國立交通大學電資學院

資訊科學與工程研究所

博 士 論 文

IP 網路與分波多工網路之最佳化選徑與資源配置演算法

Optimization-based Approaches for Routing and Resource Provisioning in IP and Optical WDM Networks

研 究 生：陳春秀

指導教授：楊啟瑞 博士

中 華 民 國 九 十 九 年 七 月

IP 網路與分波多工網路之最佳化選徑與資源配置演算法

Optimization based Approaches for Routing and Resource Provisioning in IP and Optical WDM Networks

研 究 生：陳春秀　　　　Student：Alice C. S. Chen

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

國立交通大學 資訊學院
資訊科學與工程研究所
博 士 論 文

A Dissertation
Submitted to Institutes of Computer Science and Engineering
Department of Computer Science
College of Computer Science
National Chiao Tung University
in partial Fulfillment of the Requirements
For the Degree of
Ph.D
in

Computer Science and Information

July 2010

Hsinchu, Taiwan, Republic of China

中華民國 九十九 年 七 月

IP 網路與分波多工網路之最佳化選徑與資源配置演算法

學生：陳春秀　　　　　　　　　　指導教授：楊啟瑞 博士

國立交通大學資訊科學與工程研究所

Abstract in Chinese

網路的發展由單一功能垂直整合(如: 電話網路、行動網路)逐漸走向網路融合與水平分工的下世代網路，未來多樣化的服務都可以在這個融合的 IP 下世代網路(Next Generation Network; NGN)上發展。而融合的網路平台除了必須提供高服務品質(QoS)和高存活度(Survivability)的網路服務之外，下世代網路也需要有極高的網路訊務(traffic)承載能力。而網路服務業者在提供高品質與高存活度的網路平台的同時，也必須兼顧網路的建置成本(CPEX)與運營成本(OPEX)的降低，因此如何將網路資源進行最佳化的配置成為重的課題。

光封包交換技術(Optical Packet Switching; OPS)因為可以直接在光訊號領域進行資料的高速交換，不需要將資料封包轉回電訊號處理，因此避免目前在高速路由器上所遭遇的超高速電路的技術瓶頸。未來 OPS 將能改變下世代網路基本的運作模式，提供數十 Gbps 的超高速的網路訊務傳輸功能。但目前 OPS 光封包交換系統的設計仍受到光交換器以及光儲存器功能尚未成熟的限制，因此光封包交換系統的設計仍是光通訊研究上重要挑戰。相對來說，光路交換(Optical Circuit Switching; OCS)可以提供穩定的網路傳輸服務，是目前 WDM 核心網路中最常被使用的光交換模式。要使光網路達到最佳的使用效率，光路(lighpath)最好能在要使用之前才建立。除了傳統上長時間固定使用的光路的服務之外，提供光路預約的服務模式可以使網路服務業者提升運營效率，也可以讓使用者享有更好的服務。但是如何同時考量光路的預約是否被接受

(admission) 以及光路的路徑規劃 (routing) 與波長使用 (wavelength assignment)是相當具挑戰性的問題。

另一個光路交換模式所衍生的問題是光路的容量與其所需承載的資料量需求有落差的現象，目前在核心網路(core network)與都會網路(metro network) 最常被使用的 SONET/SDH 網路上也存在類似的問題。SONET/SDH 網路的傳輸容量級距(granularity)規範是考量傳送電話網路中的語音話務所規劃訂定的，並不適合目前資料網路所產生的資料傳送的頻寬需求。例如：資料網路上最常用的 100Mbps 乙太網路與 SONET STS-3 的 155Mbps 就存在 55Mbps 的落差。再者，SONET 傳輸容量級距需要以四倍方式成長不能分割，非常不適合資料網路的傳輸需求，造成網路資源使用的沒有效率。此外，為了能夠提供資料的傳送保護，SONET 提供了 APS 保護機制，其中 1+1 protection 模式提供了最佳的保護機制，但同時也更加造成網路頻寬浪費。NG-SONET 為了改善 SONET 的在傳送數據資料傳輸容量級距過大的問題，增加了新的 VCAT 功能，讓點對點的大容量傳輸電路可以由數條容量較小的電路組成，但仍能維持資料的同步。有了這個新功能，讓兼顧網路資源的使用同時也能達成高存活度的網路傳輸服務的問題有了新的解法，這也是網路最佳化選徑與資源配置的一項值得研究的課題。

傳統上要提供高存活度的網路傳輸服務大都透過提供與工作路徑不同(disjoint)的額外的保護路徑達成，需要使用較多的頻寬來達成。新近被提出的網路編碼(Network Coding)技術，改變傳統網路資料直接轉送的模式，選定部分的節點將收到的資料進行編碼再轉送出去，接收端由不同的路徑接收到資料之後，依照原來編碼的方法反向操作，解出所需要的資料。因此若是選定某些路徑做為資料的備用路徑，將所要備用的資料與此路徑上原來傳送的資料編碼後傳送，將可以不增加頻寬的使用，但是又可以達到資料在有鏈路中斷(link failure)時，仍能順利送達的目標。

根據上述的問題分析，以及新的技術進展，我們進行下列四項下世代網路最佳化選徑與資源配置的問題研究，提升網路的運作效能與存活度。包括 WDM 核心網路中光封包交換系統的設計以及光路預約許可(admission)與路徑的規劃；以及 NG-SONET 網路以及 IP 網路，以最佳化的網路路徑規劃與資源配置來達成高網路服務存活度(survivability)所衍生的路徑規劃配置的問題。本論文相關章節內容說明如下：

在第一章，先簡要介紹下世代網路，說明在 WDM 網路、SONET 網路以及 IP Multicast 網路相關的技術進展，並指出在這些網路中有關網路資源規劃與配置最佳化的問題。

在第二章，介紹目前在 WDM 網路中重要的關鍵技術元件的功能與限制，並說明目前在多波長交換網路的光封包交換系統(OPS)所面臨

的一些研究議題。同時提出克服相關問題的新式 OCPS 交換模式，以及相關的實驗網路- OPSINET。接著提出一個新的具有 buffer 能力的 OPS 系統架構設計，運用 WDM 多波長的性質、AWG 的交換能力以及 Cyclic Demux 分單元的特性，設計出 non-blocking 的交換器，同時將其後所介接的 FDL 運用不同的波長擴充成為多個同樣時間長度的 FDL，大幅降低達成特定 packet loss probability 所需要使用的 FDL 的數量，並提出此一設計的效能分析。

在第三章，首先介紹光路預約問題的特性，目前在 WDM 網路中，靜態的光路規劃配置的問題，被稱為 RWA problem，其特性是沒有配置波長轉換功能的節點中，光路所經過的 link 上都需要使用同一個頻率的光波。因為這個同一光路上光波連續的限制(Wavelength Continuality)，RWA problem 已經被證明為是一個 NP-Complete Problem。光路預約需要考量光路許可、路徑規劃以及光波配置，想要達到最佳化的配置，必須同時這三項因素，基本上光路預約問題也是一個 NP-Complete Problem。我們運用網路最佳化方法來解決這一個網路資源配置的問題。

在第四章，先介紹 SONET 網路以及 NG-SONET 網路的新功能；並討論如何用用最少的網路資源來達成使用者對於網路存活度的期望的問題。依據數據資料傳送先天上可以因應網路頻寬變化調適的特性，提出一個在 NG-SONET 網路上的新的網路存活度需求的概念-網路存活度品質(Quality-of-Survivability)，讓使用者可以定義在網路正常運作模式以及面臨 link failure 或是 node failure 情況下所需要使用的傳輸頻寬。配合 NG-SONET VCAT 點對點的大容量傳輸電路可以由數條容量較小的電路組成，但仍能維持資料的同步的特性，同時考量傳輸電路的路徑規劃與路徑的存活度需求，將相關的傳輸電路分散配置，降低所使用的電路因為 link failure 或是 node failure 所造成的影響，同時達成運用最少的網路頻寬來達成使用者對於傳輸頻寬與存活度的需求。

在第五章，先簡要介紹網路編碼(Network Coding)技術，以及其在提升網路頻寬的使用效率與網路存活度上相關的研究成果，並介紹運用網路編碼技術來達成高存活度的網路群播(multicast)的研究基礎。我們分析了運用網路編碼以及樹狀結構(Tree-based)模式來進行網路群播服務所需要使用的網路頻寬，並探討其在疏密度不同的網路上的適用性。

在第六章，回顧本研究相關的研究成果，並提出未來可以再進一步探討的方向。

Optimization based Approaches for Routing and Resource Provisioning in IP and Optical WDM Networks

Student：Chun-Shiow Chen Advisor：Dr. Maria C. Yuang

Institutes of Computer Science and Engineering

National Chiao Tung University

Abstract

Next Generation Network ("NGN") shifts from separate vertically integrated application-specific networks to a single network being capable of carrying all services. In addition to providing a technology independent network platform for emerging services, the NGN needs to support ever-increasing traffic demands by a high efficiency and survivability way. One key issue of the next generation network is how to maintain Quality of Service (QoS) and survivability across a wide range of network services while lowering overall network costs (CapEx and OpEx).

Optical Packet Switching (OPS) allows forwarding of ultrahigh bit rate data packets directly in the optical domain and has been proposed as a solution to overcome the "electronic bottleneck". It will further bring fundamental changes in the design of the Next Generation Network. However, high-speed switching and optical buffering are challenging problems of the OPS system implementation. On the contrary, Optical Circuit Switching (OCS) offers explicit transport guarantees is an important operation paradigm for many network applications. At the current stage most WDM applications follow the OCS paradigm. To get the best network usage, an optical path should be setup just before it is needed.

Providing a lightpath reservation service to users can increase network operators' revenue and provide users with better services. How to jointly determine call admission control as well as Routing and Wavelength Assignment is a significant problem to network operators.

SONET/SDH has been dominating transport in metro and backbone networks for decades due to its superior survivability and short failure recovery time. But legacy SONET/SDH only supports contiguous concatenation transport switching over the overall path and its coarse granularity rates are not a good match to packet traffic. NG-SONET VCAT enables forming a high-order end-to-end large-size path by grouping multiple smaller lower-order paths. Based on the VCAT capability, an intelligent path provisioning algorithm can be used to achieve flexible bandwidth usage in NG-SONET networks. Conventional network protection approaches employ extra network resources and precompute backup paths to bypass the failure link or node. It consumes much bandwidth to provide protection. Network coding allows the intermediate nodes not only to forward packets but also encode/decode incoming packets using algebraic primitive operations [17]. By transmitting combinations of incoming data on a backup path enables each receiver node to recover a copy of the data transmitted on the working path if the working path fails.

According to the advances mentioned above, we do some research on the routing and resource provisioning problems of the next generation network to improve the network efficiency and survivability. We deal with four Routing and Resource Provisioning problems in next generation networks. The first two problems are related to transport functions of core networks in how to design a WDM OPS system and the Advance Lightpath Reservation problem in WDM Networks. The third one is about NG-SONET networks to find an optimal solution for Quality-of-Survivable multi-path routing and provisioning problem. The last one correlates to a survivable multicast IP network. This dissertation is organized as next described.

In Chapter 1, we first give a brief introduction to NGN and make descriptions of some technology progresses in WDM, SONET, and IP multicast networks. We also point out several routing and resource provisioning problems in these networks.

In Chapter 2, we first give a brief introduction to OPS enabling technologies, discuss the design issues of multi-wavelength optical packet switching networks and propose a new switching architecture to route packets and resolve contentions in both the wavelength and space dimensions together.

In Chapter 3, we focus on the routing and resource allocation issues of prescheduled lightpath provisioning problems and give a Lagrangean relaxation based near-optimal algorithm for advance lightpath reservation in WDM networks. The major challenge is that we need to determine request admission, as well as Routing and Wavelength Assignment jointly.

In Chapter 4, we investigate the problems of how to meet the survivability requirements which users expect while lowering network resources consumed and propose a Quality-of-Survivability concept benefit by a phenomenon that data services are tolerant of bandwidth degraded gradually as the available bandwidth reduces. The goal of routing and resource provisioning is to satisfy bandwidth requirements of different states and minimize total bandwidth consumption at the same time.

In Chapter 5 we briefly introduce the emerging network coding fundamentals first. Based on the observations, network coding has been proposed as a new technique to enhance network throughput and survivability in the literature, we study the problem of optimal routing and bandwidth provisioning for survivable multicast communications using network coding.

Finally, concluding remarks and future work are made in Chapter 6.

Acknowledgements

完成此論文，要感謝的人很多。謝謝我的良師益友李詩偉教授，給我最佳的指導與督促，以及無窮盡的耐心支持我完成學業。謝謝 Susan Lynn Clevenger，給我最好的英文指導以及很多的關懷。謝謝我的指導教授楊啟瑞老師，給我最大的包容，讓我有足夠的空間找到自己可以投入的研究方向。謝謝我的家人、朋友與公司的同事，在我專心撰寫論文的期間給我最大的支持與鼓勵。

Contents

List of Figures

Acronyms

ALR	Advance Lightpath Reservation problem
APS	Automatic Protection Switching
AWG	Arrayed Waveguide Grating
ATM	Asynchronous Transfer Mode
BTL	Bundle Tree-Based Link Protection Scheme
CAM	Content Addressable Memory
CapEx	Capital Expenditure
CoS	Class of Service
CSC	Core Switch Controller
DF	Deadline First
DSL	Digital Subscriber Line
FCFS	First-Come-First-Serve
FDL	Fiber Delay Line
FOWC	Fixed Optical Wavelength Converters
FSOB	Full shared Output Buffer
FWM	Four-Wave Mixing
GE	Gigabit-Ethernet
GFP	Generic Framing Procedure
GMPLS	Generalized Multi-Protocol Label switching
GPON	Gigabit Passive Optical Network
ICL	Information and Communications Research Laboratories
IP over WDM	IP-over Wavelength-Division Multiplexing
IPTV	Internet Protocol television
ISIS-TE	Intermediate System to Intermediate System with Traffic Engineering
ITL	Individual Tree-Based Link Protection Scheme
ITRI	Industrial Technology Research Institute
JIT	Just-In-Time
JET	Just-Enough-Time
LCAS	Link Capacity Adjustment Scheme
LGR	Lagrangean Relaxation
LMP	Link Management Protocols
MEMS	Micro-Electro Mechanical Systems
MP-QoS	Multi-Path Provisioning for NG-SONET Networks with Quality-of-Survivability Constraints
NCL	Network Coding for Single Link Failure Protection
NCN	Network Coding for Node Failure Protection

NGN	Next Generation Network
NG-SONET/SDH	Next Generation Synchronous Optical Networking
OADM	Optical Add/Drop Multiplexer
OBS	Optical Burst Switching
OCS	Optical Circuit Switching
OCPS	Optical Coarse Packet Switching
OLSR	Optical Label Switched Router
OLSP	Optic Label Switched Path
OPS	Optical Packet Switching
OPSINET	Optical Coarse Packet Switched IP-over-WDM Network
OSPF-TE	Open Shortest Path First with Traffic Engineering
OPEX	Operating Expenditure
OTN	Optical Transport Network
OXC	Optical Cross-Connect
PLP	Packet Loss Probability
QCP	QoS Control Processor
QoS	Quality of Service
QT	Quiescence Threshold
ROADM	Reconfigurable Optical Add/Drop Multiplexer
RSVP-TE	Resource ReSerVation Protocol with Traffic Engineering
RWA	Routing and Wavelength Assignment
SASK	Superimposed Amplitude Shift Keying
SASK	Superimposed Amplitude Shift Keying
SD	Source-and-Destination
SDH	Synchronous Digital Hierarchy
SHR	Self-healing Ring
SLA	Service Level Agreement
SOA	Semiconductor optical amplifiers
SONET	Synchronous Optical Networking
STS	Synchronous Transport Signal
TDM	Time-Division Multiplexing
TOWC	Tunable Optical Wavelength Converters
UB	Upper Bound
USHR	Unidirectional Self-healing Ring
WC	Wavelength Conversion/Wavelength Converter
WiMAX,	Worldwide Interoperability for Microwave Access
WDM	Wavelength Division Multiplexing
XGM	Cross-Gain Modulation
XPM	Cross-Phase Modulation

Chapter 1. Introduction

Nowadays the IP-based Internet supports various types of services, such as voice, video, interactive games, and emerging cloud computing technology. In order to provide Internet users broadband access and better Quality of Service (QoS), a new network framework called the Next Generation Network (NGN) is proposed [1]. A NGN is an enhanced IP-based network. As shown in [1], it shifts from separate vertically integrated application-specific networks such as the PSTN and the IP network to a single network capable of carrying all services. It is an NGN objective to support services and applications independently of the technologies concerning access networks and core networks. ITU-T Y.2001 provides a general definition of NGN [1], as follows:

> A Next Generation Networks (NGN) is a packet-based network able to provide Telecommunication Services to users and able to make use of multiple broadband, QoS-enabled transport technologies and in which service-related functions are independent of the underlying transport-related technologies. It enables unfettered access for users to networks and to competing service providers and services of their choice. It supports generalized mobility which will allow consistent and ubiquitous provision of services to users. [ITU-T Recommendation Y.2001 (12/2004) - General overview of NGN]

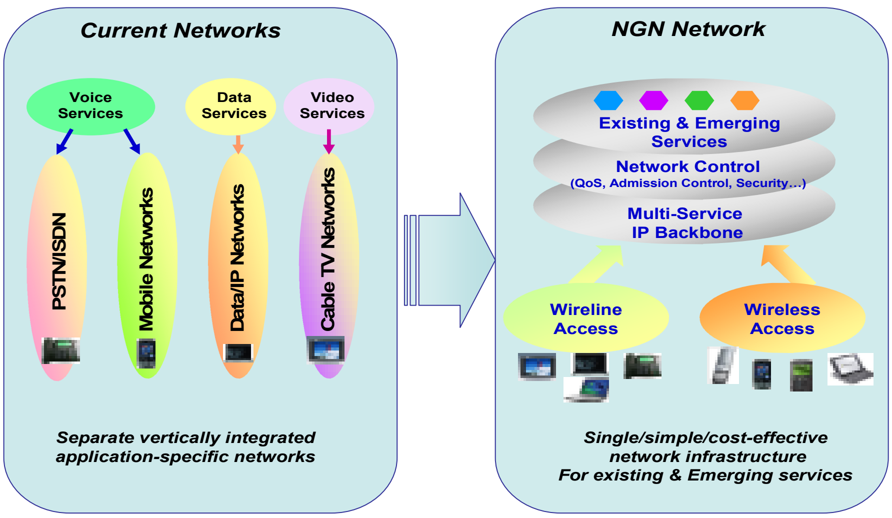


Figure 1 - Next Generation Network

The concept of Next Generation Network is to provide a new service-independent network infrastructure with QoS-enabled features and broadband transport capabilities that support the provision of value-added multimedia services over multiple and heterogeneous QoS-enabled transport technologies. The most significant change is the independence of the data transportation and the service. The NGN functions are divided into service and transport strata according to Recommendation Y.2011 [2], as shown in Figure 2. The service stratum makes requests to transport stratum to get the required network resource and service reliability. NGN transport stratum is required to use the IP protocol for general, ubiquitous and global public connectivity. The IP protocol may be carried over various underlying transport technologies of the transport stratum (e.g., cable access, xDSL, wireless access, Ethernet, optical access, or OTN) according to the operator's environment.

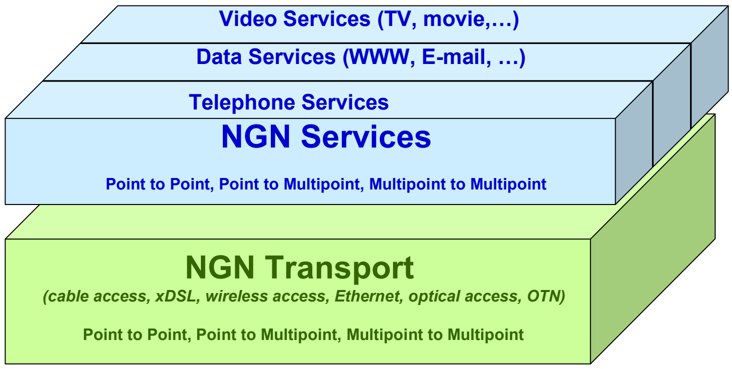


Figure 2 - ITU-T Y.2011 – Separation of services from transport in NGN

An example of a NGN network configuration is shown as Figure 3. End-user equipment may be either mobile or fixed. End-user networks can be networks within homes or enterprise networks. Access network functions collect and aggregate the traffic from end-user networks to the core network. Usually, the access network functions are performed by access networks and access transport networks. The core network function is responsible for ensuring information coming form the access networks transport throughout the core network. It links access transport networks and connects with other core networks.

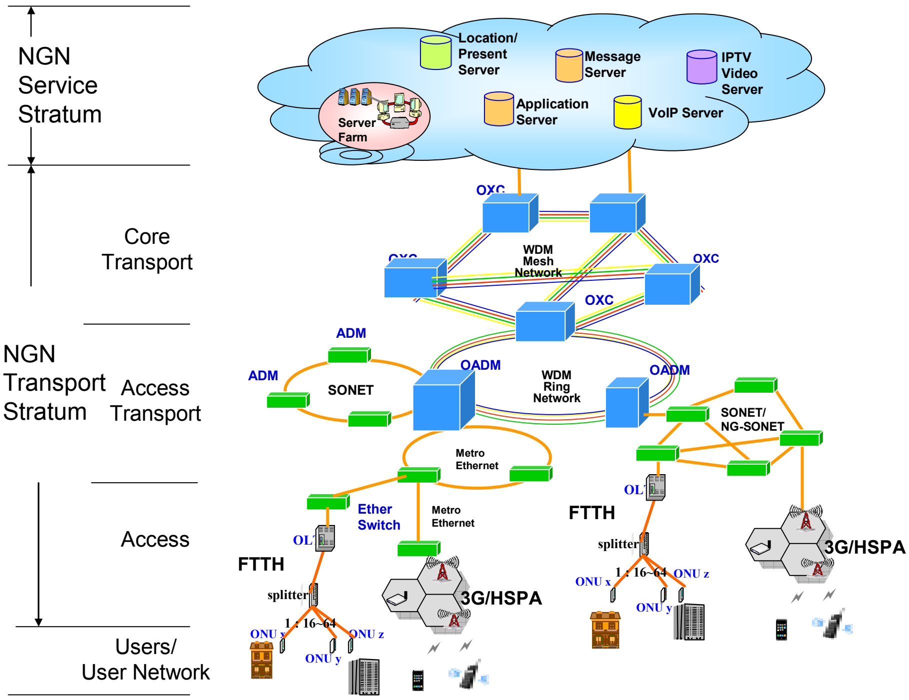
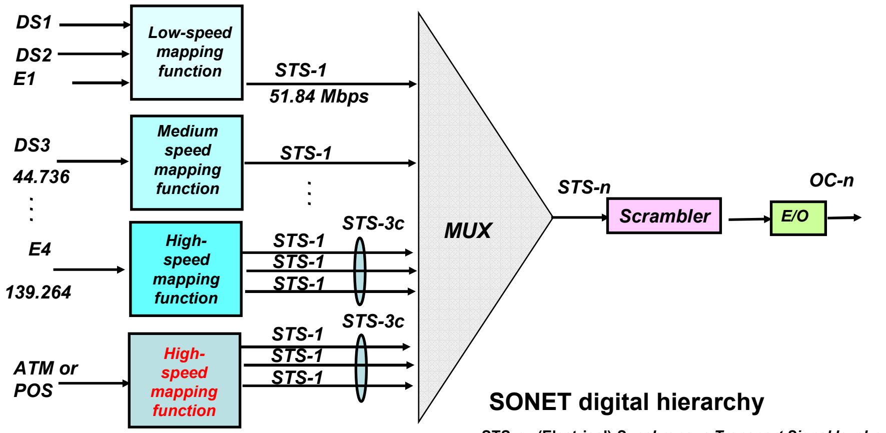


Figure 3 - An example of a NGN network configuration

The access networks connect business and residential subscribers to central offices of their service provider. It spans a distance of a few kilometers perhaps up to 20 kilometers. Diversified technologies, such as xDSL, Cable Modem, Passive Optical Network and WiMAX, are deployed to allow much more flexible use of the access network. However, numerous researchers are working on the emerging access network technologies to provide fully converged services, ubiquitous access and diverse users' devices. 4G Wireless Systems, GPON (Gigabit Passive Optical Network) and the hybrid wireless-optical network are emerging as a promising technology to provide economical and scalable broadband access.

An access transport network usually spans a city to connect those access networks in part or all of a city and covers distances of a few ten to a few hundred kilometers. The major functions of an access transport network include traffic aggregation and routing. SONET/SDH is the most common technology used in transport networks. It is capable of carrying data from different access networks through a synchronous, flexible, optical hierarchy. SONET/SDH is designed to optimize TDM-based traffic. It was initially deployed to carry circuit originated traffic (such as T1 and T3 TDM) over fiber, but it quickly evolved mapping and concatenation capabilities to also carry ATM, Frame Relay, IP and Ethernet traffic. SONET/SDH is a circuit-switched transport and supports contiguous concatenation transport switching over the whole path. The basic units of transmission in SONET are STS-1 (51.84 Mbps), STS-3 (155.52 Mbps), STS-12 (622.08 Mbps), and STS-48 (2.488 Gbps). As shown in Figure 4, multiple lower order signals can be adapted into a higher order signal.



SONET digital hierarchy

STS-n : (Electrical) *Synchronous Transport Signal level n*
OC-n : Optical Channel level n
Frame time is constant, size increases in proportion to n

North America T- System

Electrical Signal	Bit Rate((Mbps)
DS1	1.544
DS2	6.312
DS3	44.736
DS4	274.176

Electrical signal	Optical signal	Bit rate (Mbps)
STS-1	OC-1	51.84
STS-3	OC-3	155.52
STS-9	OC-9	466.56
STS-12	OC-12	622.08
STS-18	OC-18	933.12
STS-24	OC-24	1244.16
STS-36	OC-36	1866.24
STS-48	OC-48	2488.32
STS-192	OC-192	9953.28

Figure 4 - SONET multiplexing

Compared to the requirements for transport networks, SONET/SDH falls short in inefficient payload mapping and lack of framing protocol [3]. The inefficient payload mapping is attributed to concatenation, which has strict payload size restrictions and requires contiguous payload elements. One of the main problems perceived in the SONET/SDH system is its inefficient transport of current Ethernet which runs at 100 Mbps and 1 Gbps. The virtual concatenation (VACT) is a new feature of NG-SONET/SDH. VCAT enables forming a high-order end-to-end large-size path by grouping multiple smaller lower-order paths [4].

With VCAT, flexible bandwidth usage can be achieved using intelligent path provisioning in NG-SONET/SDH networks.

Ethernet is successful in local area networks. Efforts to extend its boundaries beyond LAN to the carriers' backbone networks are in progress. Metro Ethernet [5] is another new solution for access transport networks. It is based on the Ethernet standard and concerns such issues as CoS, SLAs and management. Metro Ethernet products are widely used in service provider networks, such as mobile and broadband backhaul.

The core network is the backbone of modern IP networks. It spans a distance of a few hundred to a few thousand kilometers in length. The core network provides two major functions. The first one is longhaul data transportation and the other is the exchange of information between different worldwide sub-networks. The technologies currently used in the core and backbone network facilities are WDM, OADM, OXC, and submarine cable systems.

Wavelength Division Multiplexing (WDM) [6] is basically a modern fiber optical transmission technique which multiplexes various optical carrier signals on a single optical fiber by using different wavelengths to carry different signals. The capacity of a given link can be multiplied by simply upgrading the WDM multiplexers and demultiplexers at each end. Since the WDM technique is capable of providing data capacity in excess of hundreds gigabits per second, modern transport networks increasingly employ this technology to utilize the vast transmission bandwidth of fiber to accommodate unprecedented, accelerating demand for bandwidth. With the availability of optical fiber amplifier technologies and the WDM multiplexing technique, optical networking is an immediate success owing to its obvious merits; gracious capacity is increased by adding a wavelength at a time without having to install additional fibers.

Optical add/drop multiplexer (OADM) and optical cross-connect (OXC) are two important network elements in WDM optical transport networks. Through configuring these two network elements, network operators can setup lightpaths. The network managerial and reconfiguration capabilities of OADMs and OXCs evolve from fixed to configurable continuously. There are three generations of the optical networking technique evolvement. In the initial phase, OADM or OXC are not configurable, that is, they are fixed. In fixed OADMs, the add/drop and through channels are predetermined and can only be manually rearranged after installation. The second generation of optical networking investigates the reconfigurable aspect of all-optical multi-wavelength networking and the viability of transparent networking due to no electronics element is involved in the data plane. The reconfiguration is applied to each wavelength. An end-to-end optical circuit between a node pair called lightpath can be setup through configuring the optical network elements properly. The configuration can be set through network management or based on a short optical label which includes information related to source, destination, and others. We call the whole wavelength switching paradigm as the Optical Circuit Switching (OCS) paradigm. Most IP over WDM network applications follow the OCS paradigm now. In the IP over WDM network, an optical path is a large pipe to transport data from one end to the other. This makes relatively static utilization of individual WDM channels. The packet routing proceeds in the electronic IP routers which are connected to OXCs/OADMs. These architectures rely on Optical-to-Electrical-to-Optical (O/E/O) conversions since the data transportation is the optical domain, but all packets processing and routing are done in the electrical domain.

The optical networking technology has come to the third generation in recent years. Reconfigurable OADMs(ROADMs), Reconfigurable OXCs and GMPLS (generalized multi-protocol label switching) have been proposed in order to automate lightpath setup procedures. GMPLS [7] is an extension of MPLS and used as the control mechanism for configuring not only packet-based paths, but also optical-based paths. It consists of several protocols, including routing protocols (OSPF-TE or ISIS-TE), link management protocols (LMP), and a reservation/label distribution protocol (RSVP-TE). GMPLS serves as a control mechanism for ROADMs and OXCs allowing the creation or termination of label switched lightpaths in the optical network to adapt to changing loads.

Some emerging technologies are developed for enhancing QoS-enabled features and broadband transport capabilities of the NGN network. The most significant technology improvement happening in optical communication is Wavelength Division Multiplexing (WDM). WDM, a modern fiber optical transmission technique, can scale the capacity of a single optical fiber deeply into the terabit per second range. As mentioned before, NGN transport stratum is required to use the IP protocol. However, the scalability of electronic IP routers and their ability to match the rising transmission capabilities of WDM in the optical layer is difficult. This situation led to research interest in optical packet switching (OPS) [8],[9]. In OPS, packets are directly switched in the optical domain in order to bypass the electronic switching bottleneck. OPS paradigm can advocate efficient sharing of wavelength channels among multiple connections satisfying a multitude of applications with diverse Quality of Service (QoS) requirements flexibly and cost-effectively. Current applications of WDM mostly follow the Optical Circuit Switching (OCS) paradigm by making relatively static utilization of individual WDM channels.

Within the NGN architecture, the resource and admission control functions within access and core networks determine the demand admission control, bandwidth reservation and allocation as well as priority handling upon the request from the service stratum [1]. The transport service provision is based on transport subscription information, SLAs, network policy rules, service priority, and transport resource status and utilization information. Most IP over WDM network applications follow the whole wavelength switching paradigm now. The emerging network application like cloud computing relies deeply on the ready-availability of broadband and grid computing. Cloud computing applications such as enterprise cloud may need to setup high-speed lightpaths in order to synchronize the distributed database located in diverse campuses periodically. A major feature of such applications is that traffic demands are requested to the network in advance before the connections are set up [35]-[38] and last in a pre-scheduled time period. It will be a new type of transport service request coming from the service stratum in NGN networks. One major challenge arising in these Advance Lightpath Reservation problems has been to jointly determine call admission control as well as Routing and Wavelength Assignment (RWA) [11].

One key issue of the next generation network related to resource and admission control functions is how to maintain quality of service (QoS) and survivability across a wide range of network services while lowering overall network costs (CapEx and OpEx). The survivability refers to a network's capability to provide continuous service in the presence of failures. How to prevent service interruption, and keep service loss to a minimum if a network failure is inevitable, becomes a critical issue. New technologies like NG-SONET and network coding provide new capabilities to improve service survivability. Virtual concatenation, a new function of NG-SONET, enables forming a high-order, end-to-end, large-size path by grouping multiple smaller lower-order paths. Those lower-order paths may individually take different routes to reduce the damage caused by a link failure and improve service survivability. Network coding allows the intermediate nodes not only to forward packets but also encode/decode incoming packets using algebraic primitive operations [17]. By transmitting combinations of incoming data on a backup path enables each receiver node to recover a copy of the data transmitted on the working path if the working path fails. According to the advances mentioned above, some research is needed on the routing and resource provisioning problems of the next generation network to improve the network efficiency and survivability.

In this dissertation, we deal with four Routing and Resource Provisioning problems in next generation networks. The first two problems are related to transport functions of core networks in how to design a WDM OPS system and the Advance Lightpath Reservation problem in WDM Networks. The third one is about NG-SONET networks to find an optimal solution for Quality-of-Survivability multi-path routing and provisioning problem. The last one correlates to a survivable multicast IP network. The remainder of this dissertation is organized as next described. In Chapter 2, we first give a brief introduction to OPS enabling technologies, discuss the design issues of multi-wavelength optical packet switching networks and propose a new switching architecture to route packets and resolve contentions in both the wavelength and space dimensions together. In Chapter 3, we focus on the routing and resource allocation issues of prescheduled lightpath provisioning problems and give a Lagrangean relaxation based near-optimal algorithm for advance lightpath reservation in WDM networks. The major challenge is that we need to determine request admission, as well as Routing and Wavelength Assignment jointly. In Chapter 4, we investigate the problems of how to meet the survivability requirements which users expect while lowering network resources consumed and propose a Quality-of-Survivability concept benefit by a phenomenon that data services are tolerant of bandwidth degraded gradually as the available bandwidth reduces. The goal of routing and resource provisioning is to satisfy bandwidth requirements of different states and minimize total bandwidth consumption at the same time. In Chapter 5 we briefly introduce the emerging network coding fundamentals first. Based on the observations, network coding has been proposed as a new technique to enhance network throughput and survivability in the literature [13]-[16],[43]-[46]. We study the problem of optimal routing and bandwidth provisioning for survivable multicast communications using network coding. Finally, concluding remarks and future work are made in Chapter 6.

Chapter 2. Multi-wavelength Optical Packet Switching Networks

The ever-growing demand for Internet bandwidth and recent advances in optical communication technologies brings about fundamental changes in the design and implementation of the next generation core networks. In conventional IP over WDM network, optical signals are converted into electronic ones for packet switching inside an electronic switch. The packets are transformed to optical format again for being carried in optical fiber. Such O/E/O conversion incurs high cost and technical difficulty. Furthermore, as data transmission rates are ever-increasing, it is more and more difficult for electronics switching to meet such high-speed requirements. Besides, bandwidth requirements driven by the deployment of new IP services and the increasing penetration of existing services are constantly changing. OCS paradigm is not optimally bandwidth-efficient for transporting traffic from these IP-based services. The OPS provides a packet-based optical switching solution that is, packets are directly switched in the optical domain through an OPS node from any input port to any output port. It is capable of achieving high statistical multiplexing gains, better packet loss performance, and Quality of Service (QoS) differentiation. It has been envisioned as the ultimate solution for the data-centric optical Internet.

A generic functional block of an OPS node, shown as Figure 5 - Optical packet switching system, consists of a multiplexer/demultiplexer pair, an input interface, a switching fabric, a buffer, an output interface, and a control unit. The demultiplexer separates the incoming multi-wavelength optical signal into several single wavelength optical signals. These optical signals are forwarded to the input interface where headers and payloads are decoupled. Then the headers are sent to the control unit which performs electric header processing in order to obtain routing information. The information is used to determine the routing of the switching fabric so as to deliver the signal to the right destination. Payloads of these packets are maintained in optical format inside the switching system. They are exchanged by the switching fabric and put into optical buffers if contentions occur. Finally, the new headers will be generated and combined with the original optical packets. The designs of switching fabrics and optical buffers are important problems in optical packet switching system.

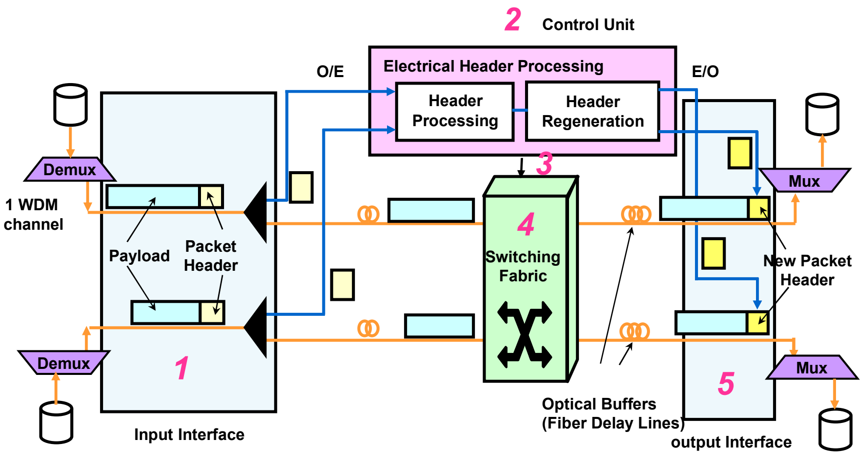
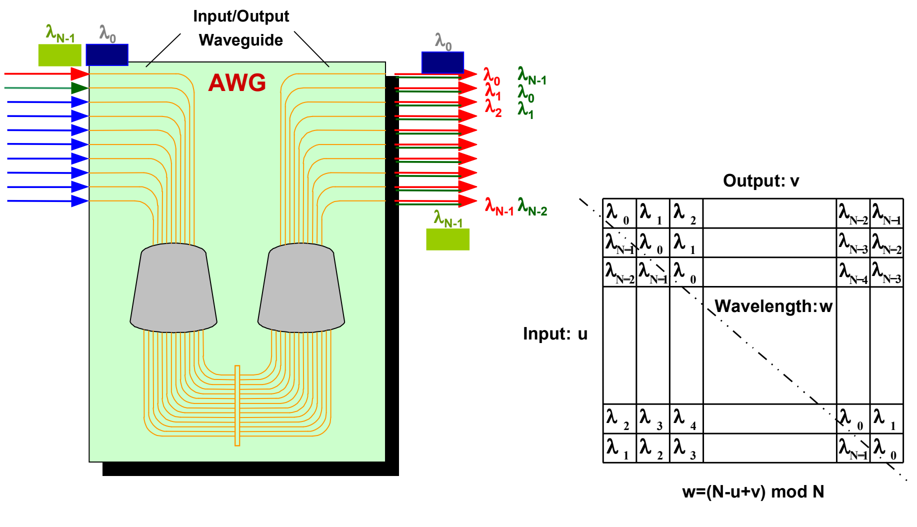


Figure 5 - Optical packet switching system

Optical switches which perform switching functionalities to route the incoming packets to the correct output ports in a very short time period are crucial to the design of an OPS system. There are several versatile technologies used to fabricate optical switches, such as micro-electro mechanical systems (MEMS) switches, thermal optical switches, electro-optical switches and others [51]. The characteristics of optical MEMS are low crosstalk, wavelength insensitivity, polarization insensitivity, and scalability. Its switching speeds range from millisecond to sub-millisecond. The advantages of thermal optical switches are polarization-insensitive operations and switching speeds on the order of milliseconds. Electro-optical switches like LiNbO3 switches and semiconductor optical amplifiers (SOA)-based switches offer relatively faster switching speeds. They can switch a packet within a few nanoseconds. Each optical switching technology has unique performance characteristics. To meet the switching requirements of an OPS system, switching speeds of optical switch fabrics for packet switching should be in nanosecond order and optical switches need to be strictly non-blocking. Arrayed waveguide grating (AWG) can switch fast, is scalable to large size and consumes little power there for it is promising for constructing high-speed large-capacity switching fabric. Using limited range wavelength converters and arrayed waveguide grating routers to construct a strictly non-blocking optical switching fabric has been proposed in the literature [56].

An AWG provides a fixed routing of an optical signal from a given input port to a given output port based on the wavelength of the signal [50]. Generally, it consists of two star couplers joined together with arms of waveguides of unequal lengths as shown in Figure 6(a). A useful characteristic of the AWG is its cyclical wavelength routing property illustrated by the table in Figure 6 - Arrayed waveguide grating (AWG)(b). Signals of different wavelengths coming into an input port will each be routed to a different output port. Different signals using the same wavelength can be input simultaneously to different input ports, and still not interfere with each other at the output ports. If the multi-wavelength input is shifted to the next input port, the demultiplexed output wavelengths also shift to the next output ports accordingly. An AWG with N input and N output ports is capable of routing a maximum of N2 connections. If an ‘‘out-of-range’’ wavelength is sent to the input port, that wavelength is simply lost or ‘‘blocked’’ from reaching any output port. Because the AWG is an integrated device, it can easily be fabricated at low cost. The disadvantage of the AWG is that it is a device with a fixed routing matrix which can not be reconfigured.



(a) Arrayed waveguide grating (b) Wavelength routing table

Figure 6 - Arrayed waveguide grating (AWG)

Wavelength conversion plays a major role in providing the wavelength flexibility in WDM networks. By the introduction of wavelength converters, the data modulated on an incoming wavelength can be transfer to a different outgoing wavelength. Thus, wavelength converters combined with AWG can construct a switching fabric. The design proposed in [49] used a single stage of AWGs is blocking. In [56], novel constructions of strictly non-blocking and rearrangeably non-blocking switching fabrics are given. Various approaches for realizing wavelength conversion have been proposed including cross-gain modulation (XGM), cross-phase modulation (XPM) and four-wave mixing (FWM). It has been shown that the XGM WC's bit rate can come close to 100 Gb/s. The major drawback of the XGM WC is the degradation of the extinction ratio when converting from shorter to longer wavelengths. However, the XGM WC is very popular due to its simplicity, polarization independence and insensitivity to input wavelength. The XPM scheme generally exhibits better conversion efficiency than the XGM scheme. It has high conversion efficiency, polarization immunity, and no increase in phase noise, but also linear signal up-conversion with a low optical power requirement. Four-wave mixing (FWM) in SOA is an attractive mechanism for wavelength conversion since it preserves amplitude, frequency, and phase information; it is generally format independent and also largely bit-rate-independent, thus offering the best transparency. It is superior owing to its ultrafast response. It is also the only approach that allows simultaneous conversion of multiple wavelengths [50].

The all-optical buffer being used for contention resolution is an enabling technology for all-optical packet switched networks. In the optical buffer, data would be kept in optical format (i.e., in the form of light) throughout the storage time without being converted into the electrical domain. All-optical buffers are currently achieved by either fiber delay lines (FDLs), or slow-light technologies. However, the slow-light technologies [54] have been shown to have limited capacity and a delay-bandwidth product; in addition, it is too sensitive for wavelength accuracy that makes it not a feasible solution to support optical buffers. The optical fiber delay line (FDL) is currently the practical way to implement optical buffering. Nevertheless the properties of fiber delay lines differ significantly from properties of electronic buffers. The FDLs can not delay packets for an arbitrary period of time but only for multiples of a basic unit, called the granularity of the FDL. That is, only a discrete set of delays can be provided for contention resolution. Feed-forward and feedback are two kinds of FDL structures in optical buffering [52]. In the feed-forward structure, the packets heading for the same output port at the same time are fed into fiber delay lines of different lengths to resolve contention. A packet coming out of the FDL should be routed to an output port immediately. In the feedback method, a packet may re-circulate in the switch several times until an output port becomes available again. However the feedback architecture leads to larger switch fabric and more crosstalk from which the signal could suffer from significant power loss and noise.

2.1 Optical Coarse Packet Switching

The OPS faces some technological limitations, such as the lack of optical signal processing and optical buffer technologies, as well as large switching overhead. In light of this, while some works [9],[10] directly confront the OPS limitations, others attempt to tackle the problem by exploiting different switching paradigms, in which Optical Burst Switching (OBS) [17]-[24] has received the most attention. OBS [17] was originally designed to efficiently support all optical bufferless [18],[19] networks while circumventing OPS limitations. By adopting per-burst switching, OBS requires IP packets to be first assembled into bursts at ingress nodes. Essentially, major focuses in OBS have been on one-way out-of-band wavelength allocations (e.g., Just-In-Time (JIT) [20], and Just-Enough-Time (JET) [18], and the support of QoS for networks without buffers [18],[19] or with limited Fiber-Delay-Line (FDL)-based buffers [21]. In particular the JET-based OBS scheme is considered most effective, wherein a control packet for each burst payload is first transmitted out-of-band, allowing each switch to perform a just-in-time configuration before the burst arrives. Accordingly, a wavelength is reserved only for the duration of the burst. Without waiting for a positive acknowledgment from the destination node, the burst payload follows its control packet immediately after a predetermined offset time, which is path (hop-count) dependent and theoretically designated as the sum of intra-nodal processing delays.

However, its just-in-time-based design results in several complications [25]. These OBS design complications have been the primary motivators behind the design of the OCPS paradigm. To circumvent OPS limitations, a new Optical Coarse Packet Switching (OCPS) paradigm is proposed. Similar to OBS, OCPS supports per-burst switching, which is labeled-based, QoS-oriented, and either bufferless or with limited FDL-based buffers. Being different from OBS which uses out-of-band control, OCPS adopts in-band control in which the header and payload are modulated and transported via the same wavelength.

An OCPS switching network comprises ingress/egress routers, Optical Label Switched Routers (OLSR), and GMPLS controllers. IP packets in an OCPS network belonging to the same loss class and the same destination are assembled into bursts. The header of a burst payload carries forwarding (i.e., label) and QoS (e.g., priority) information. A label is the network control information that is swapped at each switching node. The header and the payload of a burst are time-aligned. They are modulated based on a Superimposed Amplitude Shift Keying (SASK) technique [27]. A burst is assembled at an ingress router then forwarded along a pre-established Optical Label Switched Path (OLSP). At each switching node, the header and payload are first SASK-based demodulated. While the header is extracted and electronically processed, the payload remains transported optically in a fixed-length FDL achieving constant delay. Provided with no buffer and that there is more than one burst payload at the switch destined for the same wavelength output, contention occurs and resolution is required. Each burst payload is then SASK-based re-modulated with the new header, and switched according to the label information in the header. Finally at egress nodes, the reverse burstification process is performed and IP packets are extracted from bursts.

The ingress router simply performs burstification. It consists of five major components: Scheduler/Shaper, Gigabit-Ethernet (GE) Interface, Header/Payload Generator, Optical Transmitter, in addition to the GMPLS controller interface, as shown in Figure 7. All label and wavelength information have been downloaded in advance from the GMPLS controller to the ingress router through the GMPLS controller interface and saved.

The Scheduler/Shaper performs QoS-enabled packet aggregation. A burst is generated and transmitted either when the burst size reaches its maximum or the maximum burst assembly time expires, respectively. After having determined a burst to be generated, the Header/Payload Generation module aggregates packets and in turn performs framing for packet delineation in addition to generating header information. Through the GE interface, the header and payload ultimately pass in parallel to the Header/Payload Generator. The payload is then encoded via the 8B/10B Encoder. At the optical transmitter, the header and payload are

SASK-based modulated and transmitted via a preconfigured wavelength.

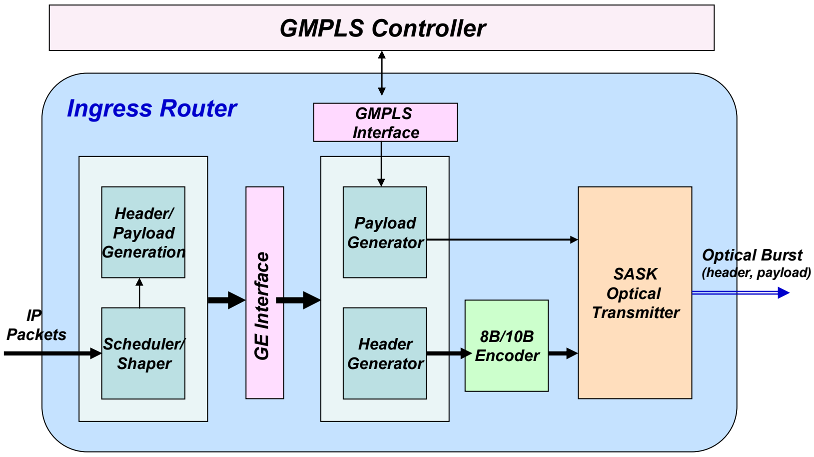


Figure 7 - Ingress router architecture

The Optical Label Switched Router (OLSR) (see Figure 8) performs packet/burst switching functions mentioned above. It consists of three major components for each input port (fiber), and one cyclic-frequency AWG switch for the entire node. The three components are: Header Extractor/Eraser, Burst Mode Receiver for Header (BMRH), and Core Switch Controller (CSC). First, the Header Extractor/Eraser extracts the header, and erases it for the payload. While the payload continues traveling optically along the internal FDL, the header is received and recovered in amplitude by BMRH. With the recovered header, CSC performs label swapping, QoS control, and laser tuning control. Notice that owing to the use of an AWG switch, once an OLSP is established, the path is determined locally via the binding from an old label to a new (label, wavelength) pair. All label and wavelength information has been downloaded in advance from a GMPLS Controller and saved in the Content Addressable Memory (CAM).

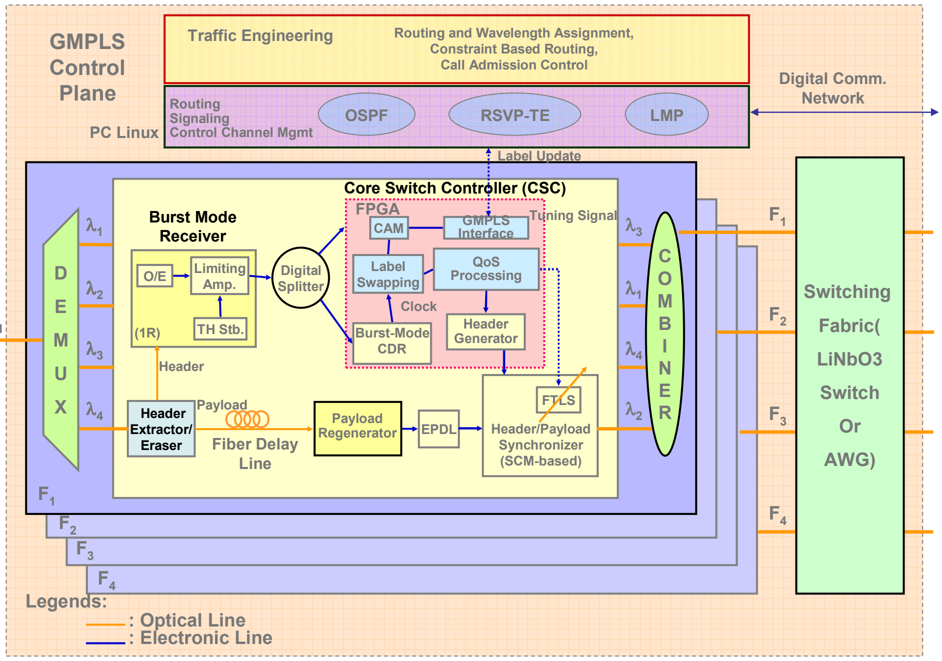


Figure 8 - Optical label switched router architecture

The QoS Control Processor (QCP) is responsible for prioritized contention resolution and header integrity assurance. It is worth noting that, due to AWG, any two bursts arriving from different input ports never contend. On the contrary, contention will occur for bursts arriving from the same input port but carried by different wavelengths, and destined for the same output port. Basically, to switch a burst to the destined output port, an idle wavelength is selected. If all wavelengths are busy, higher priority bursts receive absolute precedence over lower-priority bursts. That is, owing to bufferless, one of the lower-priority bursts being served is preempted and discarded. Finally, with the new (label, wavelength) pair read from CAM, CSC generates the new header and sends a tuning signal to the gated tunable laser. The new header is re-synchronized with the payload having traveled within the FDL.

2.2 Optical Coarse Packet Switched IP-over-WDM Network (OPSINET)

Based on OCPS, we construct an experimental optical IP-over-WDM network. It is a collaborative project between National Chiao Tung University and the Information and

Communications Research Laboratories (ICL)/Industrial Technology Research Institute (ITRI). The experimental optical IP-over-WDM network is referred to as OPSINET. The main objective is to examine and resolve fundamental OPS transport and QoS challenges from both the system and network-layer perspectives.

OPSINET consists of three types of nodes - edge routers, optical lambda/fiber switches (OXCs), and Optical Label Switched Routers (OLSRs), with multi-granularity switching capabilities, as shown in Figure 9. While lambda(λ) and fiber OXCs are layer-1 optical devices that switch on a single lambda and an entire fiber, respectively, the OLSRs are layer-3 optical nodes that route and switch packets on a label basis. The label-based routing and switching in OPSINET is managed by the control plane implemented by an out-of-band GMPLS network. The GMPLS network [26] connects a number of GMPLS controllers, each of which governs the routing/switching of an OPSINET node.

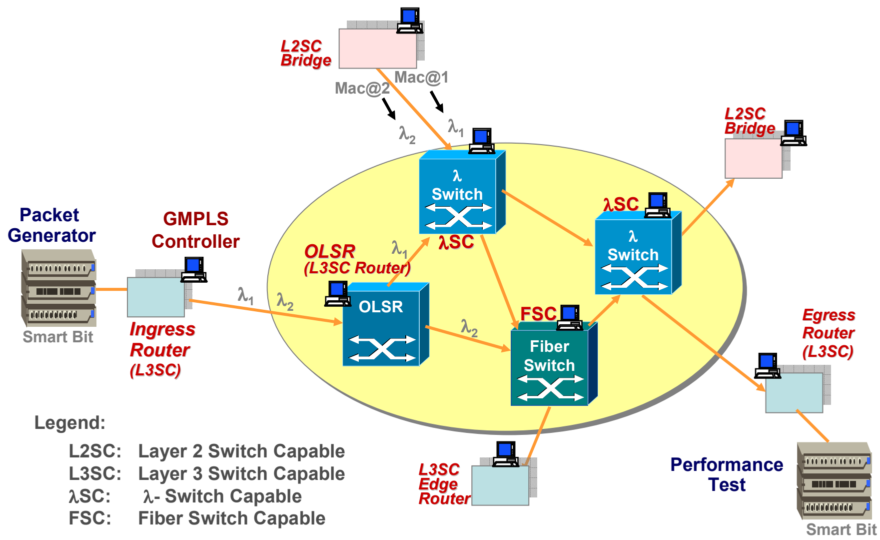


Figure 9 - OPSINET testbed configuration

A snapshot of OPSINET is displayed in Figure 10. In the basic transport, OPSINET performs efficient per-burst switching by means of the time-aligned design and SASK-based modulation of the header and burst payload. Through this experiment, we perceive that the data-centric optical Internet can become a reality based on the OPS technology.

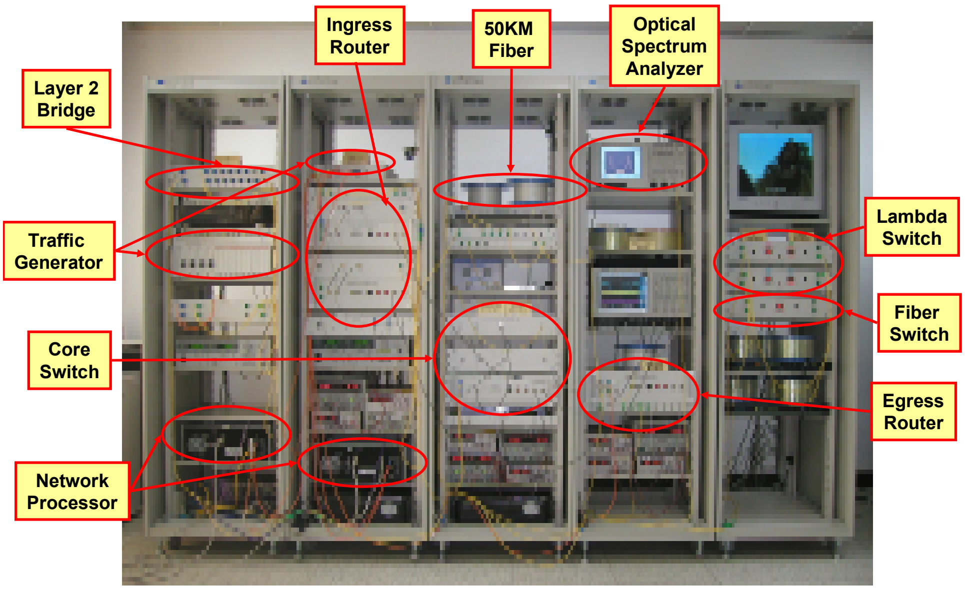


Figure 10 - OPSINET: a snapshot

2.3 Fully Shared Output Buffer Switch using Cyclic DeMUX

We have implemented the OLSR system, shown in Sec. 2.2, that could perform header/payload mux/demux and optical label swapping. We used LiNbO3 to be the switching fabric. However, $LiNbO_3$ is highly polarization dependent. In addition, due to the small port count number of $LiNbO_3$, the switch size is quite limited. That makes the architecture not scalable. In this section, we further present an AWG based switching architecture with shared output buffers aims to resolve the scalability and packet contention problems of OPS.

The basic requirements of an OPS system are capable of minimizing packet loss probability and achieving QoS differentiation. Although using a large-size non-blocking optical switch or equipped many optical buffers can reduce packet loss probability, it results in poor system scalability. Furthermore the functionality of the optical switches and optical buffers are still under development, therefore the design of the optical-buffered switch architecture and the corresponding scheduling and routing algorithms are still in its early stage.

In general, the switching subsystem can be categorized as being non-blocking or blocking. For the blocking switches, the Banyan switch is the most scalable and economic architecture but suffers for internal blocking. The non-blocking switching subsystem can always connect input and output ports without affecting other existing connections. However, the non-blocking optical switches are less scalable and economically infeasible due to using a large number of switching elements. There is another type of non-blocking switch constructed by limited range wavelength converters and arrayed waveguide grating (AWG) [53], which converts each packet to an appropriate wavelength thus establishing a path to the required output port according the routing properties of the AWG. The AWG is fast switching, scalable and low power consumption, but the control algorithm to properly decide the wavelength of each packet is a challenge.

According to the position of the buffer, buffered-packet switches are essentially classified as input buffering, output buffering, shared buffering, and recirculation buffering [29]. While input (output) buffering has a separate buffer for each input (output) port, shared buffering allows buffers to be shared among multiple inputs and/or outputs. Recirculation buffering can support dynamic buffering durations at the expense of additional hardware to maintain signal quality. Output buffering has been shown to be effective for packet switching. It is profound by its performance on low packet loss without suffering the head-of-line problem arising in an input buffered switch.

There have been several optical-buffered switch architectures proposed in the literature. Chiaroni et al. proposed a broadcast-and-select optical packet switching architecture [30] that can easily perform many-to-many switching the employment of optical splitters and couplers results in significant power loss. Danielsen et al. [31] proposed three output-buffered optical packet switching architectures which can resolve contentions in the wavelength dimension, but the scalability of space-switches makes the architectures hard to implement. Hunter et al.

[32] presented the architecture by cascading many small switches with FDLs in between them to develop large optical buffers (SLOB). The architecture can accommodate more buffers but greatly increase the power loss. In WASPNET [33], tunable optical wavelength converters (TOWCs), an arrayed waveguide grating (AWG), a space switch, and shared feedback buffers are used in the switch. The utilization of the shared feedback buffers is poor, since only one packet can appear in the output of a shared feedback buffer at a time.

The FDL is currently the practical way to implement optical buffering but its coarse granularity and large volume introduce another challenge to OPS system design. In a WDM network, wavelength is additional dimension that can be applied to reduce packet loss probability. Wavelength converters integrated with FDLs can construct a multi-wavelength FDL buffer subsystem to resolve contentions in both space and wavelength dimensions. In this dissertation, we propose a novel fully shared output buffer (FSOB) switch using cyclic DeMUX to reduce the size of optical buffers and minimize packet loss probability. The switching function is accomplished by integrating the arrayed waveguide grating (AWG) and the tunable optical wavelength converter (TOWC). Through it, multiple packets carried by different internal wavelengths are scheduled to switch to the same output port but receive different delays afterward. By incorporating a cyclic AWG DeMUX at the output of optical buffers, this system permits fully output buffer sharing in the switch. Finally, through the fixed optical wavelength converters (FOWC) and a multiplexer, the packets are reconverted to destined wavelengths and sent.

2.4 System Architecture of the FSOB Switch

The FSOB switch is an output buffered multi-wavelength optical packet switching system with single-stage FDL-based optical buffers. The FSOB architecture combines two AWGs with wavelength converters to make the switch a non-blocking one. That is, there is no packet collision inside the switch. It also provides shared output buffers by means of multi-wavelength capability, with which two or more packets can appear in the output buffer at the same time. As shown in Figure 11, it consists of two parts: the central switch controller (CSC) and the output buffered multi-wavelength optical packet switch. The CSC runs an algorithm to decide the routing paths of incoming packets inside the switch in order to forward them to their destined output ports without blocking. It controls the wavelength converters in the output buffered multi-wavelength optical packet switch to change the wavelength on which a packet is carried in order to route the packet to its destination port. The output buffered multi-wavelength optical packet switch consists of four stages: an input DeMUX interface, a packet classifying stage, a switching and buffering stage, and finally an output MUX interface, as illustrated in Figure 11.

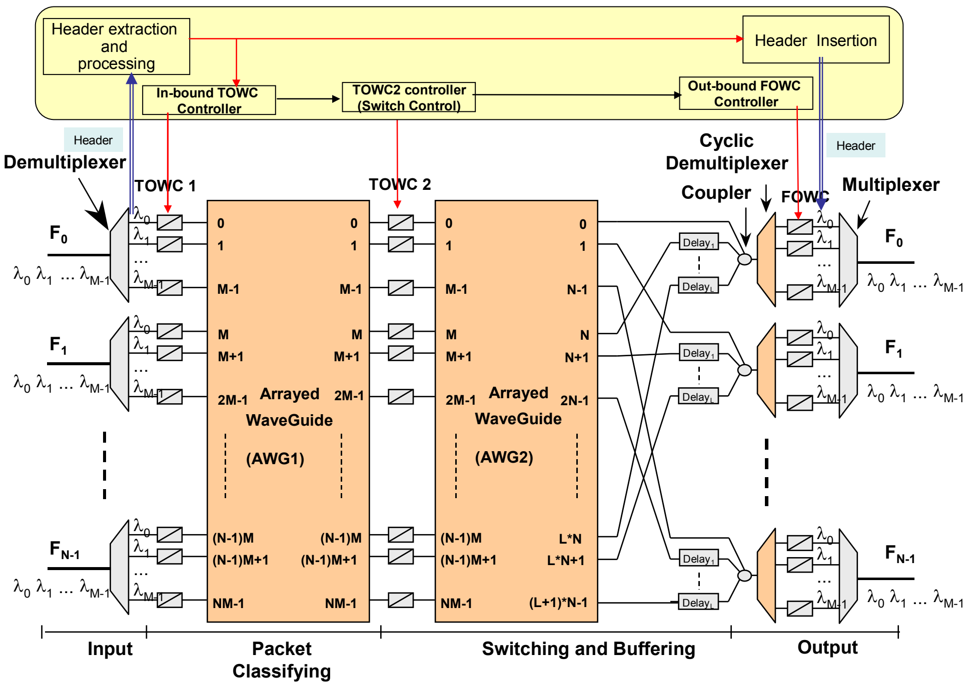


Figure 11 - FSOB system architecture.

In the input section there are N input fibers, each carrying M wavelengths, where M, N are positive integers. After demultiplexing, for each packet a TOWC converts its input wavelength to an internal wavelength according to Eq. (2-1). By taking the cyclic property of AWG on the wavelengths shown in Figure 6 - Arrayed waveguide grating (AWG)(b), those incoming packets with the same destination ports are converted to appropriate wavelengths in order to be routed to consecutive outputs of AWG1. The function of this packet classifying stage is to sort input packets according to their destination ports. This stage consists of M-by-N tunable wavelength converters, named as TOWC1, and an NM-by-NM cyclic AWG wavelength router, named as AWG1. NxM internal wavelengths are required in this stage. The internal wavelengths are only used inside the output buffered multi-wavelength optical packet switch and can be different from the wavelengths used to carry packets in the output fiber.

$$w_{ij} = (NM - (i * M + j) + t) \bmod NM \tag{2-1}$$

Where i is the input port index, j is the wavelength which the incoming packets is on, and t is the order of the packet after all packets are sorted, Where $1 \leqq t \leqq NM$

The switching and buffering stage is responsible for routing optical packets to the appropriate output buffers without contention. It comprises a wavelength converter array, a NM-by-NM cyclic AWG router, named as AWG2, and N groups of FDL buffers. For each output fiber a FDL buffer group is provided. Each FDL buffer group includes L types of delay lines to provide a delay unit from 0 to L units of packet duration. The interconnection between the AWG2 and the j-th buffer of i-th group is determined by Eq. (2-2).

$$\alpha(p, l) = p + lN \tag{2-2}$$

Where N is total number of output fibers, p is the output fiber ($1 \leqq p \leqq N$), and l is the delay unit ($0 \leqq l \leqq L$).

Every packet is converted to a new wavelength according to the required delay time and the output port for the packets. To fully utilize buffer utilization and reduce packet loss rate, each buffer can accommodate packets with different wavelengths. The buffering process is controlled to operate as a multiple wavelength First-In-First-Out (FIFO) buffer. In a heavy loaded situation, some packets are subject to being dropped due to all FIFO buffers being fully occupied. The dropping operation is performed by converting the dropped packet to a dummy wavelength that is out of the passband of AWG2.

The detailed switching process is next described. For a packet at the *t-th* TOWC2 bound for port p and delayed l units, the new wavelength of the packet is determined by Eq. (2-3). The *t-th* TOWC2 converts the incoming packet to a new wavelength w_{tp}

$$w_{tp} = (NM - t + lN + p) \bmod NM \tag{2-3}$$

Where t is the order of an incoming packet after all packets are sorted, l is the delay time, and p is the output fiber where the incoming packet is bound.

The output stage comprises N output processing modules which connect to N output fibers. An output processing module consists of a cyclic AWG DeMUX, M fixed optical wavelength converters (FOWC) and a multiplexer. The cyclic AWG DeMUX provides a modular M operation on wavelengths. To our best knowledge, this is the unique design on this FSOB switch architecture. It completely enables full output buffer sharing in the switch. The cyclic demultiplexer is a passive optical device that acts as a mathematical "modular" of the wavelengths of the packets. The $1 \times M$ cyclic demultiplexer leads the wavelengths i, $M+i$, …, $nM+i$ to output channel i of the demultiplexer. The packets from the cyclic demultiplexer are sent to corresponding fixed optical wavelength converters (FOWC) to convert them into a specific output wavelength and then multiplexed and transmitted.

The architecture combines with the wavelength conversion to make the switch, a non-blocking one. That is, there is no packet collision inside the switch. Before showing this property, we first identify that there are only two possible locations that would result in packet collision. The first one is when more than two packets with the same wavelength appearing concurrently at a buffer. The other one would be when two or more packets arrive simultaneously at a FOWC We will show that the switch is collision-free under our control policy.

Property:

For packets coming in from M consecutive outputs of the packet classification, they are collision-free in both the FDL buffers, and the output processing module.

Proof:

Without loss of generality, it is assumed that there are J packets routed to the same destination intended for fiber p with desired delay k. These packets are grouped by the classifying block and are located at inputs of AWG2 from input t to input $t+J-1$. Then based on the wavelength assignment equation (2-3), the wavelengths of these packets are $(NM-t+kN+p) \bmod NM \ldots (NM-(t+J-1)+kN+p) \bmod NM$.

(a). Since those wavelengths are distinct, there is no wavelength collision in buffer k.

(b). A packet with wavelength j will come out at output channel $j \bmod M$ of the cyclic demultiplexer. Therefore, these J packets with wavelengths $(NM-t+kN+p) \bmod NM \ldots (NM-(t+J-1)+kN+p) \bmod NM$ will come out at output channel $[(NM-t+kN+p) \bmod NM] \bmod M \ldots [(NM-(t+J-1)+kN+p) \bmod NM] \bmod M$. Since those J values are also distinct, it proves collision-free at the cyclic demultiplexer output.

Based on (a) and (b), we conclude the property.

2.5 Traffic Models

For the WDM packet switch system, packets may arrive synchronously or asynchronously. We discuss the synchronous mode. Assume that the packets arrive synchronously at different wavelength channels of the input fibers in each time slot. All incoming packets are fixed length and can be transmitted with in a time slot. According to the architecture described in the previous section, there are N input fibers and N output fibers. In each fiber, M wavelengths are used to transmit packets and L FDL buffers are allocated to an output fiber.

First we give an analytical analysis for the packet loss probability of our architecture. The notation used in the model is listed as follows. The problem is formulated as a Markov chain problem such that, given the packet arrival rate and the probability of output port, the network state probability (i.e., buffer occupancy) and packet loss probability are obtained.

Since we have N input fibers, the total input slots are NxW in a slot time. The input can be viewed followed a binomial distribution *B(NW, p)*, where *p* is the probability of a slot with an incoming packet. For an incoming packet, the probability for the packet destined for port *i* is q_i. Without loss of generality, in the following analysis we designate port *1* for observation to derive its packet loss probability. Probability q_1 is replaced with q for simplicity.

N	:	the total number of input/output fibers in the system;
M	:	the total number of wavelengths on a fiber;
L	:	the total number of fiber delay lines for an output port; the FIFO depth is equal to *LM* due to the multiplication of delay unit sharing;
p	:	the probability of a non-empty incoming slot;
q	:	the probability of an incoming packet destined to the observed output port;
r_m	:	the probability of *m* incoming packets destined to the observed output port at the same time;

For an incoming slot, the probability of *m* packets destined to the observed output port is:

$$r_m = \sum_{x=m}^{NM} \left[\binom{NM}{x} p^x (1-p)^{NM-x} \right] \times \left[\binom{x}{m} q^m (1-q)^{x-m} \right] \tag{2-4}$$

The system state depends on the buffer spaces being occupied. The system at state *i* means that there are *i* packets queued in the buffer. The state transition probability is derived as:

$$P_{ij}, \qquad (\text{for } 0 \le i, j \le LM) \tag{2-5}$$
$$= \begin{cases} 0, & \text{if } j-i > (N-1)M \text{ or } i-j > M \\ \sum_{m=(L+1)M-i}^{NM} r_m & \text{if } j = LM \text{ and } -M \le j-i \le (N-1)M \\ \sum_{m=0}^{M-i} r_m & \text{if } j = 0 \text{ and } -M \le j-i \le (N-1)M \\ r_{j-i+M} & otherwise \end{cases}$$

Let row vector π be the state probability in steady state,

$$\pi P = \pi \tag{2-6}$$

$\pi \boldsymbol{1} = 1$ (where ***1*** is a column vector with all entry equals to 1) (2-7)

Packet loss probability

$$b = \sum_{i=0}^{LM} \pi_i \left[\sum_{m=1}^{NM} r_m y_{im} \Big/ m \right], \text{ where}$$
$$y_{im} = \begin{cases} m + i - (L+1)M, & \text{if } m + i - M > LM \\ 0, & otherwise \end{cases} \tag{2-8}$$

y_{im} means the number of packets dropped as the system is in state i and there are m new incoming packets destined to the observed output port.

2.6 Performance Analysis

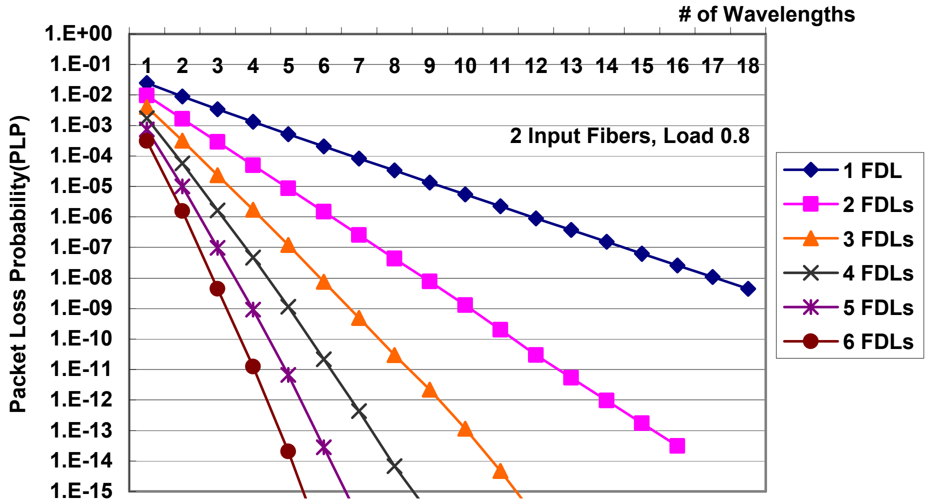
We have carried out a performance study based on the analytical analysis which was given in the previous section. In this section we examine the Packet Loss Probability (PLP) of the FSOB switch under different numbers of wavelengths, numbers of FDLs, and traffic load. Each FSOB switch handles a total of N input/output ports and M wavelengths resulting in NM connections processed by the FSOB switch. Because the traffic destination is uniformly distributed among all output ports, the probability of an incoming packet destined to the observed output port (q) is equal to $1/N$. We observe heavy load traffic first. The traffic load is set to 0.8. Consequently, the probability of a non-empty incoming slot (*p*) is equal to 0.8. In Figure 12, *p* is set to 0.8 and *q* is set to 0.5. As shown in Figure 12(a), the packet loss probability decreases as the number of wavelengths or the number of FDLs increases. For instance, to achieve a PLP of 10–7 requires 1 FDL and 15 wavelengths or 2 FDLs and 8 wavelengths. Besides, as shown in Figure 12(b), with the same number of wavelengths, applying only a few optical buffers immediately yields drastic improvement in PLP. The connection of the number of buffers and the packet loss probability can be learned from the 1 FDL case in Figure 12(c). As shown in this figure, to achieve a PLP of 10–7 requires 15 buffers. Therefore, there are different ways to construct 15 buffers, 2 FDLs and 8 wavelengths or 3 FDLs and 6 wavelengths. 2 FDLs and 8 wavelengths construct 16 buffers. 3 FDLs and 6 wavelengths can provide 18 buffers. They are both greater than 15 buffers so their corresponding PLP is smaller than 10–7. Due to the 3-FDL-6-wavelength providing 3 more buffers than the 1-FDL-15-wavelength, its PLP comes to 10-8 being far below 10–7.

Then we increase the number of input fibers from 2 fibers to 4 fibers and 6 fibers to study their PLP properties. Their PLP behaviors are shown in Figure 13 and Figure 14 and are similar to the one of the two input fibers system. We observe the connections between the packet loss probability and the buffer correlating with the systems with different input fibers under traffic load 0.8. As shown in Figure 15(a), if the system has 12 buffers, the PLP of the 2*2 (2 Fibers In and 2 Fibers Out) system is around 10-7 but the PLPs of other systems are around 10-5. When the number of input fibers and output fibers increases, the PLP also increases, but the PLP difference between two systems with continuous numbers of input fibers is getting smaller. From Figure 15(b), we observe the number of buffers needed to achieve the desired packet loss probability with different numbers of input fibers. If the PLP being smaller than 10-6 is demanded, the 2*2 system only requires 12 buffers but the 6*6 system requires 19 buffers. The same phenomenon is shown in PLP ≦ 10-7 and PLP ≦ 10-8.

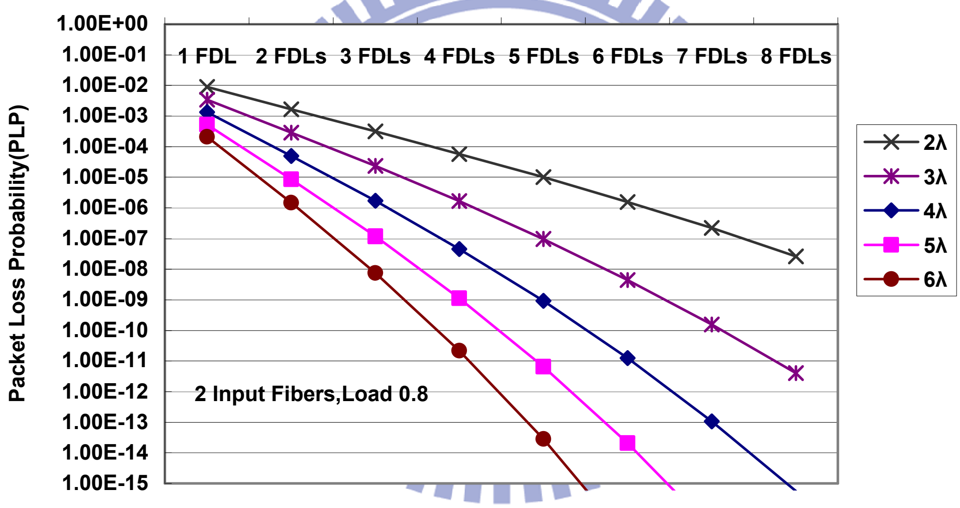
Finally, we examine the system behaviors under different traffic loads. Three traffic loads are given, the light load (0.2), medium load (0.5) and heavy load (0.8), as shown in Figure 16. To achieve the 10-6 PLP, only 3 buffers are needed in the light traffic load case, and 7 buffers are required in the medium traffic load, therefore up to 21 buffers are needed when the system is heavily loaded. We can come to a conclusion that under the same traffic load and the same demanded PLP, a system with a smaller number of inputs requires less buffer. However, the trends of PLPs corresponding to different numbers of input fibers under different traffic loads are similar. According to these numeric results shown above, a 2*2 switching system requires 12 buffers that consist of 2 FDLs and 6 wavelengths to achieve the 10-7 PLP. A 6*6 witching system requires 3 FDLs and 6 wavelengths to achieve the 10-7 PLP.

Resolving contentions in both the wavelength and space dimensions together yields immense a decline in the number of FDLs required. It means it is possible to use only a small amount of discrete FDL optical buffers combined with multiple wavelengths to provide satisfactory packet loss performance in a multi-wavelength OPS system. Thus this makes FDLs become practical to resolve output contention of an OPS system. Increasing wavelength numbers of each input fiber can enlarge the buffers on each FDL and decrease the PLP. Therefore the large number of wavelength on each fiber is preferred.

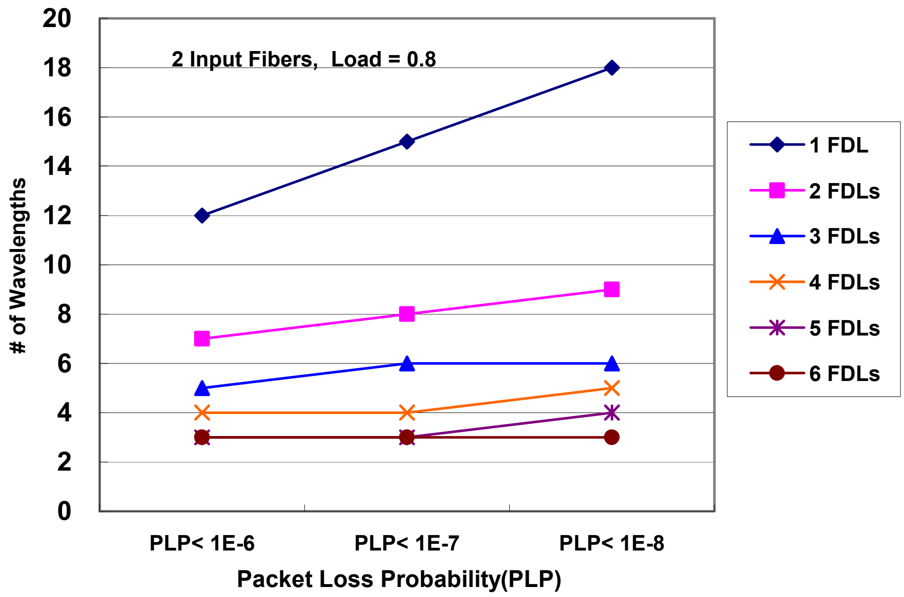
However, when in input ports increases, the internal wavelengths increased drastically due to the mass wavelength on an input fiber. Consequently the size of the AWGs used in the architecture may increase rapidly in cases of internal wavelengths being raised. To overcome this problem, we divide the FSOB architecture into several smaller blocks by partitioning the wavelengths from the input fibers into groups of fewer wavelengths. Each group of wavelengths is then processed using smaller AWGs where the number of wavelengths in a group can be determined by referencing the numerical results as shown in Figure 12 to Figure 16. Figure 17 presents an example of a 4x4 switch with four wavelengths, which can be partitioned into two 4x4 switches each with two wavelengths.



(a). Packet Loss Probability (PLP) with respect to various FDLs

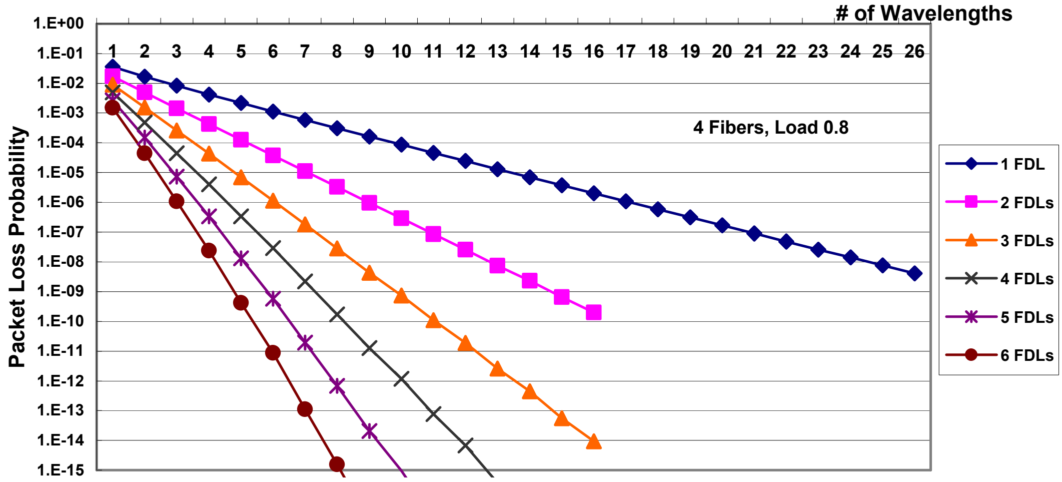


(b). Packet Loss Probability (PLP) with respect to various wavelengths

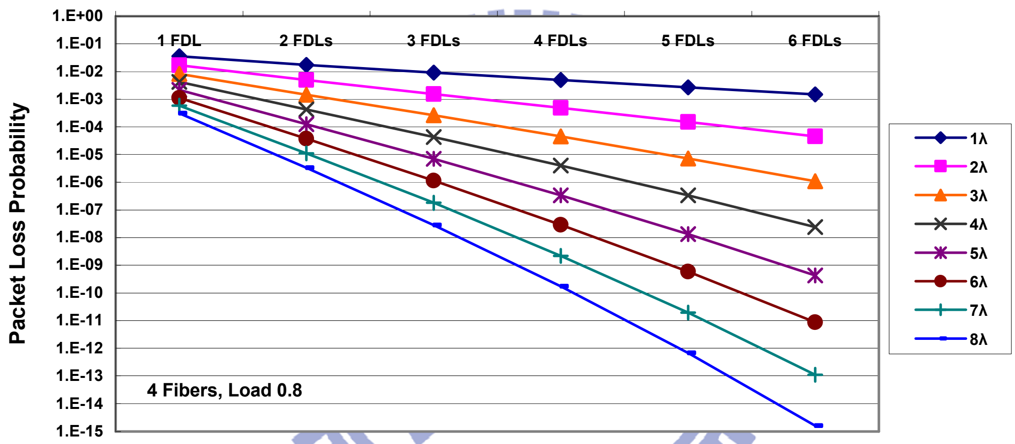


(c). Wavelengths and FDLs requirements for various PLPs

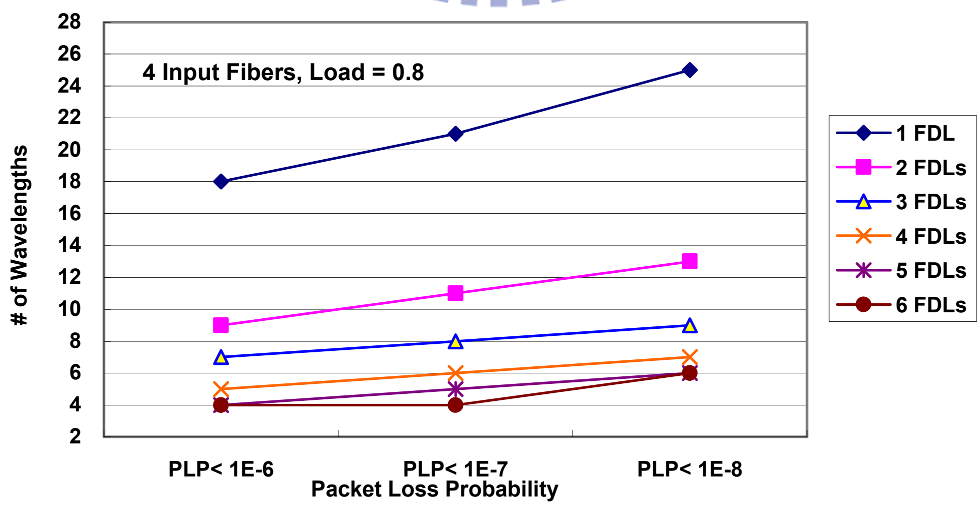
Figure 12 - Packet loss probability (PLP) of 2 by 2 system under traffic load 0.8



(a). Packet Loss Probability (PLP) with respect to various FDLs

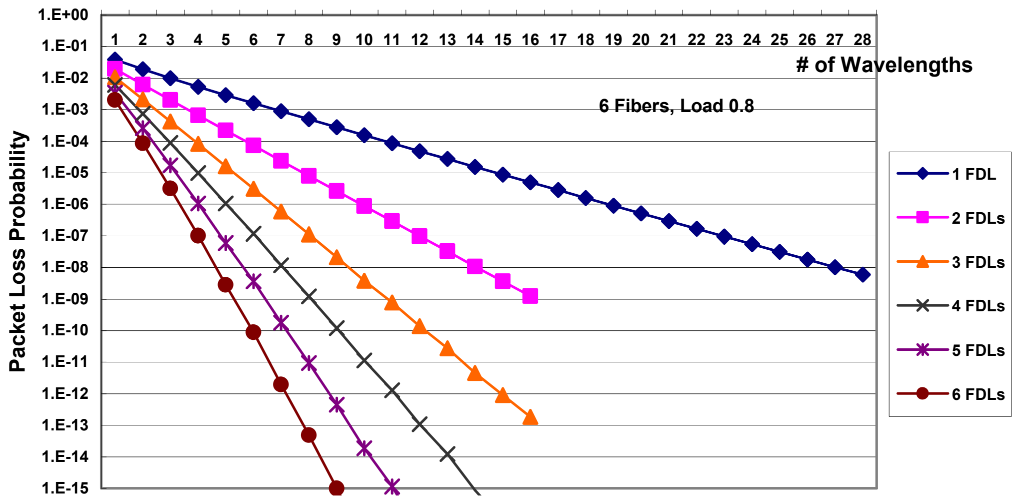


(b). Packet Loss Probability (PLP) with respect to various wavelengths

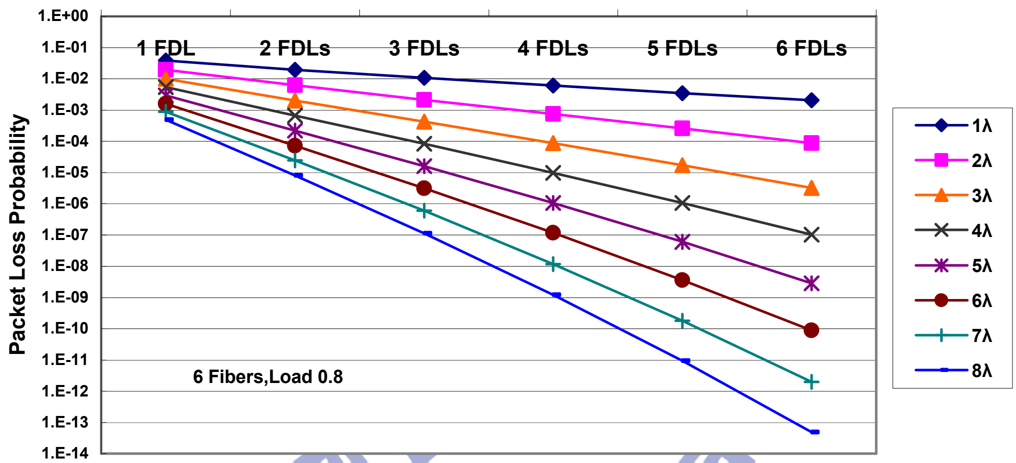


(c). Wavelengths and FDLs requirements for various PLPs

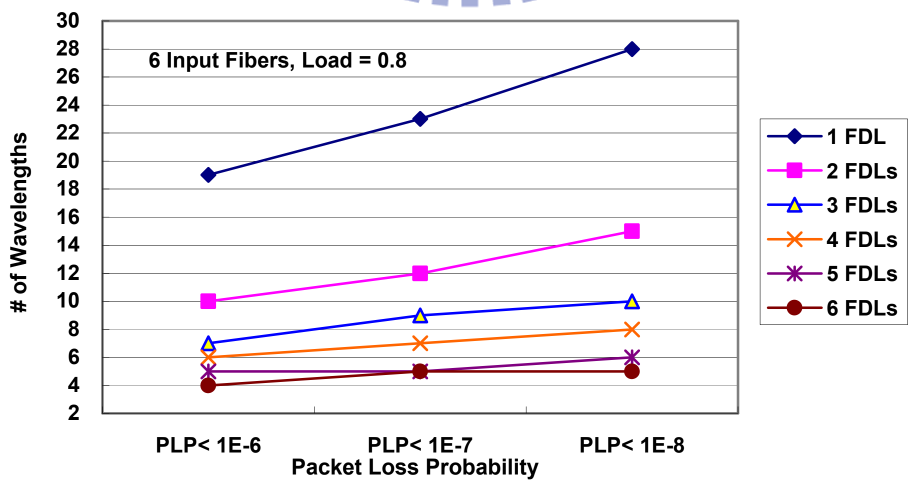
Figure 13 - Packet loss probability (PLP) of 4 by 4 system under traffic load 0.8



(a). Packet Loss Probability (PLP) with respect to various FDLs

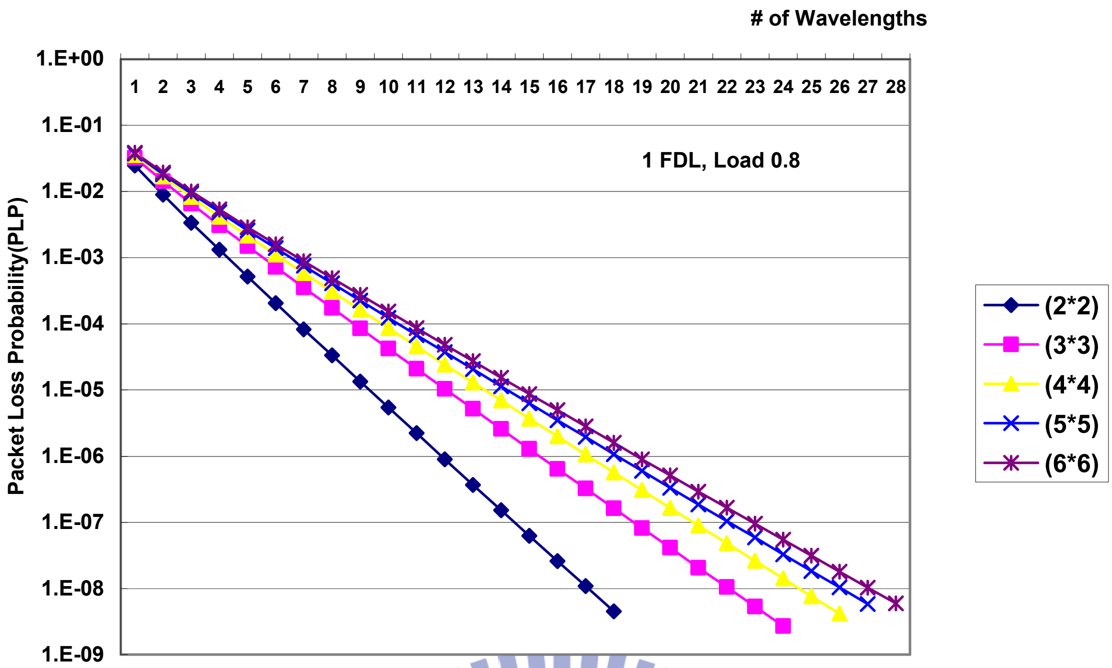


(b). Packet Loss Probability (PLP) with respect to various wavelengths

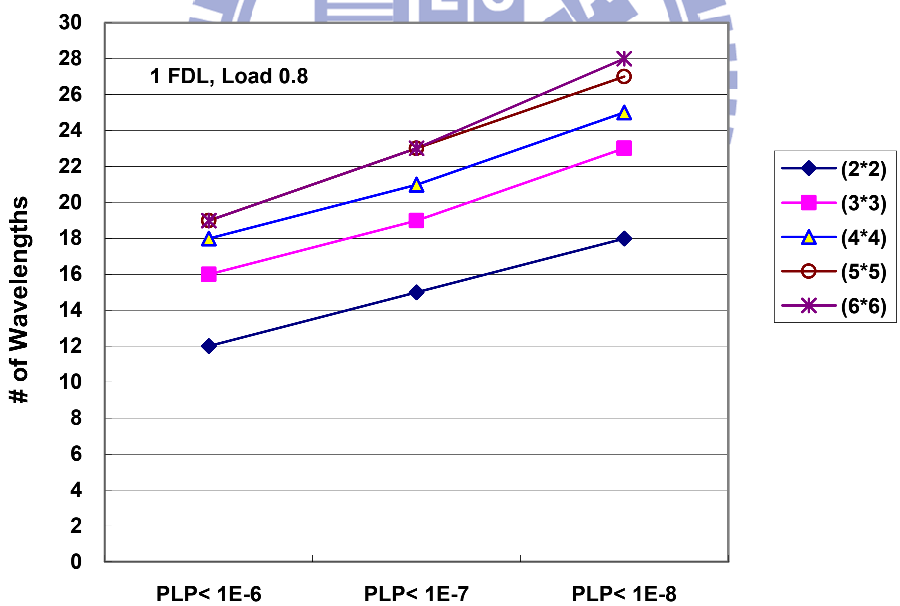


(c). Wavelength and FDL requirements for various PLPs

Figure 14 - Packet loss probability (PLP) of 6 by 6 system under traffic load 0.8



(a). Packet Loss Probability (PLP) with respect to various switching scale settings



(b). wavelengths and FDLs requirements for various PLPs with respect to various switching scale settings

Figure 15 - Packet loss probability (PLP) of various switching scale settings

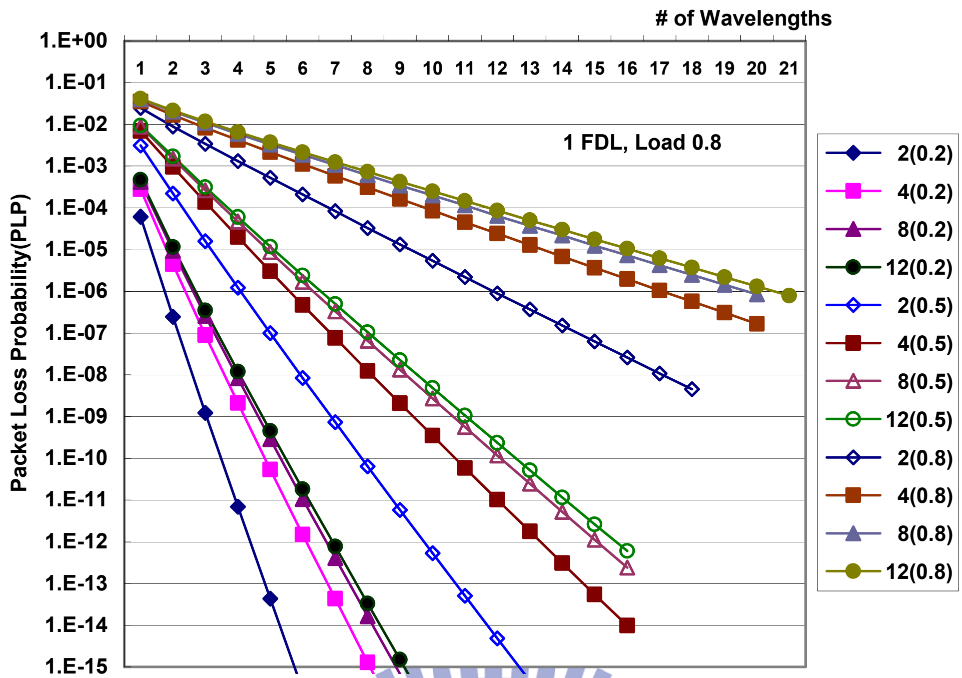


Figure 16 - Packet loss probability (PLP) of various switching scale and traffic load settings

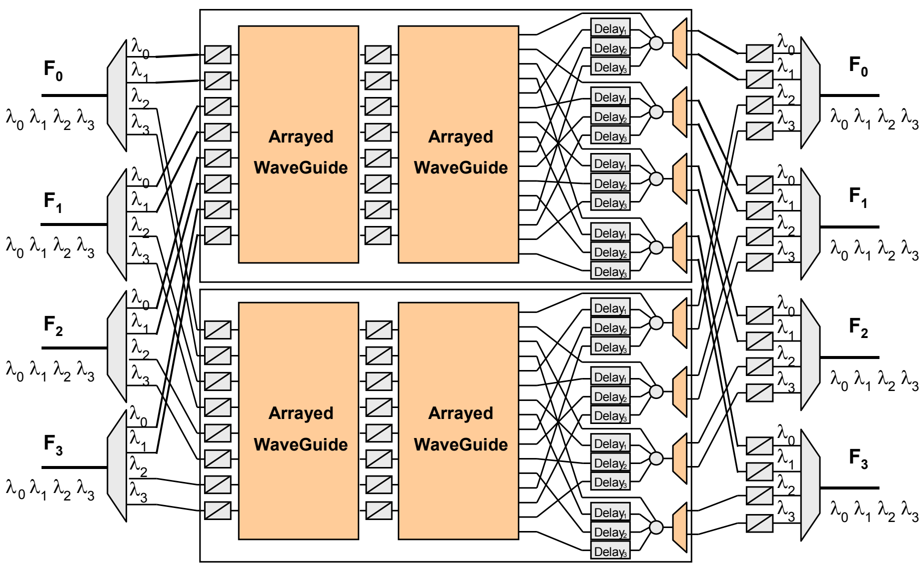


Figure 17 - A new architecture of FSOB system

Chapter 3. Advance Lightpath Reservation in WDM Networks

With advances in optical Wavelength Division Multiplexing (WDM) networks having been widely recognized as the dominant transport infrastructure for future Internet backbone network, most WDM applications follow the Optical Circuit Switching (OCS) paradigm. In some circumstances users make call requests in advance to reserve network resources for communications. Examples such as lambda grid and virtual private optical networks usually need many high-speed lightpaths for connecting computer servers in diverse enterprise campuses. A major feature of such applications is that traffic demands are requested to the network in advance before the connections are set up [34]-[37].

The Advance Lightpath Reservation problem is in short referred to as ALR problem in this dissertation. One major challenge arising in ALR problem has been to jointly determine call admission control as well as Routing and Wavelength Assignment (RWA) [11]. Particularly, for an optical network without a wavelength conversion capability, the problem deals with RWA between source and destination nodes subject to the wavelength-continuity constraint [12]. It has been shown that RWA is an NP-complete problem [12]. Therefore, the ALR problem is also NP-complete since an RWA problem is a special case of the ALR problem.

Several algorithms for resolving the ALR problem have been proposed in the literature. In [34], the authors present a basic framework for automated provisioning of advance reservation service based on GMPLS protocol suites. In [38], the ALR problem is classified into several types depending on the flexibility of call arrival time and call duration. Heuristic RWA algorithms are also demonstrated for the problems. In [35], a simulated annealing-based algorithm is proposed to find a solution for predetermined k-shortest paths. In lambda grid networks, Miyagi et al. consider how to reserve a wavelength for deadline-aware applications

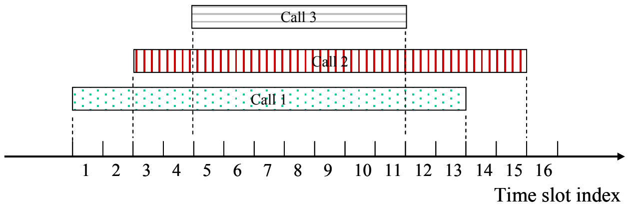
[36]. Performance for blocking probability is evaluated under greedy-based and deadline-first-based heuristic algorithms. In this dissertation, we propose an efficient Lagrangean Relaxation (LGR) approach to resolve advance lightpath reservation for multi-wavelength optical networks.

3.1 Advance Lightpath Reservation Problem Formulation

We consider a WDM network where each WDM link consists of a pair of unidirectional fiber links with a number of wavelengths on each fiber. The network is under centralized control. There is a central controller responsible for call admission control, routing and wavelength assignment so as to establish lightpaths for all connection requests on behalf of all network nodes.

The ALR problem is formulated as an integer linear programming problem stated as follows. Given a physical topology and each call information (start time, end time, revenue), determine wavelengths of lightpaths, such that the total revenue from admitting calls is maximized under the wavelength continuity constraint. The bandwidth demand is one wavelength in the context of this chapter. Throughout the chapter, we use connection and call interchangeably. For ease of illustration, we assume in the sequel that the number of available wavelengths on each link is the same.

Before describing the model, we first give an example for the ALR problem. In Figure 18, there are three calls requests. Call 1 goes from time slot 1 to time slot 13. Call 2 and call 3 start from time slots 3 and 5, and end at time slots 11 and 15, respectively. Hence we have six time events (1, 3, 5, 11, 13, and 15) that need tracking. Those six event points form the six members of the set T. Let σ_{kt} denote as a binary index to represent if call k includes event time t. Since, in this example, call 1 goes over event index 1, 2, 3, 4, and 5, we could derive that σ_{11}=1, σ_{12}=1, σ_{13}=1, σ_{14}=1, σ_{15}=1, and σ_{16}=0.



Call 1: start time =1, end time =13
Call 2: start time =3, end time =15
Call 3: start time =5, end time =11
Time event set T={1,3,5,11,13,15}
σ_{1*}={1,1,1,1,1,0}
σ_{2*}={0,1,1,1,1,1}
σ_{3*}={0,0,1,1,0,0}

Figure 18 - Example of an ALR problem

We summarize the notation used in the formulation as follows:

Input values:

L	:	set of optical links;
N	:	set of optical cross-connects;
W	:	set of wavelengths on each link; (same for all links);
$\|W\|$	:	number of wavelengths available on each fiber link;
K	:	set of connection requests;
$\|K\|$	:	number of call requests;
r_k	:	revenue for accepting call request k;
P_k	:	candidate path set for call k;
δ_{pl}	: =	1, if path p includes link l; = 0, otherwise;
T	:	index set to denote the call arrival times of all requests;
σ_{kt}	: =	1, if call k goes through event time t; = 0, otherwise;

Decision variables:

$\boldsymbol{x_{pw}}$	: =	1, if lightpath p uses wavelength w; = 0, otherwise;
$\boldsymbol{y_k}$	: =	1, if call k is accepted by the network; = 0, otherwise;

Problem (P):

$$\max \sum_{k \in K} r_k y_k$$

subject to:

$$y_k = \sum_{p \in P_k} \sum_{w \in W} x_{pw} \qquad \forall k \in K \qquad (3\text{-}1)$$

$$\sum_{p \in P_k} \sum_{w \in W} x_{pw} \le 1 \qquad \forall k \in K \qquad (3\text{-}2)$$

$$\sum_{k \in K} \sum_{p \in P_k} x_{pw} \delta_{pl} \sigma_{kt} \le 1 \qquad \forall w \in W, l \in L, t \in T \qquad (3\text{-}3)$$

$$x_{pw} = 0 \text{ or } 1 \qquad \forall p \in P_k, k \in K, w \in W \qquad (3\text{-}4)$$

$$y_k = 0 \text{ or } 1 \qquad \forall k \in K \qquad (3\text{-}5)$$

The objective function is to maximize the total revenue. Usually the revenue is proportional to the call duration. If we set r_k to be one for all requests k, the problem becomes to maximize the number of accepted calls. In that case, the problem is also equivalent to minimize call blocking. Constraints (3-1) and (3-2) require that at most one lightpath be selected for each request. If the connection of call k is rejected, in which case the corresponding variable x_{pw} is 0, zero revenue contributes to the objective function. Constraint (3-3) guarantees no overbooking on any wavelength channel at any time slot. It requires that for any wavelength on a link, there is at most one lightpath using it. Constraint (3-4) states the 0/1 binary constraint on routing variable x_{pw}. Please note that we use time event T in our model, instead of directly using a time slot index. The reason to use set T is to reduce the problem size. There are at most 2|K| members in T. That is usually far smaller than the total number of time slots. For example, in Figure 18, the total number of time slots is 16 while the total number of events is 6. By using this technique, we can reduce the total number of constraints significantly. Finally, whether a call request is accepted or not is determined by constraint (3-5).

If we set all call requests with the same duration, the above ALR problem is reduced to a general RWA problem which has been proven to be NP-complete. Therefore, it is unlikely to obtain an exact solution for realistic networks in real-time. The problem is approximated using the LGR approach presented in the next section.

3.2 Lagrangean Relaxation based Heuristic Algorithm

Lagrangean Relaxation (LGR) [39,[40] has been successfully employed to solve complex mathematical problems by means of constraint relaxation and problem decomposition. Particularly for solving a linear integer problem, unlike the traditional linear programming approach that relaxes integer into non-integer constraints, the Lagrangean-based method generally leaves the integer constraints in the constraint sets while relaxing complex constraints such that the relaxed problem can be decomposed into independent manageable subproblems. Through such a relaxation and decomposition, the LGR method is shown to provide tighter bounds and shorter computation time on the optimal values of objective functions more than those provided by the linear programming relaxation approach in many instances [40].

In this dissertation, we propose a new LGR algorithm, which is used for the first time to our best knowledge to precisely and efficiently solve the advance lightpath reservation problem. Essentially, the original primal problem is first simplified and transformed into a dual problem after some constraints are relaxed. If the objective of the primal problem is a maximization or minimization function, the solution to the dual problem is a respective upper or lower bound to the original problem. Such a Lagrangean bound is a useful by-product in resolving the Lagrangean relaxation problem. Next, due to constraint relaxation, the upper bound solutions generated during the computation might be infeasible for the original primal problem. However, these solutions and the generated Lagrangean multipliers can serve as a base to develop efficient primal heuristic algorithms for achieving a near-optimal solution.

ALR is first formulated as a combinatorial optimization problem in which the revenue from admitting call requests is maximized. The LGR approach performs constraint relaxation and derives an upper-bound solution according to a set of Lagrangean multipliers generated through subgradient-based iterations. In parallel, using the generated Lagrangean multipliers, the LGR approach employs a new primal heuristic algorithm to arrive at a near-optimal solution to the original problem. By upper bounds, we delineate the performance of LGR with respect to accuracy and convergence speed under different parameter settings and termination criteria.

3.2.1 Dual Problem and Upper Bound

In the relaxation process, constraint (3-3) is first relaxed from the constraint set. The expression corresponding to the constraint is multiplied by Lagrangean multipliers u_{wlt} and then summed with the original objective function. Problem (P) is thus transformed into a dual problem, called Dual_P, given as follows:

Problem (Dual_P):

$$
\begin{aligned}
&Z_{dual}(\boldsymbol{u}) \\
&= \max\left\{\sum_{k\in K} r_k y_k - \sum_{w\in W}\sum_{l\in L}\sum_{t\in T} u_{wlt}\left(\sum_{k\in K}\sum_{p\in P_k} x_{pw}\delta_{pl}\sigma_{kt} - 1\right)\right\} \\
&= \max\left\{\sum_{k\in K}\left(r_k y_k - \sum_{p\in P_k}\sum_{w\in W}\sum_{l\in L}\sum_{t\in T} u_{wlt} x_{pw}\delta_{pl}\sigma_{kt}\right)\right. \\
&\qquad \left. + \sum_{w\in W}\sum_{l\in L}\sum_{t\in T} u_{wlt}\right\}
\end{aligned}
$$

Subject to Constraints (3-1), (3-2), (3-4), and (3-5) where vector u (with component uwlt) is the non-negative Lagrangean multiplier vector. Problem (Dual_P) in Equation (3-6) can be decomposed into |K| independent subproblems (one for each call k). Problem (Dual_P) is then expressed as $Z_{dual}(\boldsymbol{u}) = \sum_{k\in K} Z_k^{sub}(\boldsymbol{u}) + \sum_{w\in W}\sum_{l\in L}\sum_{t\in T} u_{wlt}$, where $Z_k^{sub}(\boldsymbol{u})$ is as follows.

$$
Z_k^{sub}(\boldsymbol{u}) = \max\left\{r_k y_k - \sum_{p\in P_k}\sum_{w\in W}\sum_{l\in L}\sum_{t\in T} u_{wlt} x_{pw}\delta_{pl}\sigma_{kt}\right\}
$$

subject to:

$$y_k = \sum_{p \in P_k} \sum_{w \in W} x_{pw} \tag{s1}$$

$$\sum_{p \in P_k} \sum_{w \in W} x_{pw} \le 1 \tag{s2}$$

$$x_{pw} = 0 \text{ or } 1 \qquad \forall p \in P_k, w \in W \tag{s3}$$

$$y_k = 0 \text{ or } 1 \tag{s4}$$

By solving all the |K| $Z_k^{sub}(\boldsymbol{u})$ subproblems, we can obtain the value of $Z_{dual}(u)$. According to the weak Lagrangean duality theorem [40], Z_{dual} in Equation (3-6) is an upper bound of the original Problem (P) for any non-negative Lagrangean multiplier vector $\boldsymbol{u}$. Clearly, the lowest upper bound is to be determined. Equation (3-6) can be solved by the subgradient method, as shown as a part of the LGR approach delineated in Figure 19, which shows that the algorithm is run for a fixed number of iterations (*Iteration_Number*). In every iteration, the subproblems are solved (as described above), resulting in the generation of a new Lagrangean multiplier vector value. Then, according to Equation (3-6), a new upper bound is generated. If the new upper bound is tighter (lower) than the current best achievable upper bound (*UB*), the new upper bound is designated as the UB. Otherwise, the UB value remains unchanged. Significantly, if the UB value does not improve for a number of iterations that exceeds a threshold, called *Quiescence_Threshold (QT)*, the step size coefficient λ of the subgradient method is halved, in an attempt to reduce oscillation possibility. Specifically, in the update-step-size and update-multiplier procedures in Figure 19, the Lagrangean multiplier vector $\boldsymbol{u}$ is updated as $\boldsymbol{u}_{k+1} = \boldsymbol{u}_k + \boldsymbol{\theta}_k * \boldsymbol{b}_k$, where $\boldsymbol{\theta}_k$ is the step size, determined by $\theta_k = [\lambda_k (Z_{dual}(u) - LB)] / \|b_k\|^2$, in which λ_k is the step size coefficient, LB is the current achievable largest lower bound obtained from the Primal Heuristic Algorithm described next, and $\boldsymbol{b}_k$ is a subgradient of Zdual ($\boldsymbol{u}$) with vector size $|L*W*T|$.

3.2.2 Primal Heuristic Algorithm and Upper Bound

The primal heuristic algorithm in the LGR approach is used to find a near optimal solution. Since our problem is a maximization problem, a near optimal solution is clearly also a lower bound solution. Similar to the upper bound case, as given in Figure 19, if the new lower bound (*lb*) is tighter (larger) than the current best achievable upper bound (*LB*), the new lower bound is designated as the *LB*.

To obtain a near-optimal solution that is the highest lower bound at the end of a subgradient iteration, the LGR solution is verified whether or not it satisfies those relaxed constraints. If it does, the solution is feasible and is thus used to calculate a lower bound of the primal problem (P). If the solution is infeasible, we employ the following LGR-based heuristic algorithm, which takes advantage of Lagrangean multipliers. As shown in Figure 20, the LGR algorithm sequentially accepts connections based on the r_k - c_k values. Calls with higher r_k - c_k hold higher priority in the sequence. The routing is determined by Dijkstra's shortest path algorithm based on the link cost, $\sum_{t\in T} u_{wlt}\sigma_{kt}$, as those used in the previous section except that the cost of those links is set to infinite for wavelengths that are taken by previous calls. It prevents those calls with lower priority from using the wavelength channel taken by previous high priority ones.

If there are not enough resources for the request, the call is rejected. The algorithm runs repeatedly until all requests are satisfied or rejected.

begin
  initialize Lagrangean multiplier vector $\boldsymbol{u} := \boldsymbol{0}$
  $UB := \sum_{k\in K} r_k$ /* upper bound */
  $LB := 0$ /* lower bound */
  quiescence_age := 0
  step size coefficient $\lambda := 2$
  for each $k := 1$ **to** *Iteration_Number* **do**

```
begin
    solve subproblem for each k ∈ K
    Z_dual = Σ_{k∈K} Z_k^sub(u) + Σ_{w∈W} Σ_{l∈L} Σ_{t∈T} u_wlt   /*Eq. (6)*/
    if Z_dual < UB then
    begin
        UB := Z_dual
        quiescence_age := 0
    end
    else quiescence_age := quiescence_age + 1
    if quiescence_age ≥ Quiescence_Threshold then
    begin
        λ := λ/2
        quiescence_age := 0
    end
    run Primal Heuristic Algorithm to get lb
    if lb > LB then
        LB := lb /* lb is the new lower bound */
    run update-step-size
    run update-multiplier
  end
end
```

Figure 19 - Lagrangean relaxation algorithm (LGR)

```
begin
  Sorting Z_k^sub(u) for all calls k and put their index in priority Q
   /* Q[1] is the call with the largest Z_k^sub(u) value*/
   /* Q[|K|] is the call with the smallest Z_k^sub(u) value*/
  for each link l ∈ L, w ∈ W
     a_lw := 1 /*all wavelength channels available*/
  for (i = 1; i ≦ K; i++)
    begin
        k = Q(i) /*DeQueue the highest priority call from Q*/
        c_k := ∞
        accept := False
```

```
        for each wavelength w ∈ W do
        begin
            for each link l∈L do
        if   a_lw = 1 then
            link cost cost_l := Σ_{t∈T} u_wlt σ_kt
        else
            cost_l := ∞
      src = source(k)
      dest = destination(k)
      p' := Dijkstra-shortest-path(cost, src, dest)
      if p' is a feasible path then
      begin
        accept := True;
        /* denote c_wk as path cost of p' */
        If   c_wk < c_k   then
        begin
          c_k := c_wk
          p* := p'
          w* := w
        end
      end
    end
    If   accept = True
      begin
        Accept call k and p* is the routing path
        a_lw* := 0 for those links used by path p*
      end
    else
      Reject call k
    end
  end
  update total revenue and return as a lower bound lb
end
```

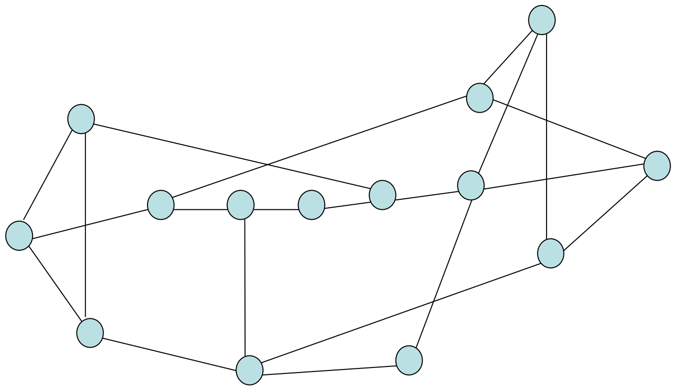
Figure 20 - Primal heuristic algorithm

3.3 Experimental Results

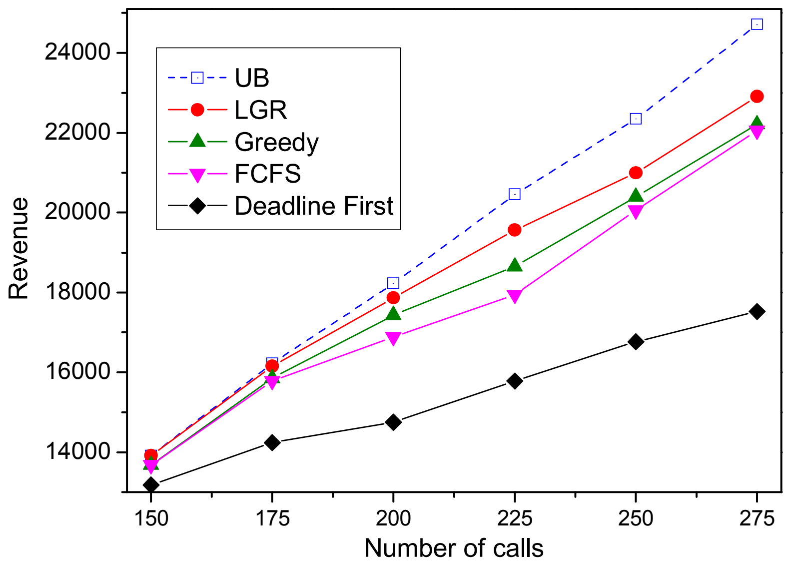
We have carried out a performance study on the LGR approach, and drawn comparisons between LGR and some heuristic algorithms via experiments over the well-known NSFNET Network. In the simulations, the start time and end time of call requests are generated randomly following uniform distribution in one day. Consequently, the mean call duration is 450 minutes. The call revenue r_k is set exactly equal to the call duration. Therefore, a call with longer duration receives more revenue than those with shorter durations.

In the computation using our LGR approach, we adopted I*teration_Number* = 3000 and *Quiescence_Threshold* = 50. The LGR algorithm can obtain near optimal results within 10 minutes of computation time operated on a PC running Windows XP with a 2 GHz CPU power. Three other heuristics are also considered in the study. The Greedy method sequentially allocates lightpaths according to connection's r_k value. Calls with larger revenues hold higher priority in the call setup process. We also consider two timing related heuristics in our experiments. The First-Come-First-Serve (FCFS) method schedules the requests according to the call arrival time while the Deadline First (DF) method instead schedules the requests according to the call finish time. The numerical results on NSFNET ranging from 150 to 275 calls are plotted in Figure 21. Figure 21(b) shows the total revenue. The LGR achieves highest total revenue followed by the Greedy method. The FCFS method and the DF method result in lower output due to the lack of taking call revenue into account. Percentage Gap (*Gap%*) is used to be the performance metric to evaluate the quality of those algorithms to a legitimate upper bound. The Percentage *Gap (Gap%)* is defined as the percentage of *(Lagrangean UB — total revenue of the considered algorithm) / Lagrangean UB*. As shown in Figure 21(c), the percentage gap between the LGR and the *UB* are within 7% for all cases.

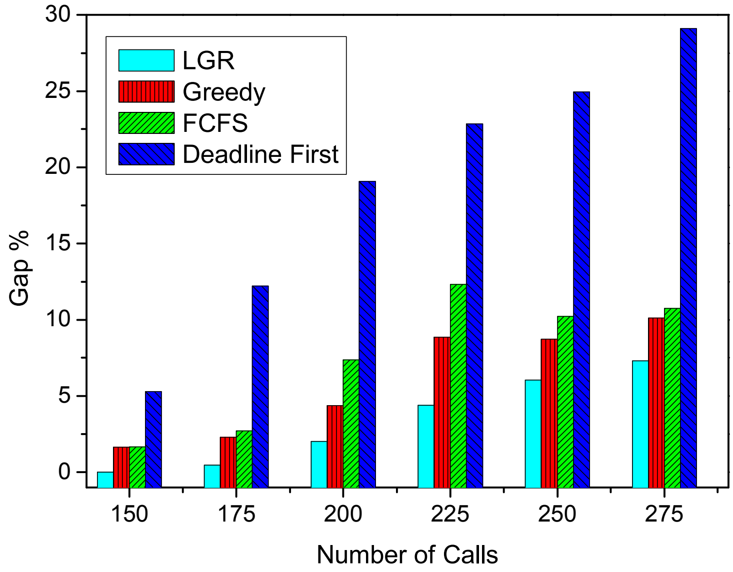
Further performance comparisons are made with respect to call blocking. As shown in Figure 21(d), the LGR outperforms the other three methods. It is interesting that the Greedy heuristic algorithm is the one with the largest number of call rejections. By closely examining the results we find that those calls rejected by the Greedy algorithm are with small call durations.



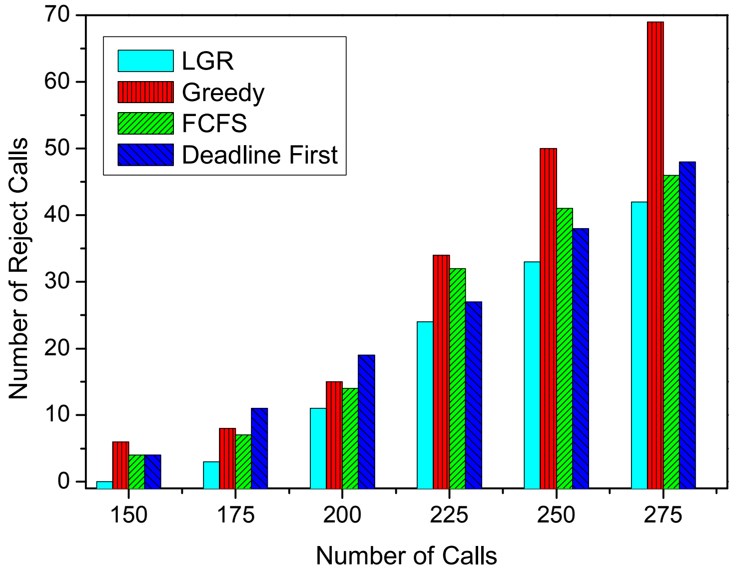
(a). NSFNET network (14 nodes, 42 links, 8 wavelengths)



(b) Performance Comparisons - Revenue



(c) Performance Comparisons – Percentage Gap



(d) Performance Comparisons – Call Blocking

Figure 21 - Simulation results of ALR

Chapter 4. Multi-Path Provisioning for NG-SONET Networks with Quality-of-Survivability Constraints

Synchronous Optical Networking (SONET) and Synchronous Digital Hierarchy (SDH) are the most popular standardized multiplexing protocols over optical fiber. Designed to optimize TDM-based traffic, SONET/SDH is very robust and reliable, containing built-in mechanisms to provide high network availability. Both SDH and SONET, with their superior survivability and short failure recovery time, have been dominating transport in metro and backbone networks for decades. However legacy SONET/SDH only supports contiguous concatenation transport switching over the overall path and is not suited to handling packet data. The SONET/SDH rates have a coarse granularity and are not a good match to packet traffic. For example, the basic unit of framing in SONET is a STS-1 (synchronous transport signal - 1), which operates at 51.84 Mbps. The next level of SONET framing, STS-3, supports triple the bandwidth, or 155.52 Mbps. Higher levels of SONET framing increase the bandwidth in successive multiples of four, up to approximately 40 Gbps. One problem with this concatenation scheme is that its bandwidth allocation is inflexible for data traffic. When the mix of data and voice traffic are carried on a SONET/SDH path, a large amount of unused bandwidth may be left over. It is caused by the fixed sizes of concatenated containers. For example, fitting a 100 Mbit/s Fast Ethernet connection inside a 155 Mbit/s STS-3c container leads to near 55Mbits bandwidth waste.

Automatic protection switching (APS) and self-healing ring (SHR) are the most common protection schemes used in SONET/SDH networks [55]. SONET SHR is a very successful technique for improving optical network survivability. SONET networks are designed to have ring architectures. The SHR can generally be divided into two schemes: Unidirectional SHR (USHR) and bidirectional SHR (BSHR). In a USHR scheme, the normal traffic goes around the ring in one direction. When the network has a failure, the traffic routed to the protection ring is carried in the opposite direction. In BSHR, working traffic flows in both directions. Networks can use the SHR technique against fiber cuts and node failures. The shortcoming of the SHR is that it spends a large amount of spare capacity to get a full protection guarantee. In response to traffic dynamics, network topologies naturally leads to mesh topologies. How to protect a link failure in a connected path becomes an important issue. APS is typically used to handle link failures. There are three APS architectures: 1+1, 1:1 and 1:N. In 1+1 APS, every working path has a protection path. The information signal from the source node is transmitted on the working and protection paths at the same time. In the normal state, the destination node receives the information signals from these two paths, compares them and selects the better one. When a link in either of these two paths fails, only one signal is received. This scheme can achieve a very short failure recovery time but it consumes double bandwidth on every link within a path. In 1:1 APS, every working link also has a protection path but only one path is working. When a link of the working path fails, the source and destination nodes rapidly switch to the protection path. In the 1:N APS scheme, N working paths share a single protection path. When a link in one of the N paths fails, the traffic on it is switched to the protection path. After the failure link is repaired, the traffic on the protection path is switched back to the repaired path to keep the protection path available for other paths.

Next-generation SONET/SDH [3], the successor of SONET/SDH, is proposed to solve deficiencies in existing SONET system. It leverages existing physical layer networking and introduces new technologies such as virtual concatenation (VC), generic framing procedure (GFP), and the link capacity adjustment scheme (LCAS). With these capabilities, both TDM and packet-oriented services are handled efficiently. Virtual Concatenation (VCAT) redefines concatenation to accept variable sizes and non-contiguous payload envelopes. Virtual concatenation provides a much finer granularity in allocating bandwidth, resulting in significant bandwidth savings compared to contiguous concatenation. GFP adopts a single approach for mapping packet data into a byte-synchronous transmission channel s, including SONET/SDH. Link Capacity Adjustment Scheme (LCAS) refers to a set of procedures for dynamically adjusting the size of virtually concatenated channels.

The working paradigm of data services is different form that of voice service. Data services can take gradual reduction as the available bandwidth reduces. It is possible to specify the bandwidth requirements for an application under different network states. VCAT, the new function of NG-SONET, enables forming a high-order, end-to-end large-size path by grouping multiple smaller lower-order paths. The network operators can enhance the bandwidth provision paradigm to adapt this new user requirement. An intelligent path provisioning algorithm based on the VCAT capability can achieve flexible bandwidth usage in NG-SONET networks.

From the users' perspective, due to the increasing reliance of our society on the trustworthy transfer of information across high-speed communication networks, it is important for a network to offer survivability, or at least graceful degradation, in the event of network failure such as link or node failures. In this dissertation, we investigate survivable multi-path provisioning problems for NG-SONET networks. Unlike the availability requirement considered in [14], we first propose the quality-of-survivability concept in bandwidth provisioning to take advantage of data services that are able to stand for bandwidth degradation as the available bandwidth reduces. Quality-of-survivability means a source-and-destination (SD) pair can specify its bandwidth requirements for different network states, i.e. normal (without failure) and failure states. We call the problem MP-QoS for abbreviation.

MP-QoS is different from 100% survivability considered in [42]. In MP-QoS, users can specify different bandwidth requirements for networks under normal (without failure) and failure states. Figure 22 illustrates an example for provisioning survivable circuits for node pair (1,5). As is shown, at least two STS-1 can survive even in the worst case scenarios (i.e., link failure on link (1,3) or link (2, 5)).

We propose a new bandwidth provision scheme that fully utilizes Virtual Concatenation (VC) capabilities of NG-SONET to provision connections over multiple paths, while ensuring its MP-QoS requirements and minimizing required network capacities.

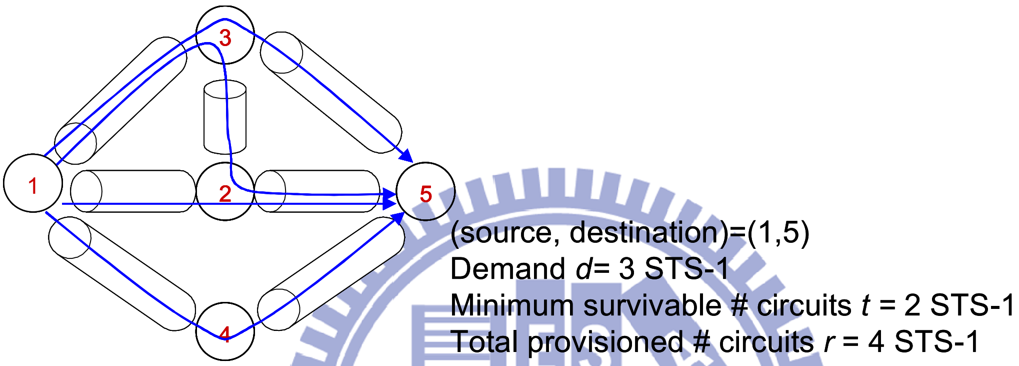


Figure 22 - Illustration of multi path provisioning with survivability

4.1 Problem Formulation

In general, three types of network failures are considered, including (1) link- failure, (2) node-failure, and (3) link-and-node failure. Link-failure usually occurs because of cable cuts, while node-failure occurs because of equipment failure at network nodes. Although node-failure is less frequent, node failure is generally more serious than a link-failure, which can totally isolate a community that is connected through the node from communicating with other places. Thus, compared to other network failures, the link-and-node failure is less common.

A new MP-QoS bandwidth provision scheme is proposed to let users specify different bandwidth requirements for networks under normal and failure states. Given a physical topology, the MP-QoS problem is to determine routing and assign capacity so as to satisfy users' requirements and minimize total bandwidth consumption. A set of ILP models is defined to obtain the optimal solutions of network resource usage under the users' different bandwidth requirements in both normal state and node-failure/ link-failure states.

4.1.1 Single Link-Failure

We assume that the links of the given network are all the same capacity, and each link is the same cost. Before describing the formulations, we first list the notations that are used in the formulation as follows:

Given that d_k is the traffic demand of request k under the network is normal and t_k is the traffic demand of request k under a link failure occurrence. δ_{pl} is the indicator function to indicate whether a path p passes through link l or not, $\delta_{pl} = 1$, if path p uses link l; = 0, otherwise. Three decision variables are introduced: x_p^0 represents the total number of STS-1 circuits flowing on path p while the network is in normal state; x_p^l represents the total number of STS-1 circuits flowing on path p while the network is in link l failure state; y_p represents the total number of STS-1 circuits provisioned on path p.

The objective function is to minimize total network resource consumption. Constraint (4-1) lets the total number of STS-1 circuits be provisioned on all paths be equal to the request d_k while the network is in a normal state. Constraint (4-2) lets the total number of STS-1 circuits provisioned on all paths be equal to the request r_k while network is in a single link fail state. Constraint (4-3) makes sure that no flow is on link l while the link l is failing. Constraint (4-4) is an integer constraint. Constraint (4-5) lets the total number of STS-1 circuits flowing on path p, while link l failure is less than the total number of STS-1 circuits provisioned on path p. Constraint (4-6) is an integer constraint. Constraint (4-7) is the capacity constraint.

MP-QoS Model I:

Input values:

L	:	set of link
C_l	:	the capacity of link l;
K	:	set of connection requests; it can also be viewed as set of SD pairs
P_k	:	candidate path set for SD pair k
d_k	:	traffic demands under the network is without any failure;
t_k	:	the minimum guaranteed survivable bandwidth under any single-link failure;
δ_{pl}	:	indicator function = 1, if path p uses link l; = 0, otherwise;

Decision variables:

x_p^0	:	represents the number of STS-1 circuits flowing on path p while the network is in normal state;
x_p^l	:	represents the number of STS-1 circuits flowing on path p while the network is in link l failure state;
y_p	:	represents the number of STS-1 circuits provisioned on path p;

Problem (P):

$$\min \sum_{l \in L} \sum_{k \in K} \sum_{p \in P_k} y_p \delta_{pl}$$

Subject to

$$\sum_{p \in P_k} x_p^0 = d_k \qquad \forall k \in K \qquad (4\text{-}1)$$

$$\sum_{p \in P_k} x_p^l \geq t_k \qquad \forall k \in K, l \in L \qquad (4\text{-}2)$$

$$x_p^l \delta_{pl} = 0 \qquad \forall p \in P_k, k \in K, l \in L \qquad (4\text{-}3)$$

$$x_p^l \in \text{integer} \qquad \forall p \in P_k, k \in K, l \in L \cup \{0\} \qquad (4\text{-}4)$$

$$x_p^l \leq y_p \qquad \forall p \in P_k, k \in K, l \in L \cup \{0\} \qquad (4\text{-}5)$$

$$y_p \in \text{integer} \qquad \forall p \in P_k, k \in K \qquad (4\text{-}6)$$

$$\sum_{k \in K} \sum_{p \in P_k} y_p \delta_{pl} \leq C_l \qquad \forall l \in L \qquad (4\text{-}7)$$

Because the MP-QoS model I consumes too many variables, a better approach would be shown as follows.

MP-QoS Model II :

Input values:

L	:	set of links;
N	:	set of SONET cross-connect nodes ;
C_l	:	the capacity of link l; (same for all links)
L_n^{in}	:	the incoming links incident to node n ;
L_n^{out}	:	the outgoing links incident to node n ;
K	:	set of each SD pair;
d_k	:	the required traffic demands under the network is without any failure;
t_k	:	the minimum guaranteed survivable bandwidth for any single-link failure

Decision variables:

x_{kl}	:	number of STS-1 circuits flowing on link l for SD pair k;
r_k	:	the total number of STS-1 circuits provisioned for SD pair k.
$\eta_{kn} := r_k$	,	if node n is the source node of SD pair k;
$= -r_k$	,	if node n is the destination node of SD pair k;
$= 0$	,	otherwise;

Problem (P):

$$\min \sum_{k \in K} \sum_{l \in L} x_{kl}$$

subject to:

$$\sum_{l \in L_n^{out}} x_{kl} - \sum_{l \in L_n^{in}} x_{kl} = \eta_{kl} \qquad \forall k \in K, n \in N \qquad (4\text{-}8)$$

$$\sum_{k \in K} x_{kl} \le C_l \qquad \forall l \in L \qquad (4\text{-}9)$$

$$d_k \le r_k \qquad \forall k \in K \qquad (4\text{-}10)$$

$$t_k \le r_k - x_{kl} \qquad \forall k \in K, l \in L \qquad (4\text{-}11)$$

$$x_{kl} \in integer \qquad \forall k \in K, l \in L \qquad (4\text{-}12)$$

Let N denote the set of cross-connect nodes; L the link set; C_l the capacity of link l; L_n^{in} and L_n^{out} are the incoming and outgoing links incident to node n. For each SD pair $k \in K$, d_k is the required traffic demands under the network without any failure while t_k is the minimum guaranteed survivable bandwidth for any single-link failure. By letting t_k be smaller than d_k, it allows a downgrade in the communication capacity under the network with a failure. For highly important applications, d_k should be set to be equal to t_k to guarantee the volume of communication under any single-link failure. Indicator function $\eta_{kn} = r_k$, if node n is the source node of request k; $= -r_k$, if node n is the destination of request k; and =0, otherwise. For output variables, decision variable x_{kl} represents the number of STS-1 circuits flowing on link l for SD pair k; and r_k is the total number of STS-1 circuits provisioned for SD pair k.

The objective function is to minimize total network resource consumption. The flow conservation law is enforced in Constraint (4-8). For each node n, if n is neither the source node nor the destination node of SD pair k, the amount of total incoming flow must be equal to the amount of total outgoing flow. If n is the source node (destination node) of SD pair k, the difference between total incoming flow and outgoing flow is r_k ($-r_k$). Here r_k is the total SONET channels provisioned for SD pair k. Constraint (4-9) is the capacity constraint. It requires the total provisioned channels on link l to be limited within the link capacity. Constraint (4-10) requires that the number of provisioned channels r_k must be equal to or larger than the demand d_k. Constraint (4-11) is the constraint to guarantee network survivability. It requires that at least t_k channels are survivable under any single-link failure state.

4.1.2 Single-Node Failure Model

By graph transformation, we can apply the link protection technique shown in the previous subsection to obtain the node failure protection. An example is depicted in Figure 23. For simplicity, only one node is transformed in the example. First we replace a node n with two artificial nodes (v_{in}, v_{out}), one for input and the other for output. An artificial link e is used to connect these two artificial nodes. All incoming and outgoing links are then connected to the artificial input and output node, respectively. By regarding a node failure as its artificial link e failure, the node failure can be viewed exactly as link failure. For each SD pair, the new source node is the artificial output node of the original source and the new destination node is the artificial input node of the original destination. After graph transformation, the MP-QoS Model for single-node failure can be described as follows.

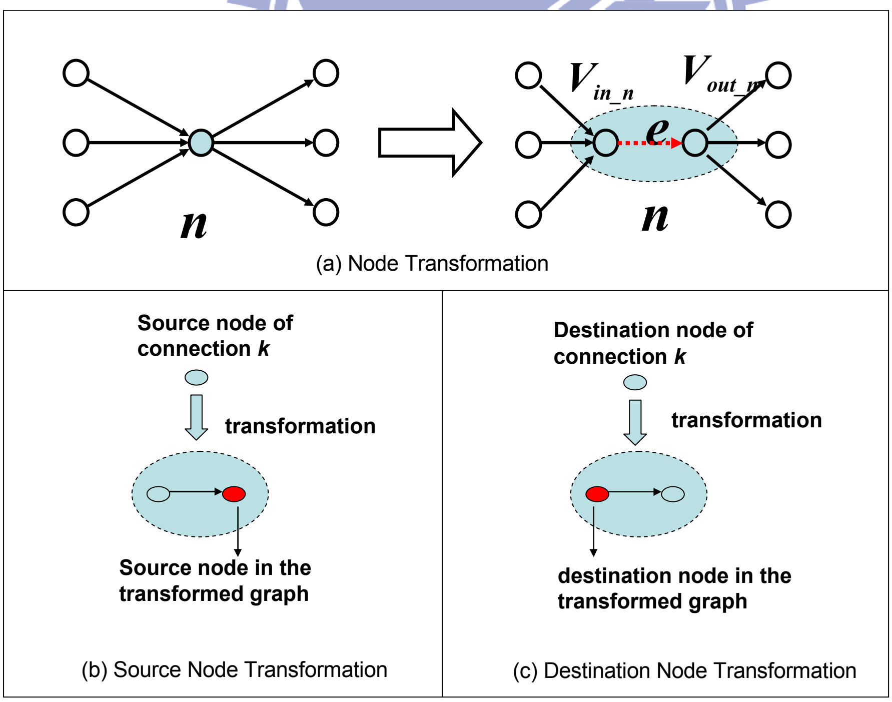


Figure 23 - Illustrations of graph transformation

MP-QoS Model for single-node failure:

Input values:

L	:	set of links ;
E	:	set of artificial edges;
N	:	set of sonnet cross connect nodes;
V	:	set of artificial vertices;
C_l	:	the capacity of link l; (same for all links)
L_n^{in}	:	the incoming links incident to node n;
L_n^{out}	:	the outgoing links incident to node n;
E_v^{in}	:	the incoming links incident to vertex v ;
E_v^{out}	:	the outgoing links incident to vertex v;
K	:	set of each SD pair;
d_k	:	the required traffic demands under the network without any failure;
t_k	:	The minimum guaranteed survivable bandwidth for any single-node failure;

Decision variables:

x_{kl}	:	number of STS-1 circuits flowing on link l for SD pair k;
r_k	:	the total number of STS-1 circuits provisioned for SD pair k.
$\mu_{kn} := r_k$	,	if node n is the source node of SD pair k;
$= 0$	,	otherwise ;
$\rho_{kn} := -r_k$	,	if node n is the destination node of SD pair k;
$= 0$	,	otherwise ;

Problem (P):

$$\min \sum_{k \in K} \sum_{l \in L} x_{kl}$$

subject to:

$$\sum_{a \in L_v^{out} \cup E_v^{out}} x_{ka} - \sum_{a \in L_v^{in} \cup E_v^{in}} x_{ka} = \mu_{kv} \qquad \forall k \in K, v \in V \qquad (4\text{-}13)$$

$$\sum_{a \in L_v^{out} \cup E_v^{out}} x_{ka} - \sum_{a \in L_v^{in} \cup E_v^{in}} x_{ka} = \rho_{kv} \qquad \forall k \in K, v \in V \qquad (4\text{-}14)$$

$$\sum_{k \in K} x_{kl} \le C_l \qquad \forall l \in L \qquad (4\text{-}15)$$

$$d_k \le r_k \qquad \forall k \in K \qquad (4\text{-}16)$$

$$t_k \le r_k - x_{ke} \qquad \forall k \in K, e \in E \qquad (4\text{-}17)$$

$$x_{ka} \in integer \qquad \forall k \in K, a \in L \cup E \qquad (4\text{-}18)$$

The objective function is still to minimize total network resource consumption. The flow conservation law is enforced in Constraint (4-13) and Constraint (4-14). For each artificial node v, if v is neither the source node nor the destination node of SD pair k, the amount of total incoming flow must be equal to the amount of total outgoing flow. If v is the source node (destination node) of SD pair k, the difference between total incoming flow and outgoing flow is r_k ($-r_k$). r_k is the total SONET channels provisioned for SD pair k. Constraint (4-15) is the capacity constraint. It requires the total provisioned channels on link l to be limited within link capacity. Constraint (4-16) requires that the number of provisioned channels r_k must be equal to or larger than the demand d_k. Constraint (4-17) is the constraint to guarantee network survivability. It requires that at least t_k channels are survivable under any single-link failure state.

4.1.3 Single Node or Link Failure Model

This model of single node or link failure is similar to the MP-QoS Model for single-node failure except that link failures here are composed of link failures and artificial link failures. Constraint (4-21) & (4-22) show the new constraints.

Problem (P):

$$\min \sum_{k \in K} \sum_{l \in L} x_{kl}$$

subject to:

$$\sum_{a \in L_v^{out} \cup E_v^{out}} x_{ka} - \sum_{a \in L_v^{in} \cup E_v^{in}} x_{ka} = \mu_{kv} \qquad \forall k \in K, v \in V \qquad (4\text{-}19)$$

$$\sum_{a \in L_v^{out} \cup E_v^{out}} x_{ka} - \sum_{a \in L_v^{in} \cup E_v^{in}} x_{ka} = \rho_{kv} \qquad \forall k \in K, v \in V \qquad (4\text{-}20)$$

$$\sum_{k \in K} x_{kl} \le C_l \qquad \forall l \in L \qquad (4\text{-}21)$$

$$d_k \le r_k \qquad \forall k \in K \qquad (4\text{-}22)$$

$$t_k \le r_k - x_{ka} \qquad \forall k \in K, a \in L \cup E \qquad (4\text{-}23)$$

$$x_{ka} \in integer \qquad \forall k \in K, a \in L \cup E \qquad (4\text{-}24)$$

4.2 Simulations and Performance Comparisons

We conduct two sets of simulations, one runs on a well-know USA benchmark network and the other runs on some randomly generated networks. Figure 24 shows the USA network that contains 24 nodes and 86 OC-48 bi-directional links. In order to evaluate the benefit of applying multi-path provisioning, simulations for traditional SONET 1+1 protection is also carried out in this study. For the 1+1 scheme, two link disjoint paths are provisioned for each SD pair. Without VCAT capability in conventional SONET network, whole traffic demand must go through the same route.

First, shown in Figure 24, we evaluate various demands and survivable requirements under different numbers of SD pairs in the USA network. In these experiments, the working demand is fixed to 3 STS-1 for all cases. In Figure 25 and Figure 26, the results of the experiments relating to the single-link failure protection scheme are shown. In Figure 25(a) for the NG-SONET network, the less stringent survivable requirement (i.e., a smaller value of t), means less bandwidth is required. In addition, the increase of total bandwidth is nonlinear for any two successive values of t. For the *(d,t) = (3,1)* case, it consumes up to 16% more resources than the *(d,t) = (3,0)* case to provide 1 STS-1 link-disjointed protection path when a link failure occurs in the network. For the *(d,t) = (3,2)* case, it consumes up to 56% more resources than the *(d,t) = (3,0)* case to provide 2 link-disjointed STS-1 protection paths when a link failure occurs in the network. The *(d,t) = (3,3)* case and 1+1 case consume more than 2 times resources than the *(d,t) = (3,0)* case to provide the full protection as there may not exist 2 link-disjointed shortest paths for a SD-pair. Due to the high traffic generated, there is no feasible solution for the *(d,t)=(3,3)* as the number of SD pairs becomes larger than 150.

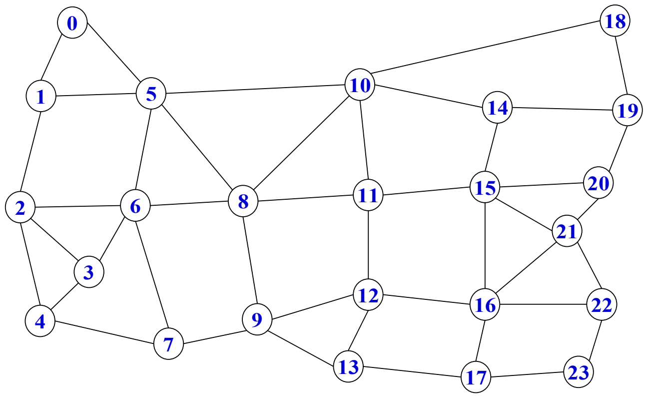


Figure 24 - The USA Network (24 nodes and 86 OC-48 bi-directional links)

To further compare the performance between multi-path provisioning in NG-SONET and in SONET, the detailed results are shown in Figure 25(b). We compare the *(3,3)* scheme with SONET 1+1. For SONET 1+1 protection, the required demand for each SD pair is set to three STS-1s. Both schemes guarantee three survivable STS-1 circuits under any single-link failure. By taking advantage of diverse multi-path routing, NG-SONET outperforms 1+1 in all cases. The NG-SONET scheme only uses 95% to 96% resources related to that are used in the 1+1 scheme. In particular, due to high bandwidth consumption, as the number of SD pairs goes over 120, there is no feasible solution for the 1+1 scheme.

In Figure 26, we study the impact of network connectivity on bandwidth consumption. The connectivity degree of a network is represented by the ratio of the number of links to the number of nodes. We make performance comparisons on several randomly generated network topologies. The total number of SD pairs is 150 and each network is equipped with 100 OC-48 links. As shown in Figure 26(a), by increasing the number of nodes from 20, 25, 30, to

40, the network connectivity changes from dense to sparse. A dense network consumes less capacity than a sparse network for all test schemes. The results reveal that node degree strongly influences the resource consumption in a survivable network. Again, for NG-SONET, the larger the *t* value, the larger the bandwidth required. SONET 1+1 protection needs the most capacity. For the *(d,t)=(3,2)* case, there is no feasible solution for networks with 40 nodes. They are infeasible for the *(d,t)=(3,2)* and SONET 1+1 protection under networks with nodes larger than 30 and 40 respectively.

In Figure 26(b), we fixed the number of nodes to be 30 and vary the number of links. The larger the number of links the denser the network is. Again, the denser network holds better performance. In this experiment, there are no feasible solutions for (3,2), (3,3), and 1+1 schemes in networks with 90 links. NG-SONET still outperforms SONET 1+1 protection schemes in all cases.

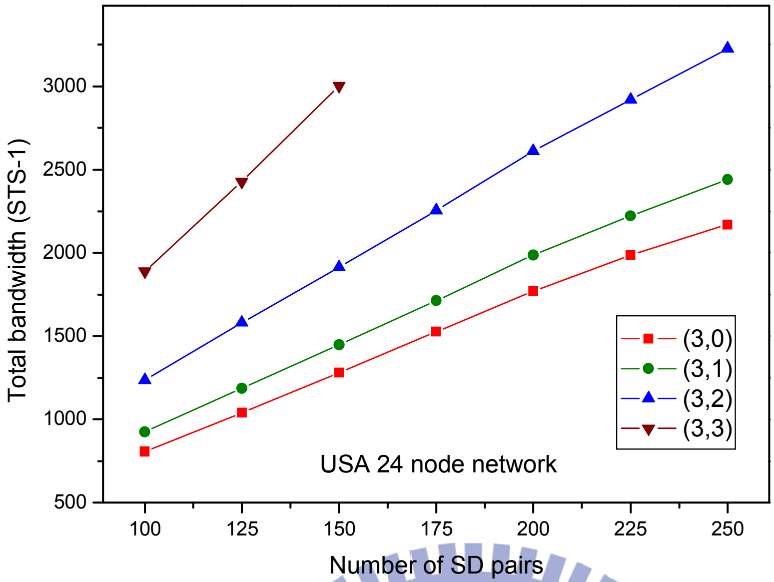
In Figure 27 and Figure 28, the results regarding the single-node failure protection are shown. The trends of the resource consumption are similar with those in single link protection schemes. In Figure 27(a) for the NG-SONET network, with the same survivable requirement (i.e., a same value of t), less bandwidth is required than it is required in the single-link failure protection scheme, but the difference between them is slight. In the node failure protection scheme, only one link is used as a SD-pair has a direct link and this reduces the resources consumed. For the *(d,t) = (3,1)* case, it consumes up to 12% more resources than the *(d,t) = (3,0)* case to provide 1 STS-1 node-disjointed protection path when a node failure occurs in the network. For the *(d,t) = (3,2)* case, it consumes up to 52% more resources than the *(d,t) = (3,0)* case to provide 2 STS-1 node-disjointed protection paths when a node failure occurs in the network. The *(d,t) = (3,3)* case and 1+1 case consume more than 2 times resources than the *(d,t) = (3,0)* case to provide the full protection as 2 node-disjointed shortest paths may not exist for a SD-pair.

The performance comparison of the NG-SONET (3,3) and SONET 1+1 protection within single node protection scheme is shown in Figure 27(b). Both schemes guarantee three survivable STS-1 circuits under any single-node failure. The NG-SONET scheme only uses 91% to 95% resources of that are used in 1+1 protection by taking advantage of diverse multi-path routing.

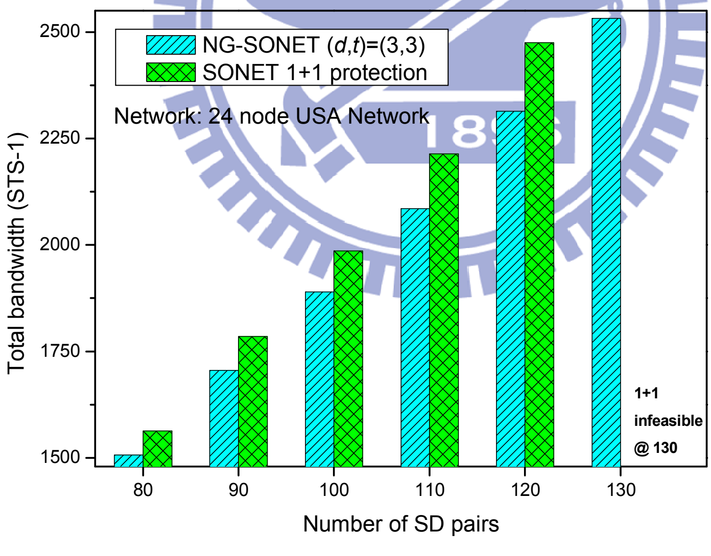
In Figure 28, we study the impact of network connectivity on bandwidth consumption. The parameters are the same as those in Figure 26. As shown in Figure 28(a), the network connectivity changes from dense to sparse by increasing the number of nodes from 20, 25, 30, to 40. We find similar movements as shown in Figure 26, that a dense network consumes less capacity than a sparse network for all test schemes. For the *(d,t)=(3,1)* case, the 30-node network needs 60% more bandwidth than the 20-node network to provide 1 STS node-disjointed path. For the *(d,t)=(3,2)* case, the 30-node network consumes 80% more bandwidth than the 20-node network to provide the 2-STS node-disjointed paths. The low node degree is unfavorable to the resource consumption in a survivable network. In Figure 28(b), we vary the number of links if the number is under 30. Again, the denser network has better performance.

The Figure 29 and Figure 30 illustrate the results regarding the networks that provide both the single-link failure and single-node failure protections at the same time. Since the protection path must satisfy link-disjoint and node-disjoint simultaneously, it is possible that the protection path will be a longer path than in the node/link failure protection scheme and consume more bandwidth. The results shown in Figure 29 and Figure 30 conform to this inference. The link-node-failure-protection scheme increases slightly more than the other schemes. The bandwidth consumption is analogous to that in the link failure protection scheme.

Through the MP-QoS multi-path provisioning scheme for NG-SONET networks, users can specify their bandwidth requirements for different network states. Numerical results reveal that the proposed scheme outperforms legacy SONET protection in all experimental cases. The other observations are: Connection requests with less stringent guaranteeing survivable requirement consume less network capacity. The required bandwidths among different *(d, t)* values increase nonlinearly. It gives a guideline for pricing connections with different class of quality of survivability requirements. Furthermore, a dense network consumes less network capacity than a sparse network. This indicates connectivity degrees should be carefully considered in designing a survivable network.

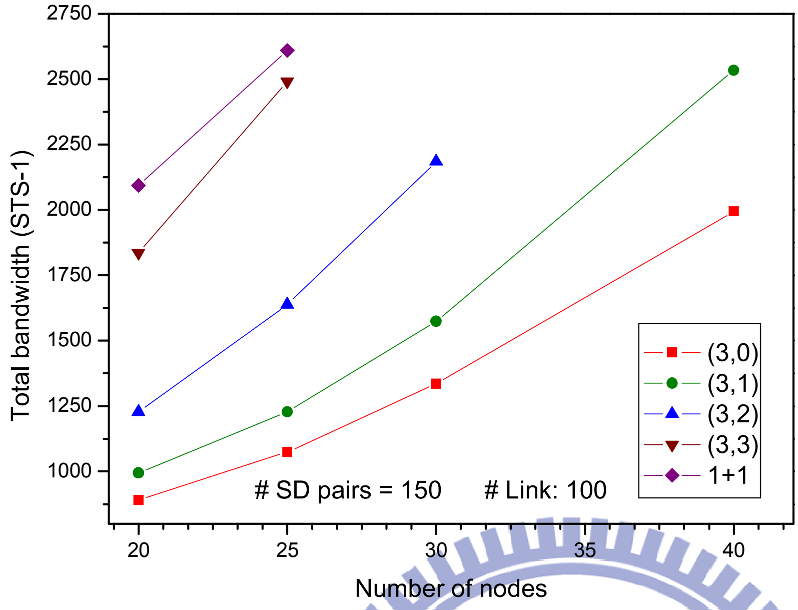


(a) Required bandwidth under different SD pair numbers

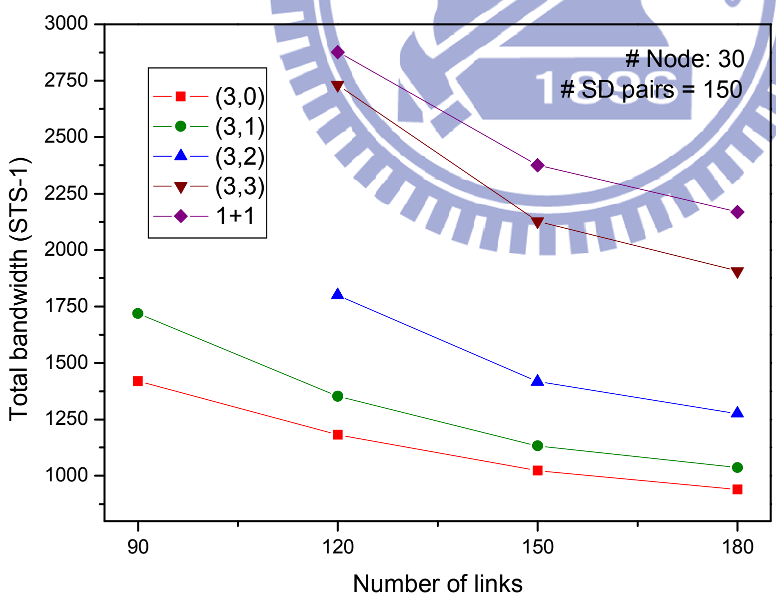


(b) Performance comparison: NG SONET and SONET 1+1 protection

Figure 25 - Simulation results of MP-QoS (on USA network, single-link failure protection)

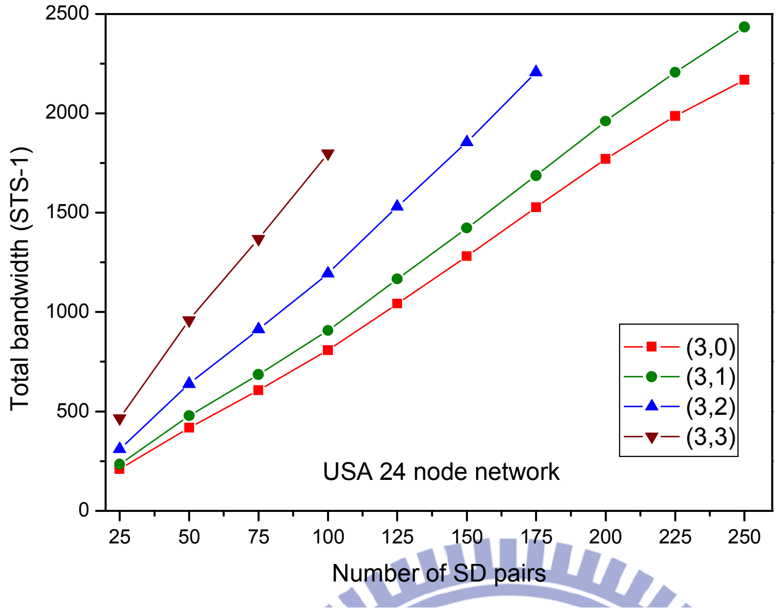


(a) Performance comparison under different node numbers

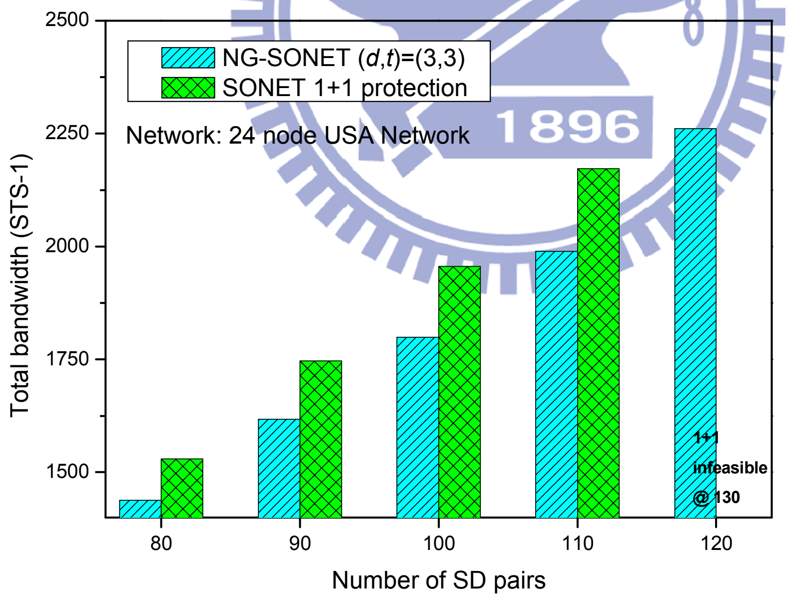


(b) Performance comparison under different link numbers

Figure 26 - Simulation results of MP-QoS (under the network with various connection degrees, single-link failure protection)

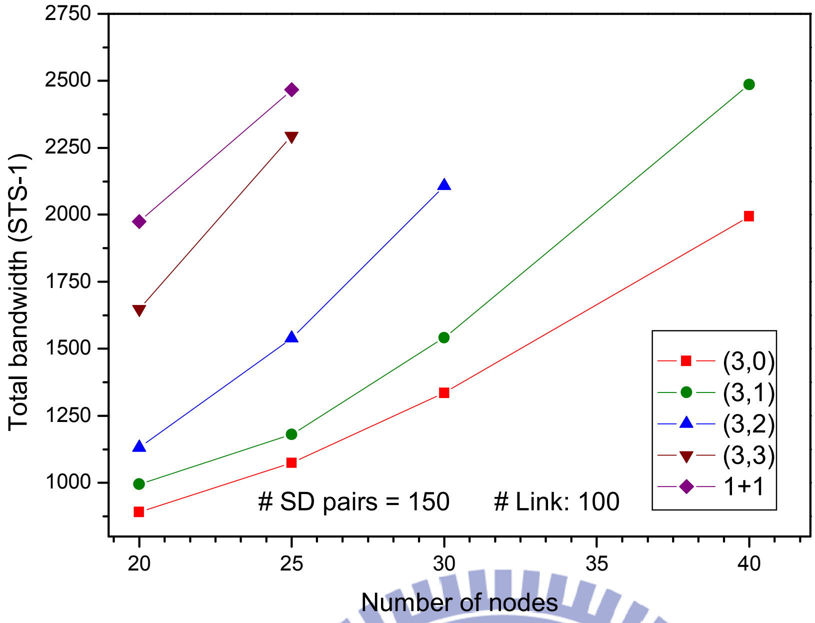


(a) Required bandwidth under different SD pair numbers

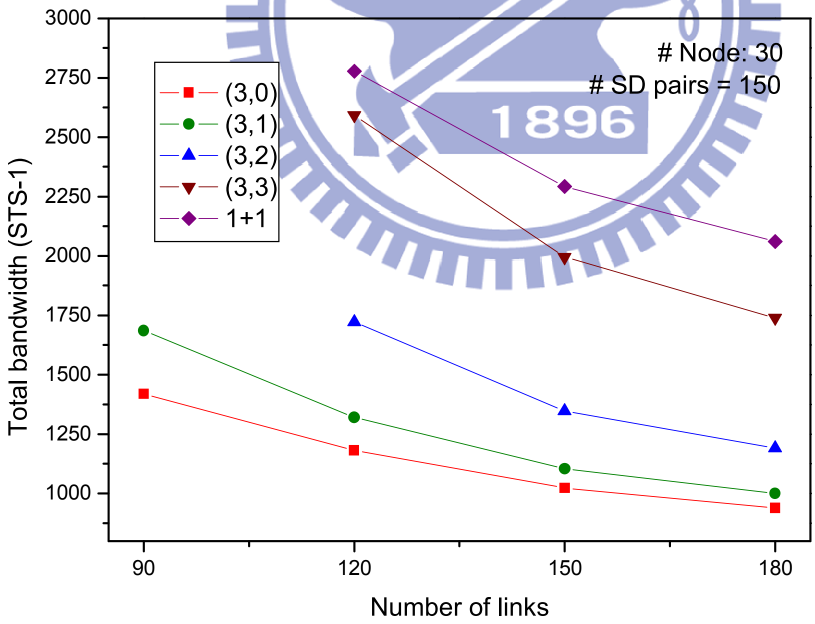


(b) Performance comparison: NG-SONET and SONET 1+1 protection

Figure 27 - Simulation results of MP-QoS (on USA network, single-node failure protection)

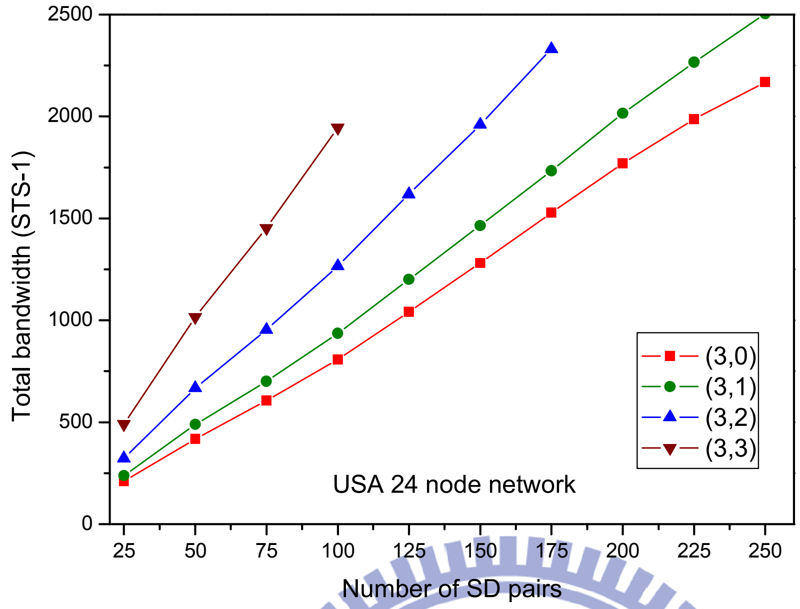


(a) Performance comparison under different node numbers

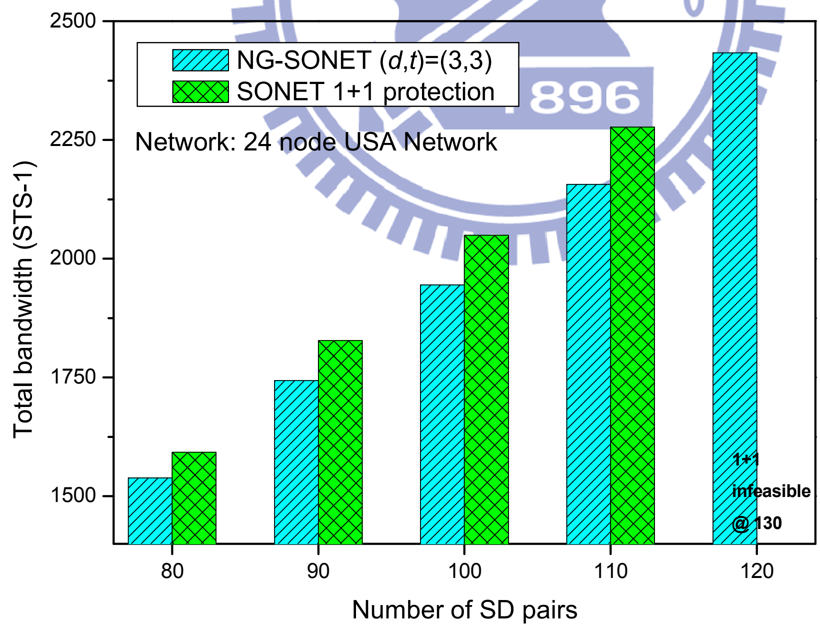


(b) Performance comparison under different link numbers

Figure 28 - Simulation results of MP-QoS (under the network with various connection degrees, single-node failure protection)

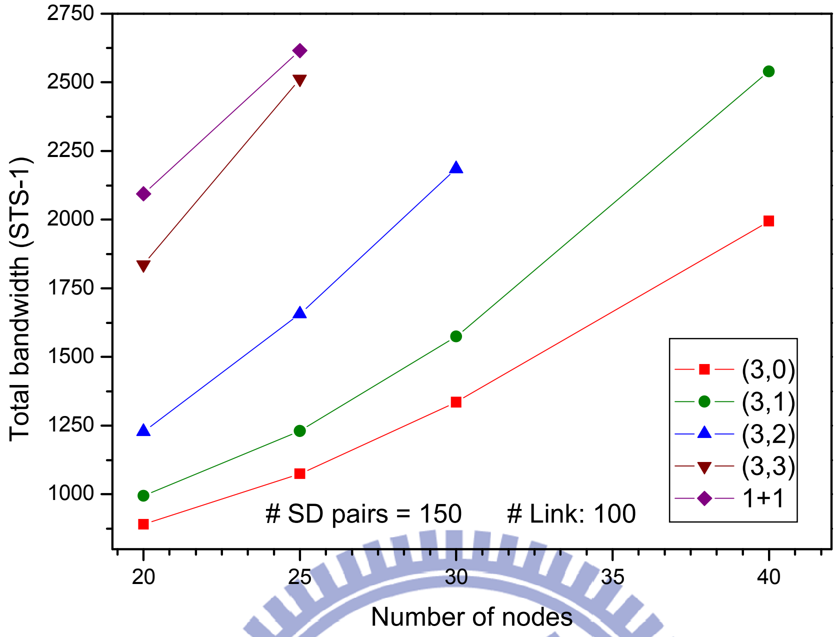


(a) Required bandwidth under different number of SD pairs

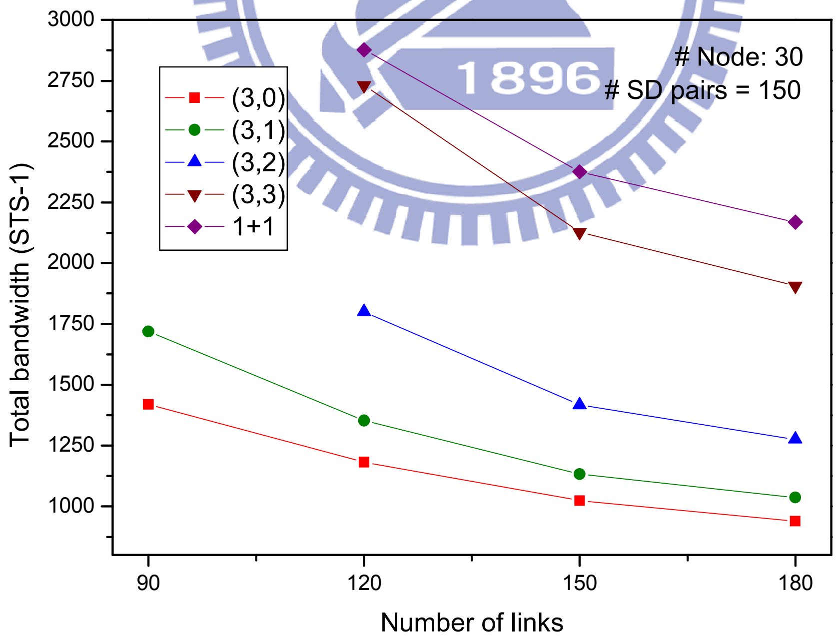


(b) Performance comparison: NG-SONET and SONET 1+1 protection

Figure 29 - Simulation results of MP-QoS (on USA network, single link/node failure protection)



(a) Performance comparison under different node numbers



(b) Performance comparison under different link numbers

Figure 30 - Simulation results of MP-QoS (under the network with various connection degrees, single link or node failure protection)

Chapter 5. Optimal Routing and Bandwidth Provisioning for Survivable Multicast Communications Using Network Coding

Nowadays networked video streaming applications like IPTV, video on demand, video conferencing, and online games are growing fast. Efficient bandwidth-saving streaming mechanisms between application servers and customers is an attractive research topic. Multicast and peer-to-peer communications are the two most important technologies. Multicast is a communication paradigm between a single sender and multiple receivers on a network. To minimize bandwidth usage, the multicast algorithm collects all Source-Destination paths of this application and forms a tree-shaped multicast connection. In Figure 31, we depict a generic model for delivering multicast videos through a packet network. In networked video streaming applications like IPTV services, a multicast connection might consist of tens to hundreds of video programs, and therefore a network failure would impact any consumer viewing these types of programs as the service quality would suddenly drop.

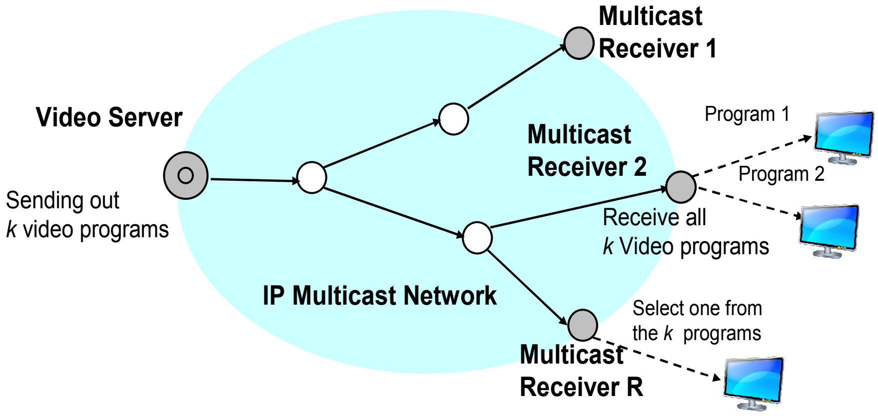


Figure 31 – An example of multicast network architecture

Conventional network protection approaches for multicast employ extra network resources and pre-computed backup paths to bypass the failure link or node. It needs to reserve spare bandwidth and backup paths via complex computation. We found that network coding can enhance network survivability [13]-[16] by electing some intermediate nodes performing packet encoding.

Network coding is an elegant and novel technique to improve the network throughput and performance. Traditionally, information flow was treated like fluid through pipes, and independent information flows were kept separate. This rule can be changed to allow intermediate nodes to not only forward but also process the incoming independent information flows. At the same time, there are ways to combine and later extract independent information. Then data streams that are independently produced and consumed do not necessarily need to be kept separate when they are transported throughout the network. Combining independent data streams allows a better tailoring of the information flow to the network environment and accommodates the demands of specific traffic patterns. Since the computational processing has become cheaper according to Moore's law, network coding utilizes cheap computational power to dramatically increase network throughput.

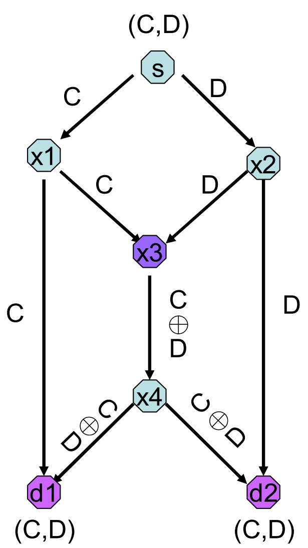


Figure 32–An example of network coding

In Figure 32, we depict an example of how network coding works. In the butterfly network, each edge can carry only single data. Source node s wants to send two pieces of data, those are C and D, to two destination nodes (d1 and d2). If we only used routing, then the central line (x3 → x4) would be able to carry C or D, but not both. If C is sent through the central line; d1 would receive C twice and can not receive D at all. If D is sent through the central line, a similar problem is posed for d2. A conclusion can be gotten that routing is insufficient because no routing scheme can transmit both C and D at the same time to both d1 and d2. By sending C⊕D through the center line, d1 receives C and C⊕D, and can find D by exclusive-or the two values. Node d2 receives D and C⊕D, and can find C by the same operation.

Network coding offers benefits along diverse dimensions of communication networks, such as throughput, wireless resources, security, and resilience to link failures. Some network coding research is focused on how to enhance throughput and provide network survivability [13]-[16] by selecting some intermediary nodes to perform packet encoding. Several network-coding-based protection schemes for point-to-point communications have been proposed in the literature, including single-link failures [43], [44], multiple link failures [45] and single-node failures [46]. Researchers conclude that both bandwidth consumption and service recovery time are reduced by applying network coding in point-to-point communications.

For a multicast communication between one source node and N receiver nodes, an algebraic network coding approach is shown in [47] for network protection. They assert the network coding problem is solvable if, and only if, the Min-Cut Max-Flow bound is satisfied for all source-destination connections. Moreover, they have proven that there exists a solution for the network coding to protect a set of failure patterns F in a finite field F_{2^m} with $m \le \lceil \log_2 |F| NR + 1 \rceil$, where the $|F|$ is the number of failure patterns and R is the information generating rate at the source. The results imply that network survivability is guaranteed if we can determine a routing and bandwidth provision that can satisfy Min-Cut Max-Flow bound for the pre-defined failure patterns.

Based on the point mentioned above, a multicast network protection problem is formulated in a mixed integer linear programming form to determine the optimal routing and bandwidth provision with the Min-Cut Max-Flow bound constraints versus any single link/node failure. Minimizing the total cost of provisioning bandwidth in the network is our objective. Through the computed optimal routing and bandwidth provision using algebraic network coding from [47], a minimum cost survivable network against any single link/node failure can be guaranteed.

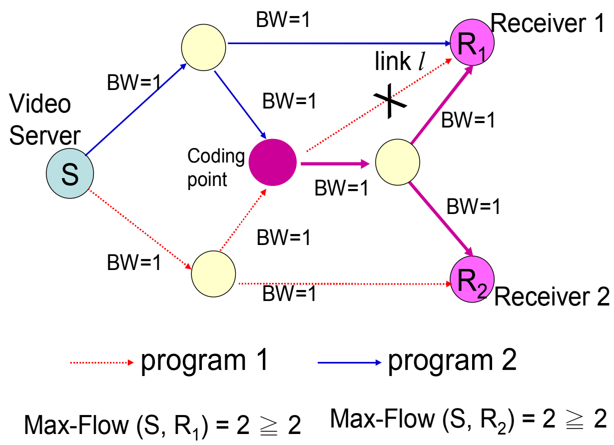
5.1 Multicast Protection Schemes

We propose four protection schemes for survivable multicast connections. The first two schemes are network-coding-based protection schemes that can provide multicast networks against any single-link failure and single-node failure. These two schemes take advantage of network coding to reduce the total bandwidth consumption, thereby lessening total cost consumption. The remaining two schemes are tree-based protection schemes that can prevent any single-link failure.

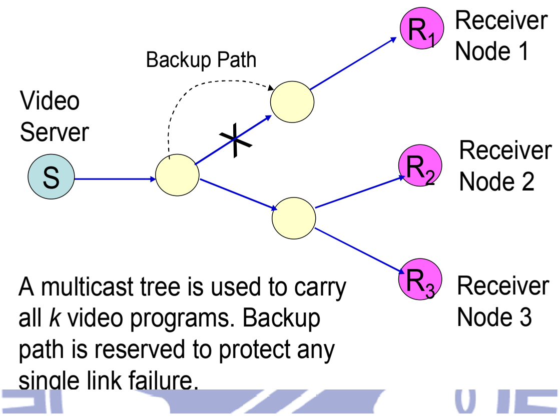
5.1.1 Network Coding for Single-Link Failure Protection (NCL)

According to [47], a network coding problem is solvable if the Min-Cut Max-Flow bound is satisfied for all multicast source-receiver (SR) connections. To explain this, we give a multicast session in Figure 33(a) where each link has one unit bandwidth (i.e., BW=1). We assume that each receiver node needs to receive two programs (i.e., the Min-Cut Max-Flow bound for all SR connections (= {(S,R1),(S,R2)}) is 2). After link l fails, the Max-Flow for (S,R1) is 2 and for (S,R2) is also 2. All SR connections are still satisfying the Min-Cut

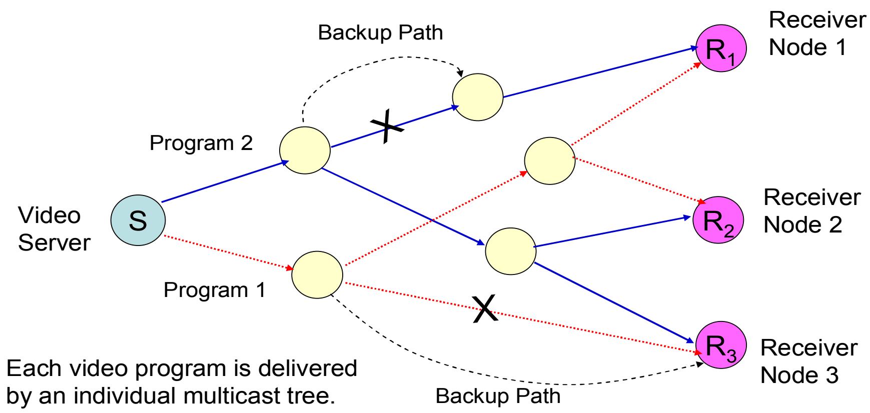
Max-Flow bound (i.e., Max-Flow (S,R1) = 2≥2 and Max-Flow (S,R2) = 2≥2). Hence, the network coding problem has a solution for this link l failure case in Figure 33(a). Through this concept, we propose a combinatorial optimization problem to determine routing and provisioned bandwidth with minimizing network cost at the same time network coding exists for any single link/node failure. The detailed formulation is shown in the next section.



(a) Network coding-based link protection scheme (NCL)



(b) Bundle tree-based link protection scheme (BTL)



(c) Individual tree-based link protection scheme (ITL)

Figure 33 - Protection schemes with/without network coding

5.1.2 Network Coding for Node Failure Protection (NCN)

By graph transformation, we can apply the link protection technique shown in the previous subsection to obtain the node failure protection. We depict an example in Figure 34. For simplicity, only one node is transformed in the example. First we replace the node *v* with two artificial nodes (v_{in}, v_{out}), one for input and the other for output. An artificial link *e* is used to connect these two artificial nodes. All incoming and outgoing links are then connected to the artificial input and output node, respectively. By regarding a node failure as its artificial link *e* failure, the node failure can be viewed exactly as link failure. Due to the space limitation, we omit the required protocol in the network to inform the node failure event and the signal for activating the network-coding-based protection.

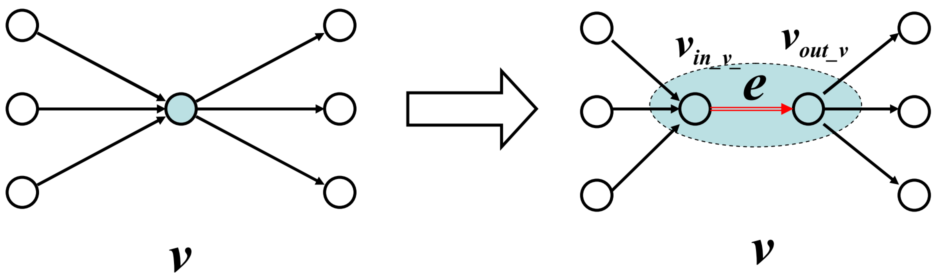


Figure 34 - Illustrations of graph transformation

5.1.3 Bundle Tree-Based Link Protection Scheme (BTL)

This scheme is similar to the link restoration scheme for multicast communication [48]. Every working link in the multicast tree is constructed with a backup route to it. Different from the work in [48], the construction of the working tree and backup path are jointly determined to further reduce total bandwidth consumption. The bundle tree-based link protection scheme (BTL) is depicted as shown in Figure 33(b). In BTL, only one multicast tree is used to carry all video programs. Backup paths are reserved to protect any single-link failure while the total cost of provisioned bandwidth is minimized.

5.1.4 Individual Tree-Based Link Protection Scheme (ITL)

The second tree-based scheme is called the Individual tree-based link protection scheme (ITL), which is shown in Figure 33(c). In ITL, each video program is delivered by an individual multicast tree. Therefore, a network operator has to determine N routing trees to carry N multicast video streams individually and pre-determines backup paths to protect any potential single-link failure. Models for the two tree-based schemes are also presented in the next section.

5.2 Optimization Models

In this section, we formulate the NCL, BTL and ITL problems as combinatorial optimization problems and determine the proposed routes and provisioned bandwidth of a multicast network to protect any single-link failure. These formulations mainly focus on the minimization of total network costs, in which the constraints are required to satisfy the demand requirement, 100% survivability constraint and physical capacity limitation. The output of the problems includes the routing paths and required bandwidth on each link.

A multicast network is modeled as a graph *G(N,L)*, where *N* denotes the set of network nodes, and *L* represents the set of physical links. The link capacity C_l is the available bandwidth on link *l*. We assume that each video program consumes the same bandwidth.

Before describing the formulations, we first list the notations that are common for all models.

Common Notations:

α_l	:	cost for one unit bandwidth on link 1 (i.e. bandwidth for one video program);
R	:	set of receiver nodes;
k	:	the number of video programs in the system;
d	:	bandwidth requirement to support one video program;

L_n^{in} : set of input links for node n;

L_n^{out} : set of output links for node n;

σ_{nr} $\begin{cases} =1 & \text{, if node n is the server node} \\ =-1 & \text{, if node } n = r \\ =0 & \text{, otherwise} \end{cases}$

δ_{nl} $\begin{cases} =1 & \text{, if node } n \text{ is the tail node of link } l \text{ (i.e. } l \text{ leaves } n\text{)} \\ =-1 & \text{, if node } n \text{ is the head node of link } l \text{ (i.e. } l \text{ enters } n\text{)} \\ =0 & \text{, otherwise} \end{cases}$

In addition to the above notations, there are some other notations used only for separate models. They are listed right before the corresponding one.

5.2.1 Single Link/Node Protection Using Network Coding

Our model is developed based on the network coding theorem shown in [47]. It guarantees that a network can survive after any single-link failure if the remaining capacity is still large enough to provide the required bandwidth after removing the failed link. To be precise, the maximum flow bound for each receiver must be greater than or equal to the demanded bandwidth in both normal and any single-link failure state. We summarize the dedicated notations used in the formulation as follows:

Decision variables:

x_l : the number of video programs carried on link l;

y_{rl} : the number of video programs carried on link l for destination node r;

z_r : the total flow amount leaving from the server node to destination node r;

Problem (NCL):

$$\min \sum_{l \in L} \alpha_l x_l \qquad \forall r \in R, n \in N$$

subject to:

$$\sum_{l \in L_n^{out}} y_{rl} - \sum_{l \in L_n^{in}} y_{rl} = \sigma_{nr} z_r \qquad \forall r \in R, n \in N \qquad (5\text{-}1)$$

$$k \le z_r - y_{rl} \qquad \forall r \in R, l \in L \qquad (5\text{-}2)$$

$$y_{rl} \le x_l \qquad \forall r \in R, l \in L \qquad (5\text{-}3)$$

$$dx_l \le C_l \qquad \forall l \in L \qquad (5\text{-}4)$$

$$y_{rl} \in \text{integer} \qquad \forall r \in R, l \in L \qquad (5\text{-}5)$$

$$z_r \in \text{integer} \qquad \forall r \in R \qquad (5\text{-}6)$$

The objective function is to minimize the total cost of provisioned bandwidth on each link. Constraint (5-1) enforces the flow conservation law in each node for flows coming from the video server to each destination node. To be more specific, the number of video programs that can leave the server node and enter destination node r is z_r. For an intermediate node that is neither the source node nor the destination node, its input flow should be equal to its output flow. Constraint (5-2) is a survivability constraint. It requires that even if link l becomes failed, the network can still provide enough bandwidth for each OD-pair to obtain bandwidth for more than k video programs. Constraint (5-3) determines decision variable x which is the bandwidth to be provisioned on link l. Constraint (5-4) is the capacity constraint.

For the single-node failure protection, we adopt the graph transformation technique as shown in Figure 34 and then apply the above model on the transformed graph. By doing so, the above model can be used to solve the single node protection problem directly.

5.2.2 Bundle Multicast Tree-based Link Protection Model

BTL requires whole traffic routing on a single working tree. It uses the link protection scheme to setup a backup path. This scheme does not apply network coding on it. We summarize the notations used in the formulation as follows:

Decision variables:

t_l : = 1, if link l is used by a working tree; = 0, otherwise;

$\bar{t}_l$: = 1, if link l is used by a backup path; = 0, otherwise;

x_{le} : = 1, if link e is on the backup path to protect link l; = 0, otherwise;

y_{rl} : = 1, if a working routing path for destination r goes through link l; = 0,otherwise;

Problem (BTL):

$$\min \sum_{l \in L} \alpha_l (kt_l + k\bar{t}_l)$$

subject to

$\sum_{l \in L_n^{out}} y_{rl} - \sum_{l \in L_n^{in}} y_{rl} = \sigma_{nr}$	$\forall r \in R, n \in N$	(5-7)
$y_{rl} \le t_l$	$\forall r \in R, l \in L$	(5-8)
$\sum_{l \in L_n^{in}} t_l \le 1$	$\forall n \in N, l \in L$	(5-9)
$\sum_{e \in L_n^{out}} x_{le} - \sum_{e \in L_n^{in}} x_{le} = \delta_{nl} t_l$	$\forall l \in L, n \in N$	(5-10)
$x_{ee} = 0$	$\forall e \in L$	(5-11)
$x_{le} \le \bar{t}_e$	$\forall l \in L, e \in L$	(5-12)
$d(kt_l + k\bar{t}_l) \le C_l$	$\forall l \in L$	(5-13)
$t_l = 0 \text{ or } 1$	$\forall l \in L$	(5-14)
$\bar{t}_l = 0 \text{ or } 1$	$\forall l \in L$	(5-15)
$x_{le} = 0 \text{ or } 1$	$\forall l \in L, e \in L$	(5-16)
$y_{rl} = 0 \text{ or } 1$	$\forall r \in R, l \in L$	(5-17)

Constraint (5-7) enforces the flow conservation law in each node for flows from the video server to each destination. Constraints (5-8, 5-9) determine decision variable t_l. Those links with t_l =1 form a working tree. Constraint (5-10) is the flow conservation law for setting up the backup path to protect link l. Constraint (5-11) is a disjoint constraint. It requires the backup path not to use the protected link. Constraint (5-12) determines decision variable $\bar{t}_l$ which is 1 if link l is used by a backup path. Finally, Constraint (5-13) is the capacity constraint.

5.2.3 Individual Multicast Tree-based Link Protection Model

The notations used in ITL are shown as follows:

Decision variables:

u_l : the number of video programs on link l assigned for working paths;

$\bar{u}_l$: the number of video programs on link l assigned for backup paths;

y_{rl}^v = 1, if link l is used to deliver video program v to receiver node r; =0, otherwise;

z_l^v = 1, if link l is used to deliver video program v; =0, otherwise;

Problem (ITL):

$$\min \sum_{l \in L} \alpha_l (u_l + \bar{u}_l)$$

subject to:

$$\sum_{l \in L_n^{out}} y_{rl}^v - \sum_{l \in L_n^{in}} y_{rl}^v = \sigma_{nr} \qquad \forall r \in R, v \in V, n \in N \qquad (5\text{-}18)$$

$$y_{rl}^v \le z_l^v \qquad \forall r \in R, v \in V, l \in L \qquad (5\text{-}19)$$

$$\sum_{l \in L_n^{in}} z_l^v \le 1 \qquad \forall n \in N, v \in V \qquad (5\text{-}20)$$

$$\sum_{v \in V} z_l^v \le u_l \qquad \forall l \in L \qquad (5\text{-}21)$$

$$\sum_{e \in L_n^{out}} x_{le} - \sum_{e \in L_n^{in}} x_{le} = \delta_{nl} u_l \qquad \forall l \in L, n \in N \qquad (5\text{-}22)$$

$$x_{ee} = 0 \qquad \forall e \in L \qquad (5\text{-}23)$$

$$x_{le} \le \bar{u}_e \qquad \forall l \in L, e \in L \qquad (5\text{-}24)$$

$$d(u_l + \bar{u}_l) \le C_l \qquad \forall l \in L \qquad (5\text{-}25)$$

$$y_{rl}^v = 0 \text{ or } 1 \qquad \forall r \in R, v \in V, l \in L \qquad (5\text{-}26)$$

$$z_l^v = 0 \text{ or } 1 \qquad \forall v \in V, l \in L \qquad (5\text{-}27)$$

$$u_l \in \text{integer} \qquad \forall l \in L \qquad (5\text{-}28)$$

$$\bar{u}_l \in \text{integer} \qquad \forall l \in L \qquad (5\text{-}29)$$

This model is similar to BTL except that each video program is delivered on an independent tree. Therefore, the decision variable y_{rl} in BTL is replaced with y_{rl}^v here in this model. Constraint (5-18) is the routing constraint. It enforces flow conservation on each node for each program. Constraint (5-19) and Constraint (5-20) jointly determine decision variable z_l^v which is 1 if link l is used for delivering video program v. The bandwidth assigned to the working link l, u_l, is determined in Constraint (5-21). Constraint (5-22) is used to determine the bandwidth assigned to the backup path. Constraint (5-23) requires that the working and backup paths must be link disjointed. Total backup bandwidth required on link e is then determined in Constraint (5-24). Finally, the capacity constraint is described in Constraint (5-25).

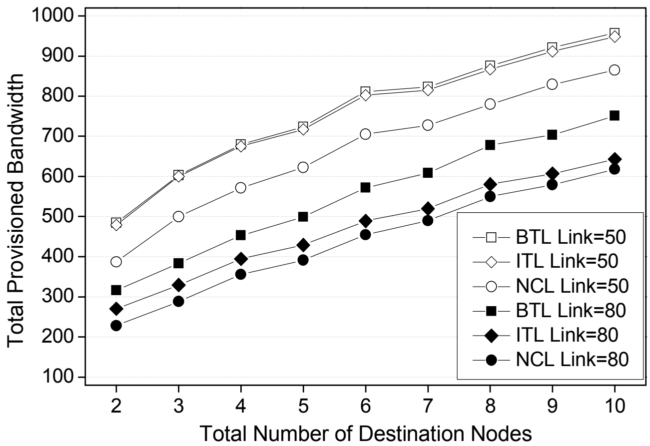
5.3 Experimental Results

We have carried out a performance study on Bundle Tree (BTL), Individual Tree (ITL), and Network Coding (NCL) approaches, and drawn comparisons via simulations over some randomly generated networks. The number of nodes in each randomly generated network is fixed to 20. Network links are being randomly determined to obtain a two-edge connected network. For simplicity, we set link cost coefficient α_l to be one for each link. The available capacity for each link is 100. In this dissertation, we assume each video program consumes unit capacity. In these experiments, server and receiver nodes are randomly selected to provide 30 video programs in each multicast connection. The models are solved using CPLEX on a PC running Windows XP with a 4 GHz CPU. Each point being plotted in the following figures is an average over 90 results.

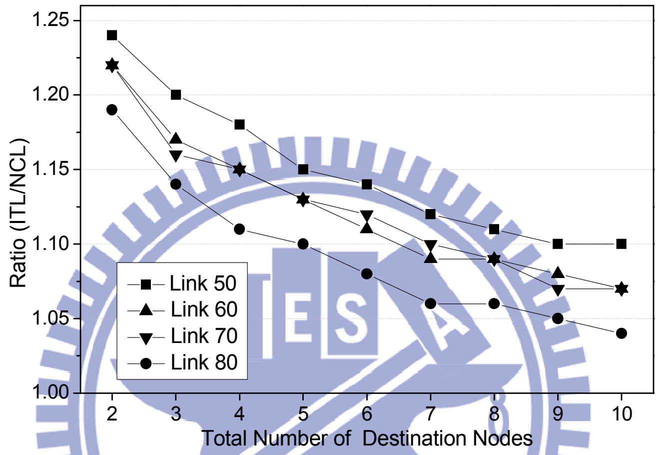
Since we set link cost coefficient α_l to be the same in these simulations, the total cost is equivalent to total bandwidth usage and the results are shown in Figure 35(a). It is clear that the total bandwidth consumption decreases for each scheme as the number of network links increases. For the number of links ranging from 50 to 80, we observe that NCL holds the best performance; BTL needs the most bandwidth; and ITL's performance is always in between NCL and BTL. We further observe that the performance of ITL strongly depends on the network degree. As the number of links is small (50 in our experiments), ITL is very similar to BTL; however, as the number of links goes up, the results move toward NCL.

We further compare the total cost of the network coding approach with the one of the individual tree approach. The comparison is made in *Ratio* which is defined as the total cost of IDL value to NCL value in percentage. As shown in Figure 35(b), we observe that the ratio between these two approaches decreases from (19% to 24%) to (4% to 10%) when receiver nodes equal 2 and 10, respectively. The gain of NCL becomes smaller and smaller as the number of receiver nodes increases; i.e., the tree-based approach is suitable for operating in a multicast network with a large number of receiver nodes.

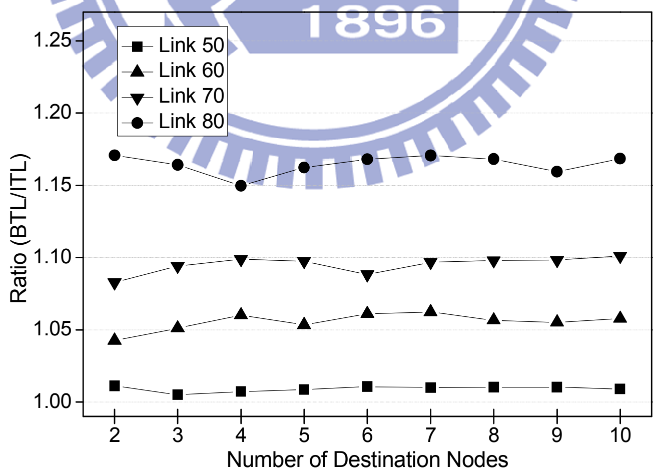
We moreover compare the total cost between these two tree-based approaches. The vertical axis is the ratio of the total cost of BTL to ITL. Numerical results are demonstrated in Figure 35(c). The ratio increases from 1%, 4% to 6%, 8% to 10%, to 15% to 17% when the network links are 50, 60, 70, and 80, respectively. We find that more network links drive the ITL scheme to a better performance. The results indicate that ITL is suitable for being applied in a dense network. However, the performance gap between BTL and ITL diminishes in a sparse network.



(a) Total bandwidth consumption



(b) Performance comparison: ratio of ITL/NCL



(c) Performance comparison: ratio BTL/ITL

Figure 35 - Experimental results of survivable multicast communications

Chapter 6. Concluding Remarks

Next Generation Network shifts from separate vertically integrated application-specific networks to a single network being capable of carrying all services. In addition to providing a technology independent network platform for emerging services, the NGN supports ever-increasing traffic demands. Considerable research work has been done on the access network and transport network so as to provide broadband transport capabilities and QoS-enabled features. The Wavelength Division Multiplexing (WDM) is a significant one. It permits better exploitation of capacity in fiber optics by simultaneously transmitting data packets over multiple wavelengths. Several switching paradigms are proposed for data transportation on WDM networks, including Optical Circuit Switching (OCS), Optical Burst Switching (OBS) and Optical Packet Switching (OPS). OBS and OPS allow switching of ultrahigh bit rate data packets directly in the optical domain and have been proposed as solutions to overcome the “electronic bottleneck” of core routers. It will further bring fundamental changes in the design of the Next Generation Network. However, high-speed switching and optical buffering are challenging problems of the OPS system implementation.

Optical circuit switching (OCS) offers explicit transport guarantees since circuit establishments are confirmed. At the current stage most WDM applications follow the OCS paradigm and in some application scenarios such as connecting high-speed computer servers in lambda grid applications, users make requests to reserve an optical path in advance. To get the best network usage, an optical path should be setup just before it is needed. Providing a lightpath reservation service to users can increase network operators’ revenue and provide better user services. How to jointly determine call admission control as well as Routing and Wavelength Assignment is a significant problem to network operators. Another problem of optical circuit switching is that its coarse granularity rates do not match the rate for packet traffic especially for those come from Ethernet network and this will cause the waste of network resources. The same problem happens to legacy SONET/SDH because it only supports contiguous concatenation transport switching over the overall path. For example, fitting a 100 Mbit/s Fast Ethernet connection inside a 155 Mbit/s STS-3c container leads to near 55Mbits bandwidth waste. Furthermore, 1+1 APS is typically used to handle link failures in SONET in order to provide superior survivability and shorter failure recovery time. But this will cause more waste of network resources. How to reduce the bandwidth waste and provide users a better service is vital to network operators.

As mentioned above, conventional network protection approaches employ extra network resources and precompute backup paths to bypass the failure link or node. It consumes much bandwidth to provide protections. Using network coding mechanism enables the intermediate nodes not only to forward packets but also encode/decode incoming packets using primitive algebraic operations [17]. By transmitting combinations of incoming data on a backup path, this enables each receiver node to recover a copy of the data transmitted on the working path if the working path fails. It can be used to enhance throughput and provide network survivability by selecting some intermediate nodes performing packet encoding.

According to the problems and technology advances mentioned above, we carry out some research on the routing and resource provisioning problems of the next generation network to improve the network efficiency and survivability.

6.1 Our Contributions

In this dissertation, we have investigated four Routing and Resource Provisioning problems in next generation networks and worked out the solutions, including: a FSOB switching system to route packets and resolve contentions in both the wavelength and space dimensions; a Lagrangean relaxation based near-optimal algorithm for advance lightpath reservation in WDM networks to determine request admission, as well as routing and wavelength assignment jointly; a quality-of-survivability concept benefited by a phenomenon that data services can take a gradual bandwidth degradation in addition to a solution to minimize total bandwidth consumption at the same time; a study of optimal routing and bandwidth provisioning problems for survivable multicast communications using network coding.

6.1.1 Multi-wavelength Optical Packet Switching Networks

We demonstrated a novel OCPS paradigm that uses an in-band-control scheme to manage per coarse packet switching. Based on OCPS, we construct an experimental optical IP-over-WDM network named as OPSINET. It consists of three types of nodes - edge routers, optical lambda/fiber switches, and OLSRs, with multi-granularity switching capabilities. OPSINET performs the OCPS paradigm and advocates the enforcement of traffic control to realize bandwidth-on-demand on a sub-wavelength basis. Through this experiment, we perceive that the data-centric optical Internet can become a reality based on the OPS technology.

A novel fully shared output buffer switch using cyclic demultiplexer is proposed to avoiding the packet loss due to output conflicts and reducing the cost of establishing the buffer. The system consists of AWGs, tunable wavelength converters, cyclic demultiplexers and FDLs. The function of TOWC1 and AWG1 is to sort the incoming packets by the destination port. The function of the TOWC2 and AWG2 is to route packets to their destination port. Through it, multiple packets with the same destination port are carried by different internal wavelengths to switch to the same output port at the same time but receive different delays. The buffering process operates as a multiple wavelength First-In-First-Out (FIFO) buffer. The packets with different destination ports are switched to their appropriate outputs without collision. Finally, a cyclic AWG DeMUX provides a modular *M* operation on wavelengths. It fully enables output buffer sharing and is the unique design of this newly proposed switch architecture.

We have done some performance studies based on the analytical analysis given in section 2.5 and examine the Packet Loss Probability (PLP) of this switch under different numbers of wavelengths, numbers of FDLs, and traffic loads. We find that it is possible to use only a small amount of discrete FDL optical buffers combined with multiple wavelengths to provide satisfactory packet loss performance in this proposed system.

6.1.2 Advance Lighpath Reservation in WDM Networks

We propose an efficient Lagrangean relaxation (LGR) approach to resolve this advance lightpath reservation for multi-wavelength optical networks. The task is first formulated as a combinatorial optimization problem in which the revenue from accepting call requests is to be maximized.

The LGR approach performs constraint relaxation and derives an upper-bound solution index according to a set of Lagrangean multipliers generated through subgradient-based iterations. The constraint to guarantee no over-booking on any wavelength channel at any time slot is first relaxed from the constraint set. The expression corresponding to the constraints, is multiplied by Lagrangean multipliers, and then summed with the original objective function. Through reorganizing the new objective function expression, we find that it can be decomposed into several independent subproblems. By carefully observing the problem, we find that the subproblem includes a shortest path problem. We apply Dijkstra's algorithm to obtain the solution of this shortest path problem. By solving all the subproblems, we can obtain the value of that new objective function.

According to the weak Lagrangean duality theorem, the solution of the relaxed problem is an upper bound of the original problem for any non-negative Lagrangean multiplier. A primal heuristic algorithm in the LGR approach is used to find a lower bound solution. We determine the lowest upper bound and highest lower bound by changing the value of the Lagrangean multiplier based on the subgradient method. The LGR solutions are verified whether they satisfy those relaxed constraints or not. If they do, the solutions are feasible and are thus used to calculate a set of bounds of the primal problem.

Finally, we assess the performance of LGR with respect to solution accuracy. We further draw comparisons between LGR and three heuristic algorithms - Greedy, First-Come-First-Serve, and Deadline-First, via experiments over the widely-used NSFNET network. Numerical results demonstrate that LGR outperforms the other three heuristic approaches in gaining more revenue by receiving more call requests.

6.1.3 Multi-path Provisioning for NG-SONET Networks with Quality-of-Survivability Constraints

We propose the quality-of-survivability concept in bandwidth provisioning to take advantage of data services being tolerant of gradual bandwidth reduction. Quality-of-survivability means a source-and-destination (SD) pair can specify its bandwidth requirements for networks under different states, i.e. normal and failure states. We first propose the MP-QoS bandwidth provision scheme to let users specify different bandwidth requirements for networks under normal and failure states. Given a physical topology, the MP-QoS problem is to determine routing and assign capacity in order to satisfy users' requirements and minimize total bandwidth consumption.

We consider three types of network failures, including (1) link-failure, (2) node-failure, and (3) link-and-node failure. A set of ILP models is defined to obtain the optimal solutions. To compare the performance between multi-path provisioning in NG-SONET and in SONET, we conduct two sets of simulations, one runs on a well-know USA benchmark network and the other runs on some random generated networks. We also carry out a simulation for traditional SONET 1+1 protection in order to evaluate the benefit of applying multi-path provisioning.

There are three observations. First, the connection which requests a less stringent guaranteeing survivable requirement consumes less network capacity. The consumed bandwidths among different requisite values increase nonlinearly. It gives a guideline for pricing connections with different quality of survivability requirements. Secondly, a dense network consumes less network capacity than a sparse network. This indicates connectivity degrees should be carefully considered in designing a survivable network. Finally, NG-SONET always outperforms SONET. The gap is even larger for a network with a higher connectivity degree.

6.1.4 Optimal Routing and Bandwidth Provisioning for Survivable Multicast Communications Using Network Coding

An algebraic network coding approach is shown in [9] for protecting multicast communications and asserts the network coding problem is solvable if, and only if, the Min-Cut Max-Flow bound is satisfied for all source-destination connections. Furthermore, they have given proof to assert a network coding solution exists for some predefined failure patterns. The results infer that we can determine an optimal routing and bandwidth provision that satisfies the Min-Cut Max-Flow bound for the pre-defined failure patterns so as to achieve network survivability of multicast applications.

We propose mathematical models for four protection schemes that can be used to obtain an optimal routing for survivable multicast communications against any single link/node failure. The first two schemes (NCL and NCN) are network-coding-based protection schemes. They take advantage of network coding to reduce the bandwidth expenditure of link failure protection, thereby lessening total cost consumption. This approach is good for network operators to provision network services with minimum costs. The remaining two schemes (BTL and ITL) are tree-based protection schemes to prevent the damage of any single-link failure. We formulate the NCL, BTL and ITL as combinatorial optimization problems and determine the routes and bandwidth of a multicast network to protect any single-link failure. These formulations mainly focus on the minimization of total network bandwidth consumption with a 100% survivability constraint on single-link failure and physical link capacity limitation. The output of the problems includes the routing paths and required bandwidth on each link.

To clarify the gain from applying network coding, we made performance comparisons between the network-coding-based protection schemes and the tree-based schemes. We observe that the network coding scheme holds the best performance for all cases we had simulated and is especially distinguished for networks with a small-to-medium-sized number of receiver nodes. The coding gain is strongly dependent on the number of receiver nodes. The results indicate that ITL is suitable for being applied in a dense network. However, ITL and BTL are indistinguishable in a sparse network. The results can provide researchers and industrial engineers with an understanding of the benefit of using network coding and furthermore for quantifying a trade-off between the network protection schemes with and without using network coding.

6.2 Future Work

The key technologies in OPS include header/payload multiplexing, nano-sec switch fabric, optical buffer, fast burst mode receiver, and synchronization. In Chapter 2, we have proposed a new switch fabric design that can share FDL based buffer efficiently. Although we have developed a high performance ASK-based header/payload multiplexing and label swapping technique in OPSINET [25], it is still limited by its high cost. An ASK-based label swapping requires two AM modulators to erase the old label and insert a new label. A new idea has been proposed in [58] where packets along a path is attached a fixed label. In the network, each node uses the label to derive packets' output port without label swapping. In other words, each node performs a function that would interpret the same label to different output port ID. In [58], the label is derived by using Chinese Remainder Theorem. Although the idea is interesting, the drawback is that the label size goes exponentially with the number of network nodes increases. How to design a new mapping function to reduce the label size is still an open question and needs to be resolved in the future.

In this thesis, we consider the light path reservation problems and have proposed a Lagrangean based algorithm to obtain a near optimal solution. The problem we considered can be viewed as a static problem where the connections requesting for services are held until a predetermined processing time point. At that time, those connection requests are processed in a batch in the network management system. The static nature of the problem enables us to apply an optimization based approach to obtain a near optimal solution such that network resources can be provisioned efficiently. Although the batch process has high efficiency, it is not suitable for some applications that request calls in advance but requires the system to respond in real-time to indicate if the requested calls are admitted or not. Such a dynamic connection request is basically in the territory of a dynamic RWA problem. However, since the calls will be serviced in the future, routing and wavelength assignment of any admitted connection can be rearranged before its service time. How to retune the admitted calls so as to make network resources and time spaces for latter connections is still an open question. Such dynamic connection request problems need to be further studied in the future.

In Chapter 5, we have proposed a mathematical model for applying network coding techniques in a survivable multicast problem. The results indicate that network coding outperforms conventional survivable path planning approaches in bandwidth usage. However, it has to pay for additional coding effort, packet synchronization and signaling delay during failure recovery. Considering the combination of these problems is still an open question and needs to be answered in the future. Another approach is to only apply coding at the sender node and decoding at the receiver node. The benefit of the approach is no signaling broadcast in the network so as to reduce the failure recovery time. We are doing research on the new topic to jointly determine the coding and path routing using optimization based approaches.

Bibliography

[1] K. Knightson, N. Morita, and T. Towle, "NGN architecture: generic principles, functional architecture, and implementation," IEEE Commun. Mag., vol. 43, no. 10, pp. 49-56, Oct. 2005.

[2] ITU-T 2011, "Series Y: global information infrastructure, Internet protocol aspects and next generation networks – frameworks and functional architecture models," Nov, 2004

[3] H. Van Helvoort, "Next generation SDH/SONET: evolution or revolution?," England: Wiley & Sons, 2005.

[4] C. Ou et al., "Survivable virtual concatenation for data over SONET/SDH in optical transport networks," IEEE/ACM Trans. Networking, Vol. 14, No. 1, Feb. 2006.

[5] R. Santitoro, "Metro Ethernet services – a technical overview," Metro Ethernet Forum, 2006.

[6] B. Mukherjee, "WDM optical communication networks: progress and challenges," IEEE J. Select. Areas Commun., vol. 18, no. 10, pp. 1810-1824, Oct. 2000.

[7] E. Mannie et al., "Generalized multi-protocol label switching (GMPLS) architecture, draft-ietf-ccamp-gmpls-architecture-07.txt, Nov. 2003.

[8] T. El-Bawab and J.-D. Shin, "Optical packet switching in core networks: between vision and reality," IEEE Commun. Mag., vol. 40, no. 9, pp. 60-65, Sept. 2002.

[9] F. Callegati, G. Corazza, and C. Raffaelli, "Exploitation of DWDM for optical packet switching with quality of service guarantees," IEEE J. Select. Areas Commun., vol. 20, no. 1, pp. 190-201, Jan. 2002.

[10] H. Dorren et al., "Optical packet switching and buffering by using all optical signal processing methods," J. Lightwave Technol., vol. 21, no. 1, pp. 2-12, Jan. 2003.

[11] B. Mukherjee, "Optical communication networks," New York: McGraw-Hill 1997

[12] D. Banerjee.and B. Mukherjee,."A practical approach for routing and wavelength assignment in large wavelength-routed optical networks," IEEE J. Select. Areas Commun. vol. 14, no. 5, pp. 903–908, Jun., 1996

[13] R. W. Yeung, S. Y. R. Li, N. Cai, and Z. Zhang, "Network coding theory," Now Publishers Inc., 2006.

[14] R. Ahlswede, N. Cai, S. Y. Li, and R. Yeung, "Network information flow," IEEE Trans. Inf. Theory, vol. 46, no. 4, pp. 1204–1216, Jul. 2000.

[15] S. Y. R. Li, R. W. Yeung, and N. Cai, "Linear network coding," IEEE Trans. Inf. Theory, vol. 49, no. 2, pp. 371–381, Feb. 2003.

[16] T. Ho, M. M´edard, R. Koetter, D.R. Karger, M. Effros, J. Shi, and B. Leong, "A random linear network coding approach to multicast," IEEE Trans. Inf. Theory, vol. 52,

no. 10, pp. 4413–4430, October 2006.

[17] Y. Chen, C. Qiao and X. Yu, "Optical Burst Switching (OBS): A New Area in Optical Networking Research", IEEE Network Mag., vol. 18 no. 3, pp. 16–23, May-June 2004.

[18] M. Yoo, C. Qiao, and S. Dixit, “Optical burst switching for service differentiation in the next generation optical Internet,” IEEE Commun. Mag., vol. 39, no. 2, pp. 98-104, Feb. 2001.

[19] V. Vokkarane and J. Jue, “Prioritized burst segmentation and composite burst-assembly techniques for QoS support in optical burst-switched networks,” IEEE J. Select. Areas Commun., vol. 21, no. 7, pp. 1198-1209, Sept. 2003.

[20] J. Wei and R. McFarland, “Just-in-time signaling for WDM optical burst switching networks,” J. Lightwave Technol., vol. 18, no. 12, pp. 2019-2037, Dec. 2000.

[21] M. Yoo, C. Qiao, and S. Dixit, “QoS performance of optical burst switching in IP-over-WDM networks,” IEEE J. Select. Areas Commun., vol. 18, no. 10, pp. 2062-2071, Oct. 2000.

[22] J. White, M. Zukerman, and H. Vu, “A framework for optical burst switching network design,” IEEE Commun. Lett., vol. 6, no. 6, pp. 268-270, June 2002.

[23] N. Barakat and E. H. Sargent, “Dual-header optical burst switching: a new architecture for WDM burst-switched networks,” in Proc. IEEE Infocom 2005, pp. 685-693.

[24] Z. Rosberg, Hai Le Vu, M. Zukerman, and J. White, “Performance analyses of optical burst-switching networks,” IEEE J. Select. Areas Commun., vol. 21, no. 7, pp. 1187-1197, Sept. 2003.

[25] M. Yuang, P. Tien, J. Shih, and A. Chen, “QoS scheduler/shaper for optical coarse packet switching IP-over-WDM networks,” IEEE J. Select. Areas Commun., vol. 22, no. 9, pp. 1766-1780, Nov. 2004.

[26] E. Mannie et al., “Generalized multi-protocol label switching (GMPLS) architecture, draft-ietf-ccamp-gmpls-architecture-07.txt, Nov. 2003.

[27] Y. Lin, M. Yuang, S. Lee, and W. Way, “Using superimposed ASK label in a 10 Gbps multi-hop all-optical label swapping system,” J. Lightwave Technol., vol. 22, no. 2, pp. 351-361, Feb. 2004.

[28] S. Yao, B. Mukherjee, and S. Dixit, “Advances in photonic packet switching: an overview,” IEEE Commun. Mag., vol. 38, no. 2, pp. 84- 94 Feb. 2000.

[29] Martin Maier., “Optical switching networks,” Cambridge University Press 2008

[30] D. Chiaroni, C. Chauzat, D. De Bouard, S. Gurib, M. Sotom, and J. M. Gabriagues, "A novel photonic architecture for high capacity ATM switching Applications," in Proc. Photon. Switching '95, Salt Lake City, Vol. 1, Mar. 1995, pp. 84-86.

[31] S. L. Danielsen, B. Mikkelsen, C. Joergensen, T. Durhuus, and K. E. Stubkjaer, "WDM packet switch architectures and analysis of the influence of tunable

wavelength converters on the performance," J. Lightwave Technol. vol. 15, no. 2, pp. 219-227, Feb. 1997.

[32] K. Hunter, et al., "SLOB: A switch with large optical buffers for packet switching," J. Lightwave Technol., vol. 16, no. 10, Oct. pp.1725-1736, 1998.

[33] K. Hunter, et al., "WASPNET: A wavelength switched packet network," IEEE Commun. Mag., vol. 37, no. 3, pp. 120-129, Mar. 1999.

[34] Zheng, J. Zhang, B. Mouftah, H. T.,"Toward automated provisioning of advance reservation service in next-generation optical internet," IEEE Commun. Mag., vol. 44, no. 12, pp. 68-74, Dec. 2006.

[35] T. D. Wallace, and A. Shami, , "Advanced lightpath reservation in WDM networks," in Proc. IEEE Infocom., 2006

[36] H. Miyagi, M. Hayashitani, D. Ishii, Y. Arakawa, and N. Yamanaka,"Advanced wavelength reservation method based on deadline-aware scheduling for lambda grid network," IEEE J, Lightwave Technol. Vol. 25, No. 10, pp. 2904-2910, Feb. 2007,.

[37] S. Naiksatam, S. Figueira, S. A. Chiappari, , and N. Bhatnagar, ," Analyzing the advance reservation of lightpaths in lambda-grids," IEEE CCGrid 2, 985–992 (2005)

[38] J. Zheng, Mouftah, and H. T.,"Routing and wavelength assignment for advance reservation in wavelength-routed WDM optical networks," in Proc. IEEE ICC. 5, 2722–2726 (2002)

[39] S. Lee, M. Yuang, and P. Tien,"A Lagrangean relaxation based approach for routing and wavelength assignment in multi-granularity optical WDM networks," IEEE J. Select. Areas Commun. vol. 22, no.9, pp. 1741–1751, Nov. 2004.

[40] R. Ahuja, T. Magnanti, and J. Orlin, "Network Flows: Theory, Algorithms, and Applications. Prentice-Hall, Englewood Cliffs, NJ, USA (1993)

[41] S. Rai et al., "Reliable multi-path provisioning for Next-Generation SONET/SDH networks with virtual concatenation," OWG2, OFC 2005.

[42] C. Ou et al., "Survivable virtual concatenation for data over SONET/SDH in optical transport networks," IEEE/ACM Trans. Networking, Vol. 14, No. 1, Feb. 2006.

[43] S. A. Aly and A. E. Kamal. "Network protection codes against link failures using network coding," in Proc. IEEE Globecom, pp. 1-6, Nov./Dec. 2008.

[44] A. E. Kamal. "1+N protection against multiple link failures in mesh networks," in Proc. IEEE ICC, Jun. 2007.

[45] A. E. Kamal. "A generalized strategy for 1+N protection," in Proc. IEEE ICC, May 2008.

[46] S. A. Aly and A. E. Kamal., "Network coding-based protection strategy against node failures," in Proc. IEEE ICC, Jun. 2009.

[47] R. Koetter and M. Médard, "An algebraic approach to network coding," IEEE/ACM Trans. Netw., vol. 11, no. 5, pp. 782–795, Oct. 2003.

[48] C. S. Wu and S. W. Lee, “Backup VP planning for multicast connections in ATM networks,” Computer Commun., vol. 22, pp. 898-906, Jun. 1999.

[49] J. Ramamirtham and J. Turner, "Design of wavelength converting switches for optical burst switching", in Proc. IEEE Infocom, volume 1, pages 362-370, 2002., New York.

[50] B. Mukherjee, “Optical WDM Networks,” New York: Springer 2006.

[51] X. Ma and G. S. Kuo, “Optical switching technology comparison: optical MEMS vs. other technologies,” IEEE Commun. Mag., vol. 41, no. 11, pp. S16-S23, Nov. 2003.

[52] L. Li, S. D. Scott, and J. S. Deogun. “A novel fiber delay line buffering architecture for optical packet switching,” in Proc. IEEE Globecom, pp. 2809-2813, Dec. 2003.

[53] M. C. Yuang, Y. M. Lin, J. L. Shih, S. W. Lee, and S. H. Lin, “A QoS optical packet switching system: architectural design and experimental Demonstration,” IEEE Commun. Mag., vol. 48, no. 5, pp. 67-75, May. 2010.

[54] P. C. Ku, C. J. Chang-Hasnain, J. Kim, and S. L. Chuang, “Variable optical buffer using slow light in semiconductor nanostructures,” Advanced Optical and Quantum Memories and Computing,” in Proc. SPIE, Volume 5362, pp. 69-80 (2004).

[55] D. Zhou and S. Subramaniam, “Survivability in Optical Networks,” IEEE Network. Mag., Volume: 14 , Issue: 6, pp. 16-23, Nov/Dec. 2000.

[56] J. Ramamirtham and J. S. Turner, “Design of wavelength converting switches for optical burst switching,” in Proc. of IEEE Infocom. 2002, vol. 2, pp. 1162–1171.

[57] D. Klonidis, C. (T.) Politi, R. Nejabati, D. Simeonidou, and M. O’Mahony, “Designing a high speed optical packet switched network,” in Proc. of IEEE ICTON. 2005, vol. 2, pp. 245–251.

[58] H. Wessing, H. Christiansen, T. Fjelde, and L. Dittmann.,” Novel scheme for packet forwarding without header modifications in optical networks,” IEEE J, Lightwave Technol. Vol. 20, No. 8, pp. 1277-1283, Aug. 2002,