single-channel DQDB, each node must issue a reservation request prior to the transmission. To ensure that packets are sent in the order they arrived at the network, DQDB requires each node to maintain a distributed queue via a Request (RQ) and a CountDown (CD) counters. DQDB was shown to achieve superior throughput and delay performance, nevertheless undergoes the unfairness problem due to long propagation delay under heavy traffic conditions.

In this WDM-DQDB line of work, the WDM Access (WDMA) [13] protocol simply extends the basic single-channel DQDB to the multi-channel case, namely, each node maintains a single distributed queue for all of the wavelengths. Due to the use of a tunable transceiver, WDMA adopts the retransmission mechanism if the receiver contention problem occurs. With such a simple design, WDMA unfortunately results in access unfairness and inefficiencies for multi-channel networks under varying traffic patterns and burstiness. The hybrid optoelectronic ring network (HORNET) [10] employs a distributed queue bidirectional ring (DQBR) protocol. Due to the use of fixed-tuned receivers, HORNET statically assigns each node a wavelength as the home channel for receiving packets. Such static wavelength

5

assignment results in poor statistical multiplexing gain and bandwidth efficiency. As a result of the home-channel design, DQBR treats wavelengths independently and requires each node to maintain a distributed queue for each wavelength. Moreover, each node is allowed to issue multiple independent single-slot requests. With DQBR, HORNET achieves acceptable utilization and fairness at the expense of high control complexity for maintaining the same number of counter pairs as that of wavelengths. Such a design gives rise to a scalability problem. Moreover, the design of permitting unlimited multiple requests with single slot granularity per request unfortunately results in unfairness problems.

A novel OPACS (optical-header processing and access control system) for a 10-Gb/s optical packet-switched [26-28] WDM metro ring network is presented in this dissertation. OPACS has two prominent features that set it apart from existing related work. First, OPACS is designed for a dual unidirectional ring network using in-band signaling control. Each control header is time-division-multiplexed with its corresponding data packet within a slot. By making use of signal gating and wavelength-time conversion techniques, OPACS enables the optical headers across all parallel wavelengths to be efficiently received, modified, and re-transmitted. Second, taking diverse traffic patterns and burstiness into account, OPACS employs a variant of the DQDB scheme, referred to as the distributed multi-granularity and multi-window reservation (DMGWR) scheme. By "multi-granularity", DMGWR permits each node to reserve different amounts of bandwidth (slots) at a time. By "multi-window", DMGWR allows each node to have multiple outstanding reservations within the window size. From numerical results that pit the DMGWR network against two other existing networks (WDMA-based and HORNET), we show that the OPACS network outperforms both networks with respect to throughput, access delay, and fairness under various traffic patterns. Experimental results

demonstrate that all optical headers are removed and combined with the data in a fully synchronous manner, justifying the viability of the system.

## *1.3 Organization of the Thesis*

The remainder of this thesis is organized as follows. Chapter 2 provides a review of existing related optical header processing techniques, experimental testbeds, existing related MAC protocols and fairness control schemes for OPS-based WDM metro ring networks. Chapter 3 presents the OPACS system architecture and node architecture. Chapter 4 introduces the design concepts and detailed operations of the MAC protocol. Chapter 5 evaluates comparatively the performance of the protocol introduced in Chapter 4. Chapter 6 presents the experimental setup and results. Finally, conclusion remarks of this thesis and some future works are provided in Chapter 7.
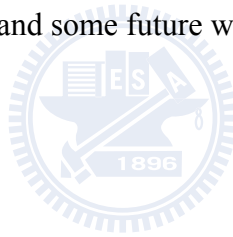
# Chapter 2. OPS-based Metro Ring Networks

This chapter provides a review of existing optical header processing techniques, experimental testbed systems, and the existing MAC protocols related to OPS based metro ring networks. Following a short introduction, in-band header control techniques are investigated. In general, the low-speed label associated with a high-speed payload is extracted, processed, and replaced at every intermediate network-switching node. Meanwhile, the high-speed payload is optically switched (controlled by the electrically processed label) to an appropriate output fiber as an entirely untouched entity. The subsequent section will assess four experimental testbed systems that are relevant to this work. The third subsequent section will provide a review of existing MAC protocols and fairness control schemes for OPS-based WDM metro ring networks. The access control techniques are investigated by means of the different buffer selection strategies to achieve high channel utilization and low access delay. Furthermore, fairness control mechanisms are described in the next subsequent sections with particular emphasis on static bandwidth allocation control schemes and explicit dynamic bandwidth allocation schemes.

## 2.1 Optical Header Processing Techniques

The header control can be in-band [25-32], where both header and payload are modulated and transported via the same wavelength (see Figure 2.1), or out-of-band [7-24], where control headers are carried via a dedicated control wavelength, as depicted in Figure 2.2. Both control methods have their merits, from a carrier's perspective, an out-of-band control system appears impractical due to the additional cost of a fixed transceiver on each node.
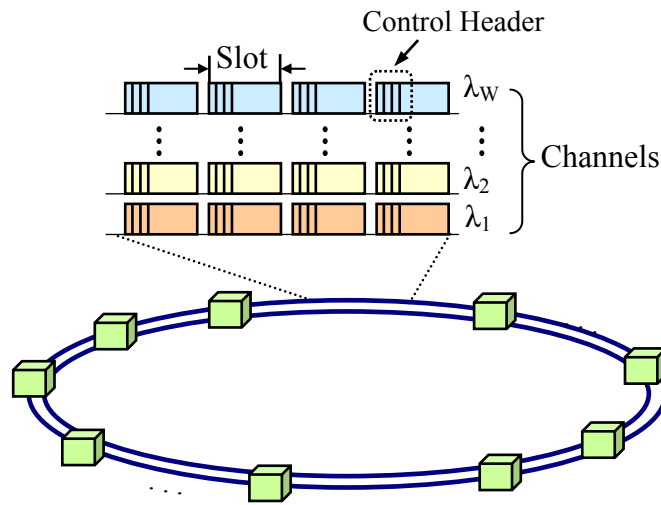
Figure 2.1.In-band header control architecture.
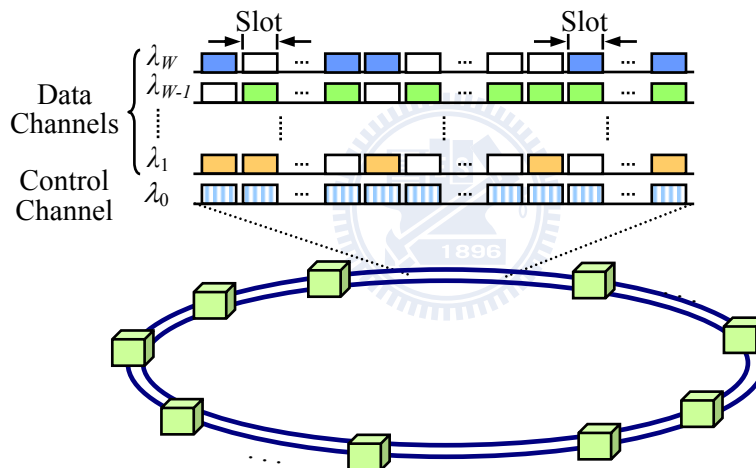


Figure 2.2. Out-of-band header control architecture.

In an in-band based OPS network [4,33], the optical data packets contain a payload and a header or label. Optical header control technique is an important aspect of OPS and it involves the extraction and processing of the headers so that the packets can be routed to their correct destinations. In an OPS network, the optical packets are first encapsulated with optical labels as they enter the network. Once they are in the network, only the optical header undergoes OE conversion in the OPS router, so that the packet's routing information and other auxiliary data, such as the wavelength and the bit rate of the payload, can be

determined. Since the payload remains in optical form from source to destination, it can be encoded at very high bit rate using any modulation format. To make the header control procedure very efficient, the header must be easily separated from the payload and it has to be processed at very high speeds.

Basically, the optical header scheme [4] should: 1) impose a minimum overhead to the data networking capacity, 2) support simple header replacement techniques, 3) maintain high signal fidelity across fiber transmission and cascaded nodes, and 4) impose low crosstalk between the header and the payload. Many in-band optical header processing techniques have been proposed and exhibit different strengths and limitations. There are three basic techniques: subcarrier multiplexed (SCM) [25-28,33-50], orthogonal modulation [51-64], and time-domain-multiplexing (TDM) [26-28,65-67].

### 2.1.1 Subcarrier Multiplexing (SCM) Header Technique

One of the most popular optical header techniques is based on subcarrier multiplexing (SCM) [25-28, 33-50] of the header with the data payload on the baseband. Initially, double-sideband SCM optical header techniques were widely utilized; however, the double-sideband SCM signal multiplexing studied in the mid 1980s [44], [45] for video transmission was known to cause RF fading, causing signal reception problems at particular distances in the network. RF fading is a result of the coherent interference between the carrier and the two sidebands, which constructively or destructively interfere depending on their relative phases determined by the initial phase conditions and the total dispersion (the product of the dispersion coefficient and the fiber transmission length).

An important issue in SCM-based optical header processing networks is the

ability to support a coarse alignment between the label and the payload, allowing a more asynchronous replacement of the SCM-encoded label [33]. This feature is significant, considering that chromatic dispersion may affect the arrival times of the header and payload signals; in addition, the low-cost electronic devices involved in the label processing may undergo frequency drifting.

To defeat the RF fading effect, a number of new techniques including carrier suppression method in optical header extraction and replacement [51] as well as single-side optical labeling technique with carrier suppression [33] have been investigated. Both techniques utilized relatively simple optical filtering techniques such as fiber Bragg gratings (FBGs) or arrayed waveguide gratings (AWG) for separating subcarrier components and showed successful penalty-free operations. Cascaded optical router operation [49], polarization and dispersion dependent fiber transmission, and first field trial across 477 km have been demonstrated [50] using the double-sideband optical header technique.

### 2.1.2 Orthogonal Modulation Header Technique

One of the most innovative optical header technologies employs the header and the data information modulated in orthogonal modulation formats of each other (e.g., in amplitude and phase domains) [55-57]. It is also possible to place two levels of labels [56] in the phase and wavelength domains for the amplitude-domain data. Demonstrated techniques involved ON–OFF keying intensity modulated (OOK-IM) data payloads with frequency shifted keying (FSK) headers [57] as well as 2.5-Gb/s OOK-IM headers with 40-Gb/s return-to-zero differential phase-shift keying (RZ-DPSK) data payloads [53]. Chi et al. [51] showed a header (label) replacement technique involving SOA, EAM, and highly nonlinear fiber for 10-Gb/s IM data

payloads and two-level optical headers (labels) at 2.5-Gb/s DPSK and the wavelength domain. In both cases of ASK labels and FSK/phase-shift keying (PSK) labels, there are tradeoff considerations related to extinction ratios in the two orthogonal domains. Likewise, crosstalk rejections in real systems become difficult since the two domains are coupled due to, for example, frequency chirping caused during amplitude modulation and vice versa. Zhang et al. [57] used integrated EAM-DFB lasers for FSK label modulation, where offsetting amplitude modulations in EAM cancel the unintentional amplitude modulation during FSK label modulation. Additional orthogonal modulation schemes include polarization modulations [59,60], wavelength-shifted keying for pulse position modulation [63], and embedded DPSK label in ASK data payload [64].

### 2.1.3 Time Division Multiplexed (TDM) Technique

Initial OPS technologies including TDM OPS technologies utilized a time-division-multiplexed (TDM) labels header technique, similar to the synchronous optical network/synchronous digital hierarchy (SONET/SDH) overhead. The label and the payload are serial in the time domain, interspaced with an optical guard time to facilitate label extraction and processing. They primarily pursued synchronous and fixed-length packet switching mainly because the TDM label already required relatively strict timing control. The KEOPS utilized the header line rate of 622 Mb/s and the flexible payload line rate of up to 10 Gb/s [65]. Both headers and payloads are led by synchronization bits to facilitate burst mode clock recovery. Upon reading the packet headers, all packets were synchronized by the optical synchronizer in the OPS system. Time domain header replacement requires time switching to remove the old header and to attach a new header to the payload. Since the header format includes a

small number of header bits (typically less than 100 bits), the overhead is very small, especially if a large data payload format (> 15 000 bytes) is adopted. Like in other OPS technologies, typical line rates for the header can be much lower than those for the data payload, and thus, relatively simple electronics can process the header and look up the forwarding table after reading the optical header content using a burst-mode optoelectronic receiver.

The bit rates of the label and the payload can be the same [65] or different [26-28,]. This method is straightforward to implement at the transmitter end, since the label and the payload are both in baseband formats and use the same wavelength channel. It also provides possibilities for all-optical regeneration and label processing [67]. However, the label receiving may require sophisticated synchronization and timing control. In some cases, accurate control signals have to be generated at each hop to inform the label processors of the temporal positions of the labels. In other cases, different power levels or coding formats (such as return to zero (RZ) and non return to zero (NRZ)) are taken to distinguish the label and the payload. Relatively complicated signaling or receiver designs make this bit-serial method difficult for practical applications. Moreover, the label and the data payload occupy separate time spaces, limiting the available data throughput.

## 2.2 Experimental Testbeds for Packet-Switched Ring Metro WDM networks

Numerous topologies and architectures for OPS-based WDM metro ring networks have been proposed and exhibit different strengths and limitations. Of these proposals, the structure of slotted rings [7-32] receives the most attention. While most of the work [13-24] is simulation driven, only a handful [8-12,25] undertakes

experimental prototypes. In this section we survey four of the most recent experimental testbed systems relevant to this work for packet-switched ring metro WDM networks: HORNET [10, 25], RingO [12], DAVID [9], and HOPSMAN [7, 8].

## 2.2.1 Hybrid Optoelectronic Ring NETwork (HORNET)

The original version of Hybrid Optoelectronic Ring NETwork (HORNET) is a unidirectional WDM ring network [25]. Each node is equipped with one fast tunable transmitter (TT) and one fixed-tuned burst mode receiver (FR). The HORNET node structure consists of three subsections: slot manager, smart drop, and smart add module. The header control of HORNET is in-band SCM-based control. The destination address of a packet is modulated onto a SCM tone using a combination of ASK and FSK. For carrier senseing, the slot manager taps off some optical power for subcarrier recovery to perform two functions in parallel. It monitors the subcarriers (carrier sense) and relays the wavelength occupancy information to the smart add. It also demodulates the subcarrier (FSK demodulation) that corresponds to the nodes drop wavelength, recovers the address, and informs the smart drop whether the incoming packet is destined for itself or for a downstream node. The smart drop module drops a fixed wavelength using a circulator and a fiber Bragg grating. The dropped wavelength is detected inside a burst mode receiver that recovers the packet bit clock. It then uses the address information provided by the slot manager to switch the received packet either to the LAN or to a retransmit queue, where it waits to be multihopped to a downstream node. The smart add module chooses a transmission wavelength depending on the destination node of the queued packet and the wavelength availability information from the slot manager to avoid collision avoidance. It then tunes the fast tunable laser transmitter to the target wavelength and

modulates the packet on to the optical carrier. Access to all wavelengths is governed by means of a Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) protocol.

As extended the first version of unidirectional TT-FR, the second version of HORNET is a bi-directional WDM slotted ring network [10]. The HORNET architecture is designed to cost-effectively scale beyond 1 Tb/s while efficiently transporting bursty and randomly fluctuating traffic. Each node is equipped with one tunable transmitter and one fixed-tuned receiver for each ring. A node contains a wavelength drop for the node's drop-wavelength on each ring, a tunable transmitter subsystem for each ring, a wavelength add and drop for the control channel wavelength on each ring, and a node controller. The node's protocols are implemented in programmable logic devices (PLDs) on the node-controller circuit board clocked at 125 MHz. A Gigabit Ethernet (GbE) chip set is used for the transmission and reception of the control channel in the testbed. Gain-clamped semiconductor optical amplifiers (GC-SOAs) are used to provide linear gain in the testbed. The header control of HORNET is out-of-band control. The extended HORNET version proposed a distributed control-channel based MAC protocol, called Distributed Queue Bidirectional Ring (DQBR).

### 2.2.2 Ring Optical Network (RingO)

The Italian Ring Optical Network (RingO) [12] project used a unidirectional slotted WDM /Time division multiplexing (TDM) architecture. The node structure based on AWGs, was first proposed because of the major flexibility given by fully demultiplexing all channels on separate fibers. In addition, the node is equipped with an array of fixed-tuned transmitters and one fixed-tuned receiver. The number of

nodes in the first RingO version is equal to the number of wavelengths. Thus, each node has its own dedicated wavelength for packet reception. All wavelengths are slotted with the slot length equal to the transmission time of a fixed-size packet plus guard time. Each node performs $\lambda$-monitoring to check the wavelength occupation state on a slot-by-slot basis to avoid channel collision. Such a simple design gives rise to a scalability problem.

The second RingO node design is based on an add–drop filter, allowing for better cascadability and less stringent physical constraints. While this structure is similar to the first version for network functionalities, it is significantly different from the physical layer point of view. The input–output optical path is greatly simplified, and consists only of a passive optical splitter and a fixed add–drop filter tuned on the wavelength that must be received locally. This setup greatly reduces node attenuation, self-filtering, and PDL effects, and allows a higher node cascadability.

### 2.2.3 Data And Voice Integration over DWDM (DAVID)

The European IST project proposed DAVID (Data And Voice Integration over DWDM) networks [9]. The DAVID MAN consists of a number of unidirectional slotted WDM rings of metropolitan dimensions, which collect traffic from several ring nodes. These nodes provide an electro/optical interface to edge routers/switches at the end of access networks via a variety of legacy interfaces (e.g., Gigabit Ethernet in business areas, PONs in mixed or residential areas, cable head-ends, or any other legacy system). The WDM rings are interconnected to other rings via a bufferless hub, and to a mesh of packet-switched OPRs in the core creating the complete optical WAN. The rings can be either physically disjoint, or be obtained by partitioning the optical bandwidth into disjoint portions.

In DAVID, a broadcast-and-select architecture, which ensures nonblocking performance, is chosen, using semiconductor optical amplifier (SOA) technology [21]. A major issue in every packet switched network is contention resolution. In electronic routers this problem is tackled using random access memory (RAM), which is unfeasible in the optical domain. Since all optical buffers today are technologically hard to realize, there seems to be a consensus that they should be avoided as much as possible or at least be limited to a minimum. As stated before, the MAN is completely bufferless in the optical domain. In the WAN, a shared recirculating FDL buffer is used to help solving contention, where exploitation of the wavelength domain does not suffice.

The hub node is used to forward optical packets between ring networks, as well as to interconnect the metro area to the backbone through an electronic Gateway. The hub is an SOA-based optical packet switch capable to cope with a very high level of traffic (Terabit/s). The lack of real optical memories is compensated through the use of an extended multi-ring MAC protocol. The optical hub is, thus, bufferless and its structure is similar to the one of the optical packet router in the backbone but with reduced targeted final capacity. The main difference between the hub and the OPR is at the control level: the optical hub is configured by a controller which exploits the control channels of each connected ring network, in order to calculate the switching permutation. The hub comprises synchronization stages, a space switching stage, a wavelength switching stage, and regeneration stages if required (depending on the power budget). Each WDM channel operates at 10 Gb/s that with 32 wavelengths per ring and a channel spacing of 100 GHz, occupy 24 nm of bandwidth per ring; this corresponds to a reasonable optical bandwidth for the introduction of a SOA-based technology. The maximum capacity of one ring becomes 320 Gb/s.

Ring access nodes in the DAVID MAN are composed of an electronic part and

an optical part. The electronic part realizes the adaptation with client layers, which. is performed in the traffic manager board (TMB). Specific burst mode transceivers (BMTs) are used to send/receive optical packets to/from the optical packet ring networks.

### 2.2.4 High-Performance Optical Packet-Switched WDM Metro ring Network (HOPSMAN)

The High-Performance Optical Packet-Switched WDM Metro ring Network (HOPSMAN) [7,8] testbed system has a scalable architecture in which the node number is unconstrained by the wavelength number. It encompasses a handful of nodes (called server nodes) that are additionally equipped with optical slot erasers capable of erasing optical slots resulting in an increase in bandwidth efficiency.

Nodes in HOPSMAN are interconnected via a single unidirectional fiber that carries multiple WDM data channels (10-Gb/s) and one control channel (2.5-Gb/s) containing the status of data channels. Channels are further divided into synchronous time slots. Nodes are equipped with one fixed transmitter/ receiver for accessing the control channel; and one or multiple tunable transmitter(s)/receiver(s) for dropping/ adding packets from/to data channels on a slot basis.

HOPSMAN has three types of nodes- POP-node (P-node), Ordinary-node (O-node), and Server-node (S-node). A P-node is a gateway between HOPSMAN and long-haul networks, and typically includes multiple tunable transmitters/receivers. An O-node is a regular node with one tunable transmitter/receiver. Finally, an S-node is an O-node but additionally equipped with a slot-eraser device, making bandwidth reusable and thus achieving greater bandwidth efficiency. Notice revealed by our study that, bandwidth efficiency improves greatly with only a few S-nodes in a network.

The HOPSMAN node architecture is shown in Figure 2.3. First, a fixed optical drop filter extracts the control channel information. In coordination with the SYNC Monitoring Module, the Channel Timing Processor is responsible to identify the beginning of a control/data slot. Notice that, the slot boundaries of the control and data channels are aligned during transmissions. With status of data channels, the MAC Processor mainly performs the MAC scheme, namely the determination of the add/drop/erase operations and the status updates of the associated control channel mini-slots. channel slot signal is transmitted via the fixed transmitter and combined with data channel slots via the optical add filter (OAF).
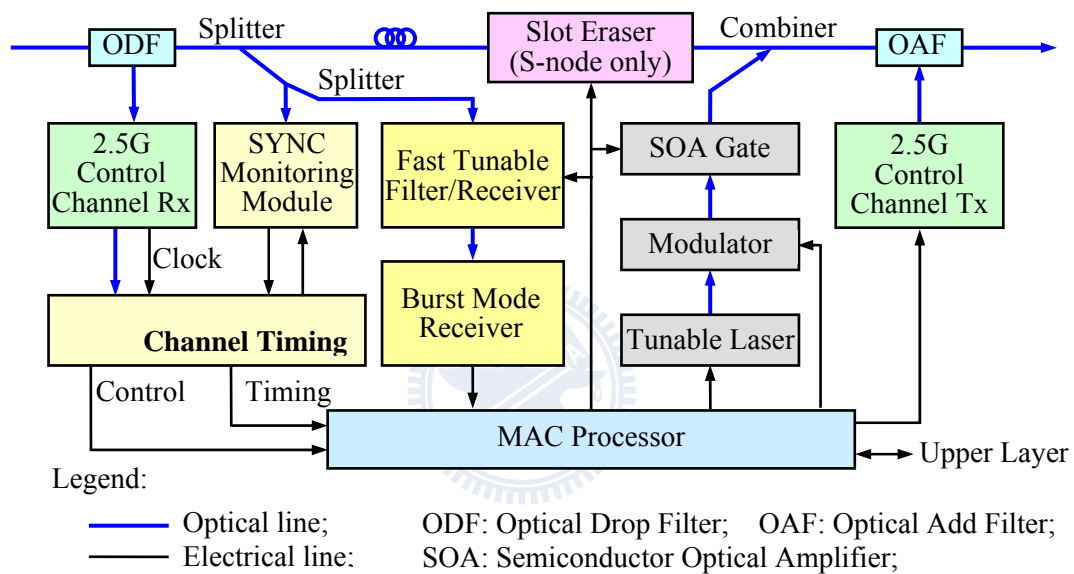
Figure 2.3. General node architecture of HOPSMAN.

## 2.3 Existing MAC protocols for OPS-based WDM Metro ring networks

Another key performance-enhancing feature pertaining to OPS-based networks is the design of the medium access control (MAC) mechanism. In general, for various node configurations of transmitters and receivers, access contentions can be either channel collisions or receiver collisions [6]. When a node inserts a packet on a given wavelength while another packet is currently passing the ring on the same wavelength, a channel collision occurs and both packets are disrupted. Receiver collisions are also known as destination conflicts when the destination node does not know about the transmission or another packet is currently received on a different wavelength. Clearly, both channel and receiver collisions have a detrimental impact on the throughput-delay performance of the network.

The degradation of network performance due to channel or receiver collisions can be mitigated or completely avoided at the MAC protocol level. Clearly, MAC protocols are required to govern, mitigate or even completely prevent access conflicts on the WDM channels shared among competing nodes. Numerous MAC protocols and fairness schemes for OPS based WDM metro ring networks have been proposed and exhibit different strengths and limitations.

### 2.3.1 MAC Protocols for OPS-based WDM Metro ring networks

This subsection describes four of the most recent MAC protocols of slotted OPS-based ring networks. Most of proposed protocols [17-24,32] provide collision-free transmission through various buffer selection policies or packet selection strategies to resolve contention conflicts among competing nodes. Basically, these protocols can achieve high channel utilization and low access delay and allow for relatively simple access schemes.

### 2.3.1.1 *Synchronous Round Robin (SRR) Ptotocol*

The Synchronous Round Robin (SRR) protocol is kind of destination stripping for a unidirectional WDM ring network i[th] fixed-size time slots. Each node is equipped with one tunable transmitter and one fixed-tuned receiver (TT-FR). In SRR [23], each node has $(N - 1)$ separate first-in-first-out (FIFO) VOQs, one for each destination. SRR adopts an a priori access strategy. Specifically, each node scans the VOQs in a round robin manner on a per-slot basis, looking for a packet to transmit. When the current slot is occupied, that is, a transmission is not possible as it would result in a channel collision, then no packet is transmitted from the selected VOQ. For the transmission attempt in the next slot, the next VOQ is selected according to the round-robin scanning of SRR. If the selected VOQ is nonempty, the first (oldest) packet is transmitted. If the selected VOQ is empty, SRR selects the first packet from the longest queue among the remaining VOQs to transmit.

Under heavy uniform load conditions, when all VOQs are non-empty, the SRR scheduling algorithm converges to round-robin TDMA. For uniform traffic, SRR asymptotically achieves a bandwidth utilization of 100 percent. However, the presence of unbalanced traffic leads to wasted bandwidth due to the nonzero probability that the a priori access strategy selects a wavelength channel whose slot is occupied while leaving free slots unused. It was shown that a posteriori access strategies avoid this drawback, resulting in an improved throughput-delay performance, albeit at the expense of increased complexity.

SRR achieves good performance requiring only local information on the backlog of the VOQs, which also avoids the well-known head-of-line (HOL) blocking problem. Owing to destination stripping, slots can be spatially reused several times as they propagate along the ring. On the other hand, slot reuse raises fairness control

problems, particularly for nonuniform traffic. A node to which a large amount of slots is directed generates a large amount of free slots, and nodes immediately downstream are in a favorable position with respect to other nodes.

### 2.3.1.2 *Multitoken interarrival time (MTIT) Protocol*

The Multitoken Interarrival Time (MTIT) access protocol is a token-based access scheme for a CC-FT$^W$-FR$^W$ unidirectional WDM ring network [32]. For each data channel, every node has one fixed-tuned transmitter, one fixed-tuned receiver, and one on-off optical switch. A dedicated wavelength is used as the control channel for the purpose of access control and ring management. The on-off switches are used to control the flow of optical signals through the ring and prevent re-circulation of the same packet on the ring. Once transmitted by the source node, the packet makes one round trip in the ring and is removed from the network by the same source node, that is, MTIT employs source stripping.

Channel access is regulated by a multitoken approach. Each channel is associated with one token that circulates among the nodes on the control channel and regulates access to the corresponding data channel. The MTIT protocol controls the token holding time by means of a target token interarrival time with value (TTIT). The TTIT is agreed upon by all nodes connected to the ring at the configuration time of the system. Upon a token arrival, the node is allowed to hold the token for a period of time equal to TTIT – TIAT, where TIAT is the actual token interarrival time between that token's arrival time and the arrival time of the token held previously. If the token holding time is up, then the node must finish the currently ongoing packet transmission and release the token. If TIAT exceeds TTIT, then the token is late and must be released immediately. In any case, if a node has no packets to transmit, then

that node must release the token immediately. Since MTIT uses source release, it can guarantee fair access to the ring if the nodes' timers operate within a certain timing tolerance and the maximum packet length is bounded.

With the $FT^W$-$FR^W$ node structure, MTIT avoids receiver collisions and allows each node to simultaneously use multiple data wavelength channels. However, the number of transceivers at each node is rather large. MTIT achieves low access delay due to the fact that a node has the opportunity to grab a token more frequently than in conventional token rings where a node has to wait one round-trip time for the next token. A unique feature of MTIT is its capability to self-adjust the relative positions of tokens along the ring circumference and maintain an even distribution of the token position. As a result, the variance of the token inter-arrival time is low, guaranteeing to every node a consistent channel access delay in support of high-priority traffic. On the other hand, the capacity of MTIT is smaller than that of destination-stripping ring networks since source stripping does not allow for spatial wavelength reuse. For uniform traffic it was shown that MTIT achieves high bandwidth efficiency and low access delay for varying packet sizes even in relatively large (thousands of kilometers) networks.

### 2.3.1.3 *Posteriori VOQ Selection Protocols*

Bengi and van As [21,22] proposed several posteriori buffer selection schemes for the HORNET architecture [25]. Each node is equipped with one fixed-tuned transmitter and one tunable receiver (FT-TR). In an empty-slot protocol, each unused slot on any wavelength channel can be used for packet transmission by a source node. However, when more than one wavelength channel carries an empty slot in the current slot period, one packet (or equivalently, one VOQ) corresponding to one of the

empty channels has to be chosen according to a prescribed selection rule. Due to the short time between channel inspection and packet transmission, the a posteriori packet selection process has to be performed at a high speed in the electronic domain, which increases the processing complexity compared to an a priori packet selection scheme. Five different a posteriori VOQ selection strategies are described and examined:

- Random Selection: The VOQ from which a packet is to be transmitted is selected randomly according to a uniform distribution.

- Longest Queue Selection: The longest VOQ is chosen upon buffer contention.

- Round-Robin Selection: The VOQ is chosen in a round robin fashion.

- Maximum Hop Selection: The packet (VOQ) associated with the maximum hop distance between source and destination node is selected when buffer contention arises.

- C-TDMA Selection: The channel-oriented TDMA (C-TDMA) scheme first attempts to select the packet according to a round-robin policy. If that selection would prevent a transmission, either due to an empty VOQ or an occupied slot, then the longest VOQ that allows for a packet transmission is chosen. This scheme is largely equivalent to the SRR scheme with a posteriori access It was found that the random and round-robin buffer selection schemes provide a satisfactory compromise between performance and implementational complexity.

### 2.3.1.4 *Source Stripping and Destination Stripping Protocols for TT-FR$^W$*

Another approach, proposed by Jelger and Elmirghani [17-19], is the use of source stripping and destination stripping protocols for the HORNET architecture [25]. Packets are buffered in a single FIFO transmit queue at each node. In the proposed

source-stripping scheme, a sender must not reuse the slot it just marked empty. The destination stripping mechanism prevented a node from reusing a slot on its assigned wavelength as soon as it was marked empty in order to prevent a single node from starving the entire network. However, this mechanism failed to achieve a complete fairness across all of the nodes sharing the same wavelength, particularly when the network traffic pattern was unbalanced.

By means of simulation it was shown that destination stripping outperforms source stripping in terms of throughput, delay, and packet dropping probability. With only one tunable receiver at each node, receiver collisions can occur. Receiver collisions can be avoided in a number of ways. In one approach, arriving packets that find the destination's receiver busy re-circulate on the ring until the receiver of the destination is free, that is, is tuned to the corresponding wavelength [17-19]. Alternatively, receiver collisions can be completely avoided at the architecture level by replacing each node's tunable receiver with an array of W fixed-tuned receivers, each operating at a different wavelength (FT-FR$^W$). Another proposal to resolve receiver contention is based on optical switched delay lines (SDLs). A destination node puts all simultaneously arriving packets except one into optical delay lines such that packets can be received sequentially.

### 2.3.2 Fairness Control Protocols for OPS based Metro Ring Networks

Although selection policies (described in the previous subsection) achieve high network efficiency, yet they do not address the inherent fairness issues among ring nodes. As several ring nodes share common channels, upstream nodes may grab all the available bandwidths, and the downstream nodes would possibly starve. In general, fairness control schemes limit the transmission of upstream nodes in an attempt to

keep enough bandwidth for downstream nodes. Existing fairness control schemes can be categorized into two classes: static bandwidth allocation control and reservation-based fairness control. Basically, the static bandwidth allocation mechanisms regulate the access by providing equal access opportunity to all competing nodes. Contrastingly, various dynamic bandwidth allocation fairness control mechanisms attempt to improve the deficiency and ultimately solve the instability of the static-based fairness protocols.

This subsection describes the fairness protocols of slotted OPS-based ring networks. To avoid starvation, the transmission rate of nodes has to be controlled in order to achieve fairness among all nodes. However, restricting nodes in their transmission decreases channel utilization. In general, there is a tradeoff between fairness and channel utilization.

### 2.3.2.1 *Multi-MetaRing (MMR) Protocol*

The Multi-MetaRing (MMR) [14, 15] algorithm adapts a mechanism originally proposed for the single-channel MetaRing [71] high-speed electronic metropolitan area network. Fairness in the MetaRing is achieved by circulating a control message, named SAT (short for SATisfied). Each node is assigned a maximum number of packets to be transmitted between two SAT visits; this maximum number of packets is the node's quota or credit. Each node normally forwards the SAT message on the ring with no delay, unless it is not SATisfied in the sense that it has not transmitted the permitted number of packets since the last time it forwarded the SAT. The SAT is delayed at unSATisfied nodes until SATisfaction is obtained, that is, either the node packet buffer is empty or the permitted number of packets has been transmitted.

In the MMR Single SAT (MMR-SS) scheme, a single SAT message regulates the

transmissions of all nodes on all wavelength channels. Each node can transmit up to K packets to each destination since the last SAT visit. Each SATisfied node forwards the SAT to the upstream node. Thus, the SAT logically rotates in the opposite direction with respect to data (although the physical propagation is co-directional). With this scheme the SAT propagation delays are very large since the SAT message has to traverse almost the entire network to reach the upstream node. Alternatively, the MMR Multiple SAT (MMR-MS) scheme uses one SAT message for each wavelength. It was shown in that this MMR-MS scheme is generally the preferable extension of the MetaRing fairnesscontrol scheme to a WDM ring.

### 2.3.2.2    *Multiple ATMR (M-ATMR) Protocol*

The access protocol discussed earlier suffers from fairness problems due to destination stripping. In [21, 22] Bengi and van As adopted an extension of the well-established Asynchronous Transfer Mode Ring (ATMR) [70] fairness protocol to the multiple channel WDM ring case. This extension is M-ATMR. In M-ATMR each node receives a prescribed number of transmission credits for each destination. When a node has used all its credits or has nothing to send, it transitions into the inactive state. In order to properly apply the credit reset mechanism, every node has to know which node was the last active node. To achieve this, each active node overwrites a so-called busy address field in the header of every incoming slot with its own address. (The busy address field may be included in the SCM header of each WDM wavelength channel.) Thus, a node receiving a slot with its own busy address knows that all the other nodes are inactive. If the last active node detects inactivity of all the other nodes, it generates a reset immediately after its own transmission. The reset mechanism causes the nodes to reset their credits to the predefined values. In this

manner, it is guaranteed that every node uses a maximum number of slots between two subsequent reset cycles. It was shown in [] that the M-ATMR fairness protocol applied for best-effort traffic provides throughput and delay fairness for both uniform and client/server traffic scenarios.

### 2.3.2.3    *WDM Access (WDMA) Protocol*

The WDM Access (WDMA) [13] protocol simply extends the basic single-channel DQDB [72] to the multi-channel case, namely, each node keeps a record of RQ and CD counters to maintain a single distributed queue for all of the wavelengths. Each node is equipped with one tunable transmitter and one tunable receiver (TT-TR) for each ring. WDMA adopts the retransmission mechanism if the receiver contention problem occurs. As shown in Figure 2.4, each control slot in WDMA consists of a Busy bit, a Request bit, a Destination Address (DS) field, and a Timestamp (TS) field. When a new packet arrives at a node, a unique timestamp is assigned to it. Similar to DQDB, each node issues a request bit in a minislot on the reverse ring. Then the node calculate the values of RQ and CD counters. As the vlaue
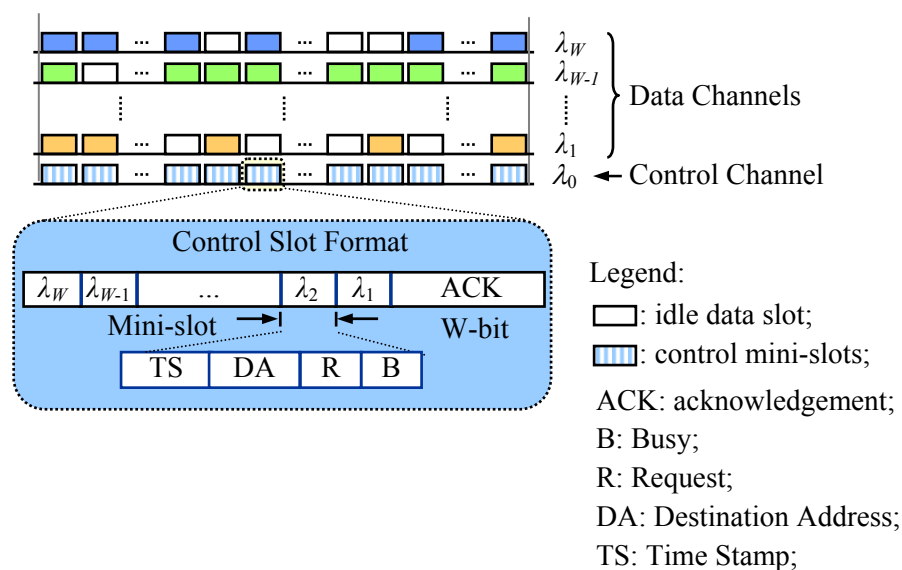


Figure 2.4 . The control slot structures of WDMA.

of CD counter changes to zero, the request reaches the top of distributed queue. The node waits for the next idle data slot to send the packet by setting Busy bit to one and writing the destination address and timestamp of the packet into the DA and TS fields. To receive data, each node constantly monitors the control channel. In case of destination conflict where more than one packet is addressed to the same destination in a slot, the packet with the smallest timestamp wins the contention. The headend exiamines each control slot by computing the results of receiver-contention according to the timestamp ordering, and writing receiver-contention into the acknowledge field in the next slot. With such a simple design, WDMA unfortunately results in access unfairness and inefficiencies for multi-channel networks under varying traffic patterns and burstiness.

### 2.3.2.4 *Distributed Queue Bidirectional Ring (DQBR) Protocol*

The extended version of HORNET [10] employs a distributed queue bidirectional ring (DQBR) protocol, which is a variant of DQDB [72] protocol. The DQBR fairness protocol works as follows. In each control-channel frame, a bit stream of length W bits, called the request bit stream, follows the wavelength-availability information. When a node on the network receives a packet in VOQw, the node notifies the upstream nodes about the packet by setting bit w in the request bit stream in the control channel that travels upstream with respect to the direction the packet will travel. All upstream nodes take note of the requests by incrementing a counter called a request counter (RC). Each node maintains a separate RC for each wavelength. Thus, if bit w in the request bit stream is set, RC w is incremented. Each time a packet arrives at VOQ w, the node stamps the value in RC w onto the packet and then clears the RC. The value of this stamp is called the wait counter (WC). After

the packet reaches the head of the VOQ, if the WC equals n it must allow n empty frames to pass by for downstream packets that were generated earlier. When an empty frame passes by the node on wavelength w, the WC for the packet at the head of VOQ w is decremented (if the WC equals zero, the RC w is decremented). Not until the WC equals zero can the packet be transmitted. The counting system ensures that the packets are sent in the order in which they arrived in the network.

Due to the use of fixed-tuned receivers, HORNET statically assigns each node a wavelength as the home channel for receiving packets. Such static wavelength assignment results in poor statistical multiplexing gain and bandwidth efficiency. As a result of the home-channel design, DQBR treats wavelengths independently and requires each node to maintain a distributed queue for each wavelength. Moreover, each node is allowed to issue multiple independent single-slot requests. With DQBR, HORNET achieves acceptable utilization and fairness at the expense of high control complexity for maintaining the same number of counter pairs as that of wavelengths. Such a design gives rise to a scalability problem. Moreover, the design of permitting unlimited multiple requests with single slot granularity per request unfortunately results in unfairness problems.

### 2.3.2.5 *Probabilistic Quota plus Credit (PQOC) Protocol*

The HOPSMAN [7,8] employs a MAC scheme, called Probabilistic Quota plus Credit (PQOC). First, a cycle (see Figure 2.5) is composed of a pre-determined, fixed number of slots. In general, PQOC allows each node to transmit a maximum number of packets (slots), or quota, within a cycle. Most importantly, even though the total bandwidth is equally allocated to every node via the quota, unfairness surprisingly appears when the network load becomes high. This is because upstream nodes can

access empty slots first, resulting in an increasing tendency for downstream nodes to encounter available empty slots that are located vertically around the back of the cycle. This issue, as well as the vertical-access constraint, gives rise to poorer delay-throughput performance for downstream nodes. To resolve the unfairness problem, the quota is exerted in a probabilistic rather than a deterministic fashion, as "probabilistic quota" implies. In other words, rather than transmitting packets immediately if there remains quota, each node makes the transmission decision according to a probability, e.g., the quota divided by the cycle length. Note that, using the probability, a node may end up making fewer packet transmissions than its quota. The problem can be simply resolved by enforcing a packet transmission in a subsequent slot time with an idle slot. Such an approach evenly distributes idle slots within the entire cycle at all times and thus eliminates unfairness against downstream nodes.

Furthermore, if a node cannot use up its entire quota in a cycle, i.e., has fewer packets than its quota, the node yields the unused bandwidth (slots) to downstream nodes. By doing so, the node earns the same number of slots as credits. These credits allow the node to transmit more packets beyond its original quota in a limited number
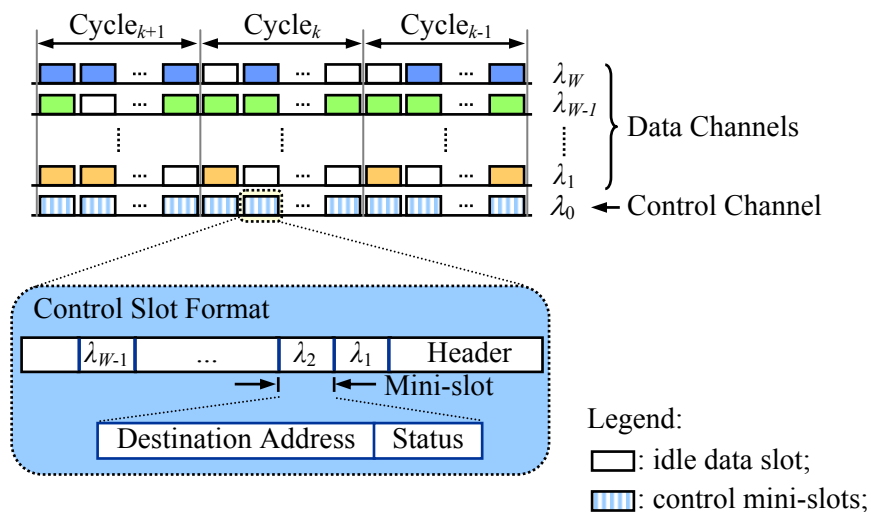


Figure 2.5. The Cycle and slot structures of PQOC.

of upcoming cycles, called the window. That is, the credits are only valid when the number of elapsed cycles does not exceed the window. The rationale behind this design is to regulate a fair use of unused remaining bandwidth particularly in the metro environment with traffic of high burstiness. Notice that there are system tradeoffs in PQOC involving the cycle length and window size. For example, the smaller the cycle length, the better the bandwidth sharing; the larger the window size, the better the bursty-traffic adaptation, both at the cost of more frequent computation. The determinations of the cycle length and window size, which are beyond the scope of this article, can be dynamically adjusted in accordance with the monitored traffic load and burstiness via network management protocols.

## *2.4 Discussions*

The testbeds described in this charpter follow the same goal that overcome the emerging metro gap between high-speed local clients (and networks) and the very-high-speed backbone networks. To overcome this metro gap, the ring networks need to efficiently use the wavelength resources, to be easily upgradeable (and scalable), and to flexibly support varying traffic loads and packet formats in a fair and cost effective manner.

Toward this end, in Table 1 we contrast the experimental testbed networks in terms of header control, node structure, scalability, and packet removal, as well as support for MAC and fairness control. We see from Table 1 that among the networks not having a control channel, the TT-FR node structure is most common. Notice that although fast tunable transmitters [68] with a laser tuning time up to several nano-seconds have emerged, fast tunable receivers [69] operating in the nano-second order remain virtually unavailable. We see from the table that all protocol-oriented and concept-oriented research efforts (as well as the HORNET testbed) allow for easy

scalability in the number of  nodes. The proof-of-concept testbed RINGO, on the other hand, are at present limited to as many nodes as there are wavelength channels.

Table 2 shows a summary of various fairness control protocols in terms of fairness category, fairness type, node structure, collision handling, burst-traffic adaptation, throughput fairness, delay fairness, wavelength sharing, as well as adapted by the testbed network. In the quota-based schemes, each node is allocated a quota that is the maximum transmission bound within a variable-length cycle. Most of the research work focuses on the dynamic adjustment of the cycle length. ATMR [70] allows the last active node to initialize a reset-signal rotating on the ring to inform all nodes to re-start a new cycle. MetaRing [71] uses a token-based signal circulating around the ring. When a node receives the token, it either forwards the token and thus starts a new cycle immediately, or holds the token until the node has no data to send or the quota of previous cycle expires. These schemes were shown to achieve high network utilization and great fairness. However, they cause cycle lengths to prolong several ring times, resulting in a large maximum delay bound and delay jitter, and thus

Table 1. Experimental testbeds comparisons

| | HORNET V1 | HORNET CC | RingO | DAVID | HOPSMAN |
|---|---|---|---|---|---|
| System architecture | Unidirectional ring | Bidirectional ring | Unidirectional ring | Bidirectional ring | Unidirectional ring |
| Header control | In-band SCM | Out-of-band | — $\lambda$ −monitor | Out-of-band | Out-of-band |
| Node structure | TT-FR | $FT^2$-$FR^{2/}$ $TT^2$-$FR^2$ | $FT^W$-FR | $FT^2$-$FR^{2/}$ $TT^2$-$FR^2$ | FT-FR/ TT-TR |
| Scalability | Y | Y | N | Y | Y |
| MAC protocol | CSMA/CA | — | — | SRR | — |
| Fairness control | — | DQBR | — | MMR | PQOC |
| Proposed by | Stanford University | Stanford University | Italy | European IST | NCTU |
| References | [25] | [10] | [12] | [9] | [7-8] |

poor bursty-traffic adaptation. As for the fairness-scheme, Multi-MetaRing [14, 15], it inherits all the pros and cons from the MetaRing. Specifically, there are $W$ numbers of tokens rotating on $W$ wavelengths for Multi-MetaRing and M-ATMR. These schemes encounters an additional problem in which a node may hold several tokens at the same time due to the fact that only one data packet can be sent per slot time. The problem results in an increase in access delay and throughput degradation.

Contrastingly, dynamic bandwidth allocation fairness protocols attempt to improve the deficiency and ultimately solve the instability of the static-based fairness protocols .WDMA [13] and DQBR [10] are all a modified version of DQDB [[72] protocol. Each WDMA node maintains a single distributed queue for all of the wavelengths. Due to the use of a tunable transceiver, WDMA adopts the retransmission mechanism if the receiver contention problem occurs. With such a simple design, WDMA unfortunately results in access unfairness and inefficiencies for multi-channel networks under varying traffic patterns and burstiness.

On the other hand, DQBR requires each node to maintain a distributed queue via a pair of counters per wavelength. With DQBR, HORNET achieves acceptable utilization and fairness at the expense of high control complexity for maintaining the same number of counter pairs as that of wavelengths. However, due to the use of fixed-tuned receivers, HORNET statically assigns each node a wavelength as the home channel for receiving packets. Such static wavelength assignment results in poor statistical multiplexing gain and thus throughput deterioration.

Table 2. Fairness schemes for OPS-based ring networks

| | Multi-MetaRing | M-ATMR | PQOC | WDMA | DQBR |
|---|---|---|---|---|---|
| Inherit from | MetaRing | ATMR | — | DQDB | DQDB |
| Category | Quota + Token based | Token based | Quota based | Explicit reservation | Explicit reservation |
| Type | Global fairness | Global fairness | Local fairness | Local fairness | Local fairness |
| Node structure | $FT^2$-$FR^2$/ $TT^2$-$FR^2$ | $FT^2$-$FR^2$/ $TT^2$-$FR^2$ | FT-FR/ TT-TR | $FT^2$-$FR^2$/ $TT^2$-$TR^2$ | $FT^2$-$FR^2$/ $TT^2$-$FR^2$ |
| Collision Handling | Collision avoidance | Collision avoidance | Collision avoidance | Collision & retransmission | Collision avoidance |
| Bursty-traffic Adaptation | poor | poor | better | poor | better |
| Throughput fairness | Y | Y | Y | Y | Y |
| Delay fairness | — | — | Y | poor | poor |
| Wavelength sharing | static | static | dynamic | dynamic | static |
| Testbeds | DAVID, RingO | HORNET V1 | HOPSMAN | — | HORNET CC |
| References | [9,14-16, 24] | [21,22] | [7-8] | [13] | [10,11] |

,

35

# Chapter 3. The OPACS System Architecture

As shown in the previous chapters, two key challenges pertaining to OPS-based WDM metro networks are the header control and MAC protocol. These requirements form the basis of the OPACS architecture. OPACS, which stands for optical-header processing and access control system, utilizes a simple and highly-efficiency in-band TDM-based header-control signaling technique and a novel MAC protocol to form an architecture that is more cost-effective at high capacities than any of its commercial predecessors.

## 3.1 System Architecture

The WDM network that is governed by OPACS [26-30] system at each node consists of a pair of unidirectional fiber rings, i.e., the forward and reverse rings, as depicted in Figure 3.1. The signal propagates on the two rings in opposite directions. More specifically, packets destined for downstream and upstream nodes are sent along the forward and reverse rings, respectively. Each ring carries a number of WDM data channels, which are further divided into synchronous time slots. In the interest of clarity, time slots and slots are used interchangeably throughout the remainder of the thesis. Each slot contains a control header field followed by a fixed-size packet payload. It is worth pointing out that compared to single ring networks, dual ring networks have several key advantages [5, 6], including a greater scalability and a higher throughput. Furthermore, dual ring networks provide an enhanced fault tolerance in terms of node/fiber failures.
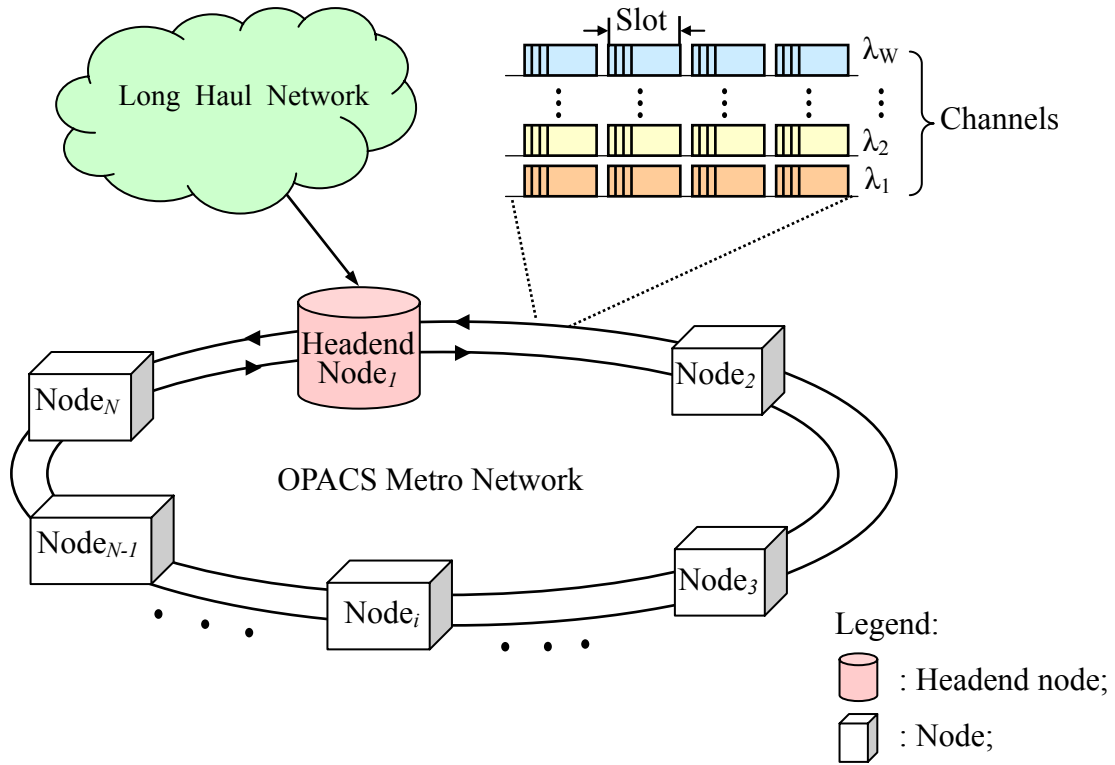
Figure 3.1. OPACS network architecture.

In general, with respect to accessing data channels, the node architecture [26] falls into one of two categories: switch-based, and broadcast-and-select-based. Basically, the switch-based architecture includes the use of a de-multiplexer and a space-switch matrix to direct all desired channels to the optical receivers. As opposed to this, the broadcast-and-select architecture uses an optical coupler to tap off a portion of the optical signal power from the ring to make all data channels available ("broadcast") to the node. The desired data channel is then "selected" via a tunable or band-pass filter. While switched-based nodes provide high channel capacity through the simultaneous access of multiple wavelengths, it becomes costly for some nodes that demand less capacity than provided. Contrastingly, the broadcast-and-select structure enables an incremental and cost-effective upgrade of the channel capacity. Accordingly, our OPACS testbed system adopts the broadcast-and-select architecture.

Notice that such a dual-ring network is logically a bus-based network (with two buses that are wrapped into two rings). Therefore, there is one server node located at

37

the beginning of the two rings, which is responsible for generating optical slots initially, and resetting all used optical slots each time after the slots have traveled one lap of the ring, as depicted in Figure 3.1. The dual-ring network interconnects a number of different access networks and transports various types of traffic from clients. Each node can serve as an access point for a local area network connects to it. An assumption is made that the number of nodes in the system exceeds the number of available wavelengths in the network.

Specifically, each header and its payload are time-multiplexed within a slot, operating at data rates of 1 Gb/s and 10 Gb/s, respectively. The rationale behind the design of using different rates is described as follows. The header in OPACS is only 26 bits long, resulting in low control overhead. Therefore, a rate of 1-Gbps is sufficiently fast and acceptable. However, the header bit-rate can be upgraded provided with a rate-compatible burst mode receiver, if the header size increases or the payload size decreases.

The header signal is RZ-encoded at a data rate of 1 Gb/s. Each header is 26 bits in length, including an 8-bit preamble, a 4-bit header control, and a 6-bit address, besides the guard time. The payload signal (250 bytes long) is NRZ-encoded at a data rate of 10 Gb/s. Since RZ pulses enables fast clock phase selection and can be erased with higher timing margin than NRZ pulses, we thereby adopt the RZ encoding format for the header signal. The payload signal is NRZ-encoded at a data rate of 10 Gb/s.

In the architecture, each network node is equipped with one tunable transmitter and one tunable receiver for each ring, i.e., each node implements the $TT^2$-$TR^2$ system with the architecture. It is further assumed that there are more network nodes than wavelengths in the system, i.e., N > W. Since the operations for accessing the two

rings are identical and independent, for the simplicity of illustration, we hereinafter only focus on the access control for the forward ring.

## 3.2 Node Architecture

The node architecture of OPACS is shown in Figure 3.2. Assume there are four wavelengths used in the network. The transmitter consists of a continuous wave laser and an EA-based external modulator. As shown in Figure 3.2, at the input as parallels slots are passing by, an optical gate/switch first separates the headers from data packets into two paths. Along the header path, the optical headers of four channels are reflected by four fiber Bragg gratings, resulting in a wavelength-to-time conversion for the headers. In other words, the channels' headers are converted from being parallel to serial in the time domain.
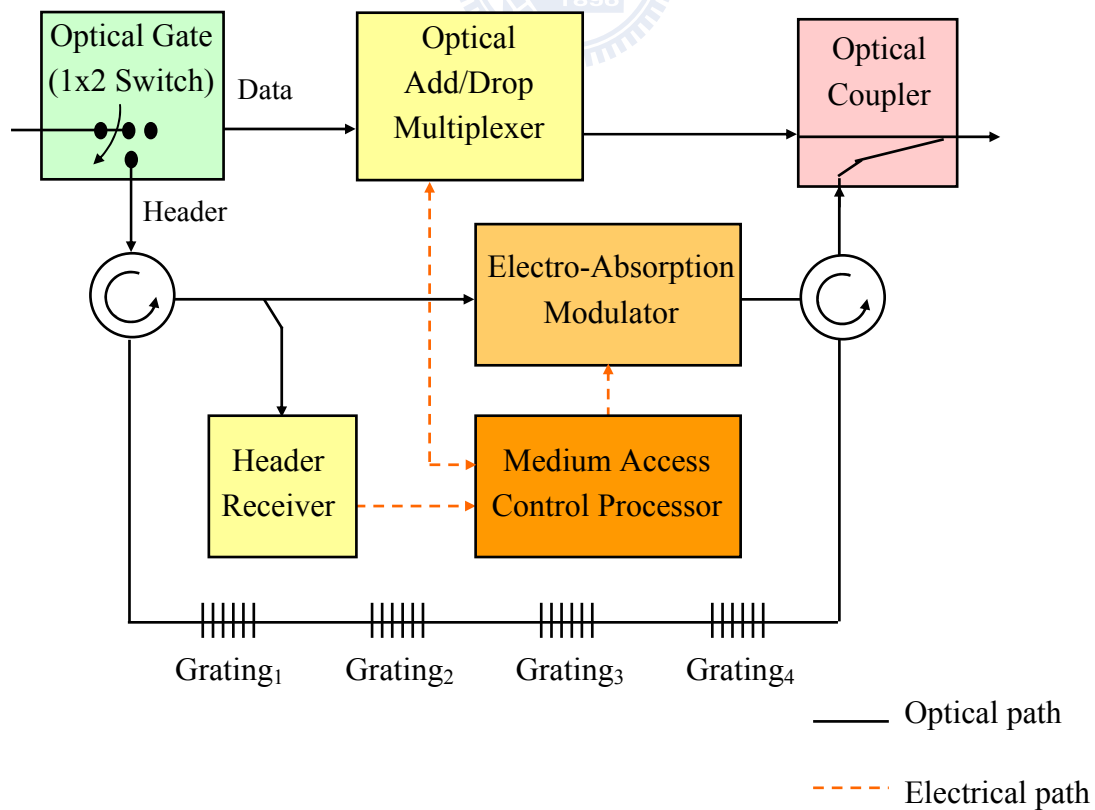
Figure 3.2.   OPACS- system architecture.

These serial headers are later tapped to an optical header receiver by a circulator. Once the MAC processor receives and processes the header information, it determines the reception and transmission of data packets, and then updates the headers according to the MAC algorithm described later. Subsequently, the optical headers are updated by the optical header rewriter, which is implemented by an electro-absorption (EA) optical modulator. After having been updated, the serial optical headers experience a reversed time-to-wavelength conversion on the outbound path through the same set of fiber Bragg gratings.

It is worth noting that, under the ideal fiber grating model (zero insertion loss and perfect channel reflection), our scheme can scale up well with greater number of wavelength. However in reality, cascading a large number of grating filters gives rise to severe signal power loss, causing a limitation on the wavelength number. The problem can be mitigated by using multiple short chains of filters, i.e., a band division parallelism, to replace a single long chain of filters. Our work with four wavelengths here is to demonstrate the proof-of-concept. Along the data path, the data packets are optically added or dropped via the optical add/drop multiplexer (ADM). Finally, the optical headers on different wavelengths are combined with their corresponding data packet via an optical coupler.

The operating principle of the EA modulator can be explained with one simple example illustrated in Figure 3.3. Note first, that the control header bits are RZ encoded. One particularly notable feature of our MAC protocol is that the header bits are always updated from 1 to 0. By taking advantage of such a feature, the header rewriter can simply be a pulse eraser that erases the pulses that are to be updated from 1 to 0. In the example mentioned above, assume that the MAC processor determines to update the fourth and seventh bits of the control header. The Pulse Generator produces the pulse train waveform with two negative pulses on the corresponding bit

positions. The EA modulator then performs a straightforward modulation (attenuation) on the incoming optical header pulses with the two negative pulses, so that the 4th and 7th bits of the header are easily updated from 1 to 0.
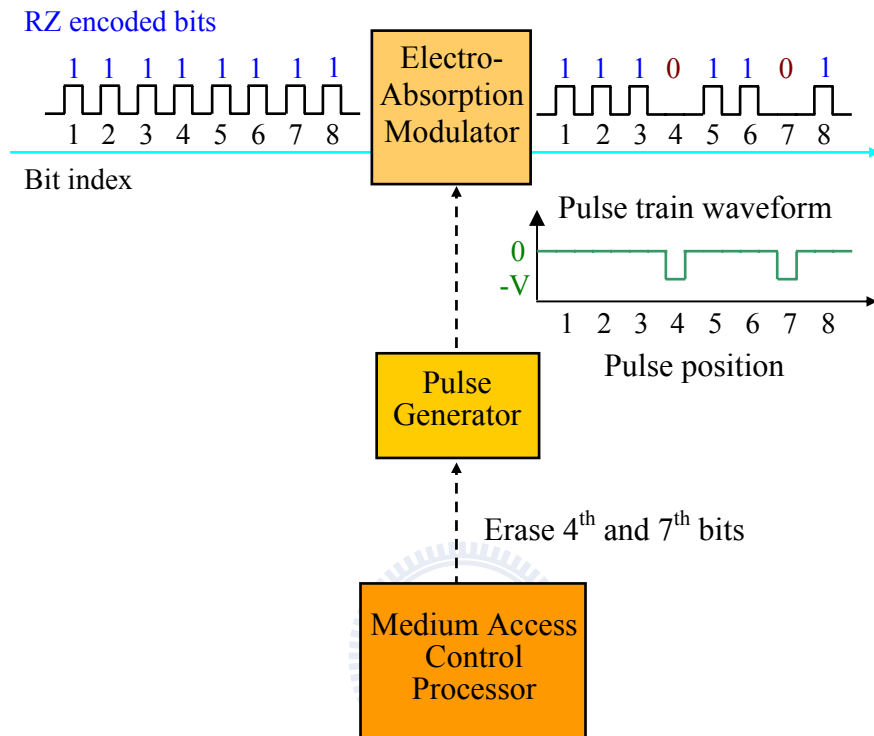


Figure 3.3. Electro-Absorption Modulator- an example.

# Chapter 4. Fairness Control for the OPACS MAC Protocol

OPACS employs a MAC scheme called the distributed multi-granularity and multi-window reservation (DMGWR) scheme. We begin this section by describing the design principles of DMGWR. We then present the detailed algorithm of DMGWR that the scheme is capable of being used in a real-world scenario.

## *4.1 OPACS Fairness Control Protocol: DMGWR Design Principles*

Before describing the scheme, it is necessary to introduce the unfairness problem of a metro ring network and network constraints that will be frequently used throughout the rest of the work. A problem common to ring and bus topologies is the different access priority given to network nodes depending on their position along the ring/bus. In general, the bandwidth of a network is shared by all nodes. Each node ready to send data should have the same opportunity to transmit data. As we have seen in the preceding section, most of the packet-switched ring WDM networks are based on a ring topology. In this architecture, each wavelength can be considered a unidirectional bus terminating at a prescribed destination (see Figure 4.1). Consider the wavelength that is received by node in Figure 4.1. When Node 1 wants to send
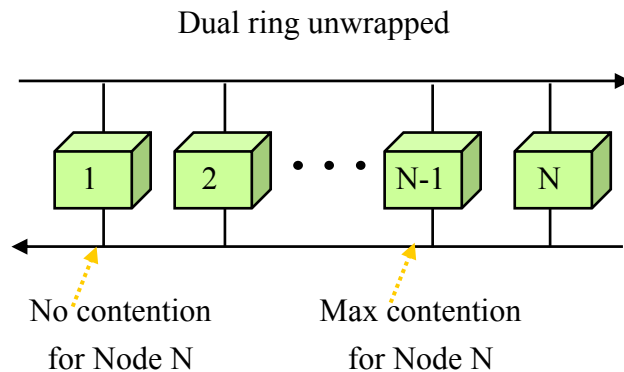
Dual ring unwrapped



Figure 4.1. The dual ring unwrapped, while focusing on the contention for Node N.

packets to Node N, it is never blocked on the wavelength received by node N. While node N-1 wants to send packets to node N, it has to contend with (can occasionally be blocked by) the packets transmitted by node 1, 2, 3,… and N-2 on the wavelength of node N. Thus, the network is biased against nodes closer to the destination. Upstream nodes have a better than average chance to receive an empty slot for transmission, while downstream nodes have a worse than average chance. At heavy traffic this can lead to starvation of downstream nodes since they "see" slots that are mostly used by upstream nodes. Clearly, fairness control is necessary for the OPS-based metro WDM ring networks to avoid this negative result.

In addition, if the network node is equipped with only one tunable receiver, the receiver-contention [6] problem occurs when there is more than one packet destined to the same node in one slot time. Accordingly, one cannot have two packets carried by different wavelengths in a time slot heading for the same destination node. Likewise, if there is only one tunable transmitter, any node is restricted to make at most one packet transmission in one slot time. Such a limitation is referred to as the *vertical-access* constraint (by vertical, we mean the access of different wavelengths within the same slot time).

Considering all wavelengths as a whole, the DMGWR scheme allows bandwidth to be allocated dynamically both in space (granularity) and in time (window). By space, DMGWR allows different bandwidth granularity, i.e., number of slots, to be reserved at a time. By time through the multi-window design, DMGWR permits a node to issue another reservation request prior to the satisfaction of previous requests in the event that new packets have arrived, as long as the total number of outstanding requests is less than a predetermined value, called the window size (*WS*).

To implement the DMGWR protocol, each node maintains a Request (RQ) counter which is used to keep track of its current downstream nodes' reservations

summary. During the procedures of dynamic bandwidth reservation, each node maintains a pending request queue (PRQ). The PRQ are comprised of multiple windows (request queues), where each window (request queue) consists of one countdown counter (CD) and one reserved packet set ($S$), where CD counter records early downstream nodes' reservation summary and $S$ records the reserved packets of the corresponding request, respectively. With the novel multiple window design, the proposed method presents two great merits. First of all, each OPMACS node can fast and dynamic reserve bandwidth particularly in the environment with bursty traffic. Second, each node will have more current information about the backlog of its downstream nodes.

Conceptually, we can divide the DMGWR protocol into three queueing subsystems: the reservation subsystem, the global distribute queue calculation subsystem, and the packet selection and transmission subsystem. With the multiple channels networks and varying traffic patterns taking into account, the beneficial properties of multi-granularity and multi-window are combined together to introduce the pipelining ability of three subsystems. We can intuitively explain the *pipelining* capability by the case of two windows in the PRQ in which the first request queue (with two packets) has reached the top of the global queue (its corresponding CD counter is zero). Basically, the following two idle data slots will be allocated to the reserved packets in top window. However, it is possible that more than two idle data slots will be vertically located at the next time slot. Owing to the *vertical-access* constraint, the OPMACS node is enforced to select only one of the idle data slots to transmit at the same time slot. To reduce the cost of metro access networks, it is important to utilize bandwidth effectively. OPMACS node allocates the remaining of idle data slots to the second request queue. Upon completing the packet transmission of the top window, it is possible that the second window reaches the top of global

distributed queue again. In other words, the whole system can be smartly done in parallel manner of three queueing subsystems and thus achieve high throughput-delay performance and efficient bandwidth utilization in various system configurations.

Without the multiple window design, e.g., with the Distributed Single-Window Reservation (DSWR), the system performance may degrade for upstream nodes. Considering again the previous example, if the node, for instance, with only one window in the global distributed queue instead. DSWR enforces the node to pass by reaming idle slots to downstream nodes. Hence, this approach would show favor to downstream nodes especially under the high-burstiness traffic condition. The novel multiple window design allows certainly the improvement of the resource utilization and delay-throughput performance with respect to the DSWR protocol.

## *4.2 Slot Format of OPACS*
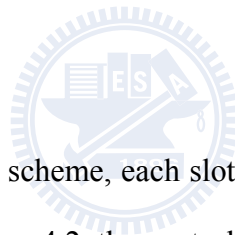
To implement the DMGWR scheme, each slot consists of a control header and a payload field. As depicted in Figure 4.2, the control header includes a one-bit Busy (B) field, a three-bit Request (R) field, a six-bit Destination Address (DA) field. The Busy field indicates whether the slot is Busy (=0) or Idle (=1). The three-bit Request field allows nodes to make a reservation request for one (=110) to seven slots (=000). The Destination Address field is used to identify the destination address of the data packet. It is worth noting that both Busy and Request control fields are designed to be always updated from 1 to 0, thereby allowing the proposed optical header replacement technique to perform a rather straightforward header rewriting operation described earlier.
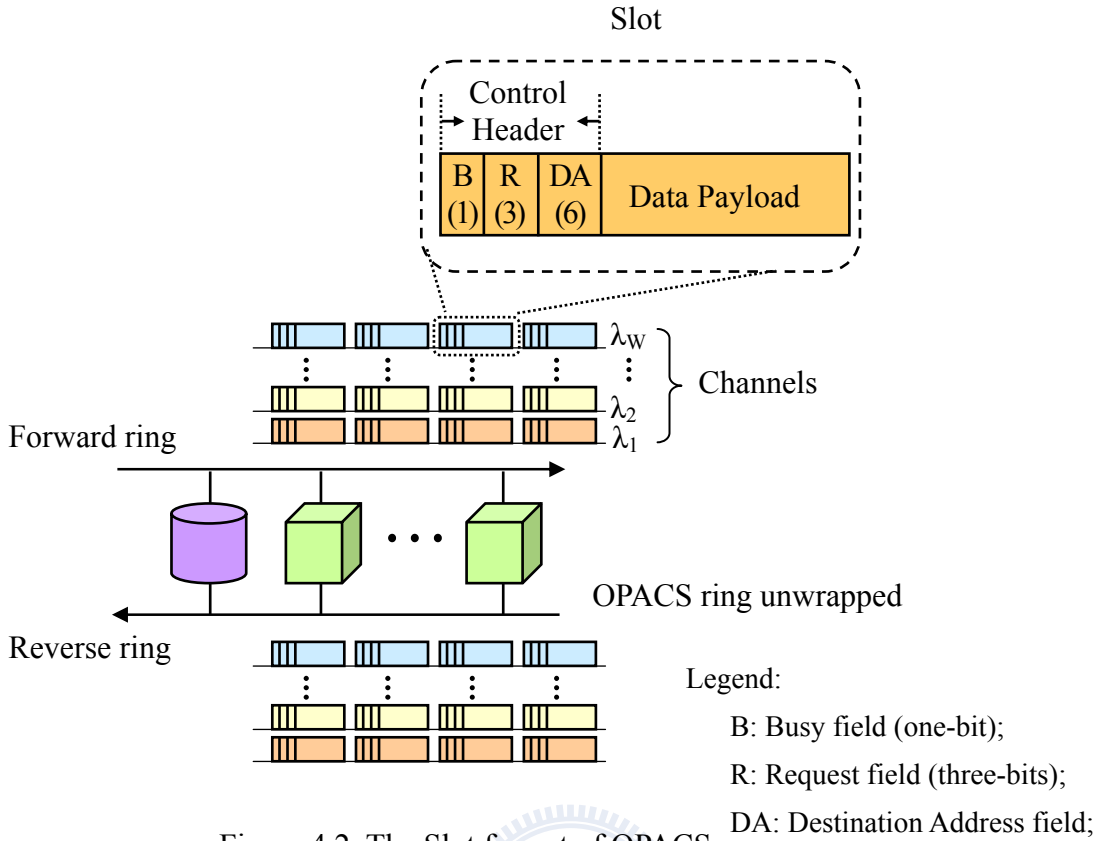
Figure 4.2. The Slot format of OPACS.

## 4.3 OPACS Fairness Control Protocol: Detailed Algorithm of DMGWR

The detailed access operation of DMGWR is described as follows. A node can be in one of the three states: idle, ready, and active. When a node has no packet to transmit, it is in the idle state. Being in the idle state, the node updates its RQ counter by adding the total number of reserved slots for each reservation request passed on the reverse ring, and subtracting the RQ counter by one for each idle data slot passed on the forward ring, as depicted in Figure 4.3. When a node has packets to transmit, it enters the ready state. Notice that, DMGWR adopts the use of virtual output queues (VOQs) [73] to buffer newly arriving packets. Namely, packets destined to different destinations are placed in different queues to prevent from throughput deterioration resulting from the vertical-access constraint.

46

Being in the ready state, the node is required to first make a reservation request by finding an available request-reservation slot on the reverse ring (see Figure 4.4). The total number of slots to be reserved ranges from one to $2^{L-1}$, where L is the length of the request field within each slot. The node then transfers the current value of the RQ counter to the CD counter and resets the RQ counter to zero. Finally, the node saves the total number of reserved slots (NRS) and the CD counter value to the pending request queue, as shown in Figure 4.3. With pending requests in the queue, the node enters the active state. Being in the active state, if the node observes more new arriving packets, the node can repeat the reservation request process as long as the total number of requests is less than or equal to the *WS*.
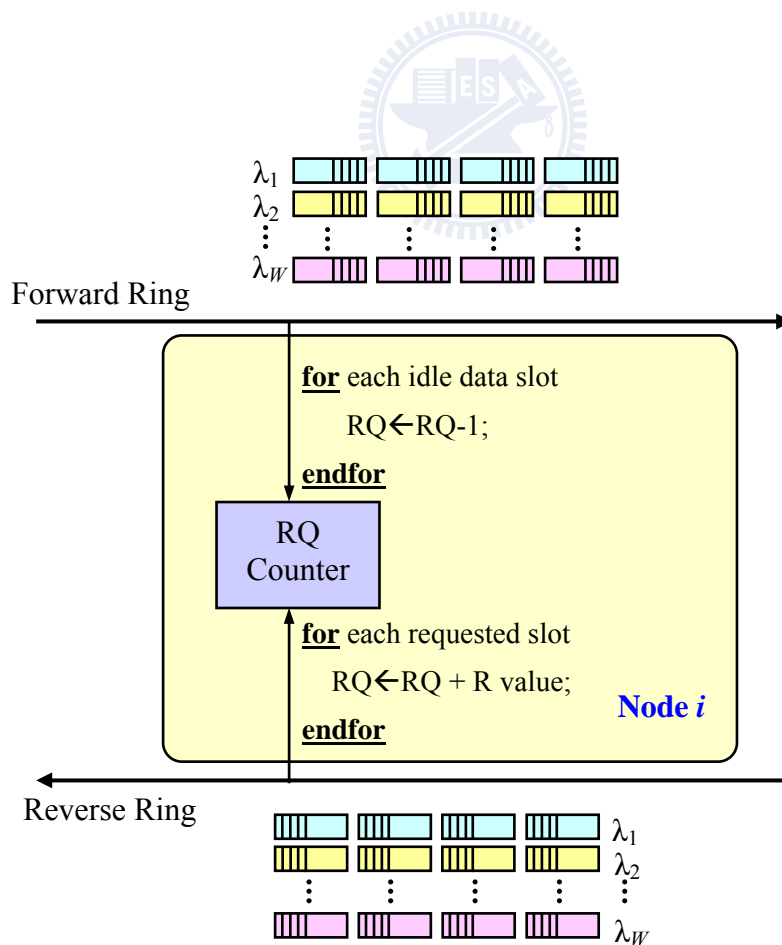


Figure 4.3. The DMGWR scheme: Idle state.

An active node updates its RQ counter by adding the total number of slots reserved for each reservation request passed on the reverse ring (see Figure 4.5). For all idle data slots (located vertically) on the forward ring within a single slot time, the node decreases the CD counter of the first reservation request by the total number of idle slots, if the CD counter is greater than the total amount of idle slots. If the total number of idle slots are greater than the CD counter, as soon as the CD counter becomes zero, the node can then transmit the next packet in the VOQ's that is free from the receiver contention problem on one of the available idle slots. The node in turn decrements the corresponding NRS value of the request by one.
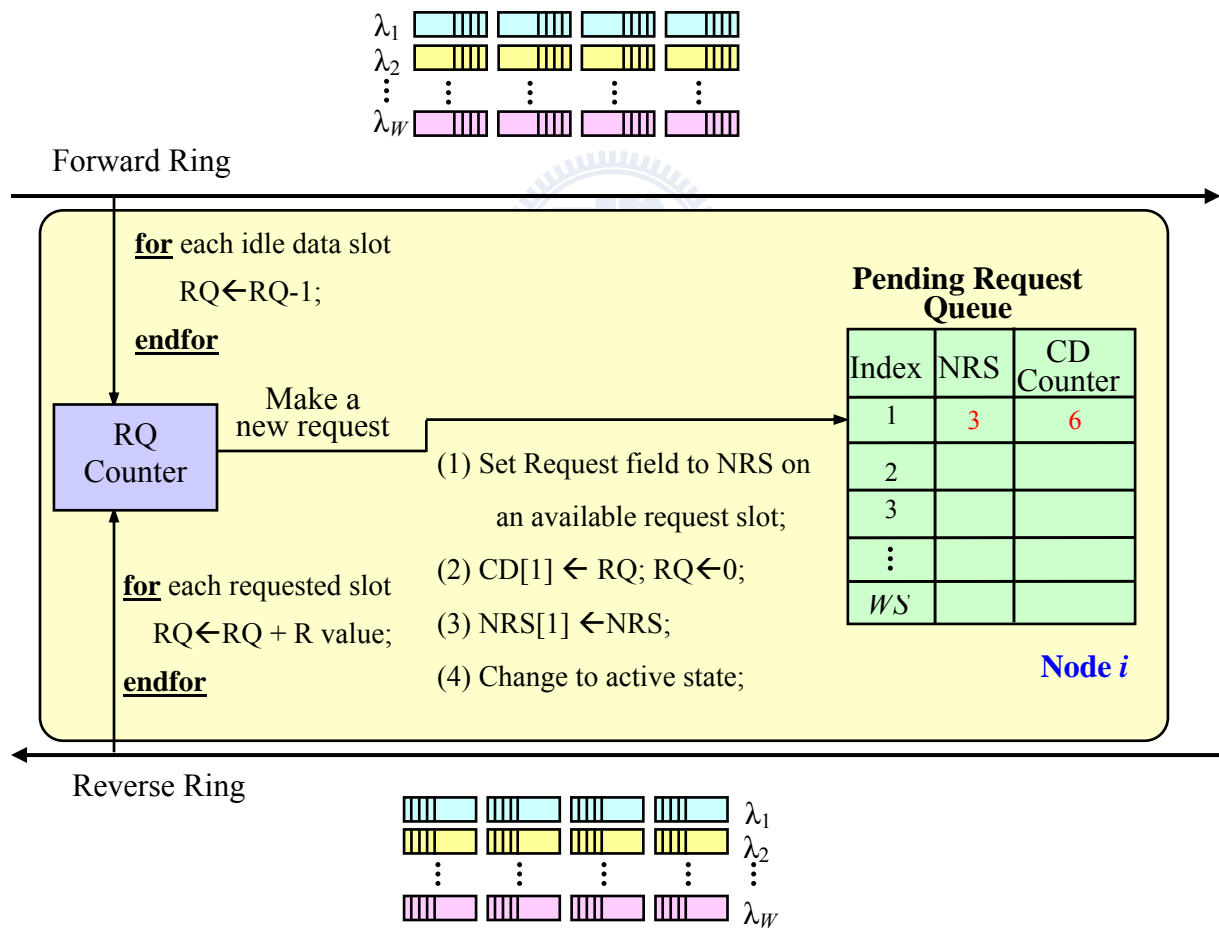


Figure 4.4. The DMGWR illustration: ready state.

When the NRS reaches zero, the pending request entry is removed from the queue. Most importantly, if the CD counter of the first request is zero but there exist no packets in the VOQ's that are receiver-contention free or there are more idle slots left, DMGWR applies the so-called pipelining operation by decreasing the next request's CD counter by the number of remaining idle slots. Such an operation is repeatedly and sequentially applied to all entries in the pending request queue until no idle slots are left uncalculated. Finally, when the VOQ's and pending request queue are empty, the node returns to an idle state.

The packet selection strategy adopts the receiver-contention-free basis methods. The detailed packet selection strategies will describe in the next section. More specfically, if the top packet violates receiver-contention-free basis, the next packet in
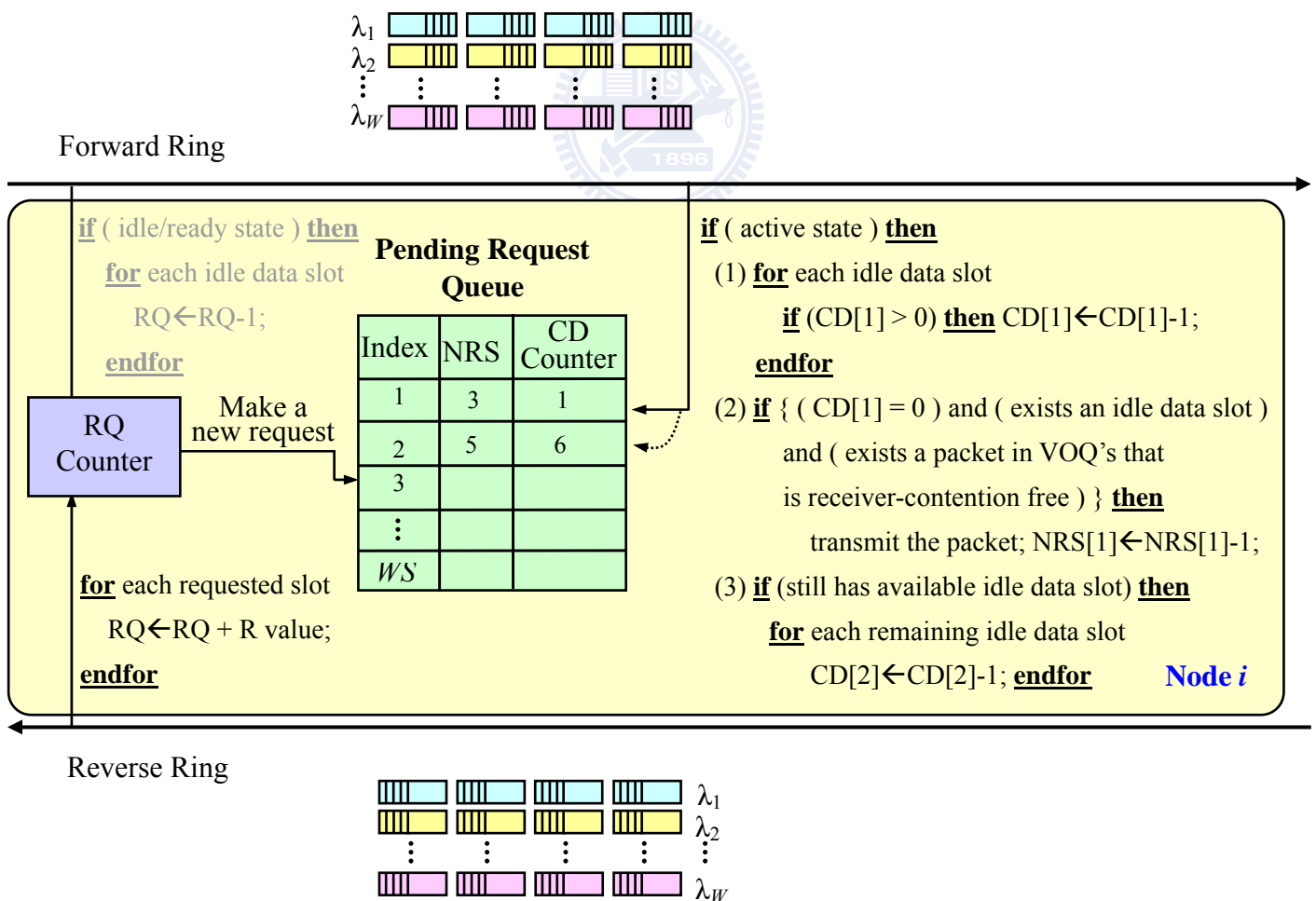


Figure 4.5. The DMGWR illustration: active state.

the reserved packet set without receiver contention will be selected to transmit. When an active node observes its pending request queue and the VOQs are all empty, the node changes its state to an idle state.

The algorithm of DMGWR is outlined in Figure 4.6. Upon arriving, packet is inserted into the VOQs. First, the node read request slots on the reverse ring to gather the current network reservation status and updates its RQ counter by adding the total number of reserved slots. The algorithm then read data slots in forward ring to gather the current network activity. If the node is in the idle state or ready state, the node updates its RQ counter by subtracting the RQ counter by one for each idle data slot passed on the forward ring. When a node has packets to transmit, it enters the ready state. The node is required to make a reservation request by finding an available request-reservation slot on the reverse ring. The node then transfers the current value of the RQ counter to the CD counter and resets the RQ counter to zero. Finally, the node saves the total number of reserved slots (NRS) and the CD counter value to the pending request queue and changes its state to active.

If the node is in the active state, for all idle data slots (located vertically) on the forward ring within a single slot time, the node decreases the CD counter of the first reservation request by the total number of idle slots, if the CD counter is greater than the total amount of idle slots. If the total number of idle slots are greater than the CD counter, as soon as the CD counter becomes zero, the node can then transmit the next packet in the VOQ's that is free from the receiver contention problem on one of the available idle slots. The node in turn decrements the corresponding NRS value of the request by one.
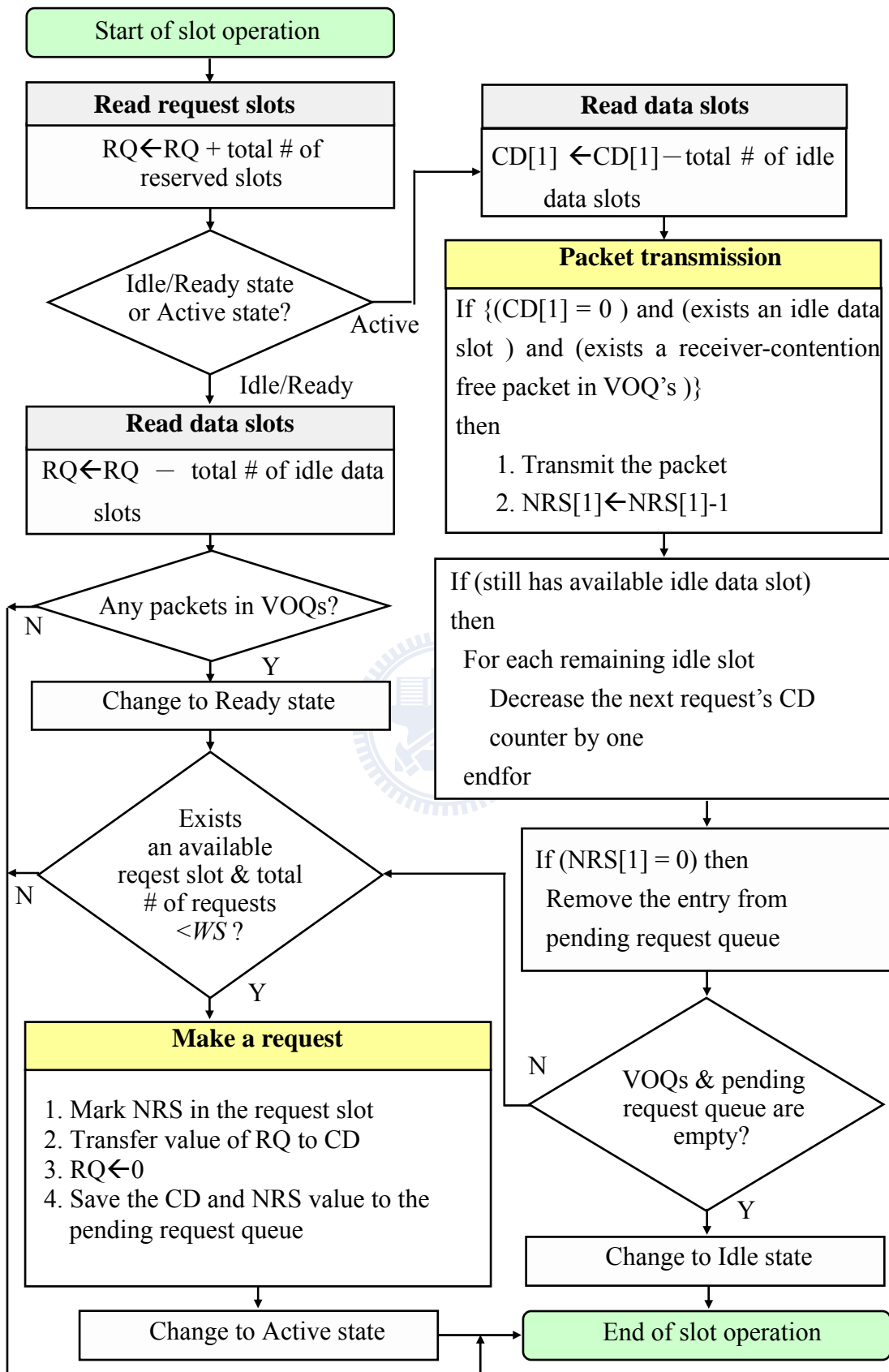
```
┌─────────────────────────────┐
│    Start of slot operation   │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐      ┌─────────────────────────────┐
│      Read request slots      │      │       Read data slots        │
├─────────────────────────────┤      ├─────────────────────────────┤
│  RQ←RQ + total # of          │      │  CD[1] ←CD[1]−total # of idle │
│  reserved slots              │      │  data slots                  │
└─────────────────────────────┘      └─────────────────────────────┘
              │                                     │
              ▼                                     ▼
        ◇ Idle/Ready state              ┌─────────────────────────────┐
         or Active state? ◇──Active──→  │    Packet transmission       │
              │                         ├─────────────────────────────┤
         Idle/Ready                     │ If {(CD[1] = 0 ) and (exists an idle data │
              ▼                         │ slot ) and (exists a receiver-contention │
┌─────────────────────────────┐        │ free packet in VOQ's )}      │
│       Read data slots        │        │ then                         │
├─────────────────────────────┤        │     1. Transmit the packet   │
│  RQ←RQ −  total # of idle data│       │     2. NRS[1]←NRS[1]-1       │
│  slots                       │        └─────────────────────────────┘
└─────────────────────────────┘                    │
              │                                     ▼
              ▼                         ┌─────────────────────────────┐
      ◇ Any packets in VOQs? ◇───N──→  │ If (still has available idle data slot) │
              │                         │ then                         │
              Y                         │   For each remaining idle slot │
              ▼                         │     Decrease the next request's CD │
┌─────────────────────────────┐        │     counter by one           │
│     Change to Ready state    │        │ endfor                       │
└─────────────────────────────┘        └─────────────────────────────┘
              │                                     │
              ▼                                     ▼
         ◇ Exists                       ┌─────────────────────────────┐
          an available                  │ If (NRS[1] = 0) then         │
          reqest slot & total ◇──N──→   │   Remove the entry from      │
          # of requests                 │   pending request queue      │
          <WS ?                         └─────────────────────────────┘
              │                                     │
              Y                                     ▼
              ▼                              ◇ VOQs & pending
┌─────────────────────────────┐        N────  request queue are
│        Make a request        │              empty? ◇
├─────────────────────────────┤                     │
│ 1. Mark NRS in the request slot │                 Y
│ 2. Transfer value of RQ to CD │                   ▼
│ 3. RQ←0                      │         ┌─────────────────────────────┐
│ 4. Save the CD and NRS value to the │  │     Change to Idle state     │
│    pending request queue     │         └─────────────────────────────┘
└─────────────────────────────┘                    │
              │                                     ▼
              ▼                         ┌─────────────────────────────┐
┌─────────────────────────────┐        │     End of slot operation    │
│    Change to Active state    │───────→└─────────────────────────────┘
└─────────────────────────────┘
```

Figure4.6. The DMGWR algorithm.

## 4.4 OPACS Fairness Control Protocol: Contention-Free Packet Selection Schemes

In this subsection, we will discuss the contention-free packet selection schemes of DMGWR. Generally, when the reserved packet reaches the top of global distributed queue, it will be allowed to be selected to transmit through any idle data slot. However, when more than one reserved packets reach the top of distributed queue and while all satisfying receiver-contention-free basis, one packet should be chosen according to certain selection criteria.

Different packet selection schemes may be envisaged at each sending node in the case that more than one reserved packet reach the top of distributed queue, listed as below. In the following, we define Ps= {P| Pi has reached the top of distributed queue and there is no receiver contention in this slot time}, hereafter called packet selection set.

**First Come First Serve (FCFS) Selection**: The packet at the head of the queue is considered for transmission. If it can be transmitted without receiver contention, the packet is sent in any selected data channel. Otherwise, the next packet in the queue without receiver contention will be selected to transmit.

**Maximum Hop (MaxH) Selection**: Under the MaxH scheme, the packet associated with the maximum hop distance between source and destination node is selected when no receiver contention arises. Assuming that node indices increase with the forward data ring direction (as in Figure 3.1). Let $s \in \{1,2,\ldots,N\}$ represent the source node index while $d \in \{1,2,\ldots,N\}$ denotes the destination node index. The individual hop count h related to a packet is calculated in the following manner: $h = d - s$, $d > s$. And the packet selection set can be given by PMaxH={Ps|max{h(Ps)}}, where h(Ps) denotes the hop count vector of selected set.

**Minimum Hop (MinH) Selection**: Under the MinH scheme, the packet associated with the minimum hop distance between source and destination node is selected when no receiver contention arises. The selection set can be given by PMinH={Ps|min{h(Ps)}}, where h(Ps) denotes the hop count vector of selected set.

# Chapter 5. Simulation Results

This chapter focuses on the performance of the protocol described in Chapter 4. First of all, it will draw the comparisons of performance among DMGWR and two prevailing schemes, HORNET DQBR [10] and WDMA [13]. Then, it will carry out a performance study on the DMGWR under various traffic distribution scenarios.

## *5.1 Simulation Model*

The simulations are event-based and written in the C language. A discrete-event simulation can be described as a system representation in which events of interest occur at discrete points in time, and the ordered occurrence of such events together with the time at which they occur described the system operations. It is important to note that all of the simulations were terminated after reaching a 95% confidence interval.

The simulations were performed under the following assumptions:

- There are $N$ nodes in the network, numbered from 1 to $N$ corresponding to the most upstream (node 1) to the most downstream (node $N$) locations, where $N$ could be 48, 60, or 40..

- All nodes are equally spaced around the ring. The distance between neighboring nodes is ten slots ($D = 10$), except when evaluating inter-distance impact on mean delay, in which case the distance will be either 1 or 5.

- To prevent packet losses at each node caused by buffer overflows, we assume that each node has virtually infinite queue size.

- Data packets have fixed lengths equal to one slot.

- The destination addresses of the generated packets are uniformly distributed

among the network nodes (assuming that a source node cannot transmit to itself).

- New packets are generated only at the moments just before the slot boundaries reach the node.

- The packet selection follows the FCFS strategy, except when comparing packet selection strategies.

- Window size is seven without specific indication.

- In the simulation for DQBR, *W* home channels are assigned to *N* nodes in a cyclic (interleaved) fashion.

To investigate the traffic under more realistic network conditions, the source traffic is modeled as a two state Markov Modulated Poisson Process (MMPP) [73], alternating between the ON and OFF states, which correspond to high and low mean arrival rates, respectively. The MMPP is characterized by four parameters $(\alpha, \beta, \lambda_H, \lambda_L)$, where $\alpha$ ($\beta$) is the probability of changing from state H (L) to state L (H) in a slot time, and $\lambda_H$ ($\lambda_L$) represents the probability of arrivals at state H (L).

For simplicity of description, the state change probability is denoted as $P_{i,j}$, i, j $\in \{H, L\}$. Namely, $P_{H,L} = 1 - P_{H,H} = \alpha$, and $P_{L,H} = 1 - P_{L,L} = \beta$. In the MMPP, the burstiness arrival is modeled with $\lambda_H$ (the high rate) and $\lambda_L = 0$. Let *B* represent the burstiness of the arrival process, we thus have $B = (\alpha + \beta) / \beta$ and mean arrival rate as $\beta \cdot \lambda_H / (\alpha + \beta)$. The Poisson arrival is modeled with $\alpha = \beta = 0.5$ and $\lambda_H = \lambda_L = \lambda$, where $\lambda$ is the probability that a packet arrives at each slot and is equal to mean arrival rate.

## *5.2 Performance Metrics*

As result of an extensive performance evaluation different performance metrics can be obtained. The most important performance measures in this thesis are listed below along with their applied definitions.

- Network throughput: number of transmitted packets/ number of received packets within the network during a specific time period

- Node throughput: number of transmitted packets/ number of received packets at a node during a specific time period

- Access delay: measured from the instant of generation of a packet to its complete transmission by the source node.

- Mean queueing delay: mean waiting time of a packet in the transmission queue until its transmission by the source node.

- Potential receiver-contention probability: the number of receiver contention occurred over the summary of the number of transferred packets and the number of receiver contention occurred.

## *5.3 Performance Comparisons*

Taking into account the assumptions made in the previous section, it will draw comparisons between the DMGWR scheme and two other existing schemes (WDMA and DQBR), and demonstrate its performance with respect to throughput, access delay, and fairness under various system settings, via simulation.

We first compare the impact the rapid growth of the number of wavelengths has on the network throughput of the three schemes (see Figure 5.1). In the experiments, the number of network nodes is 48 and the numbers of wavelengths (W) range between 2 and 24. As the total number of wavelengths increases, the total load is proportionally scaled up. When upgrading the overall network capacity to support the high bandwidth demand applications, both DQBR and WDMA unfortunately manifest a deteriorating throughput performance as the number of wavelengths is scaled up. Results confirm the limitations of the multiple distributed queues approach and highlight the performance penalty of DQBR already discussed. Note that WDMA deteriorates the most. This is due to its collision and retransmission scheme, which produces noticeable bandwidth waste. We point out that the normalized throughput of DMGWR exceeds that of DQBR and WDMA several orders of magnitude for larger network capacity particularly under extremely heavy load (0.99). Furthermore, the DMGWR scheme achieves the same degree of bandwidth efficiency irrespective of the wavelength number and load (*L*) of the network. This is a very attractive scalability feature for the next-generation commercial WDM network systems.
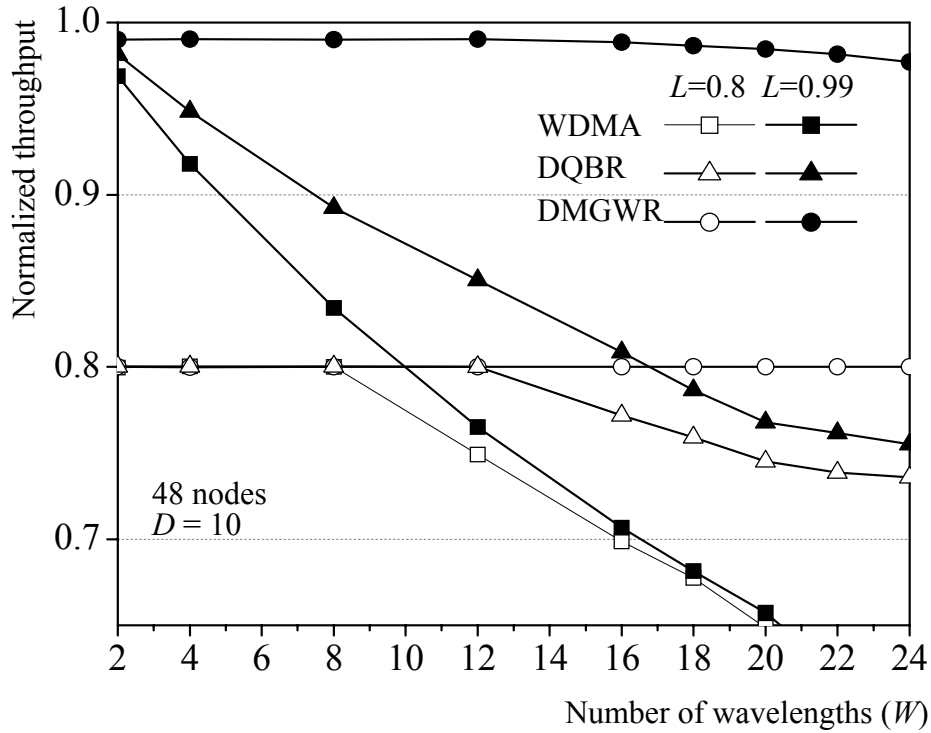
Figure 5.1. Throughput comparison.

As depicted in Figure 5.2, it compares aggregate throughput performance among three schemes when subjected to various loads and number of wavelengths. In the experiments, the number of network nodes is 48 and the numbers of wavelengths ($W$) range between 4 and 16 under various traffic load. Under the heavy load (0.75), the aggregate throughput of all of the schemes is proportionally scaled up as the total number of wavelengths increases. It observed that the aggregate throughput of DMGWR exceeds that of DQBR and WDMA several orders of magnitude for larger network capacity particularly under extremely heavy load (0.99). When upgrading $W$, both DQBR and WDMA unfortunately have a direct manifestation of the deteriorating aggregate throughput performance. This is understood by observing that a larger $W$ leads to heavier traffic, i.e., mean packet arrival rate $pr = (2 \times W \times L)/N$. As the ratio of $W/N$ increases, the $pr$ on each node increases twice and more access collisions occur. In case of the lack of statistical multiplexing gain, when the maximum transmission capacity of one protocol is reached, the congestion state is immediately deteriorated

by the almost constant arrival of packets and the node buffers are persistently filled with packets. As a direct result the throughput of either DQBR or WDMA abruptly decline when the congestion state is reached.

Furthermore, it compares the delay performance of the three schemes when subjected to various loads and burstiness. As expected, delay unfairness is manifested in both DQBR scheme and WDMA scheme. DMGWR outperforms DQBR and WDMA significantly, especially under highly bursty traffic. From Figure 5.3(a), one can see that as the burstiness increases, delays of WDMA and DQBR increase rapidly. However, DMGWR achieves a bounded delay even in highly bursty traffic. Note that in Figure 5.3(b), when upgrading the network capacity as done in Figure 5.1, both DQBR and WDMA unfortunately have a direct manifestation of the larger mean delay as the wavelength number is scaled up. This is understood by observing that as the ratio of $W/N$ increases, the mean packet arrival rate on each node increases twice and more access conflicts occur. When the maximum transmission capacity of one protocol is reached, the congestion state is immediately deteriorated by the almost constant arrival of packets and the node buffers are persistently filled with packets. As a direct result the mean access delay of either DQBR or WDMA increases abruptly when the congestion state is reached. As opposed to DQBR and WDMA, the DMGWR protocol is show to improve mean delay performance when upgrading the number of wavelength. The performance improvement is achieved by superior design of the DMGWR protocol that accommodates highly wavelength sharing among the competing nodes.

Figure 5.4 depicts the average access delay at each node under different inter-nodal distances ($D$). As expected, delay unfairness is manifested in both DQBR and WDMA. A larger ring length in WDMA leads to longer wait of upstream nodes for retransmission in the case of a packet collision. Consequently, the total delay of

59

upstream nodes drastically increases with the ring length in WDMA. We also observe that the delay of DQBR is a serrated curve on downstream side due to the impact of cyclic home channel assignment of DQBR. On the other hand, DMGWR enables fairness and efficiently shares the bandwidth between competing nodes. In addition, the mean access delay of DMGWR is independent of the ring length.

In Figure 5.5, we compare the average access delay of the three schemes when subjected to various burstiness. As depicted in Figure 5.5, DMGWR outperforms DQBR and WDMA significantly, especially under highly bursty traffic. From Figure 5.5, one can see that as the burstiness increases, delays of WDMA and DQBR increase rapidly. However, DMGWR achieves a bounded delay even in highly bursty traffic. We also observe in the Figure 5.5 that the delay of DQBR is plotted by a serrated curve (around nodes 36 and 44) on the downstream side due to the impact of a cyclic home channel assignment of DQBR. DMGWR guarantees delay under different burstiness settings. Indeed, with the ability of dynamically bandwidth allocatcation, the DMGWR can provide more efficiently resource allocation in response to bursty data traffic and time-varying traffic conditions.

We furthermore draw a delay comparison between DMGWR and DQBR over a network with malicious nodes. In the simulation, nodes 10 and/or 15 are set as malicious nodes, where each node generated an excessive load of 0.09 per wavelength, in a network with a total load of 0.85 per wavelength. As displayed in Figure 5.6(a) and 5.6(b), DMGWR causes the malicious nodes to suffer severe delays, while leaving other normal nodes completely unaffected. On the other hand, the DQBR scheme results in unexpected delay deterioration (and thus unfairness) for the downstream nodes. As the number of malicious nodes increases, the delay unfairness problem worsens, as Figure 5.6(b) demonstrates. In this case, DMGWR can still guarantee a high grade of fairness among all nodes. Thus, the DMGWR scheme is

robust and fair even when under attack by malevolent nodes.

Moreover, we observe performance comparison between DMGWR and DQBR over a network with malicious nodes under extremely heavy loads. As displayed in Figure 5.7(a) and Figure 5.7(b), under the extremely heavy loads condition, DQBR scheme results in unexpected delay deterioration for the downstream nodes. Note that in Figure 5.7(b), when upgrading the network capacity as done in Figure 5.1, DQBR scheme unfortunately have a direct manifestation of the throughput performance as the network capacity is scaled up.

Furthermore, we examine the impact of wavelength sharing on the throughput performance between DMGWR and DQBR under various loads for the same ratio of the number of nodes and wavelengths. As depicted in Figure 5.8, the static wavelength sharing scheme undergoes severe throughput deterioration toward downstream nodes under heavy loads resulting from the performance penalty of DQBR. Under the same ratio of the number of nodes and wavelengths, the larger the number of nodes, the worse the normalized throughput. In contrast to this, as a result of the full wavelength sharing and multiple-granularity-window design, DMGWR achieves 100% throughput under all loads and network configurations. Finally, we used 40/80 nodes and 8 wavelengths in the network simulations to compare the impact of wavelength sharing on the throughput performance. Also compare Figure 5.9 that increasing the number of nodes by a factor of 2, DQBR results in a lower normalized throughput and worse performance.
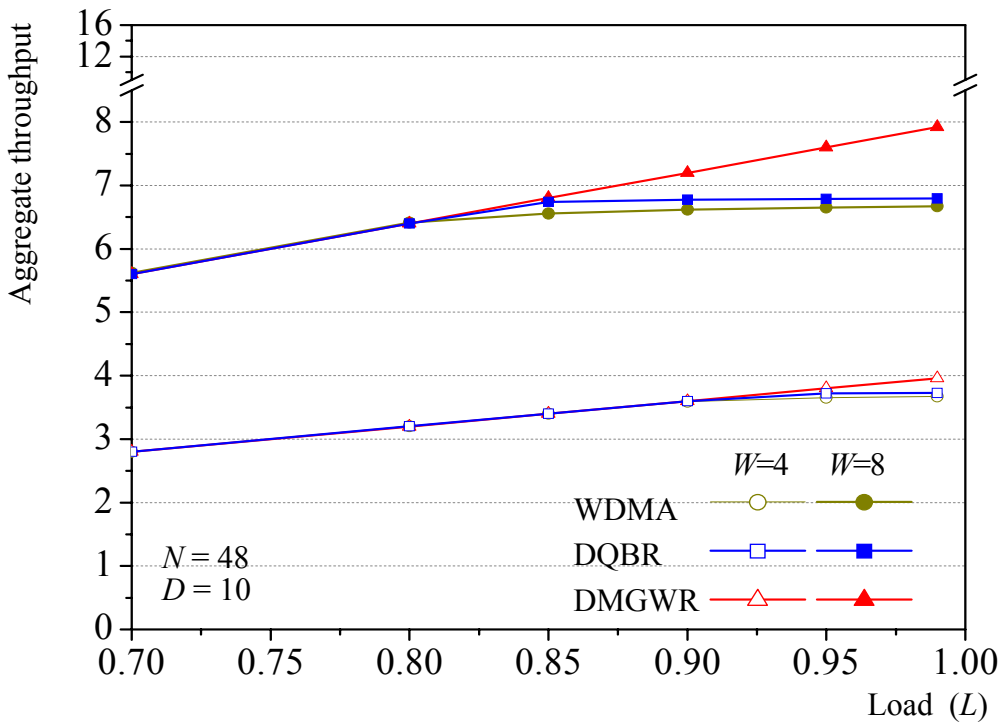
(a) Aggregate throughput comparison (W=4 and 8)
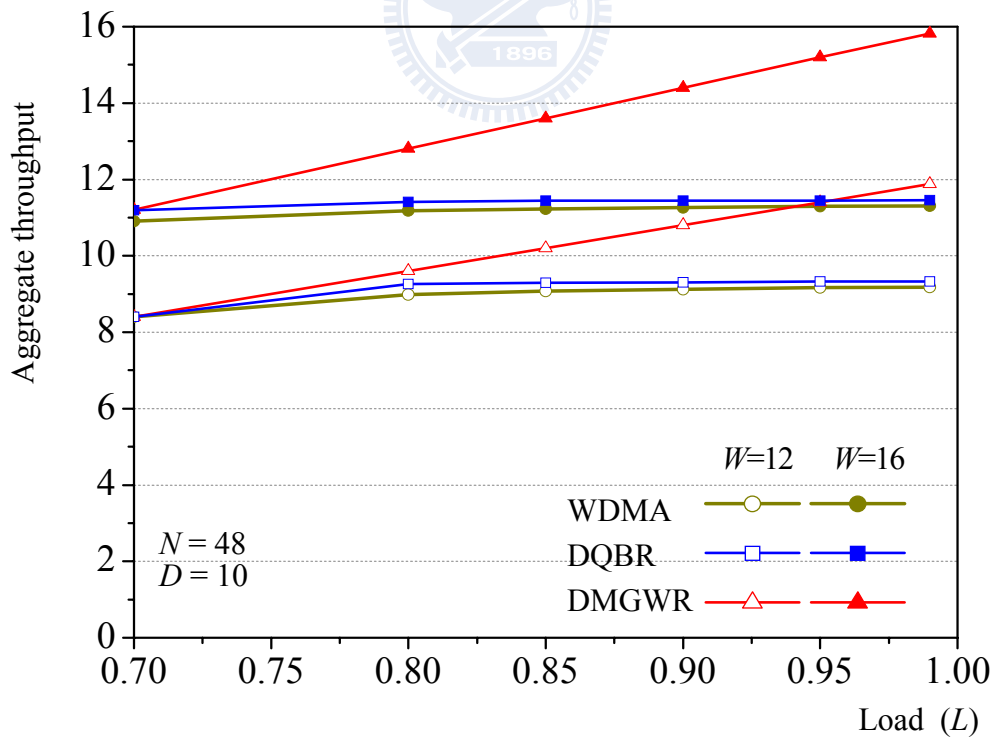


(b) Aggregate throughput comparison (W=12 and 16)

Figure 5.2. Aggregate throughput comparison under various loads and number of wavelengths.
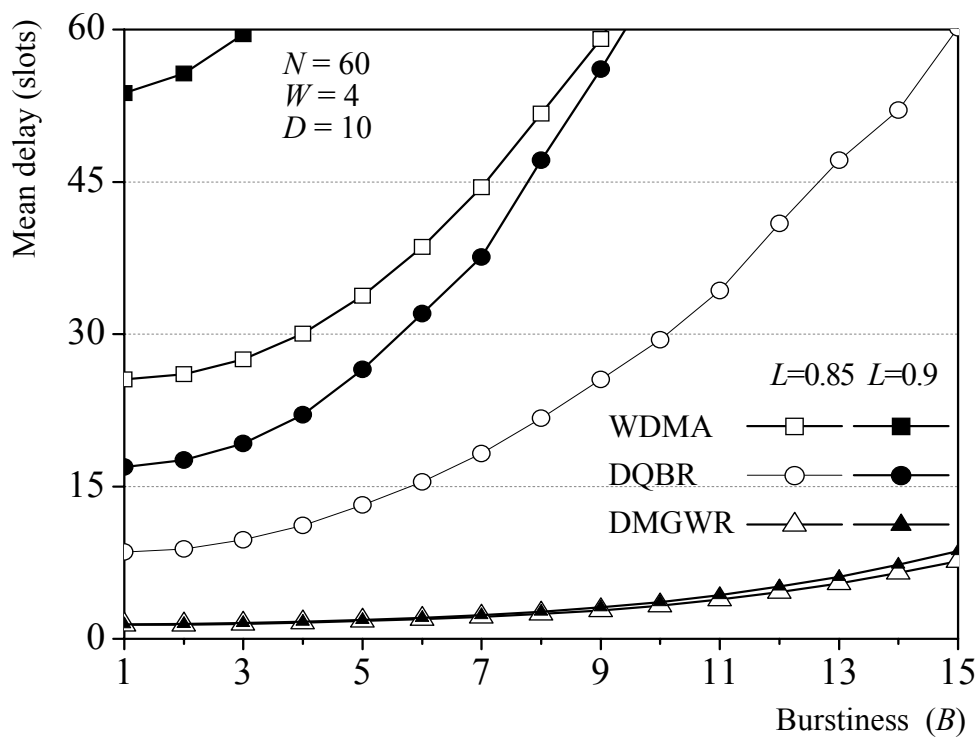
(a) Delay comparison under various bursstiness
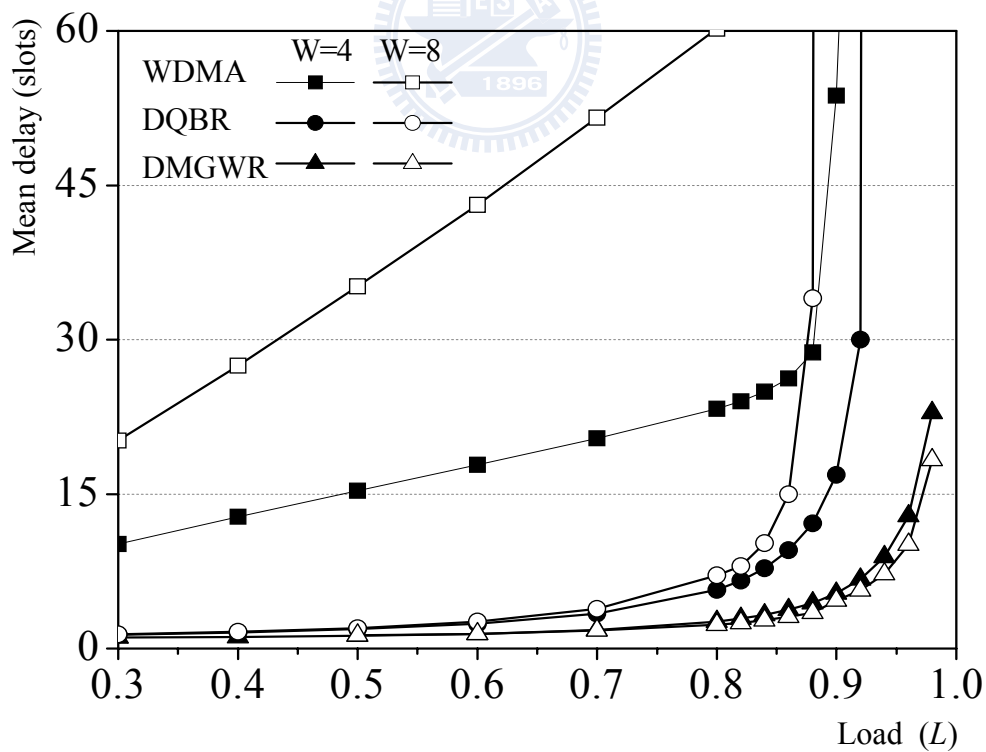


(b) Delay comparison under various load

Figure 5.3. Access delay comparison under various bursstiness and load.

63
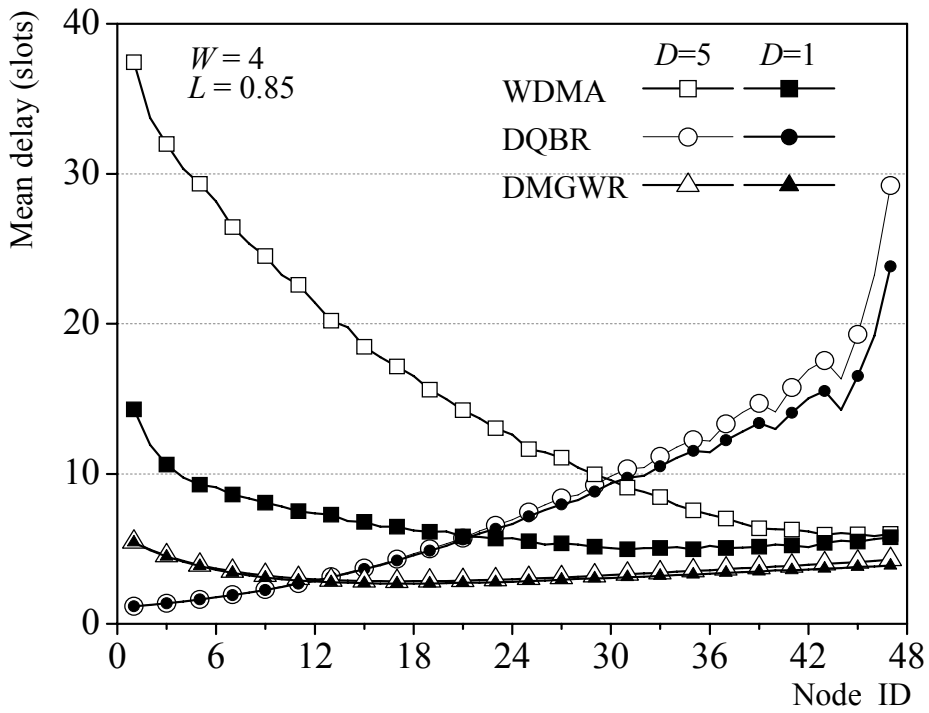
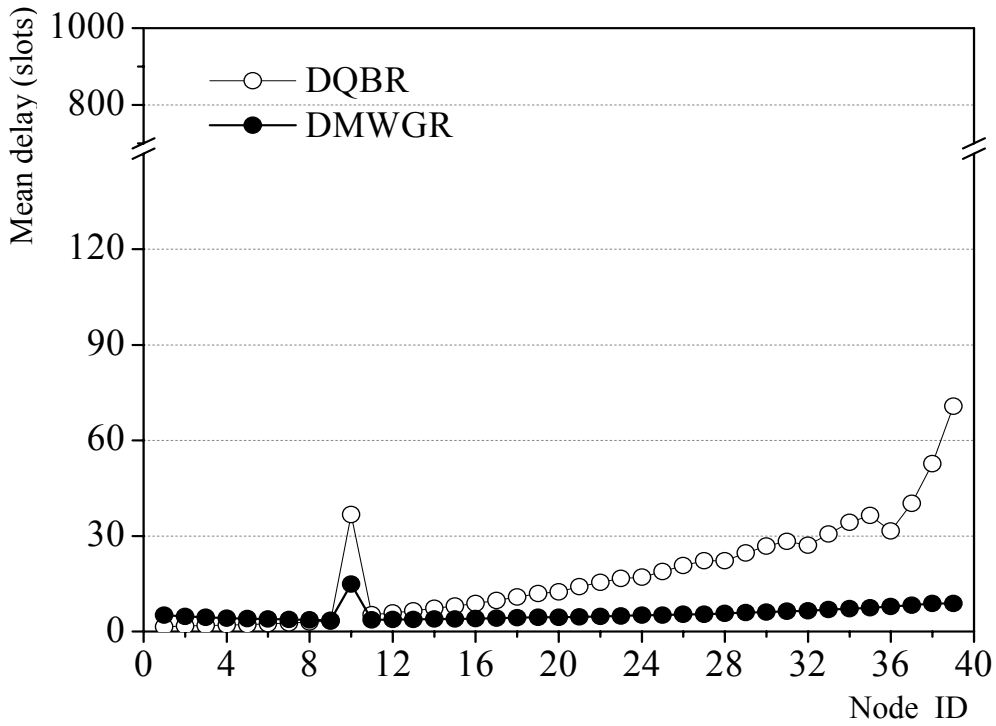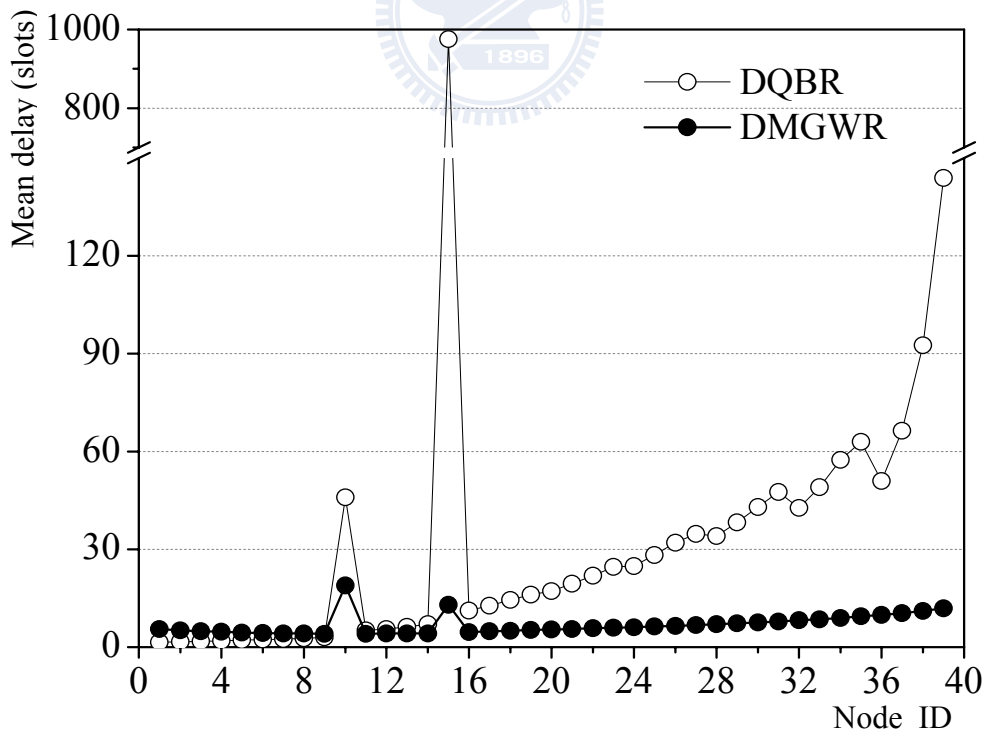Figure 5.4. Comparisons of delay Fairness under various inter-nodal distances.
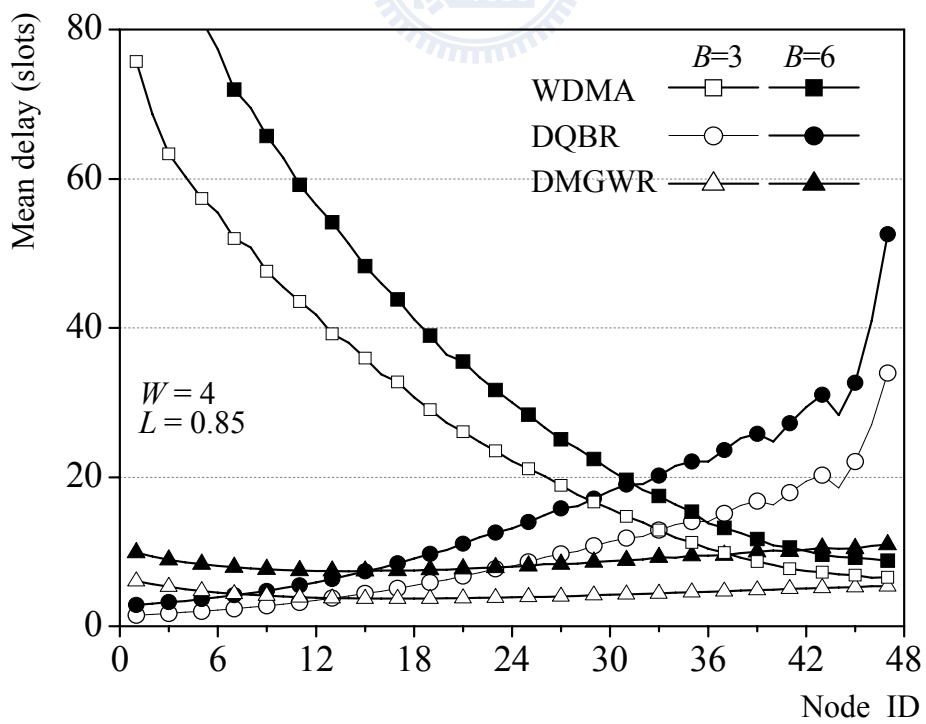


Figure 5.5. Comparisons of delay fairness under various burstiness.

(a) Delay comparisons for malicious node (Node 10)



(b) Delay comparisons for malicious nodes (Nodes 10 and 15)

Figure 5.6. Delay comparisons for netwok with malicious nodes. (a) Delay comparisons for malicious node (Node 10). (b) Delay comparisons for malicious nodes (Nodes 10 and 15).
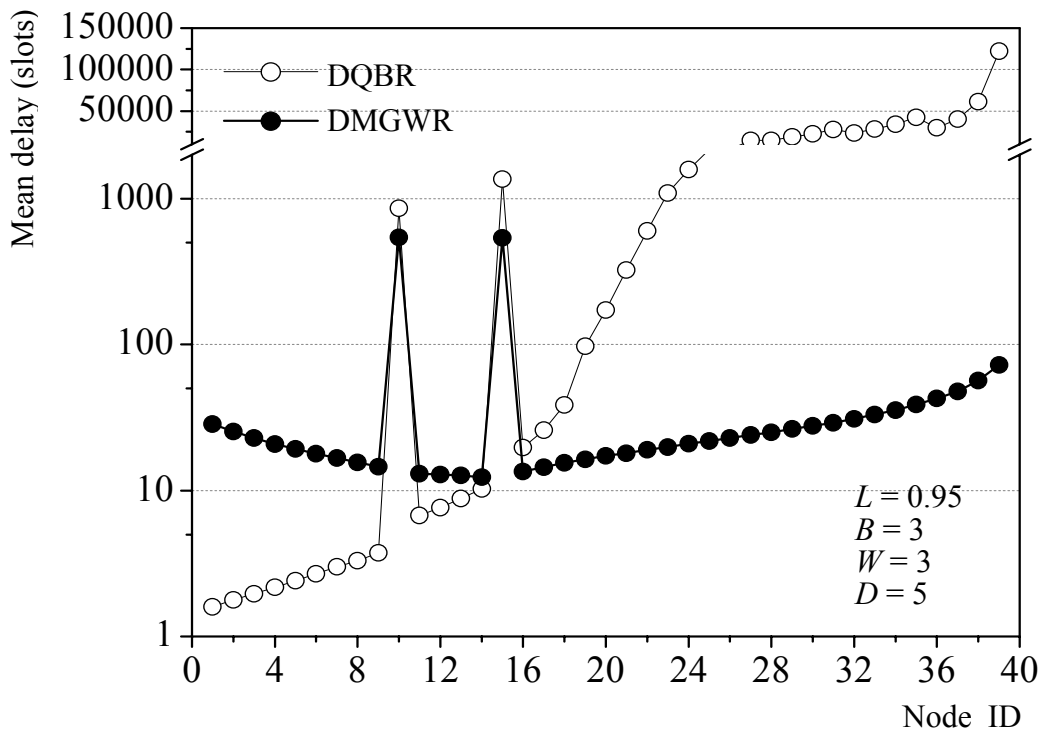
(a) Mean delay comparisons for malicious nodes (Nodes 10 and 15) under extremely heavy load
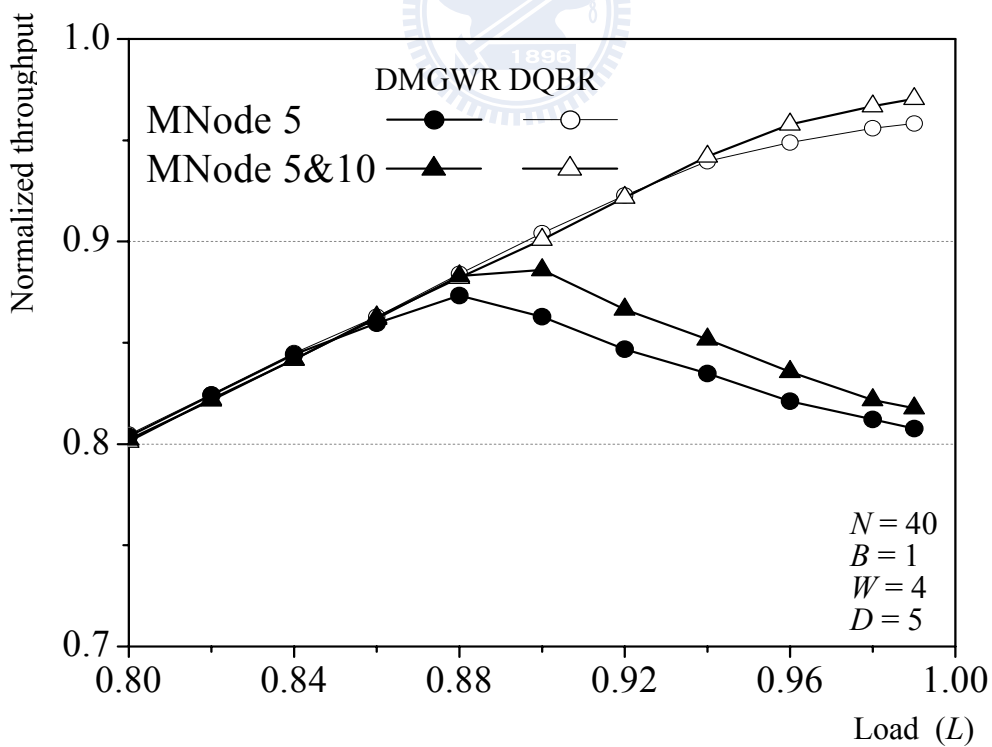


(b) Throughput comparison for malicious nodes (Nodes 10 and 15)

Figure 5.7. Performance comparisons for network with malicious nodes. (a) Mean delay comparisons for malicious nodes (Nodes 10 and 15) under extremely heavy load. (b) Throughput comparison for malicious nodes (Nodes 10 and 15).
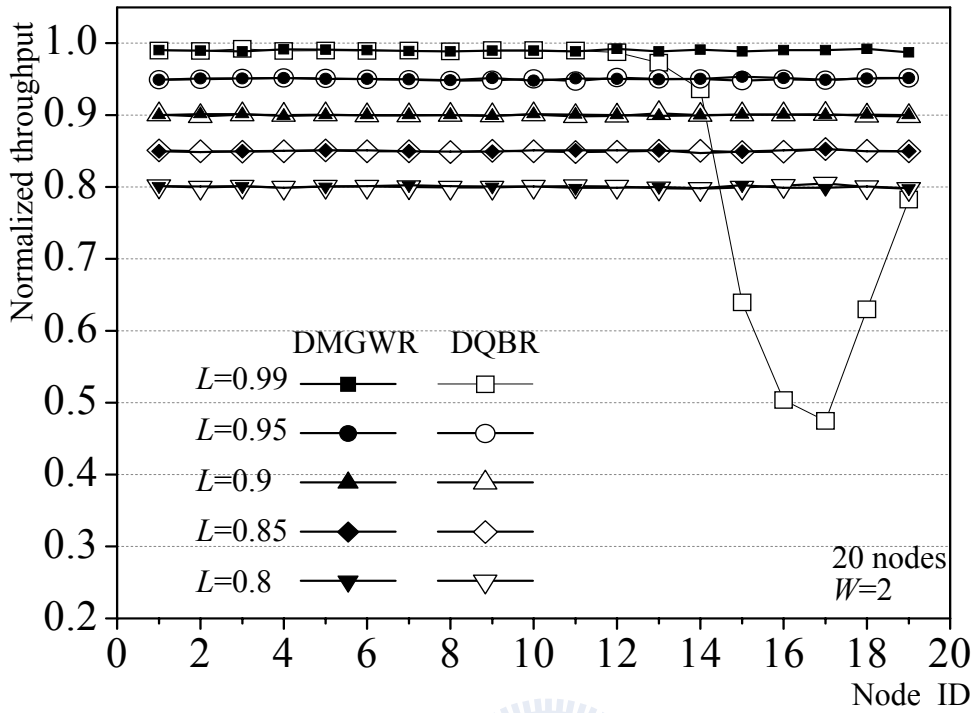
(a) Normalized throughput fairness comparisons when *N* is 20 and *W* is 2.
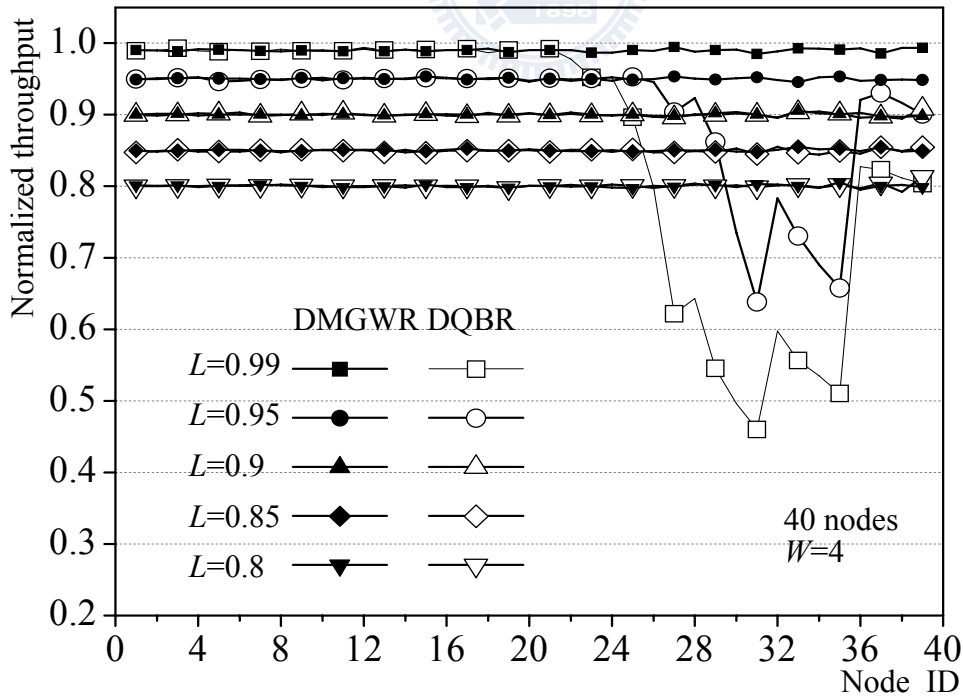


(a) Normalized throughput fairness comparisons when *N* is 40 and *W* is 4.

Figure 5.8. The impact of wavelength sharing on throughput performance under the same sharing ratio.
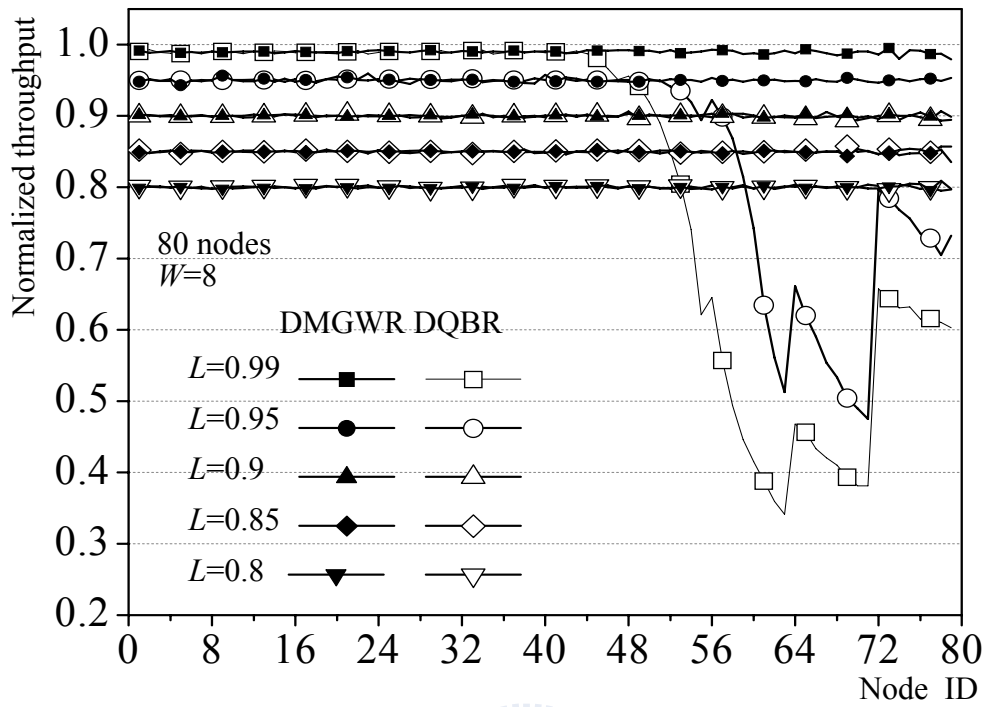
(a) Throughput fairness comparison for 80 nodes.
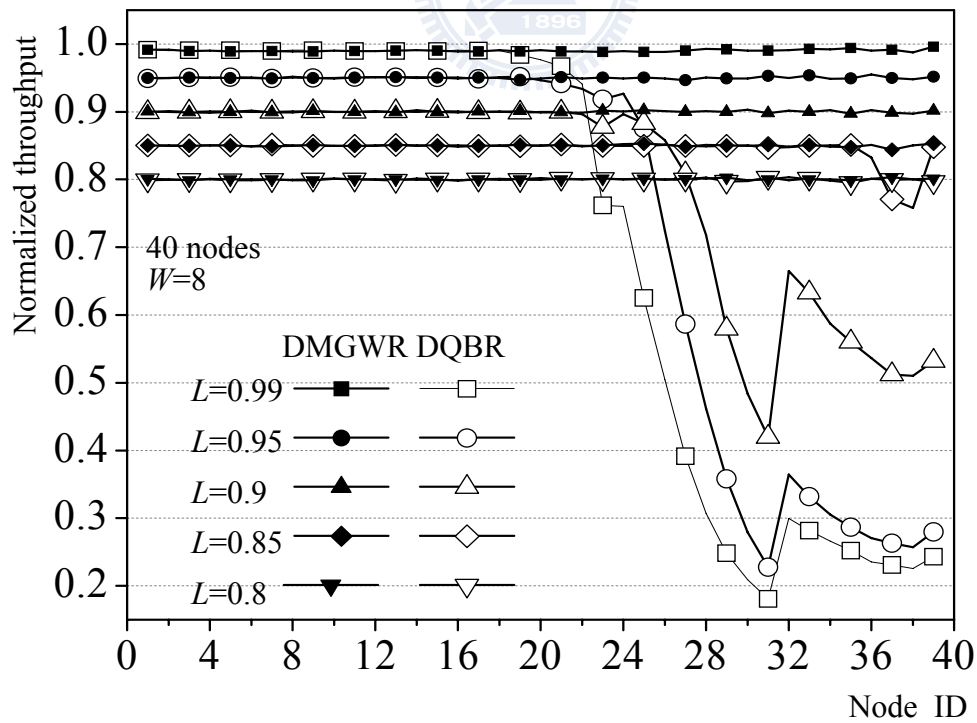


(b) Throughput fairness comparison for 40 nodes.

Figure 5.9. The impact of wavelength sharing on throughput performance with the same number of wavelengths.

## *5.4 Performance Study*

We further draw performance study between DMGWR scheme with a window size of seven and that with a window size of one on throughput, mean access delay, and potential receiver contention probability. The DSWR, referred to as Distributed Single Window Reservation, is the DMGWR with a window size of one which uses as a baseline assessment method.

First of all, we examine the impact of the multi-window design on throughput and delay performance under various traffic loads and burstiness. As depicted in Figure 5.10, under a load of 0.8 (and lower) and Poisson arrivals, both DMGWR and DSWR achieve satisfactory throughput. However, as the load increases, while DMGWR guarantees throughput fairness to all nodes, DSWR renders upstream nodes suffering from throughput unfairness. This is because under heavy loads, with the multi-granularity design the upstream nodes often encounter more slot reservations requested by downstream nodes than the average idle slots passing along the ring. In such a situation, the upstream nodes have less chances to transmit packets resulting in throughput deterioration.

We further examine the impact of the multi-window design on delay fairness performance under various traffic loads and burstiness, as depicted in Figure 5.11 and Figure 5.12. We observe in Figure 5.11(a) that as the burstiness increases, DSWR undergoes worsening unfairness and incurs rapidly deteriorating delay for upstream nodes. By contrast, the DMGWR scheme invariably achieves superior delay and fairness irrespective of traffic loads and burstiness. As depicted in Figure 5.11(b), under a load of 0.7 and Poisson arrivals, both schemes achieve satisfactory delay fairness. However, as the load increases, while DMGWR guarantees delay fairness to all nodes, DSWR renders upstream nodes suffering from delay unfairness.
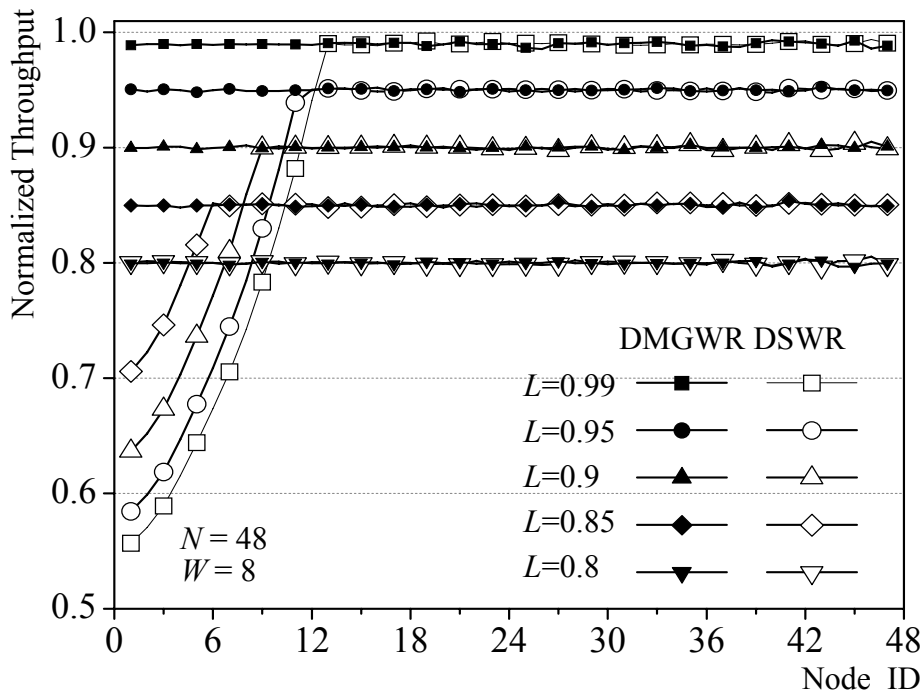
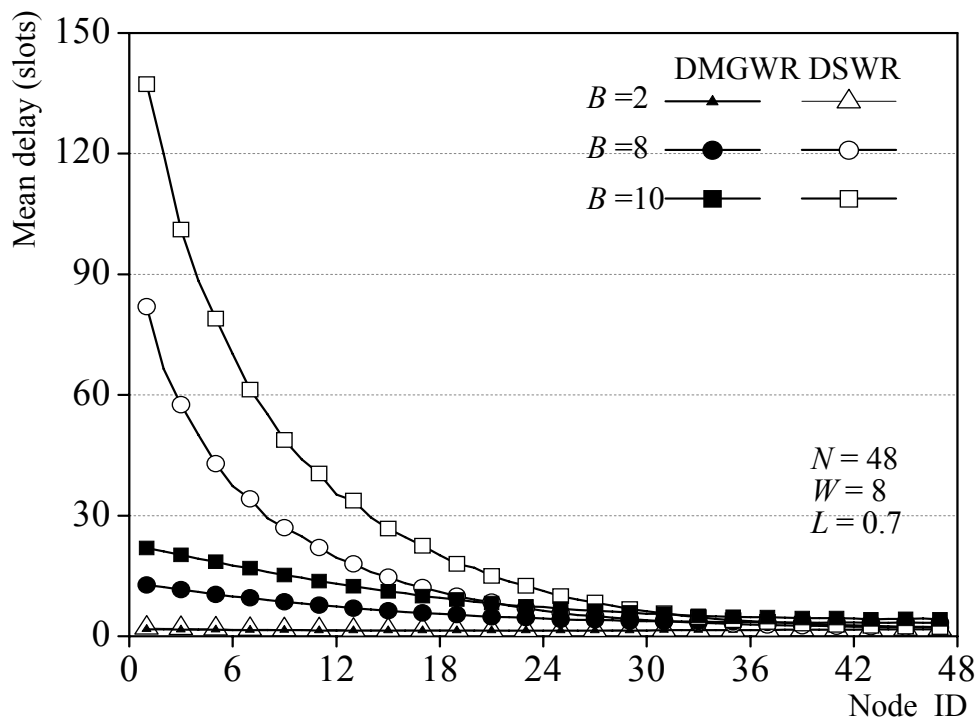Figure 5.10. The impact of multi-window design on throughput.

Moreover, we compare the impact of a multi-granularity design on the delay fairness under various loads. As shown in Figure 5.13, as the load increases, the DSWR scheme undergoes severe unfairness towards upstream nodes. In contrast to this, the DMGWR scheme achieves superior delay and fairness irrespective of traffic loads and number of granularity.

Figure 5.14 examines the mean potential receiver contention probability as a function of the offered load. In the experiment, we use the network with 48 nodes and the number of wavelengths ranges from 4 to 16. The probability of potential receiver contention is defined as the number of potential receiver contentions over the total sum of the number of transferred packets and the number of potential receiver contentions. What is called a potential receiver collision is simply a case where the top packet violates the receiver contention free basis. As one would expect for both protocols, greater values of offered load and number of wavelengths result in higher potential receiver contention probability. Finally, we discover that the DMGWR

70

protocol will yield a sudden decrease probability in conditions with a heavy load for all cases of number of wavelengths. This is due to the fact of the proposed dynamic bandwidth reservation method that a heavier network load will result in higher ratio of the number of transferred packets to the number of potential receiver contentions, thus sharply reducing the probability of potential receiver contention.

We compares the normalized throughput fairness of various packet selection strategies. In the experiment, we use the network with 24 nodes under extremely load (0.99). As depicted in Figure 5.15, we observe that the highest normalized throughput is the MaxH strategy. This can be explained by the fact that the MaxH strategy tends to uniformly distribute higher potential receiver-contention packets in each slot time, thus leading to the packet selection of downstream nodes easier to satisfy the collision-free constraint and yielding the highest normalized throughput. On the other hand, the MinH strategy tends to gather lower potential receiver-contention in a slot time, the packet selection of downstream nodes is hard to satisfy the collision-free constraint and yielding lowest normalized throughput under extremely heavy load.

Finally, we examine the mean access delay of various packet selection strategies under different loads in case of $W = 20$ and $W=10$. We observe from Figure 5.16 that the lowest mean access delay is the MaxH strategy. As expected for all of the packet selection strategies, greater values of offered load and number of wavelengths result in higher mean delay due to the higher potential receiver contention probability.
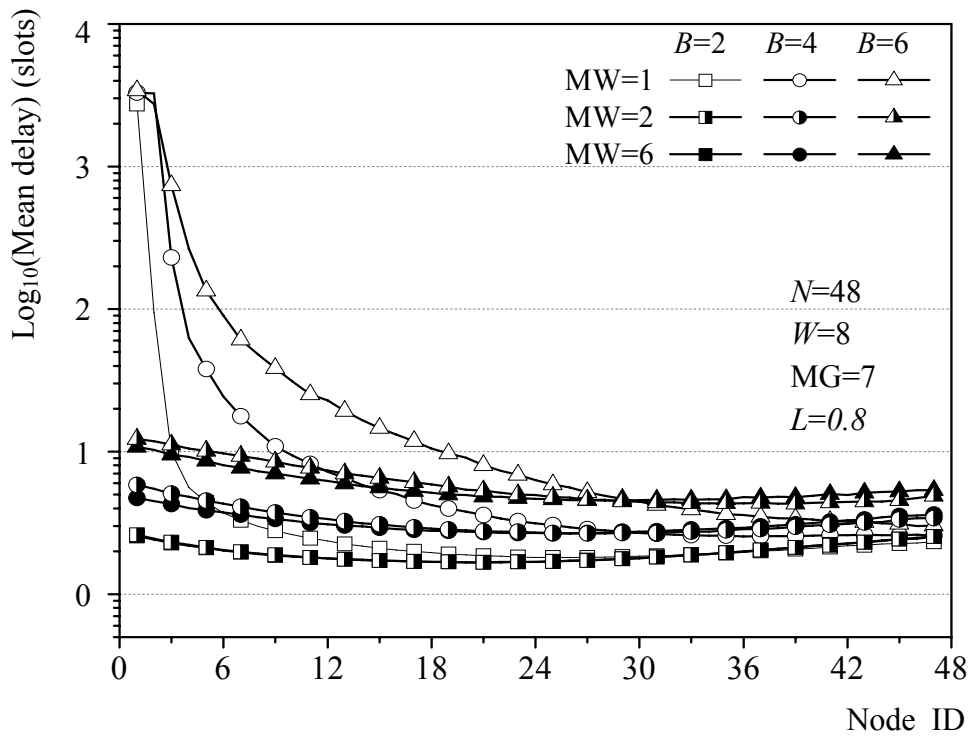
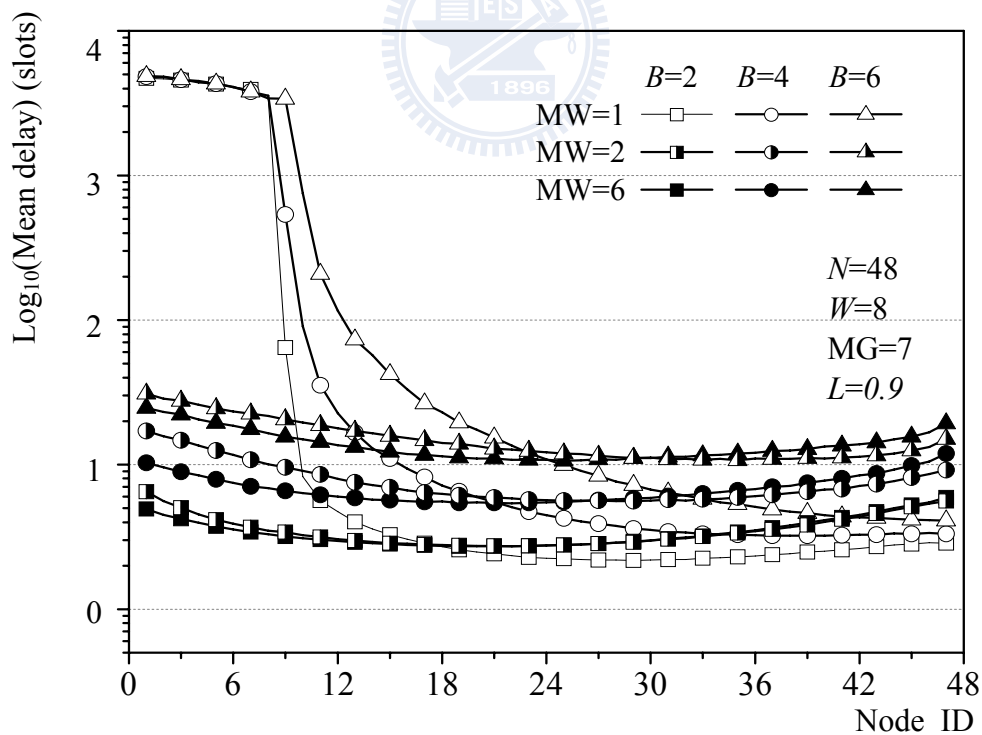(a) The impact of multi-window design on delay under various burstiness (*B*)
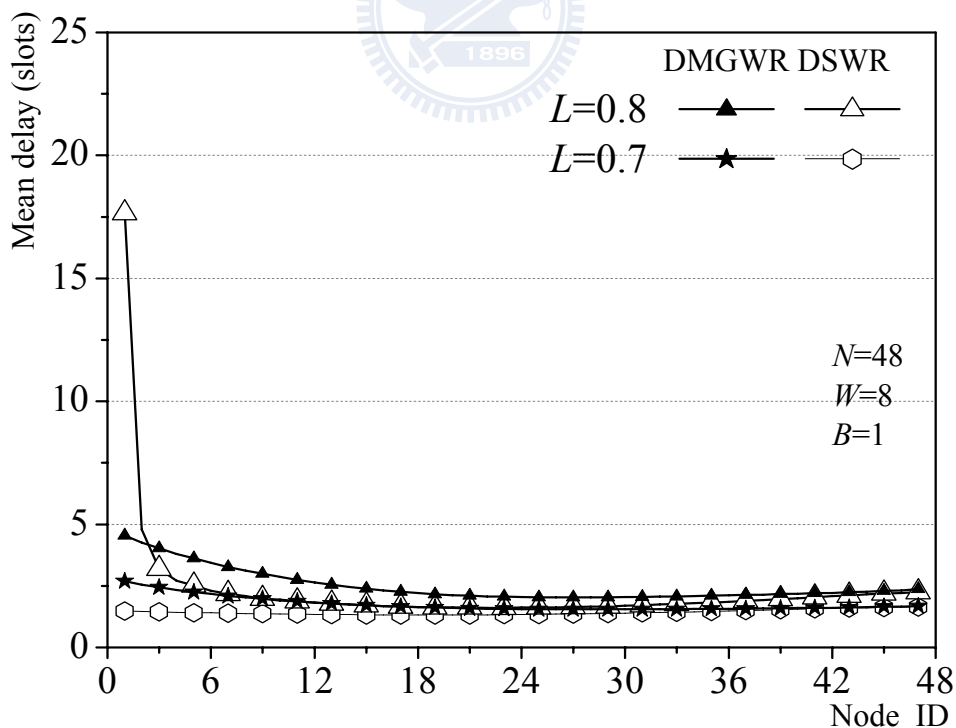


(b) The impact of multi-window design on delay under various loads

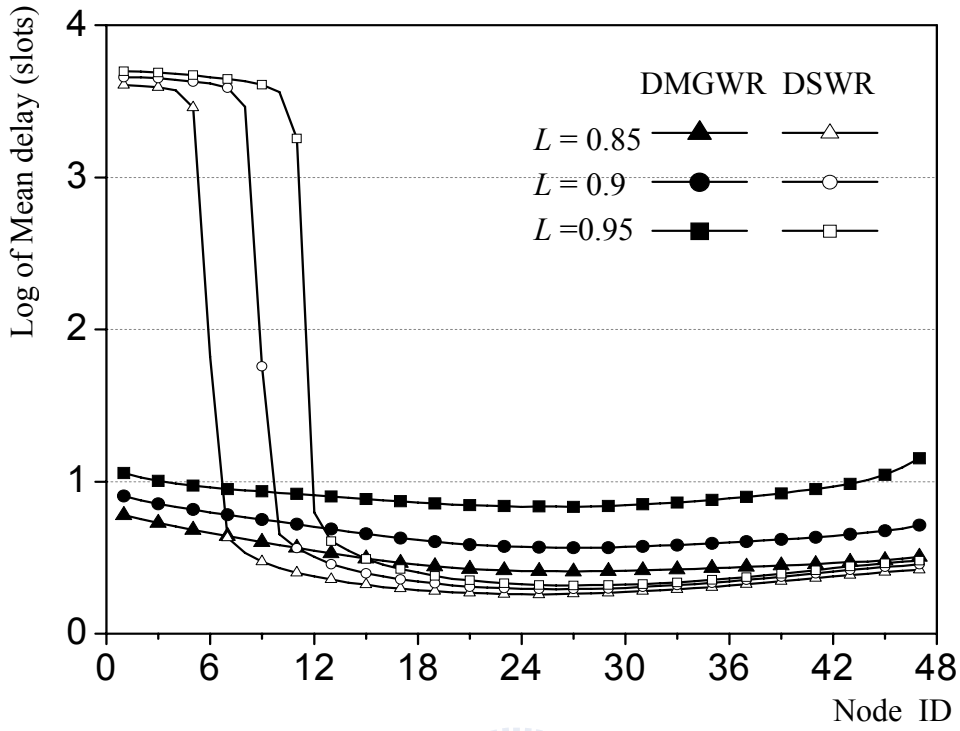Figure 5.11. The impact of multi-window design on mean delay.
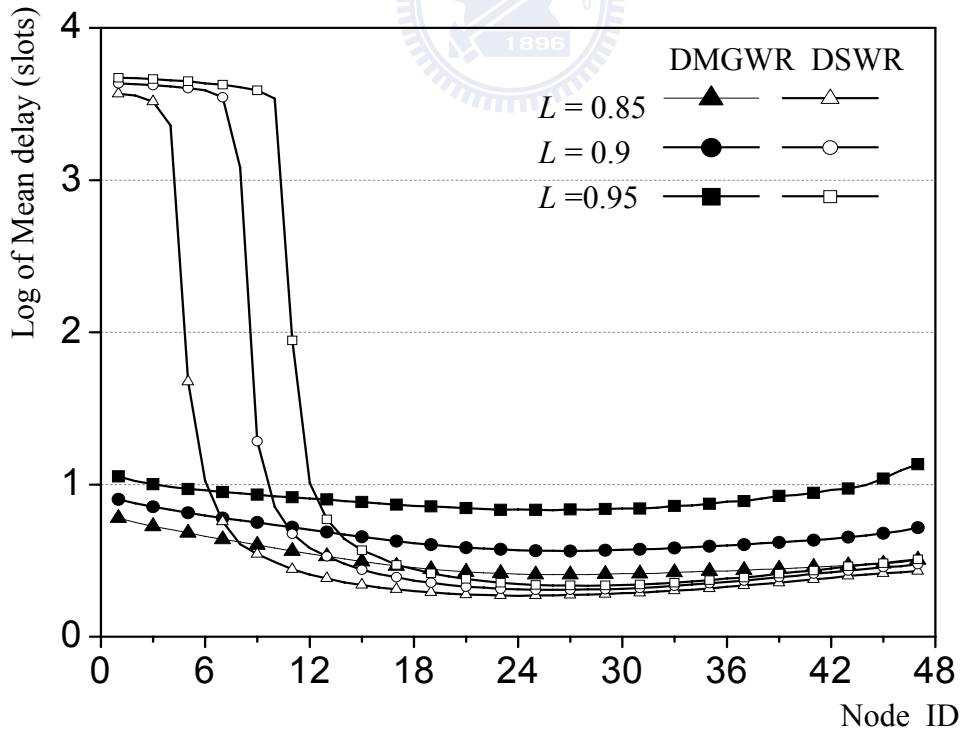
(a) Delay under various burstiness



(b) Delay under various burstiness

Figure 5.12. Delay performance comparison under various loads and burstiness.

(a) Delay fairness comparison for multi-granularity = 7



(b) Delay fairness comparison for multi-granularity = 3

Figure 5.13. Delay fairness comparison under various multi-granularity.
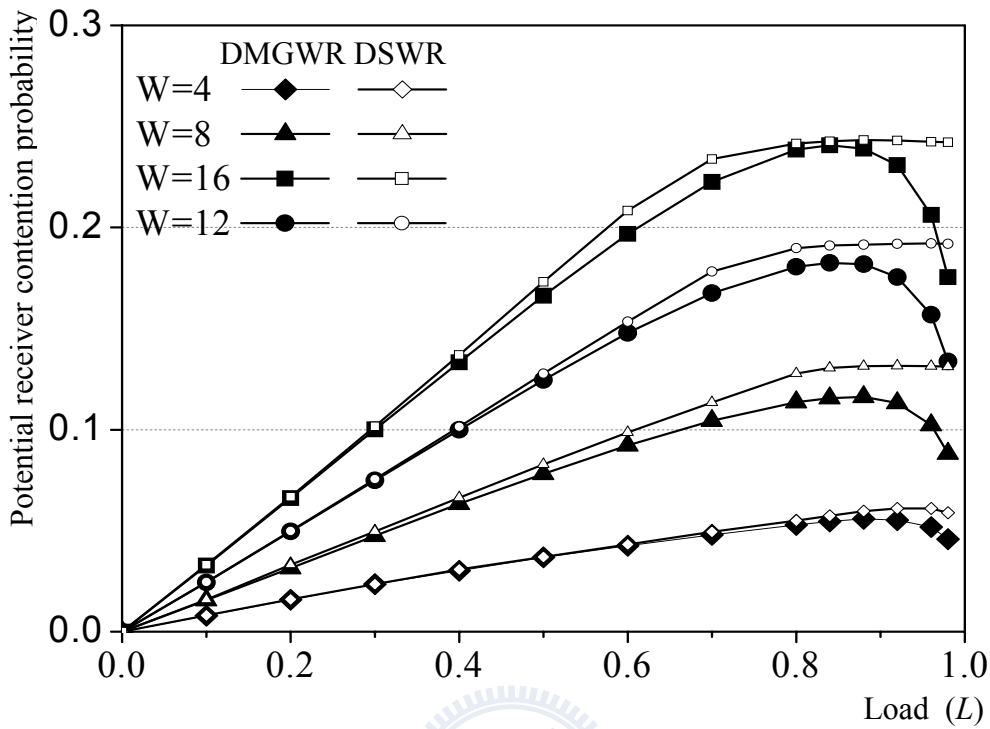
74

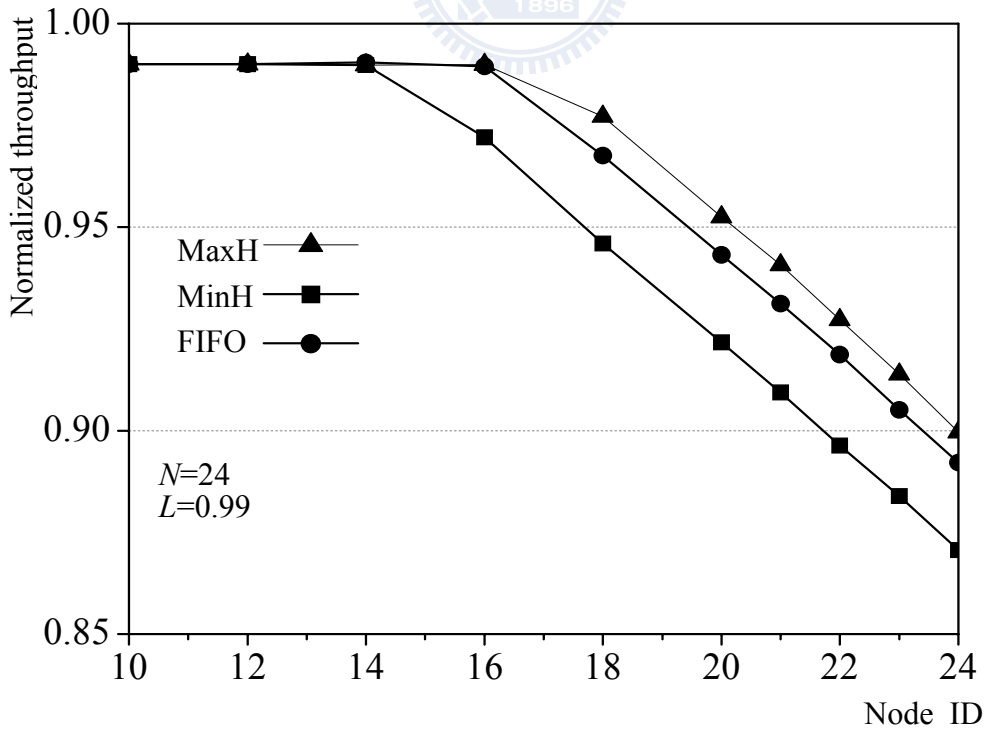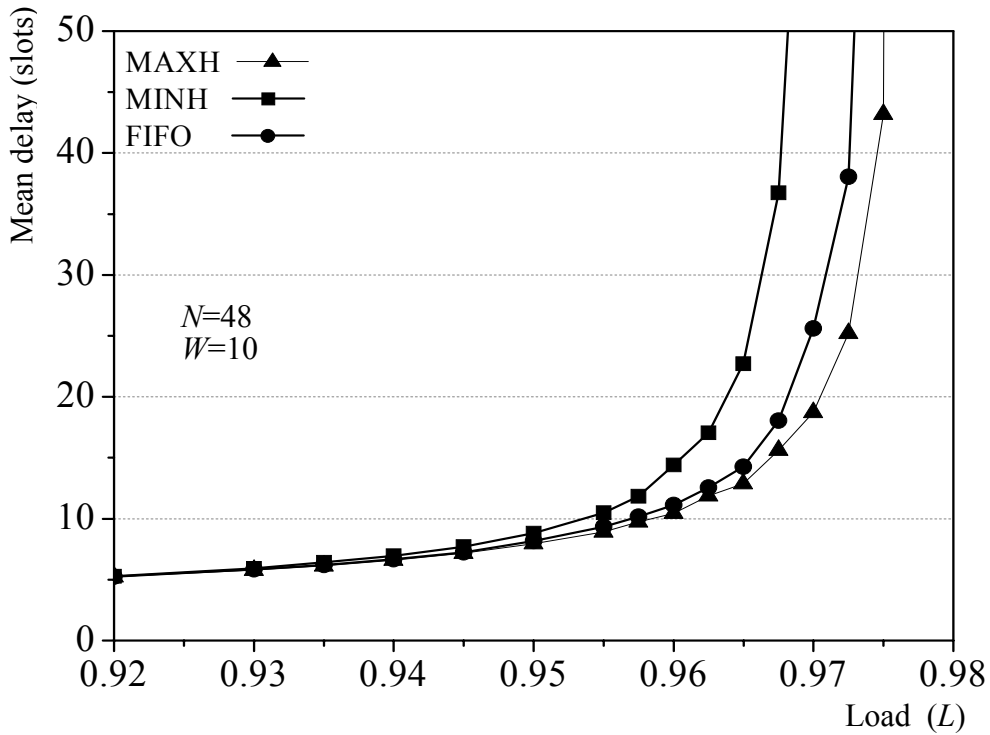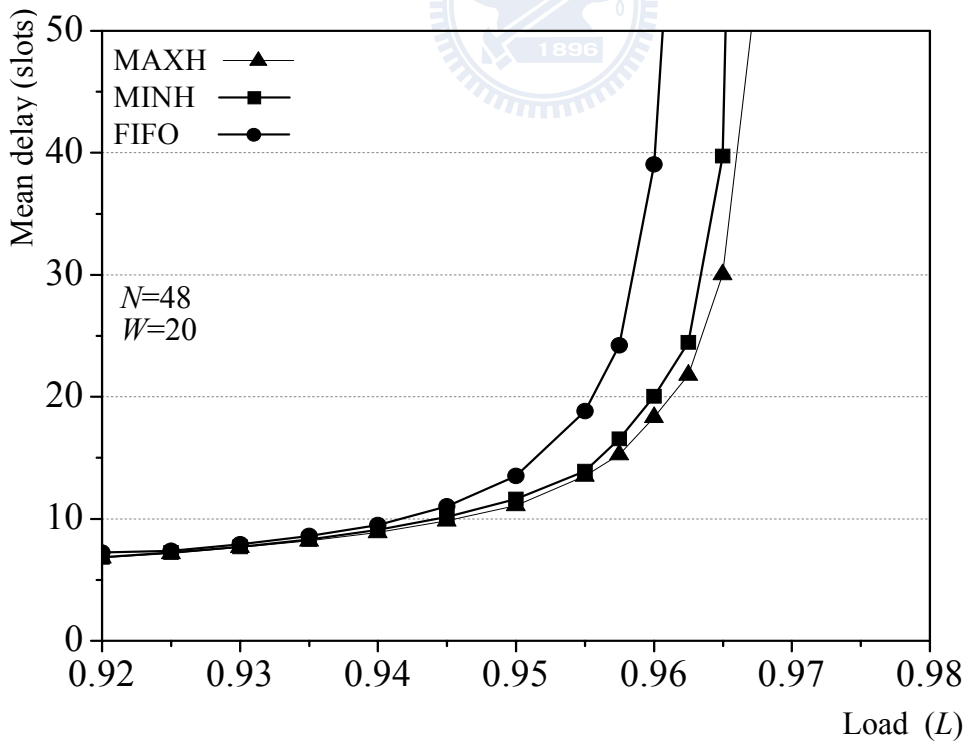Figure 5.14. Receiver contention probability comparison.



Figure 5.15. Normalized throughput fairness comparison for packet selection

strategies.

(a) Delay performance comparison under number of
wavelengths 10



(b) Delay performance comparison under number of
wavelengths 20

Figure 5.16. Delay performance comparison for packet selection strategies under various

number of wavelengths.

## *5.5 Discussions*

The main assumption made in this thesis is that the size of the generated packets is fixed and the packet size is equal to the length of a slot. While this had been assumed in order to simplify the complexity of the computer simulations, it does not correspond to a real networking environment where the packet size is variable. Therefore, one feasible mechanism could be considered in the network to deal with variable packet sizes. Under this scheme, packets larger than the slot size could be subdivided into small slices and each slice would fit in a slot precisely. As a result, only the last slice would be smaller than a slot and the bandwidth utilization is less impacted. It is worth noting that this technique and its technical implementation have already been implemented and validated in the HORNET project [10].

With the dynamic bandwidth allocation, DMGWR protocol provides a simple, fast, and fine granular resource allocation. Toward this end, in Table 3 shows a summary of multiple channel DQDB-like protocols. DMGWR exceeds these schemes several orders of magnitude in normalized throughput for larger network capacity particularly under extremely heavy load (0.99). DMGWR network outperforms these networks with respect to throughput, access delay, and fairness under various traffic patterns. More importantly, the fairness performance of DMGWR protocol is irrespective of the propagation distance and burstiness. In particular, under attack by malevolent nodes, OPMACS with DMGWR was justified robust and fair.

Table 3. Multi-channel dynamic bandwidth allocation schemes comparisons

| | WDMA | DQBR | DMGWR | DSWR |
|---|---|---|---|---|
| Node structure | $FT^2$-$FR^2$/ $TT^2$-$TR^2$ | $FT^2$-$FR^2$/ $TT^2$-$FR^2$ | $TT^2$-$TR^2$ | $TT^2$-$FR^2$ |
| Collision Handling | Collision & retransmission | Collision avoidance | Collision avoidance | Collision avoidance |
| Bursty-traffic adaptation | poor | poor | best | better |
| Propagation distance adaptation | poor | better | best | better |
| Scalability | N | poor | best | poor |
| Statistical multiplex gain | poor | poor | best | poor |
| Impact of attacking by malevolent nodes | poor | poor | best | better |
| Wavelength sharing | dynamic | static | dynamic | dynamic |
| Testbed | — | HORNET CC | OPACS | OPACS |
| References | [13] | [10,11] | [26-30] | [26,27] |

，

# Chapter 6. Testbed Experimentation and Results

An experimental laboratory testbed was conducted to model the OPACS system. The testbed serves to justify the feasibility of the header processing mechanism of OPACS discussed in Chapter 3. Notice that the MAC control part of OPACS is not included in the experiment. This is because the MAC performance, such as system throughput and access delay, can generally be best delineated via simulation results, as what presented in the Chapter 5.

## 6.1 Experimental node setup

The experimental set-up is shown in Figure 6.1. Four tunable optical transmitters are used to generate optical packets, respectively at wavelengths 1548.4 nm, 1553.8



Legend:
   TL: Tunable Laser;
   EA: Electro-Absorption;
   OADM: Optical Add/Drop Multiplexer;

Figure 6.1. Experimental node setup.

nm, 1554.8 nm, and 1556.1 nm, with an average power of 0 dBm per channel. It is worth pointing out that these wavelengths are not selected following the ITU-T WDM standards, but due to the availability of grating filters on these wavelengths. However, since there is only short fiber span inside the header processor, the system is thus free from dispersion. Thus, the location of wavelengths is irrevelant to the feasibility of the architecture.

The transmitter consists of a continuous wave laser and an EA-based external modulator. A 10-GHz pulse pattern generator with pre-stored header and payload bit sequences is used to modulate the light. The header signal is RZ-encoded at a data rate of 1 Gb/s. Each header is 26 bits in length, including an 8-bit preamble, a 4-bit header control, and a 6-bit address, besides the guard time. The payload signal (250 bytes long) is NRZ-encoded at a data rate of 10 Gb/s. Both the header and payload are generated by the 10-GHz pattern generator. Particularly, the generation of the 1-Gb/s RZ-encoded header-bit waveform is emulated through the generation of five consecutive 10-GHz NRZ pulses.
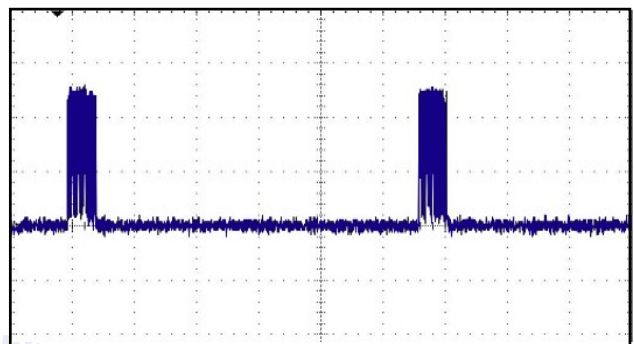
## 6.2 Experimental results

Two experiments are carried out to demonstrate the viability of OPACS. In the first experiment, the packet signal trace is first shown in Figure 6.2(a), with the eye diagram of the payload shown in the inset of the figure. At the input of the header processor, we use an optical splitter and two 2-ns-switching-time SOA-based optical gates to separate the payload (Figure 6.2(b)) from the header (Figure 6.2(c)). In the payload path, a tunable fiber delay line is employed to ensure that payloads and headers are synchronized upon departure. In the header path, the headers of different channels are reflected by fiber Bragg gratings of different distances with a total loss of
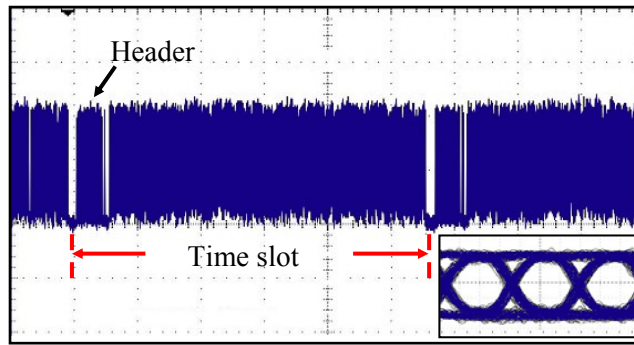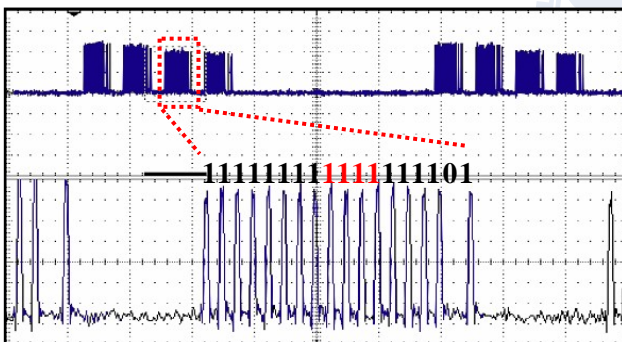
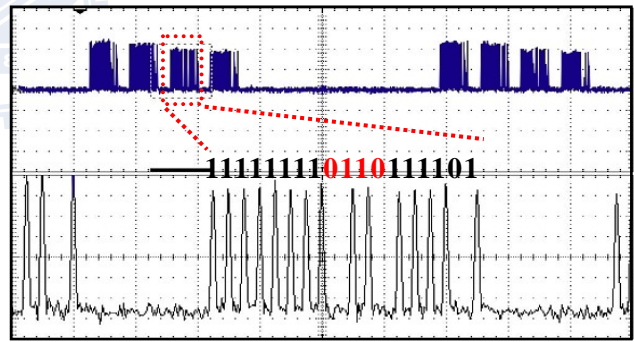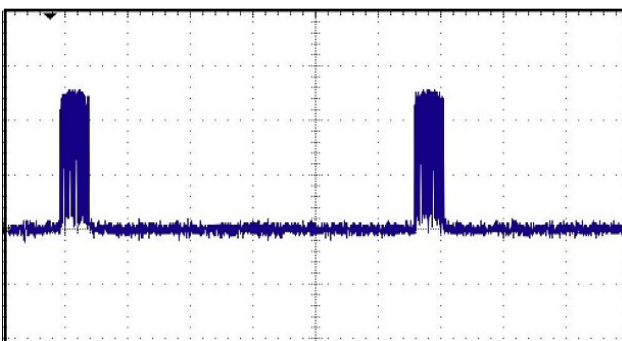(a) . Packet signal trace with eye diagram of payload
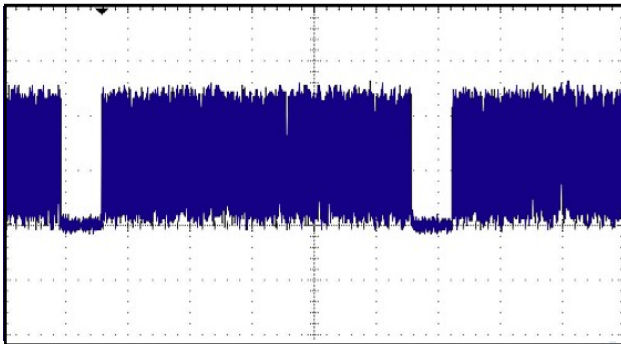


(b) . The payload part



(c) . The parallel headers
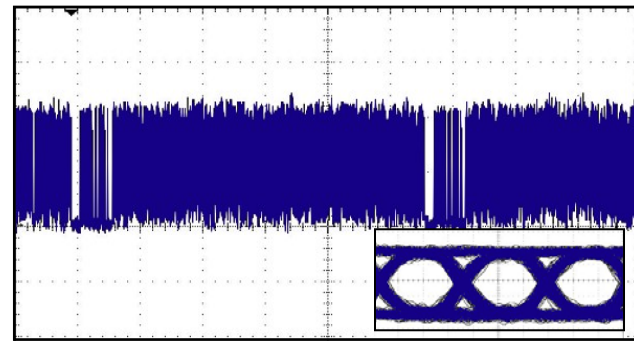


**111111111111111101**

(d). The serial headers



**111111110110111101**

(e). The serial headers after being modified by EAM



(f). The modified parallel headers.



(g). Departing packet signal trace with eye diagram of payload .

Figure 6.2. Experimental results- signal traces observed at stages (a)-(g).

3dB. The lengths of different channels' round-trip paths are given as multiples of the header duration. Specifically, if channel $\lambda 1$'s path length is T, then $\lambda 2$'s round-trip path length will be $T+\Delta T$, $\lambda 3$'s round-trip length will be $T+2\Delta T$, and so on, where $\Delta T$ is one header duration. Accordingly in the experiment, the fiber length between gratings is half of the header duration, namely 13 ns. With this timing arrangement, after being reflected by the grating array, the headers are converted from parallel to serial in the time domain (Figure 6.2(d)). Therefore, instead of using a header receiver for each channel, the system requires only one header receiver and rewriter module for the update of all headers. They are then routed to the header receiver and rewriter (EA Modulator) via a circulator.

The header signal is tapped by a 50/50 tap coupler and received by an optical burst mode receiver for recovering the header information. The bottom part of Figure 7.2(d) shows the RZ-encoded signal of the header on $\lambda 3$. Recall that the header information is always changed from 1 to 0 due to the MAC design, so that the header rewriter is designed as a pulse eraser. Since RZ pulses enables fast clock phase selection and can be erased with higher timing margin than NRZ pulses, we thereby adopt the RZ encoding format for the header signal. Although different headers are converted from parallel to serial by gratings, the clock phase of the combined header signal is not continuous. Thus, the header receiver determines the best clock phase for each individual header signal from the header's preamble pulses.

As shown in Figure 6.2(e), the control bits are modified (based on the DMGWR scheme) from "1111" to "0110" by the EA modulator with an extinction ratio of 12 dB. The headers are then time-to-wavelength converted from serial back to parallel (Figure 6.2(f)) in the time domain through the same grating array. Through such a design, the grating array in the input section can be simultaneously used in the output section with a reversing signal propagation direction. Finally, the modified header

signals are recombined with the payload (Figure 6.2(g)) via an optical coupler without any crosstalk between the headers and payload. The eye diagram of the payload demonstrates that the payload signal is penalty free throughout the system.

In the second experiment, Figure 6.3(a) shows the time domain scope trace of four overlapping headers. The average input power per channel to the header processing module is -5dBm. After the wavelength-to-time conversion, which results in 3~4 dB loss, four headers are time separated, as shown in Figure 6-3(b).
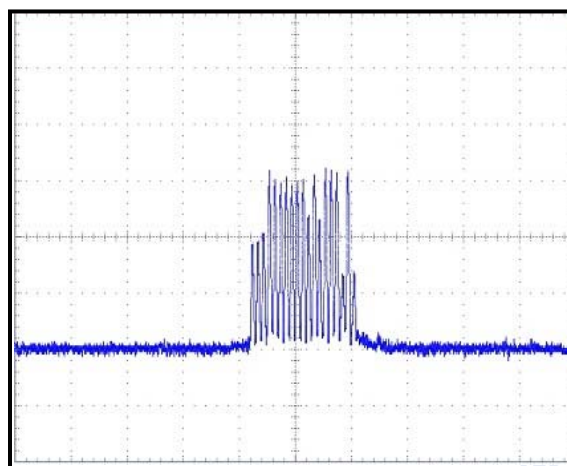


(a) Original parallel



(b) Time-divided
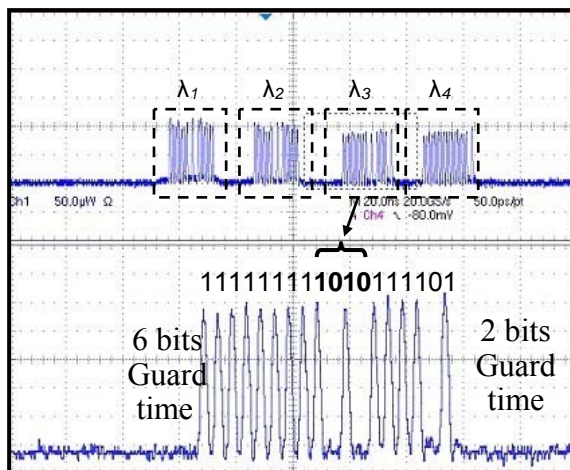


(c) Updated headers
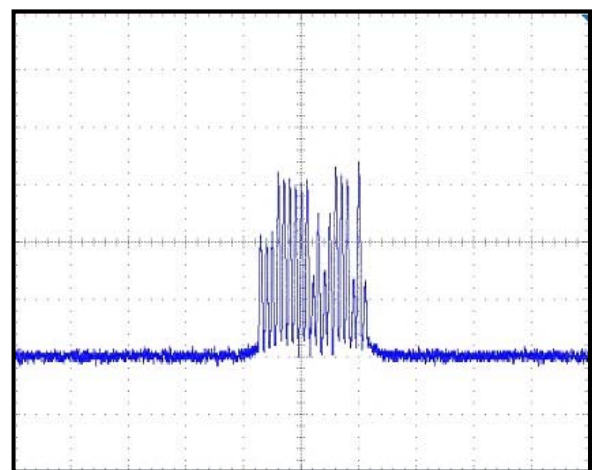


(d) updated parallel

Figure 6.3. The second experimental results.

In Figure 6.3(b), we depict the header for $\lambda_3$ (the third header) with control "1111" and address "111101". The headers were passed to the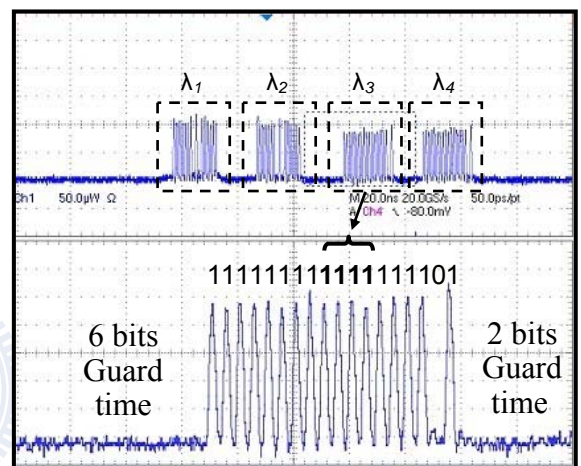 header processor by a 50/50 tap coupler and a burst mode receiver. As shown in Figure 6.3 (c), the control bits were modified according to the DMGWR scheme to "1010" by the EA modulator with an extinction ratio of 12 dB. Finally, Figure 6.3 (d) displays the four overlapping header after being successfully modified and time-to-wavelength converted through the same fiber Bragg grating.

# Chapter 7. Conclusions and Future Work

The work presented in this thesis investigates the OPACS architecture and shows that it is advantageous, viable, and practical. A diverse compilation of accomplishments that includes qualitative, quantitative, and experimental results has been presented in Chapters 3, 4, 5, and 6. The results were accomplished using a variety of tools, including a computer simulator developed for the project and a testbed constructed in the laboratory.

## 7.1 Conclusions

The thesis has presented the design and experimentation of an optical-header processing and medium access control system, OPACS, for a 10-Gb/s OPS WDM metro slotted-ring network. The system includes an in-band TDM-based optical header control subsystem and a DMGWR medium access control scheme. To perform header-modification operations, unlike traditional TDM-based approaches which generally require highly precise control timing and alignment to perform header erasing and rewriting operations, OPACS allows multiple optical headers to be efficiently and simultaneously detached and attached from/to the data payload by taking advantage of the sharing of fiber Bragg grating array between the input and output sections of the system and the particularly notable DMGWR design.

With the ability in dynamic bandwidth reservation via the multi-granularity and multi-window designs, DMGWR provides more efficient resource allocation in response to bursty data traffic and time-varying traffic conditions. Simulation results clearly demonstrate that DMGWR outperforms two existing networks (WDMA and HORNET networks) with respect to throughput, access delay, and fairness under

various traffic patterns. Essentially, DMGWR guarantees access fairness to all nodes regardless of the propagation distance and traffic burstiness. DMGWR is shown robust and fair even under attack by malevolent nodes. Furthermore, OPACS with DMGWR was shown to achieve exceedingly efficient and fair bandwidth allocation under various traffic loads and burstiness. Besides, we discussed three packet selection schemes for the DMGWR protocol. The MaxH selection scheme are found to provide a satisfactory compromise between performance and implementation complexity.

Finally through the 10-Gb/s experimental system, experimental results have illustrated signal traces observed at seven different stages within the system, demonstrating the viability of OPACS for optical packet-switched WDM metro ring networks. Due to the simplicity and highly-efficiency of the TDM-based header processing technique and notable DMGWR protocol being used, the OPMACS testbed is particularly attractive for high-capacity next-generation metropolitan access network market.

## 7.2 Future Works

As summarized in Section 7.1, the concepts of OPACS have been successfully demonstrated. Nonetheless, several interesting research avenues remain for the new network. One important networking detail is neglected in this thesis. Thus far it has been assumed that all traffic carried by the OPACS network is best-effort based traffic. As a result, the current MAC protocol is designed only for such conditions. However, the real networks face the different type of service, such as voice, video, and data. Thus, the OPACS MAC must be designed to differentiate between classes of service in the traffic it transports and to provide the appropriate level of quality of service (QoS). Second, the entire protocol suite must have the ability to accommodate

circuit-based traffic. This implies that OPACS would be transparent to the connections to the network, whether they are packet-based like Gigabit Ethernet or circuit-based like SONET. Finally, it remains to future work to investigate the design and analysis of the QoS-Enabled OPACS network.

# Appendix

## Distributed Queue Dual Bus (DQDB)Protocol

The Distributed Queue Dual Bus (DQDB) [72] scheme was created for single-channel dual-bus metro networks of the 1980s. It is also known as IEEE 802.6. As depicted in Figure A.1, there are two unidirectional fiber buses, called the forward bus and reverse bus, running in opposite directions. The signal propagates on the two buses in opposite directions. More specifically, packets destined for downstream and upstream nodes are sent along the forward and reverse buses, respectively. The dual-bus network interconnects a number of different network nodes. There is two server node located at the end of each buses, which is responsible for continuously generating fixed-length slots initially, and resetting all used slots each time after the slots have traveled one lap of the bus, as depicted in Figure A.1. Since the operations for accessing the two buses are identical and independent, for the simplicity of
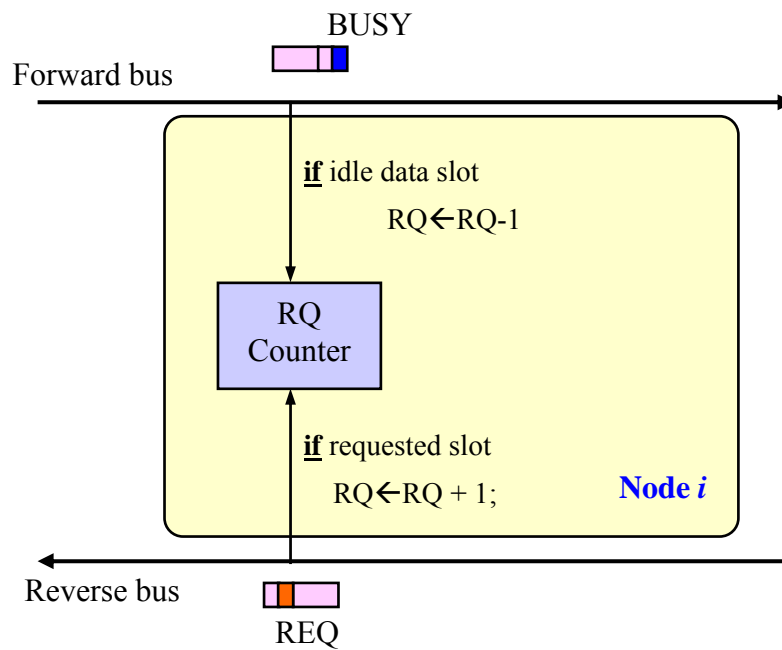


Figure A.1. The Idle state of DQDB scheme.

illustration, we hereinafter only focus on the access control for the forward bus.

DQDB [72] uses two control bits, a busy (B) and a request (R) bit in each slot to control access to the bus. To ensure that packets are sent in the order they arrived at the network, DQDB requires each node to maintain a distributed queue via a Request (RQ) and a CountDown (CD) counters. When a node has no packet to transmit, it increases the RQ counter by one for a requested slot passing by on the reverse bus, and decrease the RQ counter by one for an idle slot passing by on the forward bus (see Figure A.1).

When a node has a packet to transmit, it must issue a reservation request prior to the transmission, as depicted in Figure A.2. It will first find an available request slot on the reverse bus and set it to one, it then transfers the RQ counter to CD counter, and resets the RQ counter to zero. The node continues to increase the RQ counter by one for each requested slot on the reverse bus and decrease the CD counter by one for each idle data slot on the forward bus. When the CD counter goes to zero, the node waits for the next idle data slot to transmit the packet. If the node has more packets to
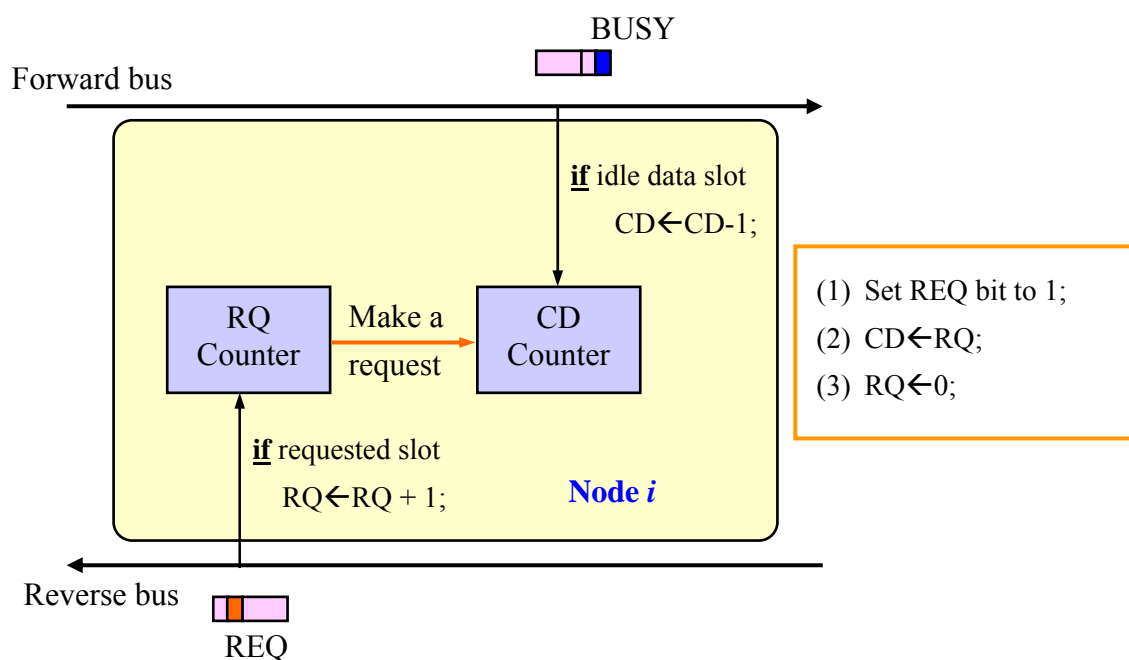


Figure A.2. The Countdown state of DQDB scheme.

89

transmit, it will issue another request reservation and then start to the countdown process; otherwise, the node goes back to the idle state. DQDB was shown to achieve superior throughput and delay performance, nevertheless undergoes the unfairness problem due to long propagation delay under heavy traffic conditions.

## **MetaRing Protocol**

The MetaRing protocol was originally proposed to address fairness in a single channel ring network [71]. In MetaRing, a control signal, called SAT (from SATisfied), is circulated from node to node in the opposite direction of data, possibly on a dedicated control channel. A node forwarding the SAT is granted a transmission quota $Q$: the node can transmit up to Q packets before the next SAT reception. When a node receives the SAT, it immediately forwards the SAT to the upstream node on the ring only if it is satisfied, i.e., if

  • no packets are waiting for transmission, or

  • $Q$ packets were already transmitted since the previous SAT reception.

If the node is not satisfied, the SAT is kept at the node until the node becomes satisfied. Thus, SATs are delayed by nodes suffering throughput limitations, and SAT rotation times increase with the network load. To be able to provide full bandwidth to a single node, the quota Q must be at least equal to the number of data slots contained in the ring: thus, $Q \geq$ RTT. To avoid throughput limitation, nodes must be able to buffer at least $Q$ packets: thus, the FIFO queue length must be larger than Q.

If a folded bus topology instead of a ring is assumed, as in the WONDER case, some issues need to be considered. First, the value of Q must be larger than in the ring case, since each time a SAT is forwarded, on average RTT slots are needed to reach

90

the next node. Therefore, the quota Q must be at least N times the RTT to avoid starvation when only one node is active: thus, $Q \geq N \times RTT$. Queue lengths must increase accordingly.

Second, due to the folded bus topology, only the last node can potentially delay the SAT. Indeed, the last node has the lowest opportunities to access the channel since all other nodes are positioned upstream to it. As a consequence, when the first node forwards the SAT to the last one, all the other nodes have already renewed their quotas and typically have also began transmission. Thus, the last node receives the SAT and delays it until the channel becomes free. In overload conditions, the SAT is delayed until all nodes run out of quota. Since the SAT propagates in the upstream direction, each node releases the SAT and is able to transmit on average for RTT time slots until it is flooded by the traffic from the upstream node who has renewed its quota. As a result, when the SAT comes back to the last node, all nodes have a residual quota approximately equal to Q−RTT. Now, it is straightforward to realize that in the worst case, the SAT will be delayed at most for $N \times (Q- RTT)$ slots. Only when the last node exhausts its quota, the SAT is released and forwarded to the other nodes. However, since all these nodes are satisfied (i.e., they run out of quota), the SAT is simply forwarded with no delay until it reaches again the last node, where it is delayed again until satisfaction is achieved.

## ATMR Protocol

Based on Orwell [70], the fairness protocol developed for ATMR achieves global fairness through a cyclic credit reset procedure and a distributed window mechanism. Each node maintains a window counter, or windows size in the ATMR terminology, that indicates how many cells that node may transmit within a fairness cycle, known

as reset period in the protocol's terminology. The initial value of the window counter is set to a pre-defined credit. A node decrements its window counter by 1 every time it transmits a cell. A node can transmit only if its window counter is greater than zero.

The ATMR protocol represents a credit allocation scheme and provides fairness control by means of a distributed credit mechanism and a cyclic reset scheme, which is based on a monitoring system. In the initialization phase, a predefined credit (termed window size) is allocated to each node. Moreover, every node maintains a window counter, which is decremented each time the node uses a free slot. When the window counter expires, the node gets into the inactive state, i.e. it is not allowed to send any data until the next reset. And also becomes inactive if it has nothing to transmit, i.e. its transmission queue is empty. In order to properly apply the reset mechanism, every node has to know about which station was the last active node. Therefore, each active node overwrites a so-called busy address field in the header of every incoming slot with its own address. Thus, each node receiving a slot with its own busy address assumes that all other nodes are inactive. If the last active node detects inactivity of all other nodes it generates a reset immediately after its own transmission. The reset mechanism causes the nodes to set up their window counters to the predefined window size representing the credit allocation. This way it is guaranteed that every node uses a maximum number of slots between two subsequent reset cycles.

Every node with a window counter greater than zero and with traffic backlog to transmit, writes its own address (busy address) into the access control field (ACF) of each incoming cell regardless of the status of that cell. A node that finds its own address knows that all the other nodes have completed their transmissions.

A node that detects that all the other nodes have completed their transmissions issues a reset cell. A reset cell rotates around the ring resetting every node's window

counter to its initial value. The node that issues a reset cell is responsible for removing the cell from the ring. The time interval between two consecutive visits of a reset cell defines a reset period.

# References

[1] B. Mukherjee, "WDM Optical Communication Networks: Progess and Challenges," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 10, Oct. 2000, pp. 1810-1824.

[2] M. Yuang, S. Lee, P. Tien, Y. Lin, J. Shih, F. Tsai, and A. Chen, "Optical Coarse Packet-Switched IP-over-WDM Network (OPSINET): Technologies and Experiments," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 8, Aug. 2006, pp. 117-127.

[3] S. Yao, S. Yoo, B. Mukherjee, and S. Dixit, "All-Optical Packet Switching for Metropolitan Area Networks: Opportunities and Challenges," *IEEE Communications Magazine*, vol. 39, no. 3, Mar. 2001, pp. 142-148.

[4] S. J. Yoo, "Optical Packet and Burst Switching Technologies for the Future Photonic Internet," *IEEE/OSA Journal of Lightwave Technology*, vol. 24, no. 12, Dec. 2006, pp. 4468-4492.

[5] C. Develder et al., "Benchmarking and Viability Assessment of Optical Packet Switching for Metro Networks," *IEEE/OSA Journal of Lightwave Technology*, vol. 22, no. 11, Nov. 2004, pp. 2435–2451.

[6] M. Herzog, M. Maier, and M. Reisslein, "Metropolitan Area Packet-Switched WDM Networks: A Survey on Ring Systems," *IEEE Communications Surveys & Tutorials*, vol. 6, no. 2, 2004, pp. 2-20.

[7] M. Yuang, Y. Lin, S. Lee, I. Chao, B. Lo, P. Tien, C. Chien, J. Chen, and C. Wei, "HOPSMAN: An Experimental Testbed System for a 10 Gb/s Optical Packet-Switched WDM Metro Ring Network," *IEEE Communications Magazine*, vol. 46, no. 7, July 2008, pp. 158-166.

[8] M. Yuang, I. Chao, Y. Lin, B. Lo, P. Tien, and S. Lee, "A High-Performance Optical Access and Control System for Packet-Switched Metro WDM Ring Networks," in *Proc. IEEE GLOBECOM'2008*.

[9] C. Linardakis, H. Leligou, A. Stavdas, and J. Angelopoulos, "Using Explicit Reservations to Arbitrate Access to a Metropolitan System of Slotted Interconnected Rings Combining TDMA and WDMA," *IEEE/OSA Journal of Lightwave Technology*, vol. 23, no. 4, April 2005, pp. 1576–1585.

[10] I. White, M. Rogge, K. Shrikhande, and L. Kazovsky, "A Summary of the HORNET Project: A Next-Generation Metropolitan Area Network," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 9, Nov. 2003, pp. 1478-1494.

[11] I. White, M. Rogge, K. Shrikhande, and L. Kazovsky, "Design of a Control Channel Based Media Access Control Protocol for HORNET," *Journal of Optical Networking*, vol. 12, 2002, pp. 460–473.

[12] A. Carena, V. Feo, J. Finochietto, R. Gaudino, F. Neri, C. Piglione, and P.

Poggiolini, "RingO: An Experimental WDM Optical Packet Network for Metro Applications," *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 8, Oct. 2004, pp. 1561-1571.

[13]  J. Lu and L. Kleinrock, "A WDMA Protocol for Multichannel DQDB Networks," *IEEE GLOBECOM*, 1993, pp. 149-153.

[14]  M. Marsan, A. Bianco, E. Leonardi, A. Morabito, and F. Neri, "All-Optical WDM Multi-Rings with Differentiated QoS," *IEEE Communications Magazine*, vol. 37, no. 2, Feb. 1999, pp. 58-66.

[15]  A. Bianco, D. Cuda, J. Finochietto, and F. Neri, "Multi-MetaRing Protocol: Fairness in Optical Packet Ring Networks," *IEEE ICC*, 2007.

[16]  C. Linardakis, H. Leligou, A. Stavdas, and J. Angelopoulos, "Implementation of medium access control for interconnecting slotted rings to form a WDM metropolitan area network", *Journal Of Optical Networking*, vol. 3, no. 11, Nov. 2004, pp. 826-836.

[17]  C. Jelger and J. Elmirghani, "A Slotted MAC Protocol for Efficient Bandwidth Utilization in WDM Metropolitan Access Ring Networks," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 8, Oct. 2003, pp. 1295-1305.

[18]  C. S. Jelger and J. M. H. Elmirghani, "A Simple MAC Protocol for WDM Metropolitan Access Ring Networks," *IEEE GLOBECOM*, vol. 3, San Antonio, TX, Nov. 2001, pp. 1500–04.

[19]  C. S. Jelger and J. M. H. Elmirghani,, "Performance of a Slotted MAC Protocol for WDM Metropolitan Access Ring Networks under Self-Similar Traffic," *IEEE ICC*, vol. 5, New York, NY, Apr. 2002, pp. 2806–2811.

[20]  E. W. M. Wong, A. Fumagalli, and I. Chlamtac, "Performance Evaluation of CROWNs: WDM Multi-Ring Topologies," Proc. *IEEE ICC*, vol. 2, Seattle, WA, June 1995, pp. 1296–1301.

[21]  K. Bengi and H. As, "Efficient QoS Support in a Slotted Multihop WDM Metro Ring Network," *IEEE Journal on Selected Areas in Communications,* vol. 20, no. 1, Jan. 2002, pp. 216-227.

[22]  K. Bengi, "An Analytical Model for a Slotted WDM Metro Ring with A-Posteriori Access," *Proc. Optical Network Design and Modelling (ONDM)*, Torino, Italy, Feb. 2002.

[23]  M. A. Marsan et al., "On the Capacity of MAC Protocols for All-Optical WDM Multi-Rings with Tunable Transmitters and Fixed Receivers," *IEEE INFOCOM*, vol. 3, Mar. 1996, pp. 1206–1216.

[24]  M. A. Marsan et al., "MAC Protocols and Fairness Control in WDM Multirings with Tunable Transmitters and Fixed Receivers," *IEEE/OSA Journal of Lightwave Technology*, vol. 14, no. 6, June 1996, pp. 1230–1244.

[25]  D. Wonglumsom, I. White, K. Shrikhande, M. Rogge, S. Gemelos, F. An, Y. Fukashiro, M. Avenarius, and L. Kazovsky, "Experimental Demonstratoion of an Access Point for HORNET-A Packet-Over-WDM Multiple-Access MAN," *IEEE/OSA Journal of Lightwave Technology*, vol. 18, no. 12, Dec. 2000, pp.

1709-1717.

[26] M. Yuang, Y. Lin, and Y. Wang, "A Novel Optical-Header Processing and Access Control System for a Packet-Switched WDM Metro Ring Network," *IEEE/OSA Journal of Lightwave Technology*, vol. 27, no. 21, Nov. 2009, pp. 4907-4915.

[27] M. Yuang, Y. Wang, and Y. Lin, "Design and Experimentation of an Optical-Header Processing and Access Control System for a Packet-Switched WDM Metro Ring Network," *IEEE GLOBECOM*, 2009.

[28] M. Yuang, Y. Wang, and Y. Lin, "A Novel Medium Access Control and Processing System for a Packet-Switched WDM Metro Ring Network," *IEEE/OSA OFC*, 2008, Paper OTuI3.

[29] M. Yuang, Y. Wang, and Y. Lin, "VMAPS: a Verstile Medium Access Control and Processing System for a Packet-Switched WDM Metro Ring Network," *IEEE/OSA ECOC*, 2007.

[30] M. Yuang, Y. Wang, and Y. Lin, "Design and Experimentation of a Novel Optical-Header Processing and Medium Access Control System for a Packet-Switched WDM Metro Ring Network," *WOCC*, 2008.

[31] A. Fumagalli, J. Cai, and I. Chlamtac, "The Multi-Token Inter-Arrival Time (MTIT) Access Protocol for Supporting IP over WDM Ring Network," *IEEE ICC*, vol. 1, Vancouver, Canada, June 1999, pp. 586–90.

[32] J. Cai, A. Fumagalli, and I. Chlamtac, "The Multitoken Interarrival Time (MTIT) Access Protocol for Supporting Variable Size Packets Over WDM Ring Network," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 10, Oct. 2000, pp. 2094- 2104.

[33] G. Chang, J. Yu, A. Chowdhury, and Y. Yeo, "Optical Carrier Suppression and Separation Label-Switching Techniques," *IEEE/OSA Journal of Lightwave Technology*, vol. 23, no. 10, Oct. 2005, pp. 3372-3387.

[34] J. Cao, M. Jeon, Z. Pan, Y. Bansal, Z. Wang, Z. Zhu, V. Hernandez, J. Taylor, V. Akella, S. Yoo, K. Okamoto, and S. Kamei, "Error-Free Multi-hop Cascaded Operation of Optical Label Switching Routers with All Optical Label Swapping," *IEEE/OSA OFC*, vol. 2, Atlanta, GA, Mar. 2003, pp. 791-792.

[35] Z. Zhu, V. J. Hernandez, M. Jeon et al., "RF photonics signal processing in subcarrier multiplexed optical-label switching communication systems," *IEEE/OSA Journal of Lightwave Technology*, vol. 21, no. 12, Dec. 2006, pp. 3155-3166.

[36] G. Rossi, O. Jerphagnon, B. Olsson, and D. Blumenthal, "Optical SCM Data Extraction Using a Fiber-loop Mirror for WDM Network Systems," *IEEE Photonics Technology Letters*, vol. 12, July 2000, pp. 897-899.

[37] Y. Lin, W. Way, and G. Chang, "A Novel Optical Label Swapping Technique Using Erasable Optical Single-Sideband Subcarrier Label," *IEEE Photonics Technology Letters*, vol. 12, Aug. 2000, pp. 1088-1091.

[38] Y. Lin, M. Yuang, S. Lee, and, W. Way, "Using Superimposed ASK Label in a 10-Gb/s Multihop All-Optical Label Swapping System," *IEEE/OSA Journal of Lightwave Technology*, vol. 22, no. 2, Feb. 2004, pp. 351-361.

[39] H. J. Lee, V. Hernandez, V. K. Tsui, and S. J. B. Yoo, "Simple, polarisation-independent, and dispersion-insensitive SCM signal extraction technique for optical switching systems applications," *IEEE Electronic Letters*, vol. 37, no. 20, Sep. 2001, pp. 1240–1241.

[40] J. J. Yu, G. K. Chang, and Q. M. Yang, "Optical label swapping in a packet-switched optical network using optical carrier suppression, separation, and wavelength conversion," *IEEE Photonics Technology Letters*, vol. 16, no. 9, Sep. 2004, pp. 2156–2158.

[41] T. Kawanishi, K. Higuma, T. Fujita, J. Ichikawa, T. Sakamoto, S. Shinada, and M. Izutsu, "High Speed Optical FSK Modulator for Optical Packet Labeling," *IEEE/OSA Journal of Lightwave Technology*, vol. 23, no. 1, Jan. 2005, pp. 87-94.

[42] M. Cerisola, T. K. Fong, R. T. Hofmeister, L. G. Kazovsky, C. L. Lu, P. Poggiolini, and D. I. Sabido, "Subcarrier multiplexing of packet headers in a WDM optical network and a novel ultrafast header clockrecovery technique," in *OFC Tech. Dig. Series. Postconference Edition*, Washington, DC, 1995, vol. 8, pp. 273–274.

[43] C. L. Lu,M. Sabido, P. Poggiolini, R. T. Hofmeister, and L. G. Kazovsky, "Cord—AWdma optical network—Subcarrier-based signaling and control scheme," *IEEE Photonics Technology Letters*, vol. 7, no. 5, May 1995, pp. 555–557.

[44] E. Park and A. E. Willner, "Self-routing of wavelength packets using an all-optical wavelength shifter and QPSK subcarrier routing control headers," *IEEE Photonics Technology Letters,* vol. 8, no. 7, Jul. 1996, pp. 938–940.

[45] R. Olshansky, "Microwave subcarrier multiplexing: New approach to wideband lightwave systems," *IEEE Circuits Devices Mag.*, vol. 4, no. 6, Nov. 1988, pp. 8–14.

[46] R. Olshansky and V. A. Lanzisera, "60-channel FM video subcarrier multiplexed optical communication system," *IEEE Electronic Letters*, vol. 23, no. 22, Oct. 1987, pp. 1196–1198.

[47] H. J. Lee, S. J. B. Yoo, V. K. Tsui, and S. K. H. Fong, "A simple all-optical label detection and swapping technique incorporating a fiber Bragg grating filter," *IEEE Photonics Technology Letters*, vol. 13, no. 6, Jun. 2001, pp. 635–637.

[48] Z. Q. Zhu, V. J. Hernandez, M. Y. Jeon, J. Cao, Z. Pan, and S. J. B. Yoo, "RF photonics signal processing in subcarrier multiplexed optical-label switching communication systems," *IEEE/OSA Journal of Lightwave Technology*, vol. 21, no. 12, Dec. 2003, pp. 3155–3166.

[49] J. Cao, M. Y. Jeon, Z. Pan, Y. Bansal, J. Taylor, Z. Wang, Z. Zhu, V. Hernandez,

K. Okamoto, S. Kamei, and S. J. B. Yoo, "11-hop operation of optical-label switching system with all-optical label swapping," *IEEE/OSA ECOC*, Rimini, Italy, 2003.

[50] V. J. Hernandez, Z. Pan, J. Cao, V. K. Tsui, Y. Bansal, S. K. H. Fong, Y. Zhang, M. Y. Jeon, S. J. B. Yoo, B. Bodtker, S. Bond, W. J. Lennon, H. Higashi, B. Lyles, and R. McDonald, "First field trial of optical label-based switching and packet drop on a 477 km NTON/Sprint link," *IEEE/OSA OFC*, Washington, DC, 2002, vol. 1, pp. 168–169.

[51] N. Chi, J. Zhang, P. V. Holm-Nielsen, C. Peucheret, and P. Jeppesen, "Transmission and Transparent Wavelength Conversion of an Optically Label Signal using ASK/DPSK Orthogonal Modulation," *IEEE Photonics Technology Letters*, vol. 15, May 2003, pp. 760-762.

[52] M. Ohm and J. Speidel, "Quaternary Optical ASK-DPSK and Receivers with Direct Detection," *IEEE Photonics Technology Letters*, vol. 15, Jan. 2003, pp. 159-161.

[53] F. Liu and Y. Su, "DPSK/FSK Hybrid Modulation Format and Analysis of its Nonlinear Performance," *IEEE/OSA Journal of Lightwave Technology*, vol. 26, no. 3, Feb. 2008, pp. 357-364.

[54] N. Chi, L. Xu, L. Christiansen, K. Yvind, J. Zhang, P. Holm-Nielsen, C. Peucheret, C. Zhang, and P. Jeppesen, "Optical label swapping and packet transmission based on ASK/DPSK orthogonal modulation format in IP-over-WDM networks," *IEEE/OSA OFC*, Washington, DC, 2003, vol. 2, pp. 792–794.

[55] V. Olmos, J. Zhang, P. V. Holm-Nielsen, I. T. Monroy, V. Polo, A. M. J. Koonen, C. Peucheret, and J. Prat, "Simultaneous optical label erasure and insertion in a single wavelength conversion stage of combined FSK/IM modulated signals," *IEEE Photonics Technology Letters*, vol. 16, no. 9, Sep. 2004, pp. 2144–2146.

[56] N. Chi, C. Mikkelsen, L. Xu, J. F. Zhang, P. V. Holm-Nielsen, H. Y. Ou, J. Seoane, C. Peucheret, and P. Jeppesen, "Transmission and label encoding/erasure of orthogonally labelled signal using 40 Gb/s RZ-DPSK payload and 2.5 Gb/s IM label," *IEEE Electronic Letters*, vol. 39, no. 18, Sep. 2003, pp. 1335–1337.

[57] J. F. Zhang, N. Chi, P. V. Holm-Nielsen, C. Peucheret, and P. Jeppesen, "An optical FSK transmitter based on an integrated DFB laser-EA modulator and its application in optical labeling," *IEEE Photonics Technology Letters*, vol. 15, no. 7, Jul. 2003, pp. 984–986.

[58] C. W. Chow, C. S. Wong, and H. K. Tsang, "Optical packet labeling based on simultaneous polarization shift keying and amplitude shift keying," *Optics Letters*, vol. 29, no. 16, Aug. 2004, pp. 1861–1863.

[59] T. Koonen, G. Morthier, J. Jennen, H. de Waardt, and P. Demeester, "Optical packet routing in IP-over-WDM networks deploying two-level optical

labeling," *IEEE/OSA ECOC.*, Piscataway, NJ, 2001, vol. 4, pp. 608–609.

[60] N. Chi, B. Kozicki, J. Zhang, P. V. Holm-Nielsen, C. Peucheret, and P. Jeppesen, "Transmission properties of an all-optical labelled signal using orthogonal IM/FSK modulation format," in *IEEE/OSA ECOC*, Rimini, Italy, 2003, pp. 300–303.

[61] C. W. Chow and H. K. Tsang, "Orthogonal label switching using polarization-shift-keying payload and amplitude-shift-keying label," *IEEE Photonics Technology Letters*, vol. 17, no. 11, Nov. 2005, pp. 2475–2477.

[62] L. Xu, N. Chi, L. K. Oxenlowe, J. Mork, S. Y. Yu, and P. Jeppesen, "A new orthogonal labeling scheme based on a 40-Gb/s DPSK payload and a 2.5-Gb/s PolSK label," *IEEE Photonics Technology Letters*, vol. 17, no. 12, Dec. 2005, pp. 2772–2774.

[63] L. Xu, T. Wang, O. Matsuda, M. Cvijetic, I. Glesk, and P. R. Prucnal, "Wavelength-shift-keying (WSK) encoded pulses for optical labeling applications," *IEEE Photonics Technology Letters*, vol. 17, no. 12, Dec. 2005, pp. 2775–2777.

[64] W. Hung, C.-K. Chan, L.-K. Chen, and F. Tong, "A novel phase encoding scheme for optical packet label processing," *IEEE/OSA ECOC*, Rimini, Italy, 2003, Paper Tu4.4.4..

[65] C. Guillemot, M. Renaud, P. Gambini, C. Janz, I. Andonovic, R. Bauknecht, B. Bostica, M. Burzio, F. Callegati, M. Casoni, D. Chiaroni, F. Clerot, L. Danielsen, F. Dorgeuille, A. Dupas, A. Franzen, P. B. Hansen, K. Hunter, A. Kloch, R. Krahenbuhl, B. Lavigne, A. Le Corre, C. Raffaelli, M. Schilling, J. C. Simon, and L. Zucchelli, "Transparent Optical Packet Switching: The European ACTS KEOPS Project Approach," *IEEE/OSA Journal of Lightwave Technology*, vol. 16, no. 12, Dec. 1998, pp. 2117-2134.

[66] C. Bintjas, N. Pleros, K. Yiannopoulos, G. Theophilopoulos, M. Kalyvas, H. Avramopoulos, and G. Guekos, "All-Optical Packet Address and Payload Separation," *IEEE Photonics Technology Letters*, vol. 14, no. 12, Dec. 2002, pp. 1728-1730.

[67] H. Teimoori, J. Topomondzo, C. Ware, and D. Erasme, "Optical Packet Header Processing Using Time-to-Wavelength Mapping in Semiconductor Optical Amplifiers," *IEEE/OSA Journal of Lightwave Technology*, vol. 25, no. 8, Aug. 2007, pp. 2149-2158.

[68] J. Simsarian, A. Bhardwaj, J. Gripp, K. Sherman, Y. Su, C. Webb, L. Zhang, and M. Zirngibl, "Fast Switching Characteristics of a Widely Tunable Laser Transmitter," *IEEE Photonics Technology Letters*, vol. 15, no. 8, Aug. 2003, pp. 1038-1040.

[69] P. Tang, O. Eknoyan, and H. F. Taylor, "Rapidly Tunable Optical Add-Drop Multiplexer (OADM) using a Static-Strain-Induced Grating in LiNbO3," *IEEE/OSA Journal of Lightwave Technology*, vol. 21, no. 1, Jan. 2003, pp.

236-245.

[70] K. Imai, T. Ito, H. Kasahara, and N. Morita, "ATMR: asynchronous transfer mode ring protocol," *Computer Networks and ISDN Systems*, vol. 26, no. 6-8, March 1994, pp. 785-798.

[71] I. Cidon and Y. Ofek, "MetaRing - A Full-Duplex Ring with Fairness and Spatial Reuse," *IEEE Transactions on Communications*, vol. 41, no. 1, Jan. 1993, pp. 969-981.

[72] Distributed Queue Dual Bus (DQDB) Subnetwork of a Metropolitan Area Network (MAN), *IEEE Standard 802.6*, Dec. 1990.

[73] W. Fischer and K. Meier-Hellstern, "The Markov-modulated Poisson Process (MMPP) Cookbook," *Performacne Evaluation*, vol. 18, no. 2, Sept. 1993, pp. 149-171.

# Biography

**Ya-Shian Wang** received the the M.S. degrees in Computer Science and Electrical Engineering from the National Central University, Jhongli, Taiwan, in 1994. She is currently working toward the Ph.D. degree in Computer Science from National Chiao Tung University, Hsinchu, Taiwan. In 1994, she joined the Telecommunication Laboratories, Chunghwa Telecom Co., Ltd., Taoyuan, Taiwan, where she is currently engaged in network control and management of broadband networks. Her current research intrerests include optical networking, broadband network management, and performance modeling and analysis.