University of Alberta

Capacity Planning and Management for Mesh Survivable Networks under Demand Uncertainty

by

Kwun Kit Dion Leung

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of *Doctor of Philosophy*

Department of Electrical and Computer Engineering

Edmonton, Alberta
Fall 2005

0-494-08674-2

Your file Votre référence
ISBN:
Our file Notre reference
ISBN:

To

my wife, Hui Kan,

my mom, Lai To,

my brother, Kwun Ho,

and *my late mentor, K.C. Chau,*

who will never be forgotten.

Abstract

This thesis presents a set of optimization-based strategies to assist network planners and operations support engineers in planning and managing the capacities of mesh-based survivable transport networks in the face of demand uncertainty. While there have been many works on network design, consideration of demand uncertainty into network design models has remained one of the least explored areas. The extent of uncertainty in planning problems in general has been already classified by others as follows: Level I: A Clear-Enough Future, Level II: Alternative Futures, Level III: A Range of Futures and IV: True Ambiguity. We have followed this schema and propose a set of new optimization models for the three levels where uncertainties are more pronounced:

- For Level II: A two-part, stochastic programming-based optimization model is developed for incorporating demand uncertainty and network survivability into a single capacity-planning formulation. While almost all published studies on the design of survivable networks are based on a specific demand forecast (i.e. Level I) and optimize capacity cost for a single target planning view, the two-part formulation explicitly incorporates a set of plausible demand scenarios and optimizes both present and future long-term capacity investment. We also extend the two-part formulation to capture the modularity and economy-of-scale effects and show significant capacity cost savings of the new models over traditional single-forecast design methods.
- For Level III: A framework, based on the concepts of Pattern Forecast Accuracy (PFA) and Servability, is designed for assessing the robustness of the ability of various survivable networks to cope with uncertainty in the demand forecast. This framework serves as an evaluation tool for network operators to effectively identify robust survivable network designs from any given sets of cost-optimal designs.

- For Level IV: We develop two operational strategies, namely, max-profit demand loading and re-optimization strategies, for managing as-built capacities of any mesh survivable transport networks. The value of the demand loading formulation is to help service providers to identify and route a set of demands that could generate the maximum profit, taking the provisioning cost and service revenue into consideration. Multiple quality-of-protection (multi-QoP) service classes (i.e. protected, unprotected and preemptible classes) are also considered in the demand loading formulation. As another valuable tool for network operators, re-optimization strategy is used to improve a network's ability to carry future traffic, through rearranging solely the existing spare capacities or with the latitude of also rearranging in-service paths.

With the fact that the expenditure on transport capacity is in the order of millions and even billions of dollars, the potential capital savings from these optimization models can be substantial.

Acknowledgements

To most people, this thesis is filled with black texts on white papers; to me, every page, every piece of work in this thesis is full of color, hard work, and good memories. I would like to express my sincere gratitude to a number of individuals and organizations.

First, I am grateful to Dr. Wayne Grover, not only for being my supervisor for the graduate research work, but also for showing me the true definitions of professionalism, dedication and passion in research. Wayne, you had me when I attended your first EE589 lecture, and I am always proud of being one of your students.

I am also deeply thankful to a number of colleagues, Dr. Matthieu Clouqueur, Kent Lam, Dr. David Mazzarese, Gangxiang Shen, Peter Giese, Govindkrishna Kaigala, Dr. John Doucette, Adil Kodian, Anthony Sacks, Dr. Yan Xin and Dr. Fengqin Zhai, for providing me a friendly and stimulating research environment at TRLabs. I would also like to thank Professor Masayuki Murata and Dr. Shin'ichi Arakawa for giving me an opportunity to experience research life in Japan and to work with you during the summer of 2004 at Osaka University. At TRLabs, I am also indebted to Luke Chong, Linda Richens and Rhoda Hayes, for not just helping me with administrative and computing issues, but also for your kindness, personal support throughout my entire research program.

My PhD final oral defense was truly enjoyable and unforgettable, and with that I am thankful to the members of my supervisory committee, Dr. Mike MacGregor, Dr. Ivan Fair, Dr. Bruce Cockburn and my external examiner, Dr. Deep Medhi, for giving me a thought-provoking, philosophical afternoon session. I am also greatly appreciated your valuable comments, improvements, corrections and clarifications to the thesis.

Finally, I must express my genuine gratitude to my mother, Lai To, my brother, Koon Ho, for your constant support and encouragement throughout my entire life, and especially my wife, Hui Kan, for your unconditional love, positive attitude, and numerous delicious home-made meals. My life would never be the same without you.

Preface

Composing this thesis has been an interesting yet challenging process. The first challenge was to identify my intended readers. My supervisor and examining committee members are, of course, the primary readers. Industry professionals such as network planners, operations support engineers, network software developers come to mind as secondary readers, for these are the people who might apply the new concepts developed in this thesis to their daily planning tools and network operations. This secondary audience influenced the content selections in the introductory chapters and explains why the core chapters are presented in a rather solution-oriented approach.

As committee members and industry often have their own preferences in terms of the level of detail and how each chapter benefits their own research environment, the materials in this thesis are organized into **three modules** that sequentially address the questions of *why*, *what* and *how*. The **first module**, covered by Chapter 2, attempts to provide cross-disciplinary and general views on the issues of telecommunications economics, network survivability and demand uncertainty. Specifically, this module presents facts and figures to show the importance of the research area and to help address open-end questions such as: Why do we care about network survivability in general? Why are we particularly interested in designing networks to protect against fiber cuts? What are the limitations of any given survivable network capacity plans? It is expected that, after reading this chapter, interesting yet unanswered questions will arise.

The **second module,** which consists of Chapters 3 and 4, sets the overall problem scope of the research problems and, particularly, pinpoints what decision variables and input parameters should be modeled in the context of transport networking. Chapter 4 goes further into each research problem and analyzes the existing literature classified by the problem objectives, problem statements and modeling methodologies. This module also identifies what thesis-related problems researchers have been trying to solve, how they solve them, and other open questions that have not yet been investigated.

Chapters 5 to 8 summarize the key research work published from September 2002 to May 2005 as part of my doctoral program. In that **final module**, each chapter is presented in a logical progression beginning with a short introduction, specific research questions, new modeling concepts, optimization formulations and, finally, concluding with experimental results. As motivational paragraphs and background information of each study have been thoroughly covered in the previous two modules, each chapter in the final module is meant to be problem-specific and solution-oriented. Mathematical formulations and results are detailed and descriptive with respect to the corresponding problem statements. By reading the final module, it is also my hope that network planners and system engineers will find these new concepts applicable to their existing research work.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

APS	Automatic Protection Switching Systems
ATM	Asynchronous Transfer Mode
ADM	Add/Drop Multiplexer
DXC	Digital Cross-connects
GbE	Gigabit Ethernet
GOF	Given Occurrence of Failures
FDM	Frequency Division Multiplexing
FICON	Fiber Connection
LP	Linear Programming or Linear Program
LTP	Long-term Planning
ILP	(Mixed) Integer Linear Programming or Integer Linear Program
ITU	International Telecommunication Union
MTP	Medium-term Planning
OADM	Optical Add/Drop Multiplexer
OD	Origin to Destination
OR	Operations Research
OEO	Optical to Electrical to Optical Conversion
OOO	All Optical (without OEO conversion)
OTN	Optical Transport Network
OXC	Optical Cross-connects
PR	Path Restoration
QoP	Quality of Protection
QoS	Quality of Service
SBPP	Shared Backup Path Protection or Shared Backup Path Protected
SCP	Spare Capacity Placement Problem
SONET	Synchronous Optical Networking
SP	Stochastic Programming or Stochastic Program
SR	Span Restoration or Span Restorable
STP	Short-term Planning
TDM	Time Division Multiplexing
SRLG	Shared Risk Link Group
WDM	Wavelength Division Multiplexing

1 Introduction

As is the case in planning and operating any public transport infrastructures, the design and management of telecommunication transport networks can be highly complex and capital-intensive. Hundreds of millions and even billions of US dollars have been spent on building and managing these "backbone" networks; researchers and engineers are constantly responding to the needs and developing new modeling tools to support network carriers in making these cost-effective planning and operational decisions.

This is also the key objective of this thesis: To develop a set of optimization models and design principles for network planners and operational support engineers to incorporate into their existing planning and network management tools. We must note that these models would not (and should not) be directly comparable to existing commercial design tools[1], but new concepts and principles discovered by these models could expand the functionality of the existing software tools, and influence how survivable transport networks could be planned and operated.

1.1 Contributions

The proposed optimization models can be classified generally into two problem areas of transport networking: capacity planning and capacity management. Considerations of network survivability and demand uncertainty add specific dimensions to this study and make this thesis unique. Based on the existing literature from universities and industry, the main contributions of this thesis include:

- Stochastic programming-based optimization models are formulated to incorporate demand uncertainty and network survivability into the capacity-planning problem of mesh-based survivable transport networks, specifically, span-restorable and *p*-cycle networks. Realistic aspects of optical networking, including modularity and economy-of-scale effects, are also captured in the models for studying the trade-off between making the capacity investment immediately and in the future.
- A unified framework is proposed for assessing the robustness of span-restorable and path-protected transport network architectures to cope with uncertainty in the demand forecast. The notion of Pattern Forecast Accuracy (PFA) is suggested for quantifying

[1] VPIsystems' VPItransportMaker Mesh™, RSoft Design Group's MetroWAND™ and Optiwave's OptiPlanner are examples for network design tools, while Cramer Systems' Cramer5, MetaSolv's Inventory Management and NetCracker's Asset Management are examples for operations system support tools.

pattern errors in the demand forecast, and Servability is introduced as a single measure used for evaluating the inherent robustness of various survivability options to forecast variations.

- A demand loading formulation is suggested for operators to identify potential demand service pairs that could generate the maximum profit, provided any provisioning cost, service revenue and demand models. The service-oriented aspects of multiple survivability or quality of protection (multi-QoP) classes are also included in the formulation. Specifically, three types of services, protected, unprotected and preemptible, are considered.
- Several re-optimization strategies are proposed to determine a capacity configuration that has a better ability to adapt to uncertain traffic. The span-restorable network is used as the basis of the formulations and the benefits of rearranging spare capacity only and with the latitude of also rearranging in-service paths, are both evaluated. These strategies suggest a way to serve demand growth while deferring unnecessary capital investment for transport capacity.

1.2 Overview

In the preface, we briefly described the organization of this thesis into three modules. Here we provide a short summary and also illustrate key highlights of subsequent chapters.

Chapter 2 covers the main themes of the thesis, namely telecommunications economics, network survivability and demand uncertainty. This chapter begins by explaining the economic implications of transport capacity design problems from an investment, decision-making perspective. Several definitions of network survivability are then presented in Section 2.3, and we also underline the importance of planning transport networks against fiber cuts. Section 2.4 first presents several demand forecasting models and then attempts to identify the root causes of demand uncertainty. Finally, we point out the different types of capacity planning problems in Table 2.3, and we end with some fundamental questions that set the scope of problems we address throughout the thesis.

Chapter 3 contains background and assumptions on modeling survivable transport networks. We begin by explaining the concepts of layering, transport node, transport demands, working and spare capacities, as well as the necessary topologies required for the capacity design of survivable transport networks. Key survivability schemes, span restoration, p-cycles and shared backup path protection, are discussed in Section 3.3.1. Additional design assumptions and design objectives are presented in next two sub-sections.

Chapter 4 presents a complete literature review on the capacity planning and management problems considered in this thesis. First, we cover the prior work on the capacity planning for mesh survivable networks, in particular on the basic spare capacity placement problem. Relevant published works on the problem of mesh capacity design under demand uncertainty are analyzed and compared in Table 4.2. Finally, related works on demand loading and the problems of reconfiguration, and re-optimizations are discussed in Section 4.4.

Chapter 5 introduces two integer linear formulations that are used to evaluate the robustness of span-restorable and path-protected networks to withstand changes in the demand forecast. The evaluative framework based on Pattern Forecast Accuracy (PFA) and Servability is also explained in this chapter. Overall, this framework suggests a descriptive approach to demand uncertainty, where we should analyze the uncertainty as a separate evaluative process, but not incorporate uncertainties into the capacity design process.

Chapter 6 presents the stochastic programming-based models from which we explicitly express demand uncertainty and survivability constraints in a unified capacity design model. An important concept to focus on is the idea of treating capacity planning as a two-part investment problem. The effects of modularity and economy-of-scale are discussed in Section 6.4. The new formulations are tested against existing models; pertinent insights and findings are offered in Section 6.7.

Chapter 7 presents an integer program model for demand loading where multiple quality-of-protection (multi-QoP) services are considered. Of special attention is how we define the provisioning cost and revenue for each service class in Section 7.3. Under any cost model and subject to any demand scenarios, the max-profit formulation can be used to reveal strategies about which specific demands would be most profitable, as well as to provide routing and protection solutions for each service demand.

Chapter 8 introduces four re-optimization strategies, devised from a master integer program formulation. Span restorable mesh networks are used in this study, and we exploit the options of rearranging only spare capacity and re-optimizing the combined working and spare capacities to achieve the best utilization. In facing random arrival of new demands, results show that all strategies can be used to improve the overall blocking performance, and also suggest deferral of new capacity addition to support demand growth.

Chapter 9 summarizes the entire thesis and publications, and presents several new topics for future research.

2 Telecommunications Economics, Survivability and Demand Uncertainty

2.1 Introduction

This chapter constitutes the first module of the thesis. Herein we aim to provide background on three core subjects, namely, telecommunications economics, network survivability and demand uncertainty. Each subject is related to the general problem of transport capacity planning. We also present facts and figures from various sources to provide motivational thoughts on specific questions. For instance, what are the implications and limitations associated with a transport capacity design? Why should we be concerned about network survivability and especially fiber cuts in a transport network design? Why should demand uncertainty be considered in the capacity planning process? It is our intention that, after reading this chapter, readers will have a good grasp of why this thesis topic was chosen. It is the combined consideration of transport network planning, network survivability and demand uncertainty that makes this thesis unique.

2.2 Economic Implications of Capacity Planning of Transport Networks

According to Oxford's Dictionary of Economics, economics is the study of how scarce resources are or should be allocated [Oxf02]. It also involves the efficient allocation of scarce resources among multiple competing ends [Lit79]. At a personal human level, the scarce resource might be personal savings that we must choose whether and how to spend them (for example, on mortgages, transportation, investments, or our children's education.) At a corporate level, a company faces similar choices: it must choose among various alternatives and make the best investments within a limited budget. Because of the scarcity of resources and virtually unlimited desires, it is not always possible to obtain all choices simultaneously. Hence, economics is often called the "science of choice."

A similar decision-making process exists in the capacity planning of transport network. Given a limited resource, network planners must determine *where* and *how much* capacity should be allocated over the network. And because of the huge amount of capital expenses involved in building such networks, decisions of capacity allocation correspondingly have much greater economic implications than choices we make at the personal level. Capital expenditures of hundreds of millions and even billions of US dollars spent on transport infrastructure alone (not including other major costs such as rights-of-way, land and property, operational cost, maintenance cost, etc.) are not uncommon [Thr00][Lev02]. Such investments are partially or completely irreversible; that is, the initial cost of investment is at least partially sunk and cannot be recovered.

The notions of "scarcity" in economics and the "science of choice" apply to *all* capacity design problems in this thesis. In other words, if one ignores these concepts in the first place, we could have designed an optical network with direct point-to-point cable among all cities, and each cable fills with maximum capacity to support future demands and network survivability. With capital investment of this great magnitude, a capacity plan should be designed with cost-effectiveness as one of the major objectives. In a detailed capacity plan, this objective should then guide us in deciding how much and where to allocate the capacity resources.

While every capacity plan might have significant economic implications, its monetary impact on the overall network design must not be over-emphasized. No matter how accurate and detailed a capacity plan might be, it can only represent a piece of the overall problem. For example, construction cost, property cost and operational cost are aspects that a capacity plan would not be able to capture and address, yet they represent considerable portions of the overall cost of the network infrastructure. The following table shows an example of a cost breakdown from Level (3) Communications' 2002 annual report [Lev02]:

	Cost (dollars in millions)
Land and Mineral Properties	$ 178
Facility and Leasehold Improvements:	
Communications	1,260
Information Services	28
Coal Mining	65
CPTC (California Private Transportation Company)	92
Network Infrastructure	4,106
Operating Equipment:	
Communications	1,386
Information Services	81
Coal Mining	78
CPTC	19
Network Construction Equipment	34
Furniture, Fixtures and Office Equipment	133
Construction-in-Progress	73
Total	7,533

Thus with reference to the cost of the network infrastructure (which comprises over 50% of the total budget), an optimally-designed capacity plan should give the benefits from the capital investment standpoint.

In this thesis, we propose a collection of new optimization-based models to help network planners in making capacity planning decisions. For example, capacity planning models can provide guidelines for planners to determine *where*, *what* and *how much* transport capacity should be allocated to create a cost-effective, survivable and future-proof network. These methods can be implemented as a set of decision-support tools to provide design insights about network planning and architecture, and also serve as benchmarks to compare against existing capacity design methodologies. For planners who might implement these new concepts, it is important to fully understand the assumptions and limitations of each model and correctly apply them to their own design environment.

While specific questions of *where*, *what* and *how much* are addressed throughout this thesis, readers who think strategically will be curious to know *why*, *whether* and *when* the capacity investment should be made. For example, should we make a capacity investment in the first place, and why? Is it worthwhile to make survivability investments in existing networks? When are the best times to make such investments, or should we defer them? These are all important and interesting questions that we cannot address in this thesis. To tackle these fundamental problems, in fact, would require a diverse knowledge of economics, finance, technology, regulation and policy [Sha00][All02][AlN99][DiP94][Lit79][NaN04]. This will be commented upon the end of Chapter 9.

2.3 Need for Network Survivability

In the previous section, we generally discussed the implications and relations of economics to transport capacity planning. In this section, we discuss another central area of the thesis – namely, network survivability.

Depending on the context and targeted audience, there are many ways to define network survivability. A standards development organization, the Alliance for Telecommunications Industry Solutions (ATIS), defines network survivability [ATI01] as (1) the ability of a network to maintain or restore an acceptable level of performance during network failures by applying various [post-failure] restoration techniques, and (2) prevention or mitigation of service outages from network failures by applying preventive techniques. The concern of network survivability begins early in the communications network for the military [WIL63][MiL80]. In the context of military data networks, survivability includes protecting assets, hiding them, and duplicating them for redundancy. It also emphasizes the "endurance" [NRC85] – the assurance that those assets that do survive can continue to perform in a battle environment for as long as needed (generally months rather than hours). "Restoral" means the ability to restore some of the damaged assets to operating status. "Reconstitution" refers to the ability to integrate fragmented assets into a surviving and enduring network [NRC85]. These networks must also function, albeit at a reduced level of performance, after many nodes and links have been destroyed [NRC85]. While our notion of network survivability is different from that in the military context, the ability of a network to protect against unexpected failures has become an increasingly important issue in today's environment where network operators, service providers and customers are constantly emphasizing the need for reliable communication. Numerous research studies have been conducted to enhance the survivability of networks and to plan for unexpected events, such as facility outages, power outage, capacity overloads and natural disasters.

Although specific definitions of survivability vary depending on the type of network considered, from a planning standpoint, there are essentially three common aspects in the process of designing survivable networks: (1) Identification of the types of attacks, failures, or accidents impacting a network; (2) Anticipation of the impact or risk due to the failure; and (3) Design of strategies to recover, mitigate, or eliminate the impact of the failures. In this section we will discuss these three aspects in general. The specifically defined scope of network survivability in this thesis will be proposed in Section 2.3.1.

The terms "attacks," "failures," or "accidents" are often used interchangeably, while [EFL99] has clearly defined the differences among them. In [EFL99], *attacks* refer to potentially damaging events orchestrated by an intelligent adversary, such as intrusions, probes, denial of service, or even nuclear bombardment [WIL63]. *Failures* are potentially damaging events caused by deficiencies in the system or in an external element on which the system depends. Examples of failures are software design errors, hardware degradation, human errors, or corrupted data. *Accidents* describe the broad range of randomly occurring and potentially damaging events such as natural disasters. Because various kinds of network failures exist, it is an important requirement to first identify and narrow down the specific types of failures we are trying to plan against.

Another important requirement of network survivability is to *qualitatively* describe, and if possible, *quantitatively* measure the impact or outage created by the failures. To most people, network outages are often associated with some costs, which can be quantified monetarily. For example, a report written by AT&T in co-operation with the Economist Intelligence Unit (EIU) indicated that the following statistics are necessary to give an enterprise some idea of losses due to network outages [ATT04]:

Table 2.1. Typical losses due to network outages. From [ATT04].

Industry	Business operation	Average cost/hour of downtime
Financial	Brokerage operations	$6.5M
Financial	Credit card/ sales authorizations	$2.6M
Media	Pay-per-view television	$1.1M
Retail	Home shopping (TV)	$113,000
Retail	Home catalogue sales	$90,000
Transportation	Airlines reservations	$89,500

Although that report and the set of values do not explicitly correlate the impacts with the causes of the failures (which might be software bugs, traffic overload, hardware failures, human errors, or fiber cable cuts), the potential revenue loss arising from a network disruption is clearly huge. As a result, today's service providers have been putting great emphasis on the importance of network survivability, from both the disaster recovery standpoint as well as the more proactive strategy for ensuring business continuity.

Financial losses from network outages might be part of the reason why network survivability is significant. At times, however, outages of the backbone networks cannot even be justified financially. The September 11 terrorist attack and the more recent Hurricane Katrina are tragic events that caused severe outage in telecommunications infrastructures, as well as access to lifeline 911 services, and such damages go beyond purely financial losses. If we think of telecommunications networks in the same way as other basic infrastructures, such as roads, water and power, failures in telecommunications networks could lead to significant societal impacts. In fact, triggered by the September 11 tragedy, both the Canadian and the US governments have subsequently launched similar programs to assess and manage risks to the society's critical infrastructures[2], and telecommunications network is considered to be one of them. These programs encourage investigation of the possible cross-disciplinary impacts among different infrastructures, in terms of the health, safety, security and economic well-being of the nations and the effective functioning of governments [MCF03] [NSE04]. The key motivation of these studies is to identify and prioritize risk protection strategies, and finding ways of prioritizing resource allocations: "Essentially the problem facing the federal government is to minimize, with a limited amount of resources, the expected impact on the nation's critical infrastructure of any future terrorist attack. Impacts could be measured in lives lost, economics dislocations, loss of

[2] In [NSE04], these critical infrastructures are grouped into ten sectors: Energy and Utilities, Communications and Information Technology, Finance, Health Care, Food, Water, Transportation, Safety, Government and Manufacturing.

military capability, loss of national morale (measured perhaps by polling), or some combination" [MCF03]. As expected, the value of the networks shall only increase as our dependency on these networks increases. The potential impacts due to the unexpected network failures, whether it is due to natural disasters, intended intrusions, or human errors, will only emphasize our societal needs for network survivability.

So far our discussions more or less describe network outages in a "qualitative" sense. The literature on network survivability standards indicates that there has been much work on characterizing outages quantitatively. The Network Performance, Reliability, and Quality of Service Committee (formerly T1A1) of the Alliance for Telecommunications Industry Solutions (ATIS) is one of the leading organizations that develop standards for quantifying network outages. They propose a framework for quantifying the severity of failure events and methods for calculating the "outage index" considering different types of networks (such as wireline, wireline, cable TV and satellite) that carry telephony services [T1A01]. Generally speaking, this outage index takes the following data items into consideration: number of customers potentially affected, outage duration, start time (i.e., time of day the outage began), and the types of services affected (i.e., intraLATA intraoffice, intraLATA interoffice, interLATA interoffice and 911 services). The aim of creating the index is to provide network carriers and users a common ground for interpreting the overall severity of an outage[3]. This idea is similar to that of the Richter scale, which provides a rough estimation of the actual impact of an earthquake while hiding all the technical details from the mass population. One must realize that the proposed index calculation is limited to voice traffic but does not take data traffic into account. Thus, to determine what kind of additional metrics should be incorporated in today's multi-service, data-centric environment might be a challenging future research topic.

Once the network failures and their potential impacts have been understood, quantified, and prioritized, the ultimate objective is to develop survivability strategies that would proactively prevent and reactively recover the failures, so that the impact from such failures could be minimized. Although this thesis considers network survivability strategies strictly from a capacity planning perspective, the reader should be informed that a *complete* network survivability strategy always require that we consider the operational issues (e.g., restoration procedures from fault identification to recovery actions, training personnel and field engineers on network-level trouble shooting, cooperation among network administrators, efficient design of operation support systems) as well as implementation aspects (e.g., types of cross-connect

---

[3] Two examples of outage index calculation are included in the Appendix D [T1A01].

employed, type of fiber used [ChV98], collaborations among multi-vendor, multi-technology network management systems [Mis04]). Only a high level of coordination between these solutions could provide truly survivable network infrastructure and sustainability of services.

2.3.1 Survivability Planning for Fiber Cuts

If we are asked to identify the causes or potential vulnerability of fiber-based backbone networks, *cable cuts* are the most frequent. In 1992, the Federal Communications Commission (FCC) required exchange and inter-exchange service providers to begin reporting outages of 30 minutes or more that potentially affected 30,000 users. Outages affecting major airports, 911 service, nuclear power plants, major military installations and key government facilities must be reported regardless of the number of customers affected. The FCC has published a report assembling approximately one year of outage statistics (failures occurred between March 1, 1992 and February 4, 1993 [NRI93]) showing that fiber cuts represent a majority of outages with great impacts. Some of the findings in that report are shown in Figure 2.1 to 2.4 inclusive [DaK95].

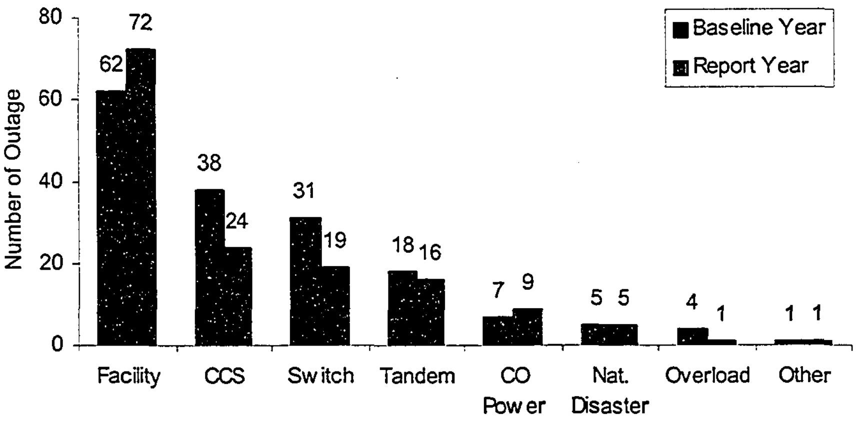


Figure 2.1. Number of outages by failure category. From [DaK95].

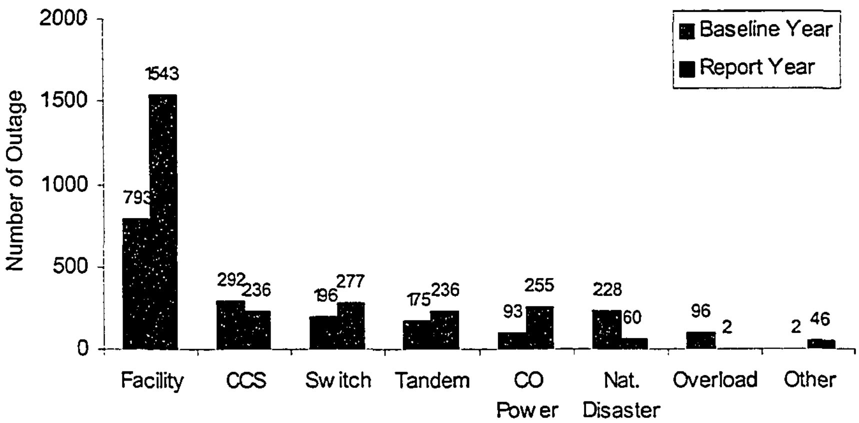


Figure 2.2. Outage index by failure category. From [DaK95].

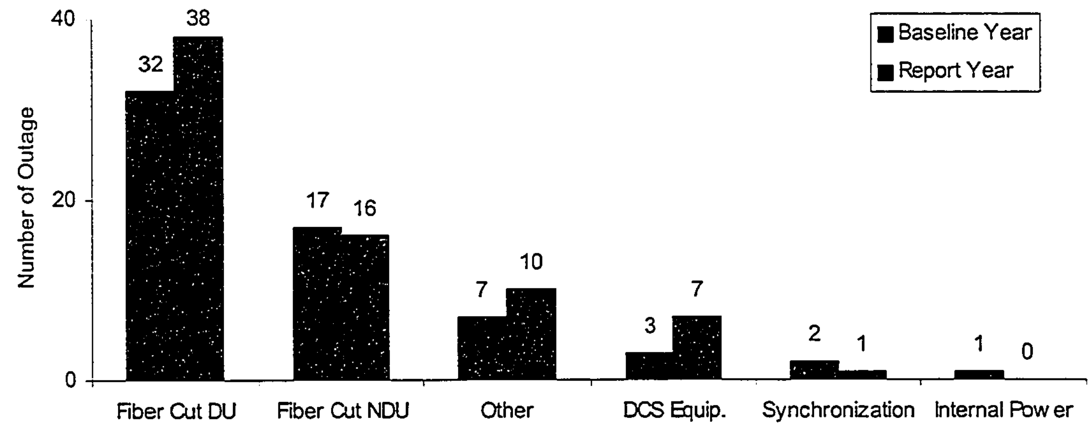


Figure 2.3. Number of outages within the "Facility" failure sub-category. From [DaK95].

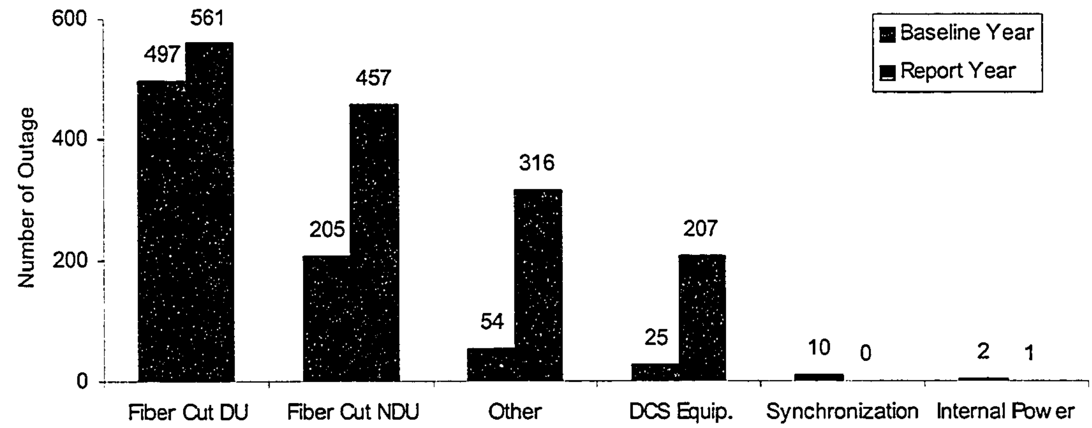


Figure 2.4. Outage index within the "Facility" failure sub-category. From [DaK95].

Figures 2.1 and 2.2 categorize the occurrences of the outages and their indices (as described in the previous section) by various types of failures. The frequency and impact of the "facility" outage are the highest among all categories. Figures 2.3 and 2.4 show the further breakdown within the facility category. We can see that fiber cut dig-ups (DU) and fiber cut non dig-ups (NDU) are the major contributors to both outage severity and frequency (as high as 54 fiber cuts within a one-year period). This is much higher than damage due to cross-connect equipment or power failures, where stringent redundancy requirement might have been enforced by FCC at that time. In the United States alone, FCC published findings that metropolitan networks annually experience 13 cuts for every 1000 miles of fiber, and long-haul networks experience 3 cuts for 1000 miles of fiber [VeP02]. Similarly, the statistics of network failures at Bell Canada has reported that cable cuts is the type of failure that occurs most frequently [FOE89]. The original report by Crawford [NRI93] provides additional details of the root causes – such as digging error, inadequate notification, inaccurate location, shallow cable – to fiber cable dig-ups and recommends best practices for reducing their occurrences. These practices cover the areas of engineering and construction, the so-called "call-before-you-dig" as standard operating procedure, effective cable location and other preventive measures.

In the past decade, despite the fact that fiber optic construction companies and network carriers have done their best to implement these best practices, fiber failures have continued to occur. Indeed, these failures were sometimes beyond the control of the network operators.

On September 21, 1999, an undersea fiber cable near Taiwan in the Asia-Pacific Cable Network (APCN) was damaged by an earthquake. Subscribers were unable to access sites outside Singapore, and several electronic commerce and portal companies, which rely on U.S. sites to provide content such as banner advertisements, were affected for several hours[4].

On November 21, 2000, one of the world's busiest and longest undersea cables, SEA-ME-WE 3 undersea communications cable, was cut 40 miles off the coast of Singapore, 80 feet underwater. The cause was unclear, but officials believed it was a ship's anchor or a sand dredge that damaged the cable. This failure caused traffic disruption on hundreds of Australian Internet service providers, including Telstra, Australia's largest ISP with more than 650,000 subscribers that rely on the cable for 60 per cent of its traffic for international access. This failure also took about a week to be fixed and for Internet traffic to return to normal[5].

On February 9, 2001, a major undersea fiber optic cable linking China to Japan and the U.S was severed by a fishing vessel. It is the only direct fiber-optic link between China and the US. The cable houses bandwidth used by international and domestic carriers in China. And because the resilient loop construction has fallen behind schedule, a cable ship was sent to repair the cable. It took about ten days before the connectivity was fully recovered[6].

On June 13, 2002, a fiber optic line owned by Ameritech Michigan was down in Berrien County, Michigan. The cause of the failure is unknown, but certainly many modem and ISDN Internet customers in the 616 and 231 area codes lost connectivity for more than 14 hours. The down line also disabled four out of five 911 dispatch centers for up to five hours[7].

Whether it is human error or an external factor that causes the fiber cuts, the economic and societal impacts due to the failures give them a top priority to be addressed. Indeed, the significance of protecting fiber cuts will grow as traffic-carrying capability (aggregate) of a fiber

---

[4] Source: IDG News Service, Singapore Bureau, www.idg.com.sg

[5] Source: Mercury News, www.mercurynews.com

[6] Source: People's Daily Online, english.people.com.cn

[7] Source: Discount Long Distance Digest, www.thedigest.com

and the value of the information being transported over these cables increases. Our scope of the thesis is to focus on the capacity planning strategies to provide network survivability against fiber cuts. As discussed previously, equally important issues related to implementation and operational procedures should also be considered to provide a complete survivability solution.

Before proceeding to the next section, let us present an example that illustrates several central ideas of the capacity planning problem of survivable networks. The intention here is to provide a preview and head start to the specific studies in Chapters 5 to 8.

Assume we are in the position of a transport capacity planner. Our goal is to determine the capacity requirement for supporting a given demand forecast (in a matrix) as shown in Figure 2.5. Each entry in the demand matrix represents aggregated, bi-direction traffic flows between any two cities. In addition, a physical network topology is provided. Currently we must solve two questions:

Table 2.2. A sample point-to-point transport demand matrix. Traffic is measured in Gbps.

Traffic	City 1	City 2	City 3	City 4	City 5	City 6	City 7	City 8
City 1	-	3	1	3	2	5	2	3
City 2	-	-	4	2	8	6	3	3
City 3	-	-	-	7	2	3	3	6
City 4	-	-	-	5	2	2	2	2
City 5	-	-	-	-	2	5	5	5
City 6	-	-	-	-	-	3	6	2
City 7	-	-	-	-	-	-	1	3
City 8	-	-	-	-	-	-	-	4

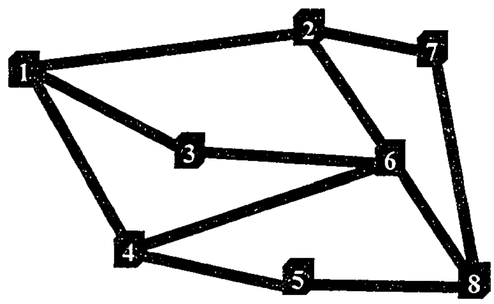


Figure 2.5. A sample network topology.

(1) If we want to support all the predicted demand, how many units of transport capacity should we allocate on each span in the network?

(2) Protection against fibers cuts is also a requirement of the capacity design. In this case, how many extra capacity units do we need to protect against any fiber cut? And where should we place them?

One approach to the first question is to determine a routing plan (e.g., based on the shortest distance route) between each city-pair, and route each demand entry onto the network map. On each span, all demand flows can then be summed up to determine the corresponding total capacity. To answer the second question, we might simply scale the existing capacity units by some constant factor X, and hopefully that will provide enough redundant capacity for re-routing the potentially failed traffic.

The above solutions could undoubtedly provide feasible answers. However, with cost-effectiveness in mind we would try to minimize the total capacity required on the network. In such a case, how do we justify that they are indeed "good" solutions? How should we determine the appropriate factor X? While we will re-visit this problem in much greater detail, this simple example enables us to illustrate the economic objectives as well as the aspect of survivability from a capacity design perspective.

2.4 Demand Forecasting and Uncertainty

Unlike network survivability, which has a relatively well-defined scope and definition, the notion of uncertainty in telecommunications is more subtle and difficult to conceptualize. In this section, we will first look at demand forecast and the possible causes of demand uncertainty. In Chapter 4, we will then classify uncertainty specifically into four levels and relate this concept to the problems of capacity planning and capacity management.

2.4.1 Demand Forecasting

Before we describe uncertainty in demand forecasts, it helps to first discuss how a forecast is developed for transport network planning. Specifically, how do we obtain Table 2.2 in the first place? Based on three application types, namely voice, transaction data and Internet traffic, Dwivedi and Wagner [DwW00] develop a traffic forecasting model for the design of long-distance optical transport networks. Each traffic type has different characteristics and has its own formula for modeling the traffic between cities i and j:

$$\begin{aligned}
&\textit{Voice traffic}(i,j) &&= \frac{K_V \cdot P_i \cdot P_j}{D_{ij}} \\
&\textit{Transaction data traffic}(i,j) &&= \frac{K_T \cdot E_i \cdot E_j}{\sqrt{D_{ij}}} \qquad (2.4.1) \\
&\textit{Internet traffic}(i,j) &&= K_I \cdot H_i \cdot H_j
\end{aligned}$$

P_i is the total population of city i, E_i is the non-production business employees, H_i is the number of internet hosts, and D_{ij} is the distance between the two cities of interest. K_v, K_T and K_I are the proportionality constants that define, respectively, the traffic levels assuming 14 minutes of long-distance voice traffic per person per day, 5 minutes of transaction modem use per non-production employee per day and 25 minutes of continuous modem access to the internet per host per day. Based on these formulas, the authors have estimated the projected demands for each application, aggregated them together to form a complete traffic forecast (in Gbps), and cross-checked with statistics published by FCC. From (2.4.1), it is interesting to note that the influence of the geographic distance on the generated traffic reduces as we move from voice to transaction data to the Internet traffic. Kazovsky et al. conducted a similar study to [KKD98] and considered four types of applications (i.e., telephony, internet, digital video distribution and digital video communications). Nonetheless, the exact mathematics of forecasting models was not reported.

The above method represents one kind of forecasting technique, and in [Lee86], Lee discusses several kinds of forecasting methods: Intuitive forecasting, Trend forecasting, Normative forecasting, Iterative forecasting and Comparison forecasting are the popular ones. We summarize only the first three here.

Intuitive forecasting, also known as the Delphi method, is a qualitative forecasting method based on independent inputs of selected experts, who contribute their subjective opinions on a particular issue. Typically, a series of questionnaires is provided to the experts, and they are encouraged to repetitively revise their answers in several round of questioning until they reach a consensus. It is believed that by going through several rounds of revisions, the group will eventually reduce the variance of the opinions.

Trend forecasting, or time-series forecasting, assumes that the future will have a predictable relationship with past statistics. An example is the use of least squares curve fitting to find a curve or a mathematical function that fits a set of data points. This function is then extrapolated to produce future demand values and to reflect a plausible relationship between demands and time.

Normative forecasting assumes that there will exist needs in the future, where these needs will directly influence the demands forecast. In other words, we do not use past demand values for prediction, but rather we use other (possibly more reliable) parameters to develop traffic forecasts. These parameters could be population of cities, gross national product (GNP), consumer index, interest rate, tariffs, disposable income per area, etc. The method [DwW00] just described belongs to this forecasting technique.

To provide some detailed examples of the above forecasting methods, readers can refer to the February 1995 issue of IEEE Communications magazine entitled "Traffic and Service Forecasts for the Years 2000 – 2005." Stordahl and Murphy [StM95] use a combination of trend forecasting and Delphi methods to predict the demand forecasts particularly for wideband and broadband services in the European residential market. Hopkins et al. [HLB95] propose several approaches, such as statistical analysis of historical data for existing services, modeling of the diffusion of demand for new services, assessment of overall telecommunications spend as a proportion of gross domestic product (GDP), to predict the trend of broadband services in the United Kingdom. Wasem et al. [WGT95] explain a two-step solution for predicting broadband demand between geographic areas. Using demographics and market research data, the authors first produce an aggregate forecast for each data service, and for each geographical area. Then, based on the assumption that demand flow among industry groups is proportional to the flow of money between industry groups, these aggregate forecasts are combined to determine forecasts of demand within and between the geographical areas. The following example can be used to illustrate the key concepts of this approach:

$$\begin{array}{c|ccc} & Edm & Cal & Van \\ \hline Edm & x & 0 & 0 \\ Cal & 0 & y & 0 \\ Van & 0 & 0 & z \end{array} \tag{2.4.2}$$

$$\begin{array}{c|ccc} & Edm & Cal & Van \\ \hline Edm & 0.3 & 0.5 & 0.2 \\ Cal & 0.3 & 0.4 & 0.3 \\ Van & 0.1 & 0.4 & 0.5 \end{array} \tag{2.4.3}$$

$$\begin{array}{c|ccc} & Edm & Cal & Van \\ \hline Edm & 0.3x & 0.5x & 0.2x \\ Cal & 0.3y & 0.4y & 0.3y \\ Van & 0.1z & 0.4z & 0.5z \end{array} \tag{2.4.4}$$

Matrix (2.4.2) is the result from step one, where the aggregated forecast demand, x, y and z, for *each* geographical area is found. [WGT95] goes into the detailed methodology of transforming demand data from number of devices currently installed (total of 12 device categories are considered), to application demand (total of 14 applications), and eventually to service demand (total of 11 services), say in Gbps. Matrix (2.4.3) comes from step two, where each entry represents the "proportion" of money flows between the node pairs. For example, the first row implies that 30% of the money flows within Edmonton, 50% to Calgary and 20% to Vancouver, and the row sums to unity. Multiplying the two matrices leads to the matrix (2.4.4), that is, the model of the demand forecast we assume throughout this thesis.

2.4.2 Macro Factors Causing Demand Uncertainty

Although practicing forecasters have proposed numerous techniques for making accurate demand prediction, only a few techniques have been proven to be successful for a specific time frame, or for a specific application [FiK02]. Indeed, both the causes and impacts of uncertainty are difficult to identify and, at times, difficult to quantify. If we examine the telecommunications industry from a broad perspective, we will see that uncertainties exist not only in the mathematically-detailed level, but also in regulatory, economic and technological levels [BeW93]. As in the case of regulation, operating a network in a deregulated, competitive environment or in a government-owned monopoly environment (e.g., as happened during the 1980's [Sha00]) would have different impacts on how customers choose between carriers and hence would affect subsequent traffic usage and pricing policies. Consequently, the pricing structure of telecommunications services and price elasticity of demand would influence user reactions and how much demand would be exchanged between cities, etc.

On the technological side, the rapid innovation and delivery of new communications services and the limited amount of historical data available have constantly imposed challenges to traditional forecasting techniques, as we explained in the previous section. In the October 2002 issue of the International Journal of Forecasting, Fildes and Kumar [FiK02], Madden et al. [MSC02] and Islam et al. [IFM02] provide reasons on why traditional forecasting techniques fail. For example, lack of historical data due to rapid developments of new technologies forms a constant barrier to adopting any kinds of time-series or normative forecasting techniques. The demise of the monopolistic service providers imposes another challenge to practicing forecasters [FiK02]. Indeed, the traffic characteristics of telecommunication services have changed dramatically over the years. Chung et al. [CCF98] have observed that call-holding time for a typical voice call has increased from 3 minutes to over 30 minutes for an Internet session; daily traffic patterns from predictable busy and off-peak hours have become busy at almost all times [CCF98]; traffic flow has changed from a geography-dependent pattern to a geography-independent pattern [CCF98][DwW00][Dwi03].

As expected, uncertainties in regulation, economy and technology will continuously affect the predictability of the demand forecast. The questions of what variables to model and how each variable affects the others and relates to the demand forecast deserve in-depth research in its own right. In this thesis, our intent is not to invent new forecasting techniques (as discussed in Section 2.4.1) to produce a more accurate demand forecast, but we treat demand uncertainty as some input (e.g., from a perfect forecast to a range of future scenarios) to the model. Our main goal is to provide a set of strategies for decision makers to choose from, and to apply each strategy in accordance with uncertain situations.

2.4.2 Capacity Planning and Capacity Management under Uncertainty

Capacity planning has always been a central problem of many businesses. It generally refers to a problem of ensuring sufficient capacity and a cost-effective plan to meet anticipated demands. Capacity management, which requires a different emphasis from an operational standpoint, considers the management of capacity resources to ensure that the capacity is well utilized for existing and upcoming demands.

In the general context of transport network planning, a time scale is an effective metric of distinguishing among different types or categories of problems. A report by Eurescom, one of the leading European telecommunications organizations, has classified planning of optical networks into three categories: long-term planning (LTP), medium-term planning (MTP) and short-term planning (STP) [EUR00a]. A LTP problem is typically characterized by a long planning period and a high level of uncertainty. Topological (e.g., determine the physical network topology, location of network nodes) and technological (e.g., select appropriate types of transport technologies, choose between mesh versus ring network architecture) as well as physical dimensioning (e.g., determine the number of cables needed, power requirement) decisions are addressed in LTP. Deciding whether to introduce a new node or a new physical span on a given network topology would be another example of long-term issues. LTP is also referred as "greenfield" or fundamental planning [Gro04], and is the most strategic and capital-intensive of the three categories. The business aspect of network planning as mentioned in Section 2.2, as well as the regulatory and organizational issues discussed in the previous section, directly affect this problem space.

MTP concerns installation, allocation, and upgrading of transport equipment systems to support demand forecast in a moderate degree of uncertainty. MTP decisions might be made in a single period or in multiple periods. To determine where to install new capacity modules in a particular fiber route to support upcoming traffic growth is one example of the MTP problem. STP departs from MTP in that routing of the transmission demands must be satisfied within the already-installed capacities, where additional capacity investment is not an option. The time period considered in STP is much less than MTP. Our earlier descriptions of capacity planning would normally fall within the scope of MTP, whereas the notion of capacity management could span the scopes MTP and STP, depending on whether we allow "add or remove" to the existing capacity in the network.

LTP, MTP and STP clearly address a variety of questions and issues, and different decision-making tools should be used to address different problems. The demand forecast is a common input parameter for all three, and the associated degree of uncertainty affects all planning processes. In LTP, the demand forecast, which is usually used as the basis of a revenue forecast, provides a key piece of information for decision makers to determine an "invest or not to invest" type of highly strategic question. Along the lines of this decision, the degree of uncertainty of a demand forecast would affect the timing of the investment, essentially posing a question of "when to invest." Combining uncertainty with certain financial values (e.g., project's current net cashflow, time over which the decision may be made, risk-free rate, variance of present value of project's cashflow), the real option technique[8] might suggest whether the capacity investment should be deferred, executed, committed in stages, or even abandoned altogether [AlN99][All02].

In MTP, we have a different problem scope. Here we need to determine where, when and how much transport capacity and equipment to install on the network, and both the *demand level* as well as the *distribution* of a demand forecast are the sensitive parameters to the network implementation. In contrast to the demand forecast used in LTP, the demand prediction of MTP is usually performed on an annual basis. Therefore, MTP is relatively more accurate and closer to the actual traffic scenario. In addition, the demand forecast of MTP contains more detail in terms of the traffic distribution between the nodes (or cities), and optimization and simulation-based techniques are typically used to support these capacity allocation decisions. Note that the decisions made in the MTP can also affect LTP process strategically. For example, a significant increase or decrease in the cost of the transport capacity could directly affect overall network cost and therefore the decisions of whether or when to make the capacity investment.

The issues of re-configuration and re-optimization of the network to adapt near-term traffic (i.e., traffic requests predicted a month or more in advance) fall within the scope of STP. Here both distribution and demand level of the forecast are usually unknown but bounded within some ranges, and the goal is to determine efficient ways to re-route or re-organize some existing connections in such a way that existing capacity can be better utilized. As in the case of the

---

[8] Further discussion on real options will be provided in Section 9.3.

relationships between MTP and LTP, decisions from STP can also affect MTP to some extent. For instance, a reconfiguration process might suggest putting more capacity on certain spans within an existing network.

Table 2.3 provides a concise summary of the three planning processes, and serves as a conceptual framework for Chapters 5 to 8. Our focus in this thesis falls directly within the MTP and STP categories, and we refer these two problems as the capacity planning and capacity management problem, respectively.

Table 2.3. Illustrating the differences among long-term, medium-term, and short-term planning.

Planning Horizon	**Long Term Planning (LTP)**	**Medium Term Planning (MTP)**	**Short Term Planning (STP)**
Forecast Period	Long (e.g., 3+ years)	Medium (e.g., 1 – 3 year)	Short (e.g., every month or week)
Forecast Uncertainty	High	Moderate	Low
Important Forecast Characteristics	Demand level or demand volume	Demand level and demand distribution	Demand level and demand distribution
Problem Nature	Business / Strategic	Strategic / Tactical	Tactical / Operational
Problem Addressed	Investment decision; go vs. no-go decision and why; buy vs. lease capacity; optimal timing to make investment; network topology, node location	Where, how much, when, what capacity modules to be installed, upgraded, removed	Demand routing decision; traffic management; reconfiguration, routing re-optimization policies within a given set of capacity

We conclude this section by re-examining the capacity planning problem that we addressed at the end of Section 2.3.1. The intention here is to show the overall problem of planning survivable transport networks under demand uncertainty.

Suppose we already had a capacity design that ensures (1) all predicted demands in Table 2.2 are served, and (2) all demands are protected against span failures, and this design costs total of *Y* dollars. Recognizing that there might be several plausible demand forecasts, instead of a single forecast we initially had expected, we must now address two additional questions besides (1) and (2):

(3) Given a set of possible demand scenarios, how much extra capacity do we need to "future-proof" our capacity design? In other words, how much contingency

(e.g., an additional of 20%, 50% or 200% of Y) should we incorporate into the initial capacity design?

(4) If the actual demand goes far below or exceeds our original design, what will be the financial or other impacts? If we build the original network, is there a way to re-configure the existing capacity configuration to improve its ability to cope with actual future demands?

An answer to question (3) would affect the initial cost of the network directly. On one hand, we do not want to put in too much extra capacity up-front; on the other hand, an under-capacitated design might lead to the expensive penalty of not serving or coping with the future demands. Therefore, there should exist some optimal balance that will determine the amount of capacity we invest now or in the future. Question (4) brings up a different issue. It almost forces us to accept the fact that we will be making an inaccurate capacity investment in any case. Here our strategy is to take whatever capacity configuration we have, and find ways to maximize its value. These four questions are fundamental to this thesis, and solutions will be treated in the following chapters.

2.5 Summary

From the fundamental capacity design problem, we have touched on and united different issues of economics, network survivability and demand uncertainty. This chapter obviously cannot cover each subject area in detail, but the facts and figures should give readers general background to justify why this thesis is a timely and interesting one.

3 Modeling Survivable Transport Networks

3.1 Introduction

Chapters 3 and 4 constitute the second module of this thesis. Containing concepts in transport networking and background on network survivability, this chapter aims to provide a modeling foundation for optimization formulations, specifically, *what* variables and parameters to model. We begin by explaining the concepts of layering, transport node, transport demands, working and spare capacities, as well as possible network topologies for the design of survivable transport networks. The key survivability mechanisms, span restoration, *p*-cycles and shared backup path protection, are also described from the capacity modeling and routing behavioral perspective. In this chapter our focus is to identify what to model, and in Chapter 4, mathematical programming techniques will be explained concerning the question of how to put the decision variables and parameters together.

3.2 Model of Transport Networks

In the most general sense, a transport network is a network that can be used to facilitate the movement of people, goods or information from one place to another. While our research only considers the transport of information over circuit-based telecommunications networks, terms such as topologies, nodes, links, routes, demands, and capacities are common in many types of facility-based networks including rail transport, road transport and air transport. In this section, we will discuss each of these terms with particular reference to optical transport networks and, based on their characteristics, we identify the key parameters and variables of general capacity planning problems.

3.2.1 The Concepts of Layering, Transport Node, Transport Demand, Working and Spare Capacity

To explain the concept of layering, it is useful to first present a general hierarchy of the access, metropolitan and long-haul networks, as shown in Figure 3.1. Starting from the end users' proximity, the **access networks** are responsible for providing a variety of services, with data rates ranging from leased lines of 1.5 Mbps to full wavelength capacities of several gigabits per second, to residential users as well as large private corporations, governments, and educational institutions. To serve the needs of a wide range of customers, access networks offer a wide range of broadband services, such as Internet, telephony, cable television, etc. to more enterprise-specific services such as FDDI (Fiber Distributed Data Interface), ESCON (Enterprise System Connectivity) and Fiber Channel.

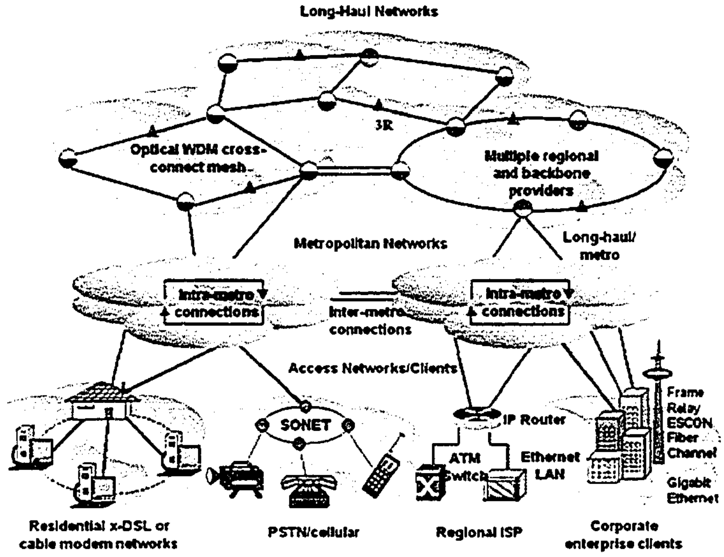


Figure 3.1. Long-haul, metropolitan, access network hierarchy. From [Sor00].

While the access network is responsible for delivering services from the service provider's facility to user's homes or businesses, **metropolitan (or metro) networks** generally provide connections between the businesses and offices within cities, as well as connect to/from points of presence of long-haul networks. In terms of transmission media, the technologies used by the metro area (and long-haul networks) are predominantly fiber-based to carry the aggregation of various kinds of services, while the media of access networks can be wireless, optical or copper.

Also commonly known as "backbone" or the "carrier's carrier" networks, **long-haul networks** connect large trans-national and global carriers, and their coverage spans both regional and international regions. Figures 3.2 and 3.3 show the network maps from two long-haul network carriers [Rea04][Glo05]. As in the case of metro networks, long-haul networks can be used to transport data services with rates up to the order of hundreds of Gbps. To efficiently and cost-effectively provide bulk connections from one point to another is always one of the main objectives of designing the long-haul networks.

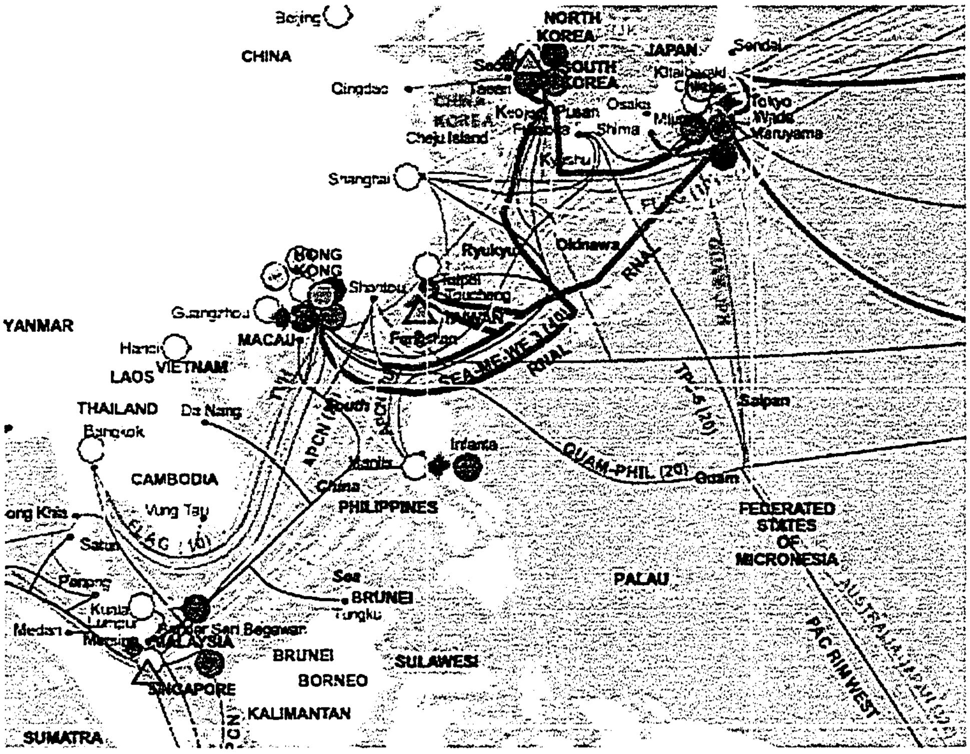


Figure 3.2. Asian network map from Reach, one of the Asian largest international carriers. From [Rea04].

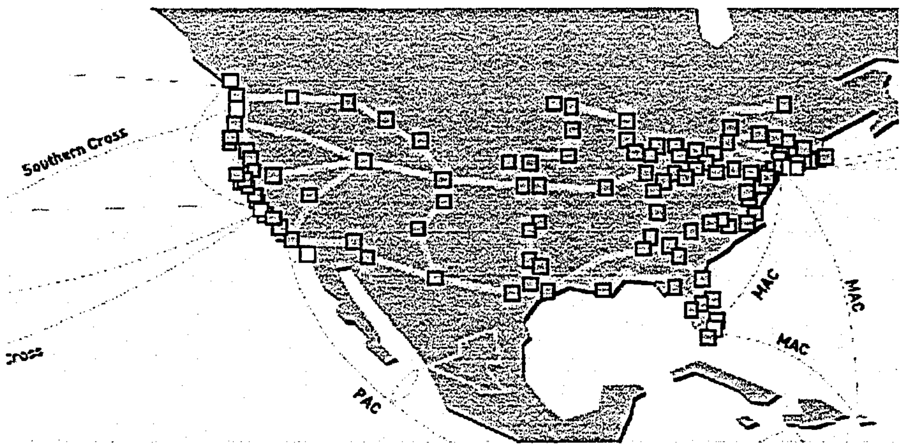


Figure 3.3. North American network map from Global Crossing, a US-based network carrier. From [Glo05].

While we can classify networks in terms of their physical and geographic coverage, from a modeling perspective, we can think of the network architecture as consisting of two layers: The bottom transport network layer, is responsible for providing connections to the upper logical layer, or service network layer [Sat96]. The service network layer (or simply service layer) consists of various kinds of data service networks, and each of which is dedicated to a specific service, from connection-oriented traffic such as voice and video, to bursty data traffic, such as the Internet. The transport layer, on the other hand, provides the fabric of transmission "pipes" that transport the *aggregated* flows of the service traffic. Represented by a set of nodes and spans, the transport network is capable of switching, multiplexing, routing and providing survivability to withstand physical failures.

From a technology-specific standpoint, we can further break down the **service layer** down to a typical three-layer, IP/ATM/SONET structure. As the top layer, today's *Internet Protocol* (IP) is probably the most commonly used wide-area networking technology, as this packet-based technology is designed to support a wide variety of applications such as the Internet. IP is capable of running over many kinds of networks below it, such as the *Asynchronous Transfer Mode* (ATM) layer [Mau99]. ATM is an international standard design for which multiple service types such as voice, video, and data can be transported over a unified platform. ATM attempts to resolve the conflict between circuit-switched networks and packet-switched networks by mapping both bit-streams and packet-streams onto a stream of small fixed-sized "cell" of 53 bytes, which is a compromise between the conflicting requirements of voice and data applications [RaS02]. To overcome the best-effort, indeterminate performance nature of IP transport, one of the key advantages of ATM is its ability to provide quality-of-service (QoS) guarantees, or guaranteed performance, throughput, latency bounds, to applications such as streaming multimedia, IP telephony and mission-critical applications.

One layer below ATM is typically the *Synchronous Optical Network* (SONET) layer. SONET is the North American standard for communicating digital information over an optical fiber, and has been used as a transport technology in metro and long-haul networks for more than two decades. One of the main characteristics of SONET is its ability to access any tributary or low-bit-rate signal without demultiplexing the entire high-bit-rate transmission signal, where an atomic (or master) reference clock is used to synchronize all sources of tributary signals. Ease of signal extraction was not always possible with the earlier transport technology based on Plesiochronous Digital Hierarchy (PDH), where different parts of the network were not perfectly synchronized. Other important advantages of SONET include its compatibility with equipment from various vendors, its ability to offer operation, administration, maintenance and provisioning (OAM&P) [9] and survivability functions, and its compatibility with any service mix including

[9] Operation, Administration, Maintenance and Provisioning (OAM&P) is a general term used to describe a group of management functions that provide system or network fault indication, performance monitoring, security management, diagnostic functions, configuration and user provisioning [AiP94].

ATM, IP and emerging services. The basic SONET signal operates at 51.84 Mbps and is designated as STS-1 or synchronous transport level-one signal. The STS-1 is the basic building block signal of SONET, and higher rates can also be time-division multiplexed (TDM) to form the next levels of the SONET hierarchy as STS-n, where n is an exact multiple of 51.84 Mbps. For further discussion on the IP, ATM and SONET technologies, readers can refer to [Sat96][Nor96][Mau99][RaS02].

In contrast to the service layer, the **transport network** is not service- or client-specific, for it serves as a united platform to carry various kinds of networks. To maximize transport efficiency, the transport layer is responsible for providing transmission paths that have the flexibility of adapting to unpredictable demand growth, and have built-in survivability functions for handling physical network failures. With the invention of wavelength division multiplexing (WDM) technology and node equipment having robust switching and management capabilities, the fiber optical technology has been adopted by the transport networks to support existing IP, ATM, SONET and emerging services such as wavelength services, fiber connection (FICON) and Gigabit Ethernet (GbE). For these reasons, the transport network is frequently referred to as the **optical transport network** (OTN).

As a core transport technology, WDM is a multiplexing scheme that allows several optical carrier signals to be simultaneously sent along a single fiber by using different optical frequencies (or wavelengths). Conceptually, this scheme is identical to the frequency division multiplexing (FDM) used in microwave radio and satellite systems. The wavelengths in both schemes must be properly spaced to avoid interchannel interference. Combined with the time division multiplexing (TDM) technique, which provides high-speed transmission on a per time slot channel basis, WDM is able to combine multiple wavelengths and transport traffic at a rate of terabits per second (Tbps) over a single fiber. Thus, when the demand exceeds the capacity in existing fibers, WDM can provide a more cost-effective solution to expand capacity than to install or bury additional fibers, especially in long-haul regions where the cost of transmission cables often dominates the overall network cost. More discussion of the WDM technologies can be found in [Muk00][Gre01][RaS02].

The International Telecommunication Union (ITU) provides a general definition of OTN. Such a network is composed of a set of optical elements connected by optical fiber links, and is able to provide the functions of transport, multiplexing, routing, management, supervision and survivability of optical channels carrying client or service signals [ITU05]. Generally, the optical network elements refer to optical cross-connects (OXC) or optical add-drop multiplexers (OADM). The main functions of these transport nodal devices include: (1) providing interfaces to the service or tributary signals, (2) multiplexing the service signals into wavelengths, (3) routing wavelengths from source to destination node, (4) switching the wavelengths of various frequencies from any input to any output ports, (5) monitoring signal performance and (6) providing fault management functions when network failures occur.

The functionalities of the switching equipment might seem straightforward to implement, yet over the past few years node equipment vendors have tried to compete by advancing their technologies to integrate all these functions at the lowest cost possible. Choosing a technology for optical switch implementation, for example, has been debated in terms of cost and switching architecture. Some vendors argue that network carriers should deploy solely optical, or optical-optical-optical (OOO) switches, so that data can be switched optically to increase the scalability of the data processing rate, while others see the advantage of implementing a more manageable optical-electrical-optical (OEO) switches where the signal undergoes electronic processing between the optical input and output and provides build-in OAM&P functions. Table 3.1 provides a snapshot of the technology tradeoff published in 2002.

Table 3.1. Comparing OEO vs. OOO switching architectures. From [JaB02].

Feature	OEO	OOO
Data format dependence	Yes	No
Cost/space/power independent of rate	No	Yes
Upgradability to higher rate	No	Yes
Subwavelength switching	Yes	Future
Waveband switching	No	Yes
Performance monitoring	Bit error rate	Optical signal degradation
Wavelength conversion	Built in	Currently electronic

As technology continues to evolve, it can be expected that new technology will provide the best of both worlds. At the time of this writing, we have already seen a new switching technology that claims to reduce the cost of OEO processing, yet providing a scalable switching architecture [Mel04]. For the reader's interest, further discussion on this debate can be found in [Sha96][JaB02][IEC04].

Therefore, to make valid assumptions for the modeling of transport networks, it is essential to first understand the implications of the type of equipment. For instance, if none of the switching equipment has OEO wavelength conversion capability, then a connection must use the same wavelength traversing different links on a network. This can limit the full use of the available capacity. Such a limitation imposes a constraint, known as the wavelength continuity constraint [RaS95][Muk97], and would ultimately affect how we formulate our capacity models. Recognizing future technological trends and knowing the capability of current cross-connect equipment, we can assume that transport nodes will be capable of (1) multiplexing various services into wavelengths, (2) switching incoming wavelength signals from any input to any output, and (3) detecting and isolating network link failures. These are also the functional requirements envisioned by ITU concerning future optical transport networks.

To model **transport demands**, we assume that all IP, ATM, FICON, and other services are aggregated into transmission-level demands, quantified by the number of transmission paths or circuits required to carry the services at suitable levels of performance. If the granularity of the demands is in wavelengths, we refer to them as **lightpaths** or **wavelength services.** The demands can also be quantified by the number of time-division connections, e.g., in OC-12 or STS-3 granularity, or in general terms of any number of managed units of transmission capacity that the aggregation of traffic requires. Therefore, from the modeling standpoint, STS-based, OCn-based or wavelength-based demands are logically equivalent as long as they are represented by discrete values. Each demand unit is also differentiated by its origin and destination nodes as well as other requirements such as the level of protection. It is also important to note that these aggregated demands have characteristics very different from those of service-layer traffic. For example, while in telephone traffic we model the stochastic behaviors (e.g., Poisson arrival rate with negative exponential holding time[10]) to determine the size of a trunk group or a server, the stochastic traffic model should not be used for modeling optical demands because each transport demand has its origin and destination, specific survivability requirement and relatively static duration.

From a provisioning standpoint, an OD transport demand can be viewed as a *circuit* whose attributes might include a unique ITU-T compliant circuit name, service order identification number, the customer who owns or leases the circuit, the date the circuit was put into service or is expected to be put into service, the customer account number to which the circuit is billed, and so on [Cis03][Mis04]. Because transport circuits have generally much longer holding time (e.g., in the orders of weeks or months) than service demands and the routing information is often available from network management systems, it is possible to reroute or re-optimize these circuits at regular intervals to enhance the utilization efficiency of the network [SDC94].

---

[10] Statistical characterization of lightpath requirements remains an open-ended issue. While many researchers assume lightpath service connections behave as traditional phone calls (i.e., Poisson arrival with negative exponential holding time) [RaS95][ZaM01][ZJS01][SSS02], others have assumed non-Poisson traffic models in their studies, especially with the objective of quantifying the benefit of placing wavelengths converters over WDM networks [SSA97][SpB98][YRL99].

As in the case of modeling transport nodes and demands, it is essential to identify properties of the **transport capacities**. Figure 3.4 shows a point-to-point WDM transport system. A **channel** is generated by a pair of transmitter and receiver and each channel is dedicated to a specific wavelength. These channels are then multiplexed for transmission, amplified for extending transmission distance and finally demultiplexed for retrieving each channel.

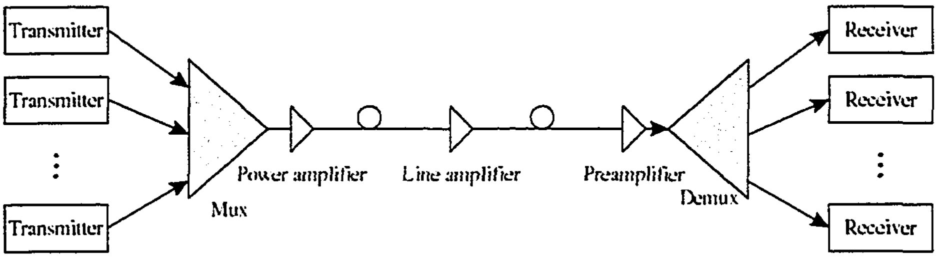


Figure 3.4. Point-to-point transport system. From [RaS02].

From a capacity planning perspective, we often model the aggregation of all point-to-point channels between adjacent cross-connects as a **span**. For the design of survivable networks, certain channels may be assigned or entirely reserved as protection or "spare" capacities, such that in the event of a fiber or node failure, these spare channels can be used to restore the failed connections. For these reasons, the terms **working** and **spare** channels are frequently used to distinguish between active links that carry actual demand units and redundant links that are used for failed demand units[11]. Akin to the idea that a **path** (or lightpath) is a concatenation of individual logical channels (or wavelength channels), a **route** is a concatenation of physical spans on the fiber network map.

The models of transport nodes, spans, demands and transmission paths described above enable us to extract details and capture network management issues such as routing, capacity management and survivability planning. Figure 3.5 summarizes the transport network model that we use throughout this thesis.

[11] In Chapter 7, we will see that both working and spare capacity can be used to carry traffic of different service classes. For now, we will treat spare capacity as the redundant, idle protection channels.

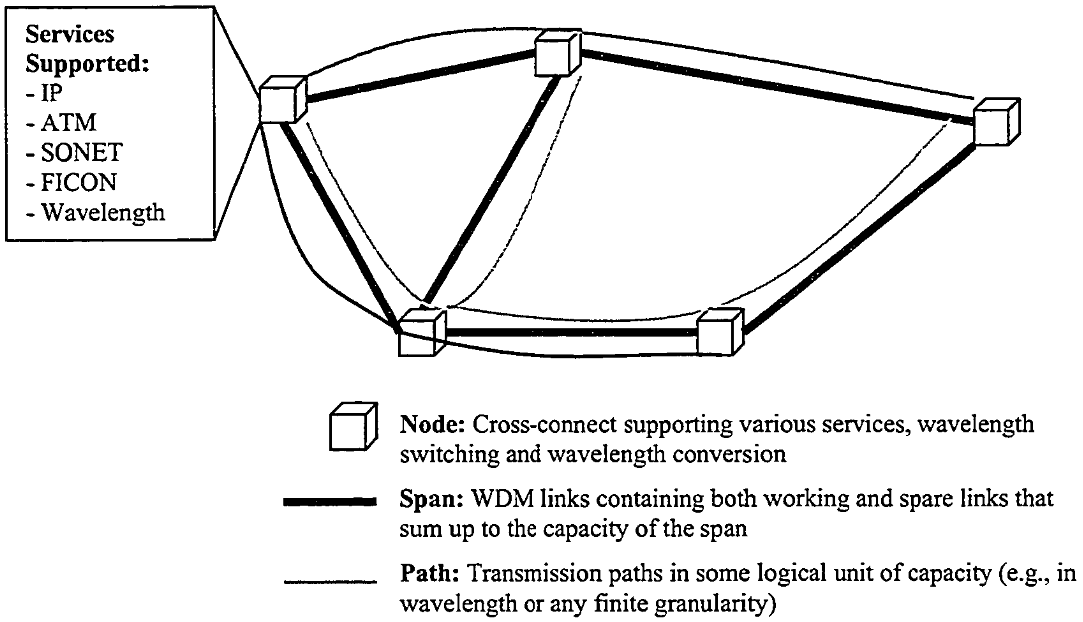


Figure 3.5. Generic survivable transport network model used in this thesis.

3.2.2 Topologies for Survivable Transport Networks

To protect against a cable-based network failure such as a fiber span cut, the physical facilities graph must be designed so that it is inherently survivable. Figures 3.6 to 3.9 illustrate four kinds of network topologies of different survivability implications.

In Figure 3.6 we have a tree-like backbone topology connecting all seven nodes. This topology cannot withstand physical failure because any node or span failure would disconnect the network into two parts. In contrast, the ring-like topology illustrated in Figure 3.7 can survive any node or span failure, because it has at least two fully node-disjointed paths between each node pair. It is often referred as a *bi-connected* topology. There is another, less restricted set of topologies, called *two-connected* topologies, wherein disjoint paths between some nodes are only span-disjoint but not fully disjointed because they might share a common node. Figure 3.8 illustrates an example of the two-connected topology, wherein a failure in node 5 can break the graph into two parts.

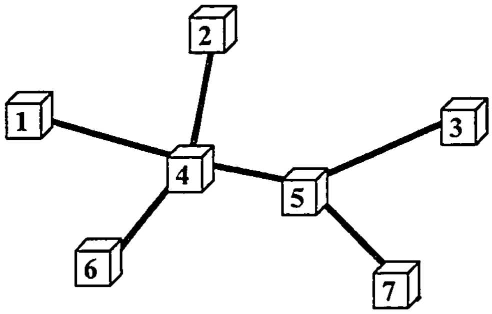


Figure 3.6. Star-like network topology.

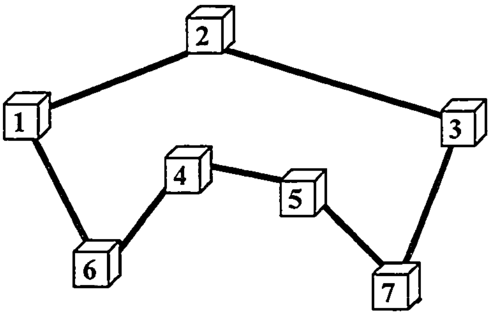


Figure 3.7. Ring topology.

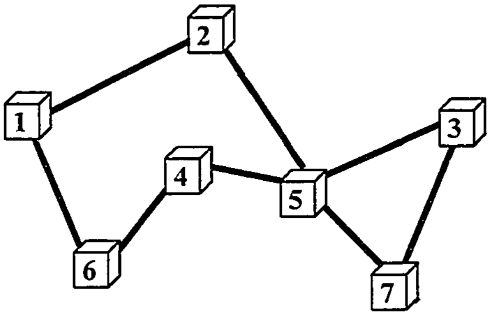


Figure 3.8. Two-connected topology.

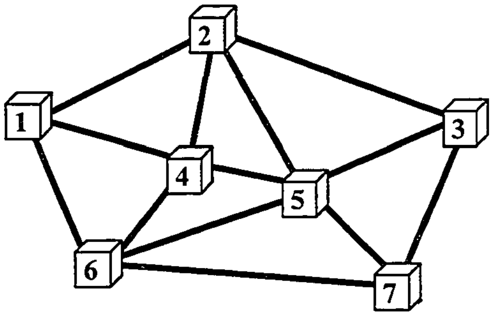


Figure 3.9. Bi-connected topology.

One can further improve the network connectivity, and therefore the inherent survivability, by adding more physical spans to form a mesh-like topology as seen in Figure 3.9. This mesh-like topology is able to provide fully disjoint path pairs between all nodes. An advantage of increasing the route diversity is that this topology allows the more efficient use of transport capacity. However, a network of higher physical connectivity also implies an increase in the overall network cost, owing to factors such as construction cost for laying additional physical cables (including rights-of-way acquisition, installation of ducting, power, etc.), physical cost for installing more intelligent nodal equipment, and operational cost for managing the more complex network. The design of the physical topology is frequently treated as a separate problem, as it includes many historical or geographical factors that cannot be captured or easily quantified by mathematical means. In this thesis, the physical network topology is thus always treated as given. In addition, all networks used in this thesis are at least two-connected in order that they be survivable under any single span failures.

Nodal degree is another common term used to differentiate topologies of various degrees of connectivity. It is equal to 2*S/N, where S is the total number of spans and N is the total number of nodes of any two- or bi-connected network. For instance, Figures 3.7, 3.8 and 3.9 have nodal degrees of 2, 2.3 and 3.7, respectively. As a general observation, the nodal degrees of today's backbone network topologies range approximately from 2.2 to 4.5. North American networks tend to be sparsely connected. In contrast, European and Asian networks are highly connected. The geographical distances between cities might explain such an observation. In the case of metro-networks, ring-based transport architectures are commonly deployed with four or five nodes [LuD89]. Networks with low connectivity seem to favor ring and point-to-point based survivability strategies, as they tend to involve the lowest overall construction and network management costs [DML94]. And as the number of nodes increases, the benefits or practicality of rings decreases [LuD89][TCK90], relative to the kind of mesh networks we consider.

3.3 Model of Mesh-based Survivability Schemes and Performances

Now we will discuss three specific survivability schemes – namely, span restoration (SR), shared backup path protection (SBPP) and *p*-cycles (*p*-cycles). Any of these schemes can be deployed over any mesh topologies with intelligent cross-connect nodes. These are the schemes we consider throughout this thesis. Our goal here is to provide a general sense of the routing behavior of each scheme upon a span failure. Such routing behavior is the fundamental factor that dictates the overall capacity requirement and allocation for a survivable network. Specific advantages, disadvantages, mathematical models and references of each scheme (along with a few other survivability schemes) will be discussed in Section 4.2.

3.3.1 Span Restoration, p-Cycles and Shared Backup Path Protection

Span restoration is one approach to survivability that occurs when a fiber cable is cut in network topology, as illustrated in Figure 3.10. Upon span failure, the end nodes of the failed span react locally and provide a set of detouring paths to reroute the failed channels. For instance, if span [4-5] containing four working channels fails, nodes 4 and 5 might provide a single or a set of paths (i.e., [4-2-5], [4-6-5], [4-1-2-5], [4-6-7-5], etc.) for diverting each failed link in the span. From the capacity modeling perspective, one of our requirements for survivability is that spare capacity links are efficiently allocated (e.g., on span [1-4], [2-4], [4-6], etc.) and sufficient to accommodate the restoration paths upon span failure.

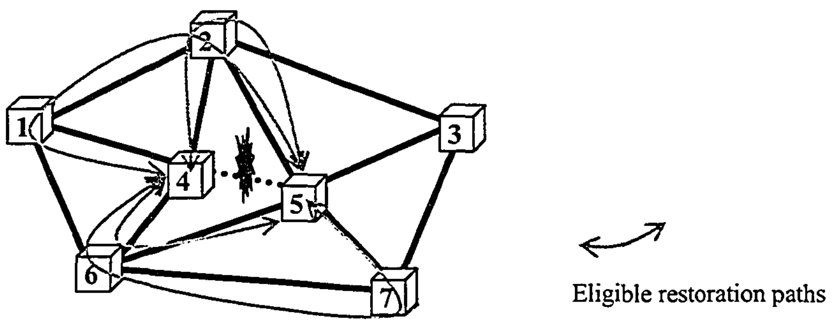


Figure 3.10. Span restoration under a span failure.

As a special type of span restoration scheme, *p*-cycles offer an alternative to SR, using pre-connected spare-capacity structures for speeding up the restoration process. Like SR, *p*-cycles is a localized recovery scheme wherein the end nodes of the immediate failed span are designed to detect, isolate and initiate the restoration process. Unlike SR principle, where spare links on each span can be *freely* used by *any* failed working connection, *p*-cycles pre-connect spare links into a set of cycles and each cycle is restricted to protecting against only a pre-defined set of link failures.

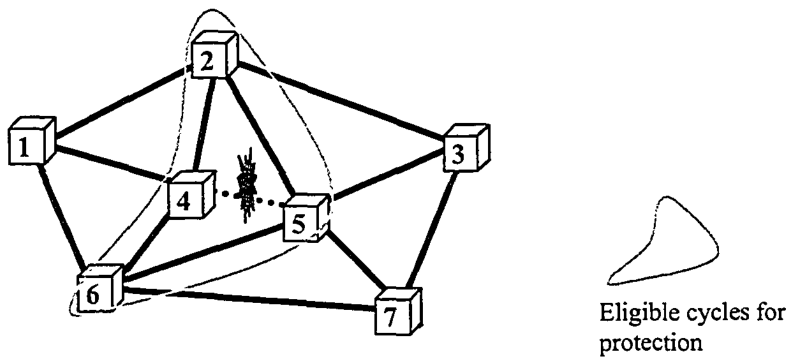


Figure 3.11. *p*-Cycles protection under a span failure.

Figure 3.11 shows a possible single-channel *p*-cycle, [2-4-6-5-2]. This cycle can provide a single restoration path for protecting failed link in either span [2-4], [4-6], [6-5] or [5-2] (i.e., the "on-cycle" failures), or provide two restoration paths [4-2-5] and [4-6-5] for protecting two working channels in span [4-5] (i.e., the "straddling" failure). From the capacity design standpoint, although the pre-connected nature of *p*-cycles may lead to slightly higher redundant designs (i.e., more spare capacities required to protect the same set of working links) than SR, the *p*-cycles has the potential to enable faster restoration process, which is a valuable aspect of any mesh-based restoration schemes.

Path-based restoration schemes provide another class of survivability options. Path restoration (PR) and Shared Backup Path Protection (SBPP) are examples. In contrast to span-based schemes, which use localized recovery actions, path-based schemes recover failed working connections from an end-to-end, origin-destination (OD) node pairs' perspective. When a span failure occurs in a path-protected network, the network management system must first identify which *specific OD pairs* (or working paths) are affected by the failure before the actual restoration process begins. Figure 3.12 illustrates path-based restoration. Upon failure of span [4-5], the end nodes (1,3) and (3,6) are responsible for providing end-to-end restoration paths to recover the failed working paths [1-4-5-3] and [6-4-5-3]. Under the PR scheme, these restoration paths can be any paths connecting the end nodes, without traversing the failed span [4-5]. In

SBPP, which is a specific case of PR, one and only one fully (both node and span) disjoint restoration (or backup) path could be used to recover each failed working path. In terms of spare capacity requirements, the SBPP is generally less efficient than PR, but more efficient than span-based restoration schemes such as SR and *p*-cycles. As far as the "routing behaviors" of these survivability schemes go, we provide additional discussions in Section 4.2.

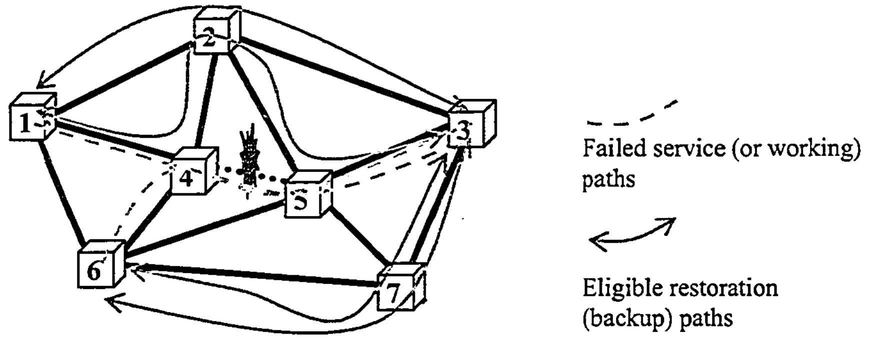


Figure 3.12. Illustration of the path restoration scheme.

3.3.2 Given Occurrence of Failure Model

The judgment of how much redundancy or spare capacity is required for a transport network generally depends on two parameters. The first is the set of anticipated failures that we would like protection from, and the second is the network survivability level that we would like to achieve. As to an insurance investment, where we pay more for a bigger coverage, investment for network survivability has the same concept. At one extreme, we can choose not to put any spare capacity on the network, as there might simply be no reason for network survivability. At the other extreme, one can allocate a huge amount of spare capacity to protect against simultaneous span/node failures or multiple network failures, provided there is a need or benefit for doing so. For the problem of capacity planning, the fundamental issue here is that *the occurrence set of given failures is always assumed*, or deterministic in the sense that the modelers can choose which specific types of failures he/she would like to plan for. In other words, it does not matter exactly how likely each failure is. Consequently, this class of models is referred to as Given Occurrence of Failure (GOF) [T1A01]. Another survivability planning model, Random Occurrence of Failure (ROF), assumes that failures can be characterized by random variables with given probability distribution functions and probabilistic survivability measures such that the availability of the network can then be determined by analytical approximations or by simulations. In this thesis, only the GOF model is considered. In all the studies and discussions, the GOF target is limited to non-simultaneous, single-span failures. Therefore, regardless of the type of survivability schemes used, a fundamental requirement of all survivable network designs is to ensure that there is enough spare capacity in the network to protect against non-simultaneous single span failures.

3.3.3 Design Objectives and Basic Metrics

There are several criteria that we use to quantify and compare the effectiveness of survivable capacity designs. In this section, we describe each of the performance measures that relate specifically to survivable transport networks in the context of capacity planning to a given demand set.

Network redundancy is a measure of architectural efficiency for a survivable network design [Gro04], and it is sometimes referred as the Efficiency Ratio [T1A01]. For a specific set of failure scenarios (e.g., all possible non-simultaneous span failures), a typical definition of network redundancy refers to the amount of spare capacity required to achieve a target level of restoration over the possible network failures, usually expressed as a ratio of total spare channels to total working channels of the network. Mathematically,

$$Redundancy \equiv \frac{\sum_{i \in Span} s_i}{\sum_{i \in Span} w_i}$$

where *Span* is the set of spans in the network, w_i is the number of working channels on span i to accommodate a given demand set under normal conditions, and s_i is the number of spare channels used for protection under failure conditions. This basic definition of redundancy can be extended to reflect various distances or costs of the spans by multiplying both s_i and w_i by a distance-specific coefficient, C_i, for each span i.

The ability to estimate the total asset value or total cost of the transport network is of interest but in academic research, exact cost data are usually not available. **Total capacity**, i.e., $\Sigma w_i + \Sigma s_i$, is therefore often used to give a coarse approximation to the total capacity cost of the network. Because installation, operation and maintenance of transmission are often related to the *distances* of network span (e.g., cost per kilometer of transmission infrastructure) [EUR00b], it is common to use the available geographic distances of the spans, and come up with a better approximation to the **total capacity cost** of the network:

$$Total\ Distance\ Capacity \equiv \sum_{i \in Span} \left[C_i \cdot (w_i + s_i) \right]$$

For the medium-term capacity planning problem (MTP), where our goal is to determine where and how much working and spare capacity we need to place into the network, a typical design objective is to minimize the above total distance-capacity expression.

In addition to finding minimum-cost capacity designs, in situations where the capacity is already in place, it is important to evaluate the robustness of such a design and to quantify the utilization of the network. **Routability and restorability** are common performance metrics used to measure the effectiveness of as-built capacity plans to a given set of demand scenarios.

Given a set of point-to-point demands and a set of working capacity, **routability** measures the percentage of demands that can be served within the fixed capacity set, without taking network survivability into the account. For example, if we have 50 demand units to be served, 80% routability means that 40 units could be served. In parallel to the network redundancy calculation, where we could assign specific cost or distance to each channel, we can assign "priority" to each OD node pair based on their route distances or other specified utility measures. Thus the definition of routability can be extended to:

$$Routability \equiv \frac{\sum_{r \in OD\,pair} A_r \cdot d\_serve_r}{\sum_{r \in OD\,pair} A_r \cdot d\_request_r}$$

where OD pair is the set of origin-destination node pair in the network, $d\_request_r$ is the number of demands requested from each OD pair r, d_serve_r is the number of demands served for each pair r, and A_r corresponds to the weighting factor (e.g., route distances) specifically in relation to each OD pair r.

Restorability [Gro04] or restoration ratio [T1A01] is another quantitative performance measure specific to survivable networks. A simple definition of restorability is the fraction of working channels that can be restored within a given survivable design, specific to a given failure scenario. For example, if a failure scenario X results in 5 (out of 50) working channels to be unrestorable by the span-restorable design, the restorability of this design under this failure scenario is 90%. The same definition can also be applied to a path-based survivable network. We can combine the measure of restorability to each specific failure scenario to a more general measure over all failure-independent scenarios. Later in this thesis, it is a strict requirement that all survivable designs must be fully restorable (i.e., restorability = 1) against any single span failure.

Servability can be thought of as an extended concept of routability, and considers all demands as "protected demands." Given a network with working and spare capacities, servability measures the fraction of demands that can be *both routed and protected*, and it can be a singie measure for comparing *any* survivable networks (including ring-based, SR, SBPP, p-cycles, etc.) in terms of their ability to withstand changes in the demand forecast. This new concept will be discussed further in Chapter 5.

3.4 Summary

In this chapter, we have identified and defined the key elements of a capacity planning design, including transport nodes, spans, demands, paths, routes, spans and links. It is crucial to note the difference between demands in the service- and transport-layer. Uncertainty in the transport demands (or aggregation of the service demands) is what we try to accommodate throughout the thesis. The routing behaviors of span-based and path-based restoration schemes have been discussed, with general comments on their relative capacity efficiencies. Finally, key design objectives – such as network redundancy and total capacity cost, plus performance measures of routability, restorability and servability – are explained.

4 Literature Review

4.1 Introduction

In this chapter, our objective is to review existing mathematical frameworks or tools used to model survivable network design. This literature review is composed of four basic topics. The first topic relates to the capacity planning problem of mesh-based survivable networks. From that, we narrow down to more specific problems of capacity planning under demand uncertainty. The conceptual modeling of uncertainty into four different levels will be given in Section 4.3.1. It is this key framework that ultimately unifies our optimization strategies under uncertainty in coherent ways. Finally, literature related to the topics of demand loading and reconfiguration are offered.

4.2 Capacity Planning for Mesh-based Survivable Transport Networks

For nearly two decades, researchers have proposed various approaches for solving the capacity planning problems of mesh survivable transport networks. The spare capacity placement (SCP) or spare capacity assignment (SCA) problem, in particular, has been a popular research topic. In a SCP problem, the objective is to minimize the amount of the spare capacity placed on transport networks to protect against an assumed set of network failures. Effective allocation of the spare capacity over the network is an essential prerequisite to the functioning of any restoration mechanism.

Sakauchi [SNH90][SOH92], Grover [GVS90][GVB91][VGM93], Doverspike [Dov91][DoW94][DML94][KDP95], Miyazaki [MCK92], Medhi [MeK95] and Herzberg [HeB94][HBU95] are some of the researchers who saw the potential benefit of the cross-connect-based mesh architecture in terms of capacity efficiency and flexibility in restoration routing, compared to traditional route-constrained, dedicated-spare survivability techniques such as Rings and Automatic Protection Switching (APS) systems. These developments inspired subsequent contributions to solving SCP problems with the use of integer / linear programming (IP/ILP) techniques.

Despite detailed differences in the proposed ILP models, all SCP optimization problems share a common objective – to minimize the total amount of spare capacity for protecting a given set of working capacities on spans. Specifically, given a network topology and a set of working capacity w_i on each span i, the goal of the ILP model is to determine *how much* and *where* to place the spare capacity s_i so that (i) the total number (or associated cost) of spare capacity is minimized and (ii) these spare capacities are sufficient to support restoration via replacement paths that reroute the failed working capacity upon a network failure. As described in Section 3.3, the restoration paths are circuit-oriented connections, and each logical path is composed of a concatenation of spare channels.

Various ILP models can be formulated for solving the SCP problem. We can identify three main classes of formulations, namely Node-Arc, Arc-Path, and the Cut-Oriented formulation [KeL01][BMS02][Gro04]. Our goal in this section is to explain the Node-Arc and Arc-Path formulations, which are the most commonly used ILP models for SCP and also most relevant to our thesis work. For the Cut-Oriented approach, readers are referred to [SNH90][SOH92] [VGM93][Gro04].

4.2.1 Node-Arc Formulation for Spare Capacity Placement Problem

The Node-Arc formulation has its origin in a class of problems termed Network Flow Problems. An important concept underlying these types of problems is the notion of a "network flow." Given a transport topology (or an undirected graph in graph theory terminology), a network flow is associated with the edge (or span) over which the flow is transported. For example, flow variable $x_{i,j}$ represents the total flow over the edge between node i and node j. Generally in transportation problems, every node in the network must either be a source node, a sink node or trans-shipment node, whose total incoming flow must equal total outgoing flow. Each flow variable has a directional attribute (i.e., $x_{i,j}$ is different from $x_{j,i}$) and is integer valued. The following integer program formulation is an example of a Node-Arc representation for the SCP problem [Gro04].

Objective: $$\textit{Minimize} \sum_{\forall (i,j)\in S} c_{i,j} \cdot s_{i,j} \tag{4.2.1}$$

Subject to:

$$\sum_{\forall j\in N|(i,j)\in S,(i,j)\neq(s,t)} (x_{i,j}^{s,t} - x_{j,i}^{s,t}) = \begin{cases} w_{s,t} \; \textit{if} \; i = s,t \\ 0 \; \textit{otherwise} \end{cases} \qquad \forall i \in N,\ \forall (s,t) \in S \tag{4.2.2}$$

$$x_{i,j}^{s,t} \le s_{i,j} \qquad \forall (i,j) \neq (s,t) \in S,\ \forall (s,t) \in S \tag{4.2.3}$$

$$x_{i,j}^{s,t} = x_{j,i}^{s,t} \ge 0 \qquad \forall (i,j) \in S,\ \forall (s,t) \in S \tag{4.2.4}$$

$$s_{i,j} \ge 0 \qquad \textit{integer} \ \forall (i,j) \in S \tag{4.2.5}$$

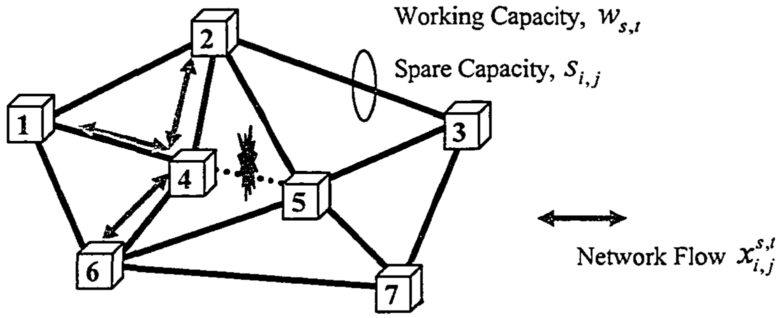


Figure 4.1. Node-arc representation of the SCP problem.

To help appreciate this formulation, we will use the same figure from Section 3.2.2. Given the network topology (with set of nodes, $i \in N$ and set of spans, $(i,j) \in S$) and a set of working capacity $w_{s,t}$, the objective (4.2.1) of the model is to minimize the total cost of the spare capacity, where $c_{i,j}$ is the cost of a unit capacity on span (i,j) and $s_{i,j}$ is the spare capacity assigned to (i,j).

To assure full restoration upon every failed span (s,t), the top part of constraint set (4.2.2) ensures the net total network flow outgoing from source node s (or incoming to sink node t) is equal to the number of working links on the failed span. The variables $x_{i,j}^{s,t}$ represent the network flow from node i to j in response to the failure of span (s,t). Suppose we have 6 working links on the failed span 4-5. We basically ensure that (1) the sum of the flows on spans 1-4, 2-4 and 4-6 must be equal to 6, and (2) the sum of the flows on spans 2-5, 3-5, 6-7, 5-7 must also be equal to 6. The bottom part of (4.2.2) is simply a condition for the tran-shipment nodes (i.e., the nodes other than 4 and 5 in this example) where each must have zero net total flow.

Constraint (4.2.3) determines the spare capacity required on each span (i,j). The inequality sign in this constraint ensures that the spare capacity is dictated by the largest restoration flows across each span over all (s,t) failure scenarios. Finally, constraint (4.2.4) asserts the symmetrical or "bi-directional" nature of restoration flows (as well as working capacity) on each span.

An important characteristic of the Node-Arc SCP formulation is that the decision variables do not directly prescribe the routes of the restoration flow. That is, upon the failure of span (s,t), the variables $x_{i,j}^{s,t}$ and $s_{i,j}$ only tell us that there is enough spare capacity to support the restoration flows on each surviving spans, but they do not explicitly tell us what direction or replacement paths should be taken. The exact restoration route information is a crucial piece of information for circuit provisioning and assurance purposes, and the Node-Arc based formulation has its limitation to capture the routing aspect of transport design.

4.2.2 Arc-Path Formulation for Spare Capacity Placement Problem

In contrast to the Node-Arc approach, where each network flow is described on a span basis, the Arc-Path formulation associates each network flow *explicitly to a given route.* Note that any Node-Arc network flow problem can be formulated in a corresponding Arc-Path version, and vice versa. Thus, the objective of the SCP problem remains unchanged, but the mathematical notions of the variables are now different. Let us use the same example to illustrate the Arc-Path representation of the SCP problem.

Objective: $$\text{Minimize} \sum_{\forall j \in S} c_j \cdot s_j \tag{4.2.6}$$

Subject to:

$$\sum_{p \in P_i} f_i^p = w_i \qquad \forall i \in S \tag{4.2.7}$$

$$s_j \geq \sum_{p \in P_i} \delta_{i,j}^p \cdot f_i^p \qquad \forall (i,j) \in S^2 \mid i \neq j \tag{4.2.8}$$

$$f_i^p \geq 0 \qquad \forall i \in S, \forall p \in P_i \tag{4.2.9}$$

$$s_j \geq 0 \qquad \text{integer } \forall j \in S \tag{4.2.10}$$

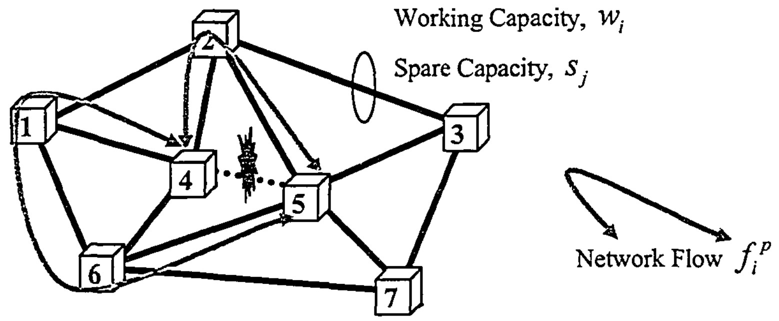


Figure 4.2. Arc-path representation of the SCP problem.

As before, a network topology and a set of working capacities are given. The difference now is that the set of nodes is no longer part of the formulation. A single index, *i* or *j*, is used to describe the set of spans *S*. Therefore, to minimize the total cost of the spare capacity, we now have c_j multiplying s_j, where c_j is the cost of a unit capacity on span *j* and s_j is the spare capacity assigned to span *j*. The notation of "eligible routes" is a special characteristic of the Arc-Path formulation, denoted by P_i, where *P* is a set of alternative detouring routes specific to the restoration flows upon the failure of span i. Hence, the variable f_i^p is simply the total restoration flow assigned to the p^{th} route upon the failure of span i.

For each non-simultaneous span failure, constraint (4.2.7) assigns the restoration flows to fully restore the failed working capacity w_i. Recall that every network flow here is associated with a route. Suppose we have six bi-directional working links on span 4-5. The f_i^p solution might indicate that 4 units are restored via route 4-2-5, while the other 2 units are routed via 4-1-6-5. In parallel to constraint (4.2.3), constraint (4.2.8) determines the size of the spare capacity on each span j. The parameter $\delta_{i,j}^{p}$ essentially provides the explicit information of the eligible routes for the restoration of span i. This binary variable has a value of 1 if span j is part of the p^{th} route for the restoration of span i; otherwise, it is zero.

Thus the Arc-Path approach provides a better route visualization and route pre-selection option than the Node-Arc approach. The downside of the Arc-Path formulation is that the pre-generation of all eligible restoration routes can become a computationally intensive process when the size or nodal degree of the network increases. Carelessly limiting the size of the eligible routes would affect the latitude of how spare capacity is shared on the network and therefore increase the total capacity of the network. With regards to this tradeoff between optimality and computational complexity, researchers such as Herzberg et al. [HeB94][HBU95] and Murakami et al. [MuK95] have proposed ways to reduce the size of the input eligible route set while still maintaining the quality of the optimal solution. The computational aspects of formulations will be discussed in subsequent sections.

4.2.3 Extensions of the Spare Capacity Placement Problem

Earlier we showed how Node-Arc, Arc-Path formulations can be used to solve the spare capacity placement problem. At times, from a research perspective, it is of value for us to identify the motivation and understand "*why*" the problem is an issue in the first place. Identifying the motivation and the reasoning behind the problems suggests a way to classify the literature. Furthermore, such a classification might develop a roadmap of future topics and new concepts that could be investigated by other researchers.

Recall the common objective of any SCP problem is to minimize the amount of spare capacity (or redundancy of transport equipment) to achieve a given survivability measure. A naturally raised question is to see if there are alternative survivability mechanisms that might further enhance the efficient use of spare capacity over the network. The goal of minimum-spare design has inspired many researchers to go beyond the original span-based survivability scheme, and to investigate other possible survivability routing principles such as path restoration, shared backup path protection, segmented-based restoration, etc., to create more efficient, cost-effective spare capacity plans.

An ILP model by Iraschko et al. [Ira96][IMG96][IMG98] and a heuristics algorithm by Veerasamy et al. [VVS95] demonstrate the capacity benefits of path-based over span-based restoration. Path-based restoration, as we briefly mentioned in Section 3.3.1, is a class of survivability schemes, which recover failed working paths from an end-to-end, per origin-destination (OD) pair basis. When we consider the overall spare capacity required to protect against non-simultaneous single span failure, this kind of restoration routing principle leads to a better sharing of spare capacity, and hence reduces the overall network redundancy. Based on five test networks of different sizes and nodal degrees, Iraschko et al. have shown that the path restoration[12] scheme can practically lead to a design with the least possible spare capacity. Veerasamy et al. also showed that the capacity efficiency of path restoration based on the test results of 50 randomly created network topologies.

Besides the span-based and path-based restoration schemes, there exist other mesh survivability schemes, and each is designed for certain specific purposes. Meta-mesh [GrD01a], for example, refines the idea of span restoration for improving the capacity efficiency on spare physical facilities graphs. By intelligently selecting some "express" demand flows to be restored at nodes outside of a degree-two chain of the network (in contrast to re-routing flows inside the degree 2 chain, where at least 100% spare capacity is required), the meta-mesh concept can save the overall spare capacity requirement over the traditional span restorable design. As another example, shared backup path protection (SBPP) can be considered to be a simplified version of path restoration, where a single disjoint backup route is preplanned for each working (or primary) route, and the spare capacity to form backup paths is shared over multiple disjoint working paths without the consideration of stub release. Similarly, segment-based survivability schemes are particularly suitable for the kind of networks (known as translucent networks) where OXC, path switchover occurs only at a few but not all the nodes. In terms of capacity efficiency, segment-based schemes are usually better than span-based schemes and comparable to path-based schemes [HoM01][HoM02][OMZ02][ShG04]. Comparative studies on the capacity efficiency of span-based, path-based, segment-based survivability schemes can be found in [DoW94][DoG01]

---

[12] To be precise, given a set of working capacity to be protected, path restoration with *stub release* gives the most efficient design. By stub release, we mean that the unaffected or remaining links of a failed working path can be used as spare capacity to restore traffic. See [Ira96][IMG96][IMG98][Gro04] for details.

[GDC02][ShG04]. They are of value to researchers and planners for comparing against other survivability metrics, such as fault management cost and speed of restoration.

One assumption of all SCP problems considered thus far is that the working capacity on each span (i.e., $w_{s,t}$ in Node-Arc formulation or w_i in Arc-Path formulation) is always given. This assumption might be true in the context of an incremental transport capacity planning problem, where the set of working capacity represents the connections that are already in place and serving some on-going traffic. For a "greenfield" capacity planning problem, instead of a given set of working capacities, a point-to-point demand forecast is usually provided. In this problem context, the set of working capacities, required to support the demands, becomes decision variables. The objective of this problem is then to determine how much and where we should allocate *both working and spare capacity* over the network in order to accommodate all the demands as well as to ensure network survivability.

This is the key idea of a "joint design" or "joint working and spare capacity placement (JCP) problem." Here we try to determine how much and where to place the working and spare capacity over the network, such that the *total* capacity cost is minimized with respect to a given demand forecast. Both Murakami and Kim [MuK95] and Iraschko [Ira96] have used the Arc-Path approach to formulate an optimization model to solve the JCP problem. Specifically, [MuK95] proposes a linear program (LP) to minimize the total capacity for a span-restorable network, and [Ira96] suggests an ILP to jointly minimize the capacity for both path-restorable and span-restorable networks. Based on a linear cost model and testing the LP over two networks, [MuK95] showed that the joint capacity plan can save 7 to 10 percent of the total capacity relative to the non-joint span-restorable design. Similarly, in [Ira96][IMG96][IMG98] Iraschko et al. test the ILP over 5 test networks and show that path-restorable networks with "stub release" can reduced by an average of 7% in total capacity cost when jointly optimizing the placement of spare and working capacity. Similar "joint versus non-joint" arguments are applicable to any kind of mesh survivable capacity design.

In addition to a predefined set of the failure scenarios, a point-to-point demand forecast is another crucial piece of input information for the survivable network capacity design. In most of the survivable design literature (including all references previously mentioned), a perfect demand forecast is always assumed. Previous research work allows us to formulate optimization models and come up with an accurate capacity allocation plan, knowing the optimal working and spare transport links to be placed on the network. However, in the presence of increasing forecast uncertainty, these single-forecast models might no longer be descriptive enough and optimal in the sense that the capacity plan might be over-built or under-built, and incur some penalty cost to cope with the unexpected.

The notion of demand uncertainty thus opens up a new dimension and presents challenges to the traditional survivable capacity design framework. How should one look ahead and take such uncertainty into consideration and come up with a more future-proof, uncertainty-aware capacity plan? How should we define forecast uncertainty and future-proof in the first place? How should we optimize as-built network capacity to cope with uncertainty? We will separate this particular problem and discuss the details in the next section. We will also review the literature illustrating the different approaches taken by researchers to date.

Table 4.1 summarizes a possible view of roadmapping the SCP problem from solving a basic span-based SCP problem, to more sophisticated path- and segment-based SCP problems. Jointly optimized capacity placement problems (JCP) were proposed to further reduce the overall capacity of the network, and lead to the solution of JCP problems considering additional aspects of demand uncertainty. This thesis attempts to advance this research area beyond the capacity modeling perspective. Note that there are other researchers who push the capacity planning problems in other directions. Examples include designing heuristic algorithms or decomposition methods to improve computational efficiency and enhance the scalability from the ILP approach [Yam95][VVS95][KeW98][KeL01][MBB01][LTS01][PiM04], and designing capacity plans with various kinds of survivability requirements or GOF targets, such as multiple span failures, node failures, or other specific shared risk link group (SRLG) environment [MCK92][LiT01][Ros02][ClG02][DoG02][ShG03].

Table 4.1. Research advances on the spare capacity placement problem.

Extensions to Spare Capacity Placement (SCP) Problem	Proposed Solutions and Concepts
Improve spare capacity efficiency	Path restoration, Shared backup path protection (SBPP), Segmented-based restoration, Meta-mesh
Reduce overall capacity requirements	Joint optimized working and spare capacity designs (JCP)
Plan for demand uncertainty in survivable capacity design	Stochastic programming, Robust optimization, Simulations, Distribution Forecast Accuracy

4.3 Capacity Planning of Mesh-based Survivable Transport Networks in Face of Demand Forecast Uncertainty

In Sections 2.4 and 4.2, we explained the limitations of current capacity design models and the motivation for a new modeling framework to deal with demand uncertainty. We will now discuss the conceptual models for classifying demand uncertainty and the different approaches for quantifying it. We will then explain the methodologies used to incorporate uncertainty into the capacity planning problems. The literature considering capacity planning of survivability and uncertainty will be reviewed, and we will briefly explain our strategies and present them in detail in Chapters 5 to 8.

4.3.1 Modeling Uncertainty in Demand Forecast

An important aspect of planning under uncertainty is to define the scope of demand uncertainty, since that would determine the strategy or mathematical technique to be used for the support of decision making. To narrow down the possible types of demand uncertainty in the capacity planning problem, we adopt a general framework (proposed by Courtney, Kirkland and Viguerie [CKV97]) that captures and classifies the notion of uncertainty into four different levels:

Level I: A Clear-Enough Future – In this case, one can develop a single forecast of the demand that is precise enough for the capacity design problem. In the past, where telephony was dominant in transport networks and exhibited (as it still does) a virtually certain 3 or 4% per annum growth, this assumption might be acceptable and traditional methods can be used to obtain optimal solutions.

Level II: Alternative Futures – Here the future can be described as one characterized by relatively few different outcomes or discrete alternate future scenarios. These unique scenarios represent the few possibilities; each of which is associated with a probability measure even though the probability might be difficult to quantify.

Level III: A Range of Futures – A range of plausible futures can be identified, and the range of possibilities should define the boundaries of the demand space in which the network is expected to serve. What distinguishes this form of uncertainty from Level II is that there may be a near continuum of finely differentiated discrete future scenarios, many times larger than in Level II.

Level IV: True Ambiguity – This is the most uncertain and often considered the most undesirable level of uncertainty to try to plan for, since multiple dimensions of uncertainty interact to create an environment that is virtually impossible to predict. There is simply no basis to forecast the future at the time of decision-making.

Level I is the simplest case of all, where a single demand forecast is considered for the capacity planning problem. Most existing work on the design of survivable networks is based on this assumption that we optimize routing, working and spare transport capacity assignment for a single target planning view. *Level II* captures uncertainty by a limited set of scenarios. In a capacity planning problem, these scenarios might correspond to a distinct set of demand forecasts or "what if" scenarios, and each is associated with a probability estimate. A more rigorous characterization of uncertainty falls into *Level III*, where a range of potential futures might be identified but there are no natural discrete scenarios. By increasing the uncertainty to *Level IV*, it is impossible to identify a range or the domain of potential outcomes.

In the context of medium-term or short-term capacity planning problems (as discussed in Section 2.4.2), the uncertainty of the demand forecast can typically be classified in Level II. This is a level of uncertainty we consider in Chapter 6. We recognize that Level III might seem to give a better description with the continuum of future demand scenarios, but in practice most planners would assume Level II uncertainty and work with a smaller number of "characteristically different" scenarios, due to the fact that the complexity of dealing with too many scenarios tends to hinder decision making [CKV97][DKR91][KOL03][HBB03] and renders the problems intractable [KaW94][Ku95][MVZ95]. Various approaches to quantify these "characteristically different" demand patterns for the evaluation of robustness of a survivable network will be discussed in Chapter 5. Another point worthy of mention is that while in the area of operations research or computing science the challenge is to improve the algorithm component to efficiently handle a large number of scenarios and periods, the business challenge is to produce a reasonable number of distinct scenarios that capture the possible tendencies of the forecast uncertainty.

Finally, Level IV uncertainty often prohibits us from vigorously *planning* but encourages us to adapt as best we can. In Chapters 7 and 8 we will propose re-optimization and demand loading models to cope with uncertainty, and use them to exploit the best use of existing capacity assets.

4.3.2 Stochastic Programming and Robust Optimization

Having identified the scope of demand uncertainty, we will now consider how to model it, or, more specifically, how to integrate uncertainty into the problem of survivable capacity planning. Recall that the aim of the capacity placement problem is to determine the minimum capacity cost of a network, given a set of input parameters such as a network topology, a demand forecast, and subject to a set of technical constraints. In the absence of demand uncertainty, the capacity design problem has traditionally been formulated as a linear program (LP) or integer linear program (ILP). The difference between the two is that some of the variables in the latter are restricted to pure integers. A general formulation of LP/ILP can be stated as follows:

$$\begin{aligned} & \text{Minimize } c^T x \\ \text{Subject to } \quad & Ax = b, \\ & x \geq 0 \end{aligned}$$

The objective is to find a set of decision variables x that minimize (or maximize) some sort of budget (or revenue). The constraint $Ax = b$ represents a variety of restrictions on our decision variables such as technical limitations, administrative policies, or other constraints. Because all input parameters A, b, c^T are assumed to be either fixed or known with certainty, LP and ILP are often classified as the "deterministic" approach [KaW94][BiL97].

Dantzig [Dan55] was probably the first to demonstrate a way of incorporating uncertainty into a two-stage linear program. This two-stage modeling framework later became known as stochastic programming with recourse [KaW94]. Unlike the traditional LP approach that optimizes resources at a single point in time, stochastic programming (SP) provides a more sophisticated framework to incorporate uncertainty into the planning process and allows a planner to deal with a situation where some of the input parameters, or essentially "the uncertainties," are characterized by probability distributions or a set of scenarios. The basic concept of SP modeling can be explained as a two-stage decision problem, where the first stage of the model determines the actions to be taken now and the second stage allows a corrective or "recourse" action to be taken after the uncertainty is realized. The general form of SP can be explained as follows.

$$\begin{aligned} \textit{Stage 1:} \quad & & & \text{Minimize } c^T x + E_w Q(x,w) \\ & & \text{Subject to} \quad & Ax = b \\ & & & x \geq 0 \\ \textit{Stage 2:} \quad & & \text{where} \quad & Q(x,w) = \text{Minimize } d_w^T y \\ & & \text{Subject to} \quad & B_w x + C_w y = e_w \\ & & & y \geq 0 \end{aligned}$$

Here E_w is the expectation, and w denotes a scenario of all possible outcomes $w \in \Omega$. The variables x are called the first-stage variables since they must be decided before the actual outcome w is observed. The variables y are the second-stage (or recourse) variables that are determined based on knowledge of the actual outcome w and the first-stage decisions x. When a discrete probability distribution $p(w)$ (where $\sum p(w) = 1$) is available for the discrete variables, i.e.,

$$E_w Q(x,w) = \sum_{w \in \Omega} p(w) Q(x,w)$$

we can formulate a large-scale LP representing a deterministic equivalent problem of the two-stage problem:

Deterministic Equivalence: Minimize $c^T x + \sum_{w \in \Omega} p(w) d_w^T y_w$

$$\begin{aligned} \text{Subject to} \quad & Ax = b & \\ & B_w x + C_w y_w = e_w & \forall w \in \Omega \\ & x \geq 0, y_w \geq 0 & \forall w \in \Omega \end{aligned}$$

Note that unlike the basic LP model, the objective here is no longer a single one but an optimal decision x that minimizes the *sum* of first-stage costs and *expected* second-stage costs. The above constraint sets comprise a deterministic portion (i.e., the first constraint) and a portion with stochastic parameters (i.e., the second constraint).

In the context of transport network planning, the uncertainty of demands is often represented by a set of demand scenarios $w \in \Omega$. In the first stage, a decision x is selected which minimizes the network costs under constraints imposed by the network structure and the probability of meeting demand. Once the allocation of available capacity x is found and a demand outcome e_w is realized, the second-stage decisions y_w can be interpreted as the extra cost to satisfy the expected demand scenario w.

Once the SP problem is transformed into a large-scale linear program, existing LP optimization software and decomposition techniques, such as Benders decomposition and Lagrangian Relaxation, might be used to solve these kinds of problems. However, in cases where the problem size (e.g., the size of scenarios) becomes large, the deterministic equivalence of SP problem can become enormous. For these cases, sampling methods, such as stochastic quasi-gradient method, importance sampling and stochastic decomposition, might be used to reduce the complexity of the problems and to make them computationally tractable. Finding ways to break large problems down to manageable sub-problems – thereby reducing the computation time – has always been an active topic in the operations research (OR) community [DaW60][SwM79] [AHK80][BMO84][HiS91][Ran92][KaW94][MuR95][BiL97][MaS98][SAG05]. We do not directly address the computing issue in this thesis, but instead focus on the modeling aspect specifically for the capacity planning problem of survivable mesh networks.

A well-known alternative to stochastic programming is Robust Optimization (RO). Like SP, RO explicitly incorporates uncertainty into the modeling framework and has been used to solve problems relating to financial asset allocation and electric power capacity planning [PKR91][MaZ94][Mul96][Mul96][BCM97]. Unlike SP, RO allows modelers to address risk aversion directly by using some kind of utility function $U(.)$ which can be non-linear or piecewise-linear. The critical difference betwen SP and RO is their objective functions, where SP models only the first moment of the distribution of the objective value $y(w)$ while RO characterizes higher moments and the decision maker's attitude toward risk:

$$\textbf{S.P.}:\ \text{Min} \sum_{w \in \Omega} p(w) y(w) \qquad\qquad \textbf{R.O.}:\ \text{Min} \sum_{w \in \Omega} p(w) U(y(w))$$

If we compare these objectives in the context of capacity planning, SP would minimize the number of unserved demands, while the RO approach might suggest minimizing the impact caused by the unserved demands. Less apparent difference is the notion of recourse: the solution of SP would *explicitly* tell us that we need Y units of capacity to augment the initial decision X when demand scenario D arises, but for RO, solely minimum-penalty (also termed "regret" in [KOL03]) capacity design is found. The sense of "correcting" or "augmenting" the initial design through future decisions is not modeled in RO.

For problems where it is not possible or straightforward to alter (e.g., adding or removing capacity from) the current outcome after it is decided, RO does offer more flexible ways to describe the penalty (due to uncertainty) than SP. Thus, depending on whether the present decisions can be corrected and how we measure the impact of future consequences, both SP and RO have their merits in each inherent framework to deal with uncertainty. Mulvey et al. provide some motivational examples where RO is preferred over SP [Mul96][MVZ95][BCM97].

The use of SP and RO to deal with uncertainty is well recognized. They have been applied to solve problems in the electric utility [SSM84][PKR91][MaZ94][GaR99][MaS98], semiconductor [HBB03], finance [Mul96][Dup02] and logistic industries [SeG80][SAG05]. Yet the application of these approaches to the capacity design of transport network in face of uncertainty has been minimal. What follows are some related prior works on the capacity design problem that consider demand uncertainty, and with a limited work that take network survivability into account.

Ouveysi et al.

In [OuT95][OSW98], Ouveysi et al. propose a LP formulation to design a network to support a given multi-hour traffic profile. The traffic profile, modeled by a set of traffic matrices or scenarios, represents the usage patterns over several time periods of a day. The objective of the multi-hour network dimensioning problem is to minimize the total capacity cost while the capacity set satisfies *all* demand matrices, at any time during the day. Note that this is a related problem and the problem of demand forecast uncertainty. The multi-hour model typically considers capacity design that serves *all* demand matrices, whereas the capacity design with uncertainty might not (and should not) satisfy all demand forecasts. Otherwise, the solution would become a "fat solution" [KaW94], that serves all possibilities and is the most expensive. Other approaches to multi-hour network design problems can be found in [Dut94][Med95][MeT98][PiM04], and none of these studies consider the aspect of network survivability.

Sen et al.

In [SDC94], Sen et al. formulate a two-stage SP for the capacity allocation design of private line services network with random demand [SDC94]. Given a total capacity budget, network topology and the random demand, which is characterized by a set of point-to-point demand matrices, the objective of the SP is to determine a capacity allocation plan such that the expected number of unserved requests (with equal priority) is minimized. Mathematically, in the standard form of the SP structure, only $\sum_{w\in\Omega} p(w)d_w^T y_w$ is considered in the objective function. Because the number of demand scenarios considered is large or might be considered as *Level III* uncertainty (e.g., 5^{82} involving approximately 82 OD pairs with each pair having 5-10 possible outcomes), a sampling-based algorithm called stochastic decomposition [HiS91] is chosen to solve this large-scale linear program. Although this study uses the SP as the fundamental framework to deal with demand uncertainty, the solution of this problem does not actually provide any recourse or corrective actions to be taken when the uncertainty unfolds. In other words, although we know how many demand requests are unserved for a given demand scenario, we have no idea of what corrective actions should be made to the current capacity plan. Another observation of [SDC94] is that this study addresses a capacity allocation problem with a known total budget, in contrast to the capacity sizing or dimensioning problem that we will consider later in Chapter 6. Other network planning problems using SP can be found in [Gai95][LOV99][RiA02]. These studies put great emphasis on how to solve the SP problem efficiently from an algorithmic perspective (e.g., [RiA02] suggests the use of L-shaped solution procedure while [LOV99] describes the advantages of using the Analytic Center Cutting Plane Method), but less weight on what variables should be modeled from a capacity design standpoint. Network survivability has not been considered in any of these studies.

<u>**Kennington et al.**</u>

Closely related to capacity planning with demand uncertainty and network survivability, Kennington et al. is probably the only group that proposes optimization models for explicitly capturing uncertainty and survivability in a capacity design problem.

In [KLO01][KOL03], Kennington et al. adopt Mulvey's idea of RO [MVZ95] and formulate integer linear programs for solving routing and provisioning problems over DWDM networks with uncertain demands. Given a fixed budget, a network topology, a few sets of demand scenarios, their model is to determine the "least regret" design that also minimizes total equipment used to support the demands. By "regret," the authors refer to the expected penalty caused by the mismatches between the infrastructure created and the actual demand for services. For example, if X number of unexpected demand pairs cannot be served, it might imply a penalty or regret value of Y. This regret-mismatch relation is a key input to the RO model. A two-phase procedure is proposed to find the min-regret design. In the first phase, an ILP is used to determine a design that gives the minimum regret. At this stage, an overall budget is given and the design does not necessarily have the minimal cost. Once the optimal regret target is obtained and set as a fixed parameter, a second phase ILP is then used to determine the minimum cost design subject to this fixed regret value. This procedure ensures that no other design can achieve this min-regret value with lower cost, and alternatively, any lower cost design must have higher regret.

As will be seen from our formulations in Chapter 6, Kennington et al.'s approach to demand uncertainty is different from ours in several significant ways. First, the objective of their ILP formulation is to *minimize regret*, whereas in our SP approach, the goal is to *minimize the overall design cost* (i.e., the cost of initial design construction and the expected cost of possible augmentations or "recourse" actions required in the future, adapting the network to accommodate various actual future demands). Thus a min-regret solution can still lead to a highly expensive design, as reported in [KLO01][KOL03]. In addition, the RO model does not consider corrective actions to cope with the future but to solely find a one-time solution that gives the minimum regret from the undesirable outcomes.

The regret-versus-demand mismatch function is relatively difficult to obtain in practice (i.e., how should one quantify the demand pair-specific penalty cost Y of not serving X units of demands?). In contrast to our approach, the notion of a capacity link-related corrective cost has a more explicit and direct meaning (e.g., the corrective cost of adding X channels next year on span A-B is Y times the current cost).

Recent work by Birkan and Kennington et al. [BKO03] combines demand uncertainty and network survivability into a single optimization model. Building upon the same two-phase robust optimization framework, Birkan and Kennington extend it to include network survivability aspect with protection schemes, including 1+1 dedicated protection, shared backup path protection, and *p*-cycles protection. They demonstrate that integer programming modeling techniques and optimization software can be used to solve difficult, real-world DWDM design problems, despite the fact that many designers have been reluctant to use optimization because of its reputation of having excessively long run times.

As we discussed earlier, while both the SP and RO approaches provide the mathematical framework for incorporating uncertainty into the decision modeling, we believe that a *minimum-cost, recourse-based SP* approach is capable of reflecting the capacity planning problem more realistically and precisely from a network operator or planner's standpoint. Operators would generally prefer a minimum cost solution (especially for a greenfield design, where initial design cost is required upfront.) In face of demand uncertainty, it is also beneficial to know where and how much link capacity to augment in order to serve the growing number of customers.

Multi-period Optimization

Another area of survivable network planning strategy is called Multi-period Planning. It considers incremental capacity and/or topology expansions over a period of years. In most of these studies, the demand forecast for each period is assumed to be known with certainty. By taking the entire time evolution of known traffic demand and the cost data into account, the multi-period planning approach can provide more comprehensive and cost-effective solutions than solving a series of single-period problems. Multi-period models can be used to address capacity expansion decisions over a period of time (e.g., up to 10 years) while addressing issues such as discounting and capacity deferral. A typical objective is to minimize the present worth of the total network cost along a given planning horizon.

Nevertheless, in the presence of uncertainty, the argument that multi-period planning is superior to sequential single-period planning might no longer hold true since the multi-period planning technique is as dependent on assumed perfect future forecasts as other traditional single-period methods. In fact, the effect of demand uncertainty can only increase as the length of the planning horizon increases, and that somewhat defeats the purpose of using any kind of multi-period optimization models under a long planning horizon. For these reasons, although our SP capacity planning framework can be extended and formulated into a multi-period planning problem, in this thesis we do not consider this approach to address *which period* transmission capacity should be installed. The questions of which period (or *when* if we consider a continuous timeline) should be considered with other strategic factors (e.g., the ones mentioned in Section 2.4) that might not even be quantified by mathematical models. For general survey and discussion on the multi-period capacity expansion problems, readers can refer to [Yag73][Zad74][Lus82][Che88] [DuL92][ChG95][PiM04]. For specific studies that also consider network survivability, please see [WCB91][PiD99][GAD01a][PiM04].

4.3.3 Descriptive Approach to Capacity Planning under Demand Uncertainty

What we present next offers a more "descriptive" or responsive view of handling uncertainty. Instead of explicitly incorporating the defined uncertainty into the mathematical model and trying to find an optimal design, the goal of the descriptive approach is to *evaluate* the effect due to such uncertainty and to possibly identify a single or several robust designs. As a complementary approach to prescriptive methods (e.g., SP and RO), descriptive approaches provide a different philosophical and strategic treatment to cope with demand uncertainty and make the overall study more complete.

The RO and SP approaches to the capacity planning problem discussed previously are all "prescriptive" in nature, i.e., these approaches lead to well-structured problems with unique objectives, data requirements, plus technical and physical constraints. In the context of capacity planning for survivable networks, solutions from these approaches provide us with clear directions on the location and quantity of the capacity placement, as well as working and restoration routing details. One might notice that such complete prescriptions of these models, however, come at a price and impose computational limitations. As the dimension of the demand uncertainty increases, so too does the size and complexity of the optimization problem. Solving a capacity design problem with thousands or more demand scenarios, for example, would require rigorous decomposition procedures and parallel computing power to make the problem manageable. The key to using these prescriptive approaches is to identify and to strike a balance among exhaustive descriptions and computational tractability. Another point to remember concerning these prescriptive techniques is that the robustness of the optimal solutions is still highly dependent on the scope or the defined set of uncertainty in the problem. These mathematically rigorous techniques enjoy the precision and comprehensiveness of modeling frameworks; paradoxically, such frameworks have their own limitations on the degree of uncertainty that can be modeled.

In contrast to the prescriptive methods, "descriptive" or "evaluative" [CHS98] methods such as simulation can provide a "softer," more flexible approach to analyzing the capacity planning problem under uncertainty [Ku95]. Unlike optimization techniques, simulation is driven by a different set of goals; it is not to find the *optimal* solution but to run many *exploratory* experiments with randomly generated values until some statistical patterns can be obtained. Scenario analysis, sensitivity analysis, and Monte Carlo simulation are some examples of descriptive tools. Generally, simulation is considered a complementary tool to optimization, when the model or nature of uncertainty becomes too complex to be analyzed or captured in a single analytical model (e.g., combined uncertainties due to cost and demand, or uncertainty due to network topology changes.) Unlike optimization, the routing algorithms used in simulation are usually straightforward and require fewer computational resources. Because of the relaxation of the problem, simulation usually can only produce non-optimized, non-minimum cost capacity designs. Nevertheless, simulation does suggest a valuable means for comparing given designs in terms of their robustness to cope with uncertainty and to identify non-robust ones, as we will see in Chapter 5. In the following, we will review some representative studies that use simulation to tackle the capacity design problem considering both demand uncertainty and network survivability aspects. For those that do not consider network survivability, the reader can refer to [CHS98][Mau02a][Mau02b].

Geary et al.

To evaluate the robustness of optical network designs under uncertainty, Geary et al. first quantify demand forecast uncertainty or forecast error in terms of volume and distribution [GAD1b][GAM03][Gea03]. The former corresponds to the total traffic growth in time, while the latter refers to the distribution changes in a demand matrix. To quantify the distribution pattern errors, the authors introduce a metric called Distribution Forecast Accuracy (DFA), which is determined by taking a linear correlation between the actual and the expected demand matrices. Thus, if the two matrices are identical, we would have a DFA of 1. As the actual matrix departs from the expected, the value of DFA decreases from 1 to –1. An in-depth discussion of Geary's DFA measure and some examples will be presented in Section 5.3.

In [GAD1b], Geary et al. explain the procedures to assess the robustness of a given capacity-protected design. In step one, a minimum-cost capacity design is created, in relation to an estimated or nominal demand forecast. In the second step, random variations of demand forecast or a set of actual demand patterns is generated, and each demand pattern is associated with a total demand volume (i.e., the sum of point-to-point demand entries) and a DFA measure. Finally in step three, a heuristic-based routing algorithm is used to route each demand pattern over the already-dimensioned network, and to record the number of demands that could not be routed.

To assess the impact of DFA on the robustness of a given network, the actual matrices are grouped into three DFA bands (i.e., $0.95 < DFA < 0.98$; $0.75 < DFA < 0.85$; $0.65 < DFA < 0.75$) and the probability of unrouted demands is measured for each one. The results have shown that when we have the actual matrices whose total volume is close to the expected one, the number of unrouted traffic varies significantly with their DFA values. The same 3-step procedure is repeated to test the robustness of the original test network and the one with two added spans. As might be expected, the better routing performance favors the network with a higher connectivity. [GAM03] is a similar to [GAD01a], except that the recourse cost (i.e., the extra investment cost required to carry the unrouted demands) is used as a measure of robustness, and both studies only consider 1+1 path protection, where the minimum-cycle algorithm is used to find the primary and backup paths.

Readers should note that none of these simulation-based studies has addressed uncertainty as explicitly as the prescriptive SP or RO approach, nor are they meant to. Demand uncertainty is only considered *after* the initial network is designed. In contrast, SP and RO approaches explicitly incorporate demand uncertainty, recourse costs, and penalty factors into the optimization models. These factors influence the initial design and specifically the amount of "future-proof" capacity to be allocated in the first place.

Verbrugge et al.

Verbrugge et al. have shown the advantage of using probability theory for handling the demand uncertainty in a capacity planning problem. In this approach, each demand entry in a matrix is represented as a random variable with an associated probability distribution. With a sufficient number of demand samples and a known confidence level, a probabilistic approximated value of the network capacity can be determined.

In [VHT02], the probabilistic model is compared with a model based on possibility theory, a theory of uncertainty closely related to fuzzy set theory. Under the assumption that demand values have a Gaussian distribution, the results have shown that the probabilistic approach is able to come up with a more cost-effective capacity plan than the possibility approach. The capacity requirements between 1:1 dedicated path protection and shared backup path protection (SBPP) of each approach are evaluated on a 27-node, 40-span network. The results show that the total (i.e., working and spare) capacity requirement of a SBPP design is about 25% less than that from a 1:1 dedicated path protected design[13].

In a similar study [VCP03], Verbrugge et al. contrast the same probability model with *a posteriori* adjustment and *a priori* adjustment approaches (commonly known as the "safety margin" approaches to uncertainty). *A posteriori* adjustment simply adds a safety margin to a deterministic result, while *a priori* adjustment adds a safety margin to the demand input prior to the capacity calculation. Based on the same network topology used in [VHT02] and the Gaussian distributed demands, the results have shown that it is difficult to come up with an appropriate value for the safety margin for both a posteriori and a priori approaches, which are widely used in practice today. A wrong choice of this safety value can easily lead to over-capacitated network designs. For the case of survivable designs, the over-dimensioning effect was even worse.

Unlike Geary's studies [GAD1b][GAM03] discussed earlier, Verbrugge et al. have suggested ways to incorporate uncertain demand parameters in a capacity design. These simulation-based approaches are different from the RO or SP modeling framework where they might not be used to generate optimized, minimum-cost capacity plans. Table 4.2 summarizes related studies on the capacity planning problem under demand uncertainty. Each study is associated with the technique used, survivability consideration and the way demand uncertainty is defined. Based on the literature, Geary's [GAD01b][GAM03] are the most relevant to our work in Chapter 5, but we use ILP approaches to compare the robustness of span-restorable and SBPP networks. The robust optimization models by Kennington et al. [KLO01][KOL03][BKO03] are comparable to our stochastic programming models in Chapter 6, and we have already explained the differences of the two approaches.

---

[13] Note that since only the shortest path routing algorithm is used to find both primary path and the corresponding backup path [VHT02], none of these designs is optimal. The network redundancy and spare capacity requirement were not reported in the publication.

Table 4.2. **Summary of prior work on capacity planning under demand uncertainty.**

Authors	Modeling Technique	Survivability Considered	Uncertain Demand Forecast Characterization	Reference
Ouveysi and Tham	Proposed LP, solved by heuristic	No	Multi-hour profile, Scenario-based	[OuT95]
Ouveysi et al.	ILP	No	Multi-hour profile, Scenario-based	[OSW98]
Dutta	IP with decomposition techniques	No	Multi-hour profile, Scenario-based	[Dut94]
Medhi and Tipper	ILP with decomposition techniques	No	Multi-hour profile, Scenario-based	[McT98]
Sen et al.	SP with decomposition techniques	No	Statistical-based	[SDC94]
Gaivoronski	SP with decomposition techniques	No	Statistical-based	[Gai95]
Lisser et al.	SP with decomposition techniques	No	Statistical-based	[LOV99]
Riis and Andersen	SP with decomposition techniques	No	Statistical-based	[RiA02]
Kennington et al.	RO	No	Scenario-based	[KLO01] [KOL03]
Birkan et al.	RO	Yes, SBPP, 1+1 path protection, *p*-cycles	Scenario-based	[BKO03]
Wu et al.	Heuristic	Yes, ring and point-to-point systems	Multi-period scenarios, demand uncertainty not considered	[WCB91]
Pickavet and Demeester	Proposed IP, solved by heuristic	Yes, 1+1 path protection	Multi-period scenarios, demand uncertainty not considered	[PiD99]
Geary et al.	Heuristic	Yes, 1+1 path protection	Multi-period scenarios, demand uncertainty not considered	[GAD01a]
Carpenter et al.	Simulation	No	Statistical-based	[CHS98]
Mauz	Simulation	No	Statistical-based	[Mau02a] [Mau02b]
Geary et al.	Simulation	Yes, 1+1 path protection	Statistical-based, but uncertainty is not considered in the design	[GAD01b] [GAM03]
Verbrugge et al.	Simulation and by safety factors	Yes, 1:1 path protection and SBPP	Statistical-based	[VHT02] [VCP03]

4.4 Capacity Management Strategies of Mesh-based Survivable Transport Network to Cope with Demand Uncertainty

In this section, we will now shift the focus from capacity planning to an operation-related problem for dealing with uncertainty. First, let us point out some key distinctions between the planning and operational problems. Recall from our earlier discussions in Sections 4.3 and 2.4.2 that the link capacities in MTP capacity planning problems are typically variables, and the goal is to determine a minimum-cost capacity placement strategy subject to a demand forecast. In contrast, the link capacities (or capacity inventory) in the operational context are usually known or given as parameters, and these problems are also of operational or business support relevance. For example, with an arbitrary set of demands, how should a service provider determine and select an optimal demand subset that would lead to a maximum profit (i.e., service revenue minus the cost of provisioning)? Given a poorly utilized network, how should a network operator re-configure existing demands and their corresponding routes in order to improve the network's potential to carry future traffic? These problems, also generally referred as capacitated problems [PiM04], will be addressed in Chapters 7 and 8.

The first question relates to the class of **demand loading** problems. These problems involve the selection of a subset of demands that give the maximal profit (or revenue if provisioning cost is not considered), as well as the respective routing and protection within a mesh network that has a finite capacity. Insights from the demand loading problems can be useful for operators to design admission control systems or for business planners to reveal strategies for service pricing and promotions. The second question belongs to the more general issue of **reconfiguration** or **re-optimization**. Reconfiguration considers the various aspects of transforming a current logical (or virtual) topology to a new one with better capacity utilization, subject to the capacity resource constraint and the constraint that limits the degree of route changes from existing connections. Combined with a single-vendor network management or the so-called multi-vendor "back office" [Lev03][Oke04] inventory management support systems, an effective reconfiguration strategy can provide the direction and alleviate the "stranded" capacities on the network, and possibly defer the capital expenses to support demand growth.

In the face of a completely uncharacterized demand profile, both demand loading and re-optimization strategies encourage one to make the best use of existing capacity assets. A well-designed demand loading strategy can provide valuable operations support for service providers to ensure maximal profit under *any* demand circumstances. Likewise, a capacity re-optimization strategy suggests a means to better utilize existing assets and defer unnecessary additional capacity investment to support incremental demand growth. In the next two sections, we will survey and analyze some of the prior work that is related to the problems of demand loading and reconfiguration. In particular, we pay careful attention to those that also consider the network survivability aspect.

4.4.1 Related Work on the Demand Loading Problem

Before we analyze the literature on demand loading, it would be helpful to first describe the various dimensions of the problem, namely demand modeling, revenue and cost modeling and optimization criteria, so that we could identify essences and differentiate it from previous work. Demand modeling relates to the types of parameters used to describe a demand or the attributes of a transport service. For example, a wavelength or private line service is, at minimum, associated with its origin node, destination node, service distance, bandwidth requests, service duration and the level of protection[14] [CoW03][CCF98][Dri04]. The revenue of the service, from a service provider standpoint, is closely related to these parameters, plus some other so-called "nonrecurring" charges, which are one-time charges that apply to specific operational activities (i.e., installation of new service, moves and rearrangements of installed services, administrative charge, design and central office connection charge and customer connection charge) [Dri04]. Setting up these services also comes at a cost or the cost of provisioning, which basically consists of fixed components (e.g., administrative charge per order), distance-related components (e.g., total channel mileage required for unprotected or protected services) and quantity-related portion (e.g., optical amplifiers required per location, regenerators required per circuit).

In addition, existing work on demand loading differ with respect to the optimization objectives and constraints considered in the models. Lee et al. [LMS89] propose an integer linear program formulation whose objective is to minimize the total costs of routing a given set of demands onto a capacitated network. Since the cost is associated with each working path, this formulation tends to load demands with the higher costs and leaves lower costs demand unserved. Lee et al. also extend this model into a multi-period formulation where the incremental demand matrix for each time period is also given. In other closely related set of literature on static routing and wavelength assignment (RWA) problems, the objective is typically to maximize the number of connections to be established within given capacitated networks [RaS95][ZJM00], to minimize the number of blocked wavelengths from a fixed set of demands [CGK92], or to minimize the number of wavelengths used under some wavelengths and optical hardware constraints

---

[14] Later we refer to it as the Quality of Protection (QoP).

[BaM96][BaB97]. But none of these studies has considered network survivability, the demand revenue and cost modeling altogether as described above. Our study in Chapter 7 is comprehensive and unique from this modeling perspective. The studies covered below are some exceptions to the literature that selectively consider the aspects of revenue, cost and network survivability models.

Anand et al.

Anand et al. [AKQ00a] propose an ILP and heuristic that solve the routing and wavelength assignment problem, with an objective of maximizing profit. In this study, the revenues (defined per each connection) and the costs (defined per each link) are assumed to be normally distributed, and do not relate to service distances and topological span distances in any way. In terms of demand modeling, only unprotected services are considered.

Anand et al. show that while the prominent minimum-cost RWA models can indirectly lead to some demand loading and demand selection solutions, the max-profit approach can explicitly indicate which subset of profitable demands to be served. The authors extend this study to consider incremental traffic in [AKQ00b], where the connection requests come in one-by-one and, once a connection is setup, it stays in the network for a long period of time. Protected services were not considered in this study.

Kabranov et al.

Similar to [AKQ00a], Kabranov et al. propose an ILP for solving a problem with a max-profit objective [KMC01]. But this work differs in that the ILP is formulated in such a way that it can only be used as a simulation tool, rather than a demand selection decision support tool. Given a set of demands and its associated revenue, the objective is to maximize the profit, i.e., revenue minus the cost of serving *all* demands, and satisfy the demand routing and capacity constraints. In other words, if the set of demands exceeds the available capacity, the formulation would become infeasible. Consequently, the authors tested the ILP against different demand scenarios and analyzed the revenue generated and cost associated with each feasible and infeasible demand set.

A unique aspect of this study is how service revenue is modeled based on the concept of "demand elasticity." In economics, elasticity, E, is a measure of the responsiveness of demand (or supply) to changes in price, p. Specifically, it is calculated as the percentage change in demand quantity, d, in response to a percentage change in price [Oxf02], or $E = -(\Delta d/d) / (\Delta p/p)$. For example, if a fall in the price of a transport service by 10% would cause an increase in service demand by 20%, we would have an elasticity value of 2, or –2 to be mathematically precise. For a constant elasticity value, we can relate the demand quality *d*, pricing of the service *p*, and elasticity *E* by a general function $d = A * p^{-E}$, where the constant *A* is known as the demand potential [KMC01][CoW03].

From this general function, assuming the elasticity is identical for all demand pairs $i \in K$, we can derive a formula for the total revenue, *R*, and incorporate this into the objective function of the ILP:

$$R = \sum_{i \in K} p_i \cdot d_i = \sum_{i \in K} \left(\frac{d_i}{A_i} \right)^{-\frac{1}{E}} \cdot d_i = \sum_{i \in K} d_i^{\frac{E-1}{E}} \cdot A_i^{\frac{1}{E}}$$

Note that in solving the ILP model, all *E*, *A* and d_i are known parameters. The unknown of this model is to determine how each demand is routed over the network, under the condition that the capacity must be sufficient to accommodate all demands.

Although this study does not consider the demand selection aspect, the max-profit optimization model has led to some interesting conclusions. First, compared to using the shortest path algorithm to route the demands, the proposed method suggests more effective routing solutions with higher profit. Second, the profit from operating a network is not necessarily proportional to the number of demands served. In a subsequent study [KaM02], the authors show that from elasticity $E = 1.0$ to $E = 2.0$, the formulation leads to very different RWA solutions. Network operation under high elasticity values tends to serve all possible demands and remain profitable, whereas in the case of low elasticity values, only a small set of the demand scenarios is profitable.

Sridharan and Somani

Sridharan and Somani [SrS00] are probably the first researchers that consider protected, best-effort and unprotected services in a max-revenue demand loading problem. The shared backup path protection (SBPP) is assumed for the protected demands, and each is guaranteed with a primary and backup path. For the best-effort services, each demand might be assigned a backup path only if capacity resources are available.

Given the network topology, a demand matrix of multi-class services and a set of “already existing” connections, the authors propose an ILP to find a subset of demands that maximize the revenue[15], which consists of three terms. The first and second terms denote the revenue generated from primary paths and backup paths, respectively; the last term denotes a penalty term for disrupting the currently working connections. We should point out that the revenue modeling in this ILP is rather simple since all primary paths (or all backup paths) are assumed to have the same revenue. The cost modeling or the cost of utilizing capacity is not included in this study. Because of the complexity of the SBPP formulation, a three-stage decomposition approach was implemented to solve a 14-node, 21-span network example and to find a feasible solution. The results show the potential gain in revenue, by serving additional best-effort demands after the fully protected and unprotected services have been served. Note that our study in Chapter 7 differs from all we discussed thus far in the sense that the multi-QoP demand service framework considered in our modeling framework covers a more complete service mix (i.e., protected, unprotected and preemptible services) with more realistic revenue and cost modeling details (e.g., revenue and cost of services are distance- and class-dependent). As will be seen later, our demand loading model is highly applicable to an environment even if demand, revenue and cost are uncertain, yet we are able to provide the optimal demand selections that maximize potential profit.

4.4.2 Related Work on Re-configuration and Re-optimization Problems

In this final review section, we will discuss the literature related to the problems of reconfiguration and re-optimization. As in the previous sections, the intent here is to identify and classify the dimensions of the problem so that they become a basis of comparison to our work in Chapter 8. Let us first group the existing works by the problem inputs and objectives:

<u>Type A:</u> Given a current and a targeted logical topology, the objective is to determine the *proper procedures* of migrating from one topology to another, so that the disruptive effects due to the reconfiguration process can be minimized.

[LHA94][BaR99][IAM03] are examples that consider the transition phase from transforming a current logical topology A to a new one A'. In the context of lightwave network [LHA94], Labourdette et al. indicate that simultaneously retuning all transmitters and/or receivers involved in the logical topology is not always possible (e.g., limited by the speed at which a laser

[15] While the authors refer the objective criterion as revenue, it might be more appropriate to think of it as *net* revenue, i.e., total revenue from services minus the service disruption cost.

or a receiver cannot be re-tuned to a new wavelength within a short period of time). Therefore, to reduce potential performance degradation (such as packet delay, packet loss and packet de-sequencing), the transition from topology A to A' must be carefully orchestrated and managed so to minimize disruption. Labourdette et al. propose an operation, called branch-exchange, that ensures only two links can be changed during a single operation. Three different algorithms are also suggested for searching the sequence of branch-exchange operations that lead to the minimal reconfiguration time.

In [IAM03], Ishida et al. propose five procedures (i.e., switch, append, backup, release and delete operations) to reconfigure logical topologies in SBPP-based WDM mesh networks. The idea of this study is to allow the use of backup lightpaths to accommodate traffic and to minimize the number of delete operations (also referred as the number of traffic loss occurrences) during the reconfiguration process. Note that the demand matrix is typically not part of the Type A problem, and the targeted topology A' is also known as an input parameter.

Type B: Given a current topology and a targeted demand profile, the objective is to determine an *optimal topology that serves the demand* while minimizing the changes required to obtain the new topology from the current one.

[BaM97][RaR00] describe problems of this kind as optimization problems. Given a targeted demand matrix, Banerjee and Mukherjee [BaM97] propose a 2-step procedure that first finds a min-cost[16] logical configuration and then tries to minimize the number of reconfiguration steps involved. To help explain this procedure, let us denote the old and new topologies as A and A', original and new demand matrices as D and D', and the objective value of a min-cost problem on the two matrices as P and P'. In step 1, two separate min-cost capacity planning problems are solved, based on D and D', resulting in objective values P and P'. In step 2, instead of solving a min-cost capacity planning problem, a new optimization model is used. The new objective is to minimize the number logical connection changes (i.e., lightpaths that are added or removed) from A to A', and the objective value P' (from solving D') is fixed and used as a constraint to this problem.

Ramamurthy and Ramakrishnan [RaR00] modify the approach from [BaM97] and introduce two metrics to quantify the 'changes' in reconfiguration, referred as the number of

[16] In [BaM97], minimizing cost refers to minimizing the average hop distance in the network. One can also interpret this as the average hop-weighted capacity required to accommodate the target demands.

reconfiguration steps (Δv) and disruption (Δp). The former is the same as the one used in [BaM97]. It indicates the number of lightpaths that would have to be removed or new lightpaths that have to be added to establish the new topology. Δp measures the change in *routing* of existing lightpaths. In practice, it is also this route changing aspect of reconfiguration that is relatively difficult to realize (compared to adding or dropping a lightpath along the same route.) Four different objective functions are considered (including the one minimizing the average number of hops [BaM97]), and two new constraints on (Δv) and (Δp) are added to the formulation by [BaM97]. Basically, these constraints explicitly specify the upper bound values of how many reconfiguration steps and how much disruption is allowed during the reconfiguration process. These constraints would allow traffic engineers to directly control the amount of change due to reconfiguration and to evaluate the trade-off between the reconfiguration penalty Δv (or Δp) and benefits gained from better objective values.

Work by Bouillet et al. also falls into this problem type, except that the targeted demand remains unchanged throughout the entire re-optimization process and the objective is to reduce the capacity required to support the same demand set [BML02][BLR05]. Shared backup path protected networks are considered in these studies and the authors propose two heuristic algorithms to perform re-optimization. The first algorithm performs a complete re-route on both primary and backup paths, while the second considers re-routing the backup paths only for minimizing service interruption. In terms of cost saving (or the total distance-weighted channels required to support the demand set), simulation results have shown that the complete re-optimization could achieve a cost saving from 3% to 5%. Most of the improvement can be achieved by re-optimizing the spare capacities alone. In [BLR05], the authors extend the work of [BML02] by considering the scenarios of changing infrastructure, such as adding new spans in physical topology and new links to existing spans. Experimental results on real carrier's networks have shown that more substantial cost savings from re-optimization can be achieved.

Type C: Only a current topology is given; the objective is to determine a new logical topology that enhances the network's *potential traffic carrying ability* based on some statistical demand traffic assumptions. A single targeted demand matrix is not given and the logical topology is the unknown in this problem.

In contrast to the previous two problem types, reconfiguration problem of this type is more forward-looking. In this setting, the target demand scenario is not precisely defined, but rather some statistically-characterized demand sets are used as the inputs to simulate the demand uncertainty and to evaluate the effectiveness of a given reconfiguration strategy.

Herschtal and Herzberg pose a reconfiguration problem in this context. In [HeH95], reconfiguration is defined as a strategy that "switches traffic between paths in order to put the network in a better state for acceptance of future demands." Three LP-based reconfiguration models were proposed, and two principles were used to devise the objective functions. The first principle is to keep the overall network usage as low as possible by routing demands via short-distance paths; the second is to keep the minimum unused capacity over all spans as high as possible. Based on an 11-node, 23-span network tested under fairly loaded dynamic traffic scenarios[17], all reconfiguration strategies have shown that about 1% to 3% improvement in blocking (calculated as the total bandwidth of all blocked demands divided by the total bandwidth of all offered demands) can be achieved. Results show that the exact improvement in blocking depends on how frequently the reconfiguration procedure is performed. Generally, the more frequently the reconfiguration is performed, the lower the blocking rate. A good, capacity-state aware routing algorithm used for dynamic traffic requests (with certain statistical arrival rate and holding times) also minimizes the blocking and reduces the number of global reconfigurations required during the life of network operation.

The study presented in Chapter 8 can be classified as a problem of this type. One major difference from [HeH95] is that we include the aspect of network survivability in our study. Motivated by the Bouillet et al.'s work [BML02], we see the needs and benefits of re-optimization as a strategy to relieve the accumulation of "stranded" capacities by rearranging existing working and spare capacities. But unlike [BML02], our model does not assume any expected or known traffic patterns and our objective is to determine a new logical topology that has a better readiness to adapt to different kinds of incremental traffic.

A final note before we leave this section is that other studies are also concerned with reconfiguration policies and address the question of *when* reconfiguration algorithms should be executed. Gençata and Mukherjee [GeM03], Golab and Boutaba [GoB04] provide comprehensive reviews to this type of problem, which we will not discuss in this thesis. Answers to the problems discussed, on *what* logical topology should be targeted, *how* to migrate from one topology to another, *when* and *how often* reconfiguration should be triggered, should suggest a comprehensive solution to improve the utilization and performance of any given network.

[17] Initial blocking percentages (i.e., benchmark case before reconfiguration) of 14% to 16% are assumed.

4.5 Summary

This chapter reviewed the literature on four relevant problems: the capacity planning of mesh-based survivable networks, the same planning problem in the presence of demand uncertainty, the demand loading problem, and issues relating to reconfiguration and re-optimization. Previous work on these problems have motivated us to move the research on these topics forward. Classification of the literature helps us not just to find new mathematical methods to solve these problems, but more importantly, to identify what research problem should be solved in the first place. Identifying new problem dimensions and adopting new ways to solve the problems highlight the originality of this thesis.

5 Evaluation of Inherent Robustness of Survivable Transport Designs against Uncertainty in the Demand Forecast

5.1 Introduction

Chapter 5 is the first chapter of the final module, and it summarizes one of my initial studies conducted in 2002 [LeG02]. In this study, we developed two optimization models that can be used to evaluate the robustness of span-restorable (SR) and path-protected (SBPP) networks to withstand changes in the demand forecast.

As mentioned in the preface, this chapter (as well as the following three chapters) follows a results-oriented and common "problems-formulations-results" structure. We try not to repeat the motivation and background portions of the studies covered by the first two modules (e.g., we will not comment on the general routing behaviors of SR or SBPP, or discuss the basics of the arc-path formulation of the spare capacity placement problem), and the original publication.

Two new concepts are introduced in this chapter, namely, *Pattern Forecast Accuracy* (PFA) to characterize the distribution errors arising in demand forecasts, and the notion of *Servability*, a single measure that can be used to evaluate the inherent robustness of different survivability options to forecast variations. The SR servability formulation serves a basis for the demand loading formulation to be discussed in Chapter 7.

As SR and SBPP have been the two leading schemes in mesh-based survivable transport networking, simulation studies have been conducted on these two network types to compare their servability performances in the sense of how well they withstand departures of the actual demand pattern from the forecast demand to which they were designed. In over 90% of the 5000 test cases, we found that SR networks have the same or slightly better servability (with average about 3%) than the SBPP designs.

5.2 Research Questions

In previous chapters, we have seen that a limitation of all optimization methods is that an optimal capacity design is only "optimal" with respect to an assumed demand forecast or a given set of demand scenarios. Recognizing the high probability that the actual demand pattern will be different from that predicted, network planners should be equipped with testing tools to evaluate the robustness, in terms of the sensitivity to changes in the demand forecast, of any given capacity designs of different survivability architectures. This study attempts to address two specific questions in this regard:

(1) How can we mathematically characterize the error of an actual demand pattern relative to the forecast on which a design was based?

(2) Given the optimal SR or SBPP capacity design with an original demand forecast, how can we measure, and if possible compare, the robustness of these networks to forecast errors?

The next three sections explain a possible solution framework to address these questions. Section 5.3 is devoted to proposing a metric called Pattern Forecast Accuracy (PFA) for quantifying pattern errors in the demand forecast, and explaining its behavior relative to correlation-based measures first proposed by Geary et al. in [GAD01b][18]. In Sections 5.4, we define the notion of servability and comment on its relevance to the measures of routability and restorability. The key optimization formulations of measuring servability will be explained in Section 5.5, and the experimental results are offered in Section 5.6.

5.3 The Concept of Pattern Forecast Accuracy (PFA)

We now explain the metric called *Pattern Forecast Accuracy* (PFA) to quantify the degree to which an actual future demand pattern differs from that which was forecast. We refer to the forecast as the *nominal* demand matrix and the one that actually occurs as the *actual* demand matrix. A perfect forecast means that the actual and nominal demand patterns are identical. In addition, for better understanding of the impact of specific types of forecast changes, we choose to separate scalar errors in forecasting the total growth *volume* from the effects of errors in forecasting the *pattern* of the growth. It is the change in *patterns*, not a simple uniform overload on the whole network, that we try to analyze in terms of its impact on the capacity designs.

To isolate pattern error from the volume error, we assume that the nominal and actual demand matrices have the same volume, V. Any bulk scalar error is normalized to reveal only the structural difference in demand patterns or distributions. To measure this, we define the PFA of the actual matrix relative to the nominal as:

$$PFA \triangleq 1 - \frac{\sum_{i \in D} | d_i - \widetilde{d}_i |}{[V - \min(d_i)] \times 2} \qquad (5.3.1)$$

[18] Readers might refer to Section 4.3.3 for a complete review of [GAD01b].

where D is the set of (unidirectional) demand values on node pairs, d_i is the demand quantity of the i-th pair in the *nominal* demand matrix, $\tilde{d}_i$ is the demand quantity of the i-th pair in the *actual* demand matrix, $V = \sum_{i \in D} d_i$ is the total volume of the nominal (and the actual) demand matrix, and $\min(d_i)$ is the minimum value of d_i from the nominal demand matrix, including zero.

PFA is based on a certain notion of the worst possible pattern of errors in a forecast and, then asks how much total pair-wise error the actual pattern embodies relative to this worst possible error pattern. Table 5.1 illustrates the concept with a nominal forecast and four actual demand matrices in rows. Each matrix is composed of three bi-directional demand pairs and has a total volume of six demand units.

Table 5.1. An example to illustrate the concept of Pattern Forecast Accuracy (PFA).

Nominal demand matrix	OD pair 1	OD pair 2	OD pair 3	PFA
(Total volume, $V = 6$)	2	3	1	
Actual matrix 1	2	3	1	1
Actual matrix 2	2	2	2	0.8
Actual matrix 3	1	2	3	0.6
Actual matrix 4	0	0	6	0

In the first actual matrix, every OD pair demand is identical to nominal, so the numerator of (5.3.1) is zero and the PFA is unity, representing the perfect forecast. At the other extreme, actual matrix 4, by construction, is the "least-similar" actual demand pattern of the same total volume. This worst-case forecast is conceptually the case where all the demand is on only one pair and every other demand pair thus experiences the most discrepancy (i.e., $|d_i - \tilde{d}_i|$) that it could possibly generate under a common total volume. The only non-zero value is generating the maximum error. PFA reflects this as the least accurate forecast with a PFA value of zero[19]. Thus, PFA considers the ratio of total absolute error on demand quantities relative to that of the singular-value, least-match demand pattern. Since we consider the absolute total error, the corresponding factor of two is needed in the denominator. Note that the basic definition of PFA treats all OD pairs equally; different topology-dependent impacts among OD pairs can also be simply reflected by attaching a parameter a_i to the demand values d_i in (5.3.1). For example, weights can be assigned according to the distances of the OD pair.

[19] In the example, the numerator of (5.3.1) is |2-0|+|3-0|+|1-6| = 10 and in the denominator, we have [(2+3+1) - min (2,3,1)] x 2 = 10.

Recall from Section 4.3.3 that Geary et al. introduced a correlation-based metric called the Distribution Forecast Accuracy (DFA) [GAD01b]. Using the following example, we show the difference between the two measures of pattern errors.

DFA, by Geary's definition, is calculated based on the correlation coefficient between the actual (A) and forecast (F) demand matrices, or mathematically, $DFA = \mathrm{cov}(F, A) / \sigma(F) \bullet \sigma(A)$ where *cov()* is the covariance function and $\sigma()$ is the standard deviation. In Table 5.2, we have a nominal demand matrix and three actual demand patterns. Each is composed of 15 OD pairs and has a total volume of 20. We compute both the PFA and DFA values (or correlation coefficient) for each pattern.

Table 5.2. An example comparing the PFA and the correlation metric introduced by Geary et al.

	N1	N2	N3	N4	N5	N6
N1	-	0	4	1	1	1
N2	-	-	2	0	2	0
N3	-	-	-	2	1	1
N4	-	-	-	-	1	1
N5	-	-	-	-	-	3
N6	-	-	-	-	-	-
Nominal demand matrix from [GAD01b]						

	N1	N2	N3	N4	N5	N6
N1	-	0	4	1	1	1
N2	-	-	2	0	2	0
N3	-	-	-	2	2	0
N4	-	-	-	-	1	1
N5	-	-	-	-	-	3
N6	-	-	-	-	-	-
Actual pattern 1						

	N1	N2	N3	N4	N5	N6
N1	-	1	3	1	1	0
N2	-	-	2	0	2	0
N3	-	-	-	2	3	2
N4	-	-	-	-	1	1
N5	-	-	-	-	-	1
N6	-	-	-	-	-	-
Actual pattern 2						

	N1	N2	N3	N4	N5	N6
N1	-	0	12	0	0	0
N2	-	-	0	0	0	0
N3	-	-	-	0	0	0
N4	-	-	-	-	0	0
N5	-	-	-	-	-	8
N6	-	-	-	-	-	-
Actual pattern 3						

Indicates a mis-forecast demand quantity

Method	Actual pattern 1	Actual pattern 2	Actual pattern 3
PFA	0.95	0.80	0.35
Correlation (DFA)	0.95	0.61	0.81

In comparing the results of the first, second and third patterns with respect to the forecast, we see a PFA indicating a progressively worse pattern error, whereas DFA first drops to 0.61 on pattern 2 and then rises back to 0.81 on pattern 3. And yet, it seems apparent by the shaded area that actual matrix 3 is the most severe case of forecast error, and PFA is able to more strongly distinguish the pattern error. When the most individual values are mis-forecast, PFA shows the lowest value, whereas the DFA value actually rises in the example. This illustrates what we found in general while working with correlation measures, namely, that pure mathematical correlation can be rather at odds with a more intuitive notion of how far the actual matrix is from the forecast. Other minor differences are that PFA can be used to reflect OD pairs' priority and is {0,1} bounded rather than {1,-1} bounded. Thus, we find PFA to be preferable to correlation measure for its more intuitive and direct responsiveness to the notion of pattern mismatch.

While it is appealing to analyze the mathematical differences between PFA, DFA, or any other possible measures, readers should note the underlying reason for defining such measures: there is a huge number of ways in which an actual pattern can deviate from the nominal forecast. Therefore, we always need some form of compact measure that is representative for large groups of actual patterns that are different in detail, but equivalent in the range of forecast error they represent.

An interesting aside of PFA is that we have been able to show that for a nominal matrix of D node pairs and a total volume of V demand units (where D and V are both integers), the total number of distinct matrices with the same volume is:

$$No.of\ Possible\ Scenarios = \prod_{i=1}^{D-1}\left(\frac{V}{i}+1\right) \qquad (5.3.2)$$

In the example of Table 5.2 ($V = 20, D = 15$), there are ~14 billion different possible demand patterns. One of these is the "PFA = 1" case. Another case (or a few cases) represent "PFA = 0." In between, many different actual matrices will share the same characteristic PFA values. Obviously, given the number of individual cases, any test for design robustness must rely on representative samples of characteristic forecast error levels.

5.4 The Concept of Servability

In Section 3.3.3, we provided definitions of routability and restorability, as they are commonly used to measure network utilization given a capacitated network. Now we define a new measure called *servability*, a generic metric that can be used to compare any kind of survivable capacity design.

Servability is defined as the fraction of all *actual* demands for which it is feasible to both route *and* protect within a capacitated network. Conceptually, it helps to think of a situation where a survivable network is given, and we ask the question of how many *protected* demands we would be able to serve within the set of as-built capacities. In this study, a protected demand is referred to as a demand unit that is fully restorable to any single span failure. In addition, we assume that the demand unit of any given OD pair is identical or has the same priority to be served.

Independent of the underlying survivable architecture and capacity designs, whether it is SR, SBPP, *p*-cycles or rings, servability therefore can be a single measure for comparing robustness of any kinds of survivable architectures to forecast errors. It is this basic concept of servability that inspired us to come up with the multiple quality-of-services classes (multi-QoP) demand loading formulations to be presented in Chapter 7.

5.5 Optimization Models for Servability Measures

Two integer programs were developed for the evaluation of maximal servability. Unlike most existing methods that use heuristics (e.g., [CHS98][GAD1b][Mau02a][GAM03]) as a performance evaluation technique, our max-servability approach aims to provide fair, optimal and repeatable comparisons. Such rigorously defined formulations help us obtain basic insights about the inherent relative abilities of SR and SBPP to cope with the forecast error. It also enables us to collect specific statistics, such as the most frequently used restoration route sets, the most frequently unserved OD demand pairs, etc., if necessary.

5.5.1 Maximum Servability for Span Restorable (SR) Networks

What follows is an explanation of the two integer linear program formulations. The first is for SR networks (SR-MS) followed by that for SBPP networks (SBPP-MS). The objective of each is to minimize the total number of "unservable" demand units, or equivalently, to maximize the total number of protected demands served. In both cases, the servability is maximized within the as-built capacities, from the corresponding minimum-cost capacity design on the nominal demand.

Sets:

- S Set of all spans in the network, indexed by j or i
- D Set of all origin-destination (OD) pairs in a demand matrix, index r
- Q^r Set of pre-determined eligible working routes for OD pair r, index q
- P_i Set of pre-determined eligible restoration routes available upon the failure of span i, index p

Parameters:

- $\zeta_j^{r,q}$ Equal to one if the q^{th} eligible route for demands between node pair r uses span j, zero otherwise

$\delta_{i,j}^p$ Equal to one if the p^{th} eligible restoration route for span i uses span j, zero otherwise

T_j Total as-built capacity for span j from the minimum-cost (or any given) design to serve the nominal demand matrix

d^r Number of demand units of OD pair r in an actual (target) demand matrix

a^r Optional parameter for setting priority among different OD pair r

Variables:

u^r Number of unserved demand units for OD pair r in the actual demand matrix

w_j Number of working capacity units required on span j

s_j Number of spare capacity units required on span j

$g^{r,q}$ Working flow assigned on the q^{th} working route to serve OD pair r

f_i^p Restoration flow assigned on the p^{th} restoration route upon the failure of span i

SR-MS: Minimize $$\sum_{r\in D} a^r \cdot u^r \qquad (5.5.1)$$

Subject to:

$$\sum_{q\in Q^r} g^{r,q} = d^r - u^r \qquad \forall r \in D \qquad (5.5.2)$$

$$\sum_{r\in D}\sum_{q\in Q^r} \zeta_j^{r,q} \cdot g^{r,q} = w_j \qquad \forall j \in S \qquad (5.5.3)$$

$$\sum_{p\in P_i} f_i^p = w_i \qquad \forall i \in S \qquad (5.5.4)$$

$$s_j \ge \sum_{p\in P_i} \delta_{i,j}^p \cdot f_i^p \qquad \forall (i,j) \in S^2; i \ne j \qquad (5.5.5)$$

$$s_j + w_j \le T_j \qquad \forall j \in S \qquad (5.5.6)$$

$$0 \le u^r \le d^r \qquad \forall r \in D \qquad (5.5.7)$$

$$\text{Servability (\%)} = 1 - \left\{ \sum_{r\in D} u^r \Big/ \sum_{r\in D} d^r \right\} \qquad (5.5.8)$$

Given the total as-built capacity of each span, T_j, and the current actual demand pattern, d^r, constraint set (5.5.2) allocates the demand flows $g^{r,q}$ of OD pair r on different working routes in Q^r. In this study, we will assume that each demand is routed via a single shortest route, i.e., there is only one eligible working route for each OD pair r in Q^r, and the demand flow $g^{r,q}$ is simply equal to the demand value d^r. Constraint set (5.5.3) determines the working capacity w_j needed on each span to simultaneously serve the demand flows. Constraints (5.5.4) and (5.5.5) correspond to the generation of restoration flows f_i^p and spare capacity s_j needed to support all restoration scenarios. Constraints (5.5.6) ensure that the sum of the working and spare capacities on each span j is within the capacities T_j, while constraints (5.5.7) ensures that the number of unserved demand units u^r of each OD pair r does not exceed its demand value. Once we obtain the optimal solution from the formulation, the overall servability is computed by equation (5.5.8). As already mentioned, the objective function can be easily extended to account for different priorities amongst demand pairs simply by weighting u^r with multiplicative factors, a^r, corresponding to the priority class of the demand. In this study, the parameters a^r are set to unity.

5.5.2 Maximum Servability for Shared Backup Path Protected (SBPP) Networks

Sets D and S, parameters a^r, d^r and T_j, and variables u^r, w_j, s_j are not restated as we have previously defined. The following additional sets, parameters and variables are defined for the SBPP case.

Additional Input Sets and Parameters:

- P_r Set of eligible end-to-end backup routes, indexed by b. These backup routes are also span-disjoint from the corresponding working routes of OD pair r.
- β_i^r Equal to one if span i is on the working path for OD pair r, zero otherwise
- $\delta_{j,b}^r$ Equal to one if the b^{th} backup route uses span j for protecting OD pair r, zero otherwise

Additional Variables:

- x_b^r Equal to one if the b^{th} backup route is used to protect OD pair r, zero otherwise
- z_b^r A positive fractional number which indicates the portion of demand units of OD pair r that is being served and protected by the b^{th} backup route

SBPP-MS: Minimize $$\sum_{r\in D} a^r \cdot u^r \qquad (5.5.9)$$

Subject to:

$$\sum_{r\in D} \beta_i^r \bullet (d_r - u_r) = w_i \qquad \forall j \in S \qquad (5.5.10)$$

$$\sum_{b \in P_r} x_b^r = 1 \qquad \forall r \in D \qquad (5.5.11)$$

$$s_j \geq \sum_{r \in D} \sum_{b \in P_r} \beta_i^r \bullet \delta_{j,b}^r \bullet d^r \bullet z_b^r \qquad \forall (i,j) \in S^2; i \neq j \qquad (5.5.12)$$

$$z_b^r \geq x_b^r - (u^r / d^r) \qquad \forall r \in D; \forall b \in P_r \qquad (5.5.13)$$

$$z_b^r \geq 0 \qquad \forall r \in D; \forall b \in P_r \qquad (5.5.14)$$

$$s_j + w_j \leq T_j \qquad \forall j \in S \qquad (5.5.15)$$

$$0 \leq u^r \leq d^r \qquad \forall r \in D \qquad (5.5.16)$$

Under SBPP, the given capacity of each span, T_j, is conceptually the same as above but comes from a prior cost-optimal SBPP design for the nominal demand pattern. The T_j values and the total capacity are, therefore, not necessarily identical for SR and SBPP servability problems. This is so because we want to measure servability of the corresponding capacity and underlying survivability architecture as it would have been built for the same nominal forecast.

The working capacity constraint (5.5.10) plays essentially the same role as that for (5.5.2) and (5.5.3), and assumes that each demand is pre-routed via a single shortest route. Unlike span restoration that allows multiple detour paths to be used for each span failure, an operating principle of SBPP is that there is only *one* span-disjoint, end-to-end backup route for each working path, which is reflected in (5.5.11). Constraints (5.5.12) allocate spare capacity s_j on each span j only for those failures where (i) the working route of demand r is affected by the failed span i (i.e., $\beta_i^r = 1$); (ii) the span j on the backup route b is used to protect the same demand r (i.e., $\delta_{j,b}^r = 1$), and (iii) z_b^r has a non-zero value. The role of z_b^r is to allow partial restorability of the demand on a certain OD pair if that is possible. The value of z_b^r is determined by (5.5.13) and (5.5.14). If backup route b is not used to protect OD pair r (i.e., $x_b^r = 0$), the combined condition of (5.5.13) and (5.5.14) will force the variable z_b^r to zero so that no spare capacity is required in (5.5.12). On the other hand, if backup route b is indeed used to reroute all or part of demand for the OD pair r (i.e., $x_b^r = 1$), the value of z_b^r will become a positive fraction, or unity (if $u^r = 0$), representing the portion of demands that requires spare capacities allocation. Constraints (5.5.15) and (5.5.16) are identical to (5.5.6) and (5.5.7) respectively. Equation (5.5.8) once again evaluates the overall servability of the SBPP network.

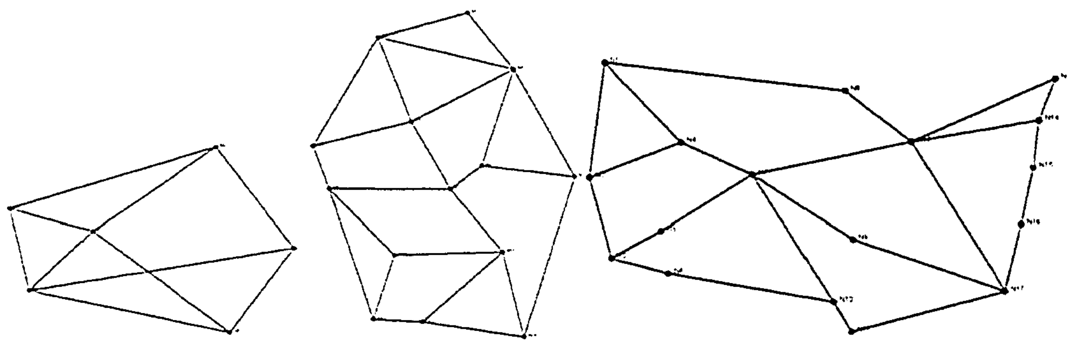
5.6 Experimental Design and Results

This section explains the experimental comparison of SR and SBPP servability under various amounts of forecast error. First, for each nominal demand forecast, we obtain the minimum-capacity SR and SBPP network designs. These solutions are obtained by solving separate spare capacity placement problems as discussed in Section 4.2 (or see [DoG00][GIZ00] for the SR design methods and [DoG01] for the SBPP capacity designs).

With these optimal capacity design models, we can produce the reference networks that set the T_j capacities, representing the as-built capacities based on the forecast. A series of PFA test cases are required and generated for the input to the simulation studies. Based on these random test demand patterns, servability of each test case can be evaluated using the SR-MS and SBPP-MS formulations as we have just explained. Once we obtain the set of servability results, we will compare the robustness of the two survivability schemes to demand uncertainty.

5.6.1 Test Networks and Reference Capacity Designs for the Nominal Demand Forecasts

Three test network topologies and the initial nominal demand forecasts are shown in Figure 5.1 and Table 5.3, respectively. The *Metro* network is a simple 6-node 10-span artificial network created for manual validation of the servability formulations. The *Germany* network is a German backbone network provided in [BaK00]. The *US* network is from [RBS01] and has the lowest degree of connectivity. Detailed descriptions of these topologies can also be found in Appendix A.



(a) Metro (b) Germany (c) US

Figure 5.1. Test network topologies for servability study.

Table 5.3. The nominal demand forecasts' characteristics for servability study.

Network	No. of (nodes, spans)	Avg. nodal degree	Avg. span distance	No. of O-D pairs	Demand units (sum, max, min)	Constant in (5.6.1)
Metro	(6, 10)	3.33	191	15	(38, 6, 1)	41
Germany	(14, 24)	3.43	110	43	(126, 8, 2)	30
US	(17, 24)	2.82	119	60	(208, 10, 2)	60

The nominal forecast is created based on a gravity-based demand model by Doucette et al. [DoG00] as shown in Equation (5.6.1). The gravity-based demand model mimics a situation where major centers tend to attract demand, but with some tendency for demand to be inversely dependent on distance. The span distances are proportional to the Euclidean distances as drawn with the mean span distances and demand generating constants detailed in Table 5.3.

$$demand\,(a,b) = \text{int}\left[\frac{nodal\ degree_a\ x\ nodal\ degree_b}{distance_{a-b}} \bullet constant\right] \qquad (5.6.1)$$

In the SBPP reference designs and servability trials, we used five eligible span-disjoint backup routes (P_r) for each OD pair. To control the computational times for the SBPP-MS formulation, it was necessary to reduce the number of OD pairs in the two larger networks. For instance, approximately half of the total pairs (i.e., 43 out of 91 OD pairs) are removed from the original matrix in the *Germany* network. Similarly, we reduced 136 OD demand pairs to 60 demand pairs for testing the *US* network. The reduced demand sets were then used in both SR and SBPP servability trials.

Six reference networks were created: two for each topology corresponding to the SR and SPBB reference designs. The total capacity of each reference network and other design details are summarized in Table 5.4. In all these designs and in the servability solution, each working demand pair is restricted to route over its single shortest route so that the working capacities of the SBPP and SR reference designs are identical. In addition, the same sets of eligible restoration (or backup) routes used in the SR (or SBPP) reference designs are used for the subsequent servability problems, such that the servability of the reference designs under the nominal forecast matrix is always 100%. Note that, in one case, the total capacity of the SR design is 11% more than that of the SBPP design. This difference in initial capacity arises because the intent is to measure the servability loss of each architecture relative to its *own* nominal (and optimal) design. If the differences were much larger, however, we would have to consider additional factors in the study design, since it would be reasonable to expect that a much larger initial investment in capacity should produce some corresponding benefit in retention of servability.

Table 5.4. Capacity requirements and routing details of the reference networks.

Optimal design	Network	Eligible restoration/ backup routes in optimal capacity design models	Total working capacity	Total spare capacity	Total overall capacity (relative % to SBPP design)
SBPP	Metro	All possible routes for each OD pair	46	30	76 (100%)
	Germany	5 shortest backup routes per OD pair	185	137	322 (100%)
	US	5 shortest backup routes per OD pair	334	238	572 (100%)
SR	Metro	All possible routes for each span	46	27	73 (96%)
	Germany	All possible routes for each span	185	148	333 (103%)
	US	All possible routes for each span	334	301	635 (111%)

5.6.2 Generation of Test-case Demand Patterns

Once we have the reference networks, the next step is to generate random demand patterns with progressively worse PFA values. One thousand test demand patterns were randomly generated for the *Metro* network while two thousand test patterns were created for each *Germany* and *US* network. The actual demand patterns were generated as deviations from the nominal matrices by random swap and random add/subtract operations on elements of the nominal demand matrix, so that total volume remains constant. PFA is not directly controlled in the synthesis of different forecast error cases. Rather, random-walk sequences of evolution away from the nominal demand patterns are generated and actual patterns are sampled during the process. The PFA of each is calculated to quantify the amount of forecast error they embody.

5.6.3 Results and Discussion

The SR and SBPP maximum servability models were implemented in AMPL [FGK93] and solved with Parallel CPLEX 7.1 MIP Solver [ILO04] on a four-processor Ultrasparc at 450 MHz and 4 GB of RAM running Sun Solaris 8 OS. For each actual demand pattern, we obtained the optimal servability result to a MIPGAP of under 10^{-4} (i.e., within 0.01% of optimal). Getting the whole set of 2000 data points for SR servability took less than 20 minutes on the *US* network, but as long as three days to solve for the corresponding SBPP servability solutions. This follows the same pattern as the reference design problems. Obtaining optimal SBPP designs has always been a challenge to researchers, especially on large networks[20].

The servability results on the three test networks are shown in Figures 5.2, 5.3 and 5.4. Each figure contains two sets of curves corresponding to the servability results from the SR and SBPP designs. For each design, there are three curves indicating the mean, and the 95th and 5th percentile servability values over all test cases of each PFA value. The actual PFA values were

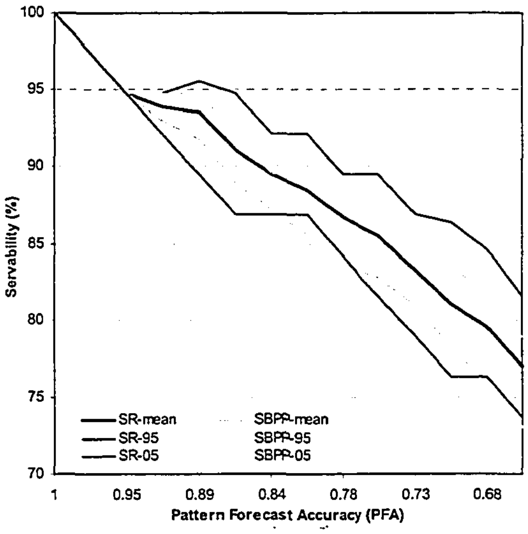


Figure 5.2. Servability vs. PFA results from the *Metro* network.

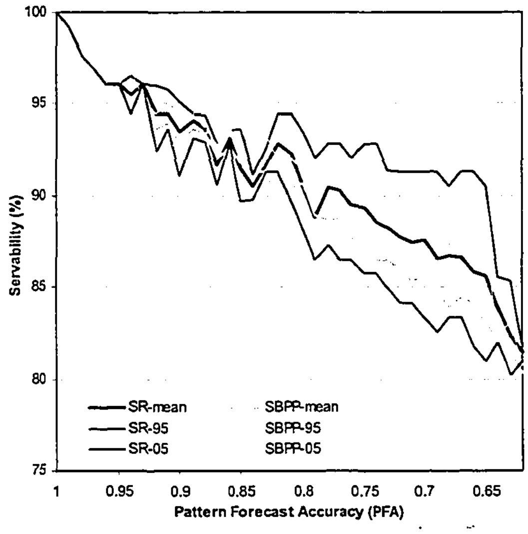


Figure 5.3. Servability vs. PFA results from the *Germany* network.

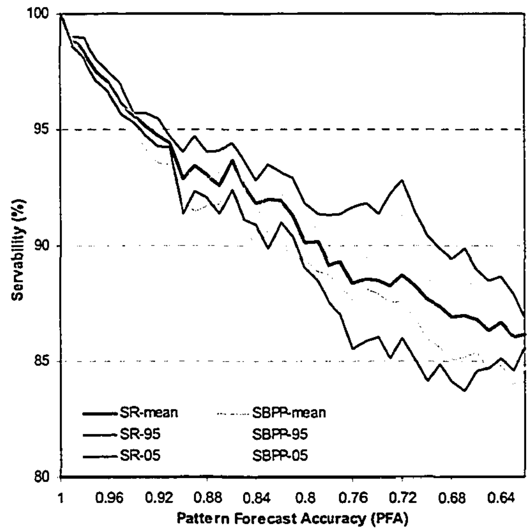


Figure 5.4. Servability vs. PFA results from the *US* network.

[20] The large number of x_b^r 1/0 decision variables is one of reasons that make solving the optimal SBPP design problems so difficult [KeW98][Gro04].

put into bins within +/- 0.05 of exact 2-digit PFA values for the percentile calculations. For example, the mean, 95th and 5th percentiles for PFA= 0.78 are actually formed over all individual PFA values from 0.775 to 0.785.

As generally expected, the servability in all cases drops as the PFA decreases. What was less expected is the almost linear average-case loss of servability of all schemes with decreasing PFA. Based on the comparison of mean servability curves, it appears that there is no significant difference between the two schemes in terms of their average servability over the sets of 1000 or 2000 PFA trials. On average, the servability of the SR designs was about 3% higher than that of the SBPP designs.

Besides the mean values, we can analyze the top and bottom envelope curves of Figures 5.2 to 5.4 and consider the range of best- and worst-case outcomes that could arise from the mis-forecasts. In the PFA region from 0.82 to 0.65, we see that SR networks can generally retain higher servability than SBPP networks. The worst individual cases of servability loss all arise under SBPP. To further examine the body of statistical trials from this viewpoint, we generate a scatter plot of the individual differences between SR and SBPP servability over the all (i.e., 5000) actual demand trials represented in Figures 5.2 to 5.4.

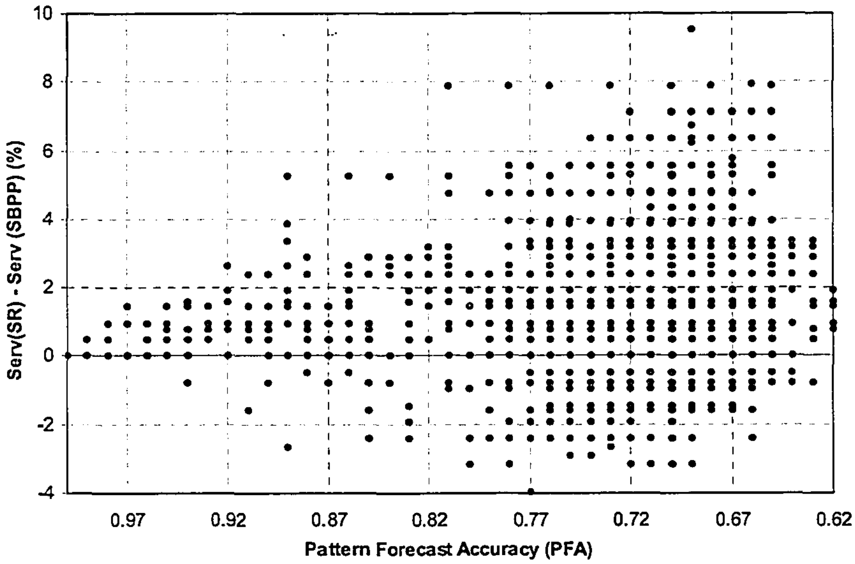


Figure 5.5. Scatter plot of *Serv*(SR) - *Serv*(SBPP) versus test case PFA over all 5000 trials.

The scatter plot in Figure 5.5 is skewed above zero, indicating there are more individual cases where SR servability remained higher that that of SBPP, especially in the low PFA region. In over 90% of the 5000 test cases, SR has the same or slightly higher servability than the SBPP design, where the difference in servability averages approximately 3%.

5.7 Summary

The concepts of Pattern Forecast Accuracy (PFA) and Servability have been explained. Based on the two concepts, we proposed a general framework for assessing the robustness of the ability of various transport network architectures to cope with uncertainty in the demand forecast. Two integer linear programs are formulated to determine the optimal servability solutions for both span restorable (SR) and shared backup path protected (SBPP) networks. We found that in over 90% of all test cases, the minimum-capacity SR designs are able to serve more mis-forecast protected demands than the corresponding minimum-capacity SBPP designs. However, because the largest differences are under 10% and average only about 3%, we can reasonably conclude that SR and SBPP minimum-capacity designs are essentially equally "future-proof."

While the basic framework, using PFA to characterize forecast error and maximum servability to assess robustness, is suggested as a general methodology for the study of robustness of various survivable transport architectures, researchers are encouraged to develop other metrics and frameworks for future contributions to this problem. Instead of using servability to assess robustness, for example, one might adopt the idea of recourse cost (e.g., the extra cost required to carry the mis-forecast demands) or the notion of regret (described in Section 4.3.2) as a metric to quantify robustness. Other non-PFA, non-DFA concepts could also be investigated to better characterize errors in the demand patterns. Advances on this research topic would provide additional tools for network planners to evaluate and compare possible designs and effectively identify those that are robust.

In the next chapter, we will introduce a complementary approach, based on the modeling framework of Stochastic Programming, to incorporate demand uncertainty into a single design formulation. The combined use of this method with the evaluative PFA-Servability framework might provide the most complete methodologies for decision makers to plan against and to evaluate the negative effects due to uncertainty in the demand forecast.

6 Capacity Planning of Mesh-based Survivable Transport Networks Under Demand Uncertainty

6.1 Introduction

This chapter presents a unique approach to the capacity-planning problem on mesh survivable networks under uncertainty. Related works [LeG04a][LeG04b][LeG05] have been published, the key findings of which are summarized in this chapter.

In Sections 4.2 and 4.3 we provided a comprehensive literature review on the capacity planning problem of survivable mesh networks. We have seen that almost all published studies on the design of survivable networks is based on a specific demand forecast, to which one optimizes routing and transport capacity assignment for a single target planning view. Generally, these single-forecast models are used repetitively by a planner to consider a range of different scenarios individually, and to develop intuition about how to proceed. In coping with demand uncertainty, these models can only treat the present capacity investment and future corrective actions as two separate problems. We presented such a possible "descriptive" approach in Chapter 5 to evaluate the servability of survivable networks to withstand changes in the demand forecast. However, this is not the same as having a planning method that inclusively and explicitly considers a range of possible futures all at once.

The key objective of this chapter is to develop new capacity planning models based on a well-known Stochastic Programming framework, which allows us to incorporate demand uncertainty and network survivability into a single design formulation. Two integer program formulations are developed, both of which try to minimize the cost of initial design construction *and* the expected cost of possible augmentations or "recourse" actions required in the future, adapting the network to accommodate different actual future demands. In practice, these recourse actions might include lighting up a new DWDM channel on an existing fiber, pulling-in additional cables, or leasing additional capacity from third party network operators, and so on. Realistic aspects of optical networking, such as modularity and economy-of-scale effects, are considered in one of the proposed formulations. These are not only important practical details to reflect in planning, but also give the "future-proof" design problem for such networks some unique aspects.

6.2 Research Questions

Recognizing that virtually all traditional methods for the capacity design of survivable networks, for both ring- and mesh-based transport designs, are based on a specific target of an assumed future demand matrix and motivated by the needs for a more complete tool to plan under uncertainty, we try to address these specific questions in the following sections:

(1) How can we incorporate the aspects of demand uncertainty, modularity and economy-of-scale of capacities into existing survivable capacity designs, such as span restorable (SR) and other capacity planning models?

(2) What are the benefits and limitations of adopting such a design approach, rather than using the conventional single-forecast, least-cost capacity design techniques?

The reminder of this chapter is arranged as follows. Section 6.3 introduces the concept of viewing capacity planning as a two-part investment problem, while section 6.4 reviews the aspects of modularity and the economy-of-scale from the capacity planning perspective. In Section 6.5, we present the two integer programs: one aims at the design of future-proof SR without the modularity aspect (TP-SR), and the other considers both modularity and economies-of-scale (TP-MSR). Finally, Sections 6.6 and 6.7 present the experimental results and discusses the significance of the new formulations.

It is important to note that since we have already explained the general difference between Stochastic Programming (SP) and Robust Optimization (RO), as well as made detailed comparisons between our work to the closely related work [KLO01][KOL03][BKO03] by Kennington et al. in Section 4.3.2, we will not repeat the justification of *why* SP is a more preferable technique than RO in modeling these research problems.

6.3 The Notion of Capacity Planning as a Two-Part Investment Problem

An important concept that sets this work apart from traditional survivable design methods is the modeling of capacity planning as a two-part investment problem, based on the mathematical framework of stochastic programming (SP) with linear recourse [KaW94][21]. The concept of SP with linear recourse can be explained as a "two-part" investment decision process. Note that while the term "two-stage" is generally used in the Stochastic Programming literature, here we choose to use "two-part" to avoid confusion with the predominance of other work in network design, where "two-stage" implies that two successive computational "stages" are used. As we will see later in the formulation, it is solved in a single computational stage or step.

---

[21] Readers might refer to Section 4.3.2 for more complete explanations.

The first part considers the budget X to be invested *at present* and the second part represents the corrective or "recourse" action Y to take place *in future* when uncertainty unfolds. Compared to the traditional approach, the two-part model better reflects the complete life-cycle investment costs associated with capacity planning by facilities-based service providers[22] today. For the simplest assumption, the recourse costs might refer to the cost of "lighting up" (i.e., fully equipping and commissioning) new fiber system and / or additional single channels on those systems needed for either protection or working capacity. Time value of money can also be reflected by discounting the recourse costs with respect to the present cost. Other less-obvious future costs such as the construction cost associated with pulling in new cables, labor-related operational and maintenance costs required to support new services, and the penalty cost of leasing capacity from third-party network operators, etc ... could also be captured by the recourse cost parameters. In this regard, economy of scale (EoS) effects can be appreciated as important factors in future-proof planning. Without the EoS effects, for example, we may prefer many small capacity modules (e.g., OC-48s and/or single wavelengths) to minimize the present costs. However, if certain OD demands increase unexpectedly, the extra cost of adding more capacity in small modules in future may exceed the cost of having large-capacity modules (e.g., OC-192s and/or whole multi-wavelength waveband equipment) in the first place.

One form of future recourse that we do not consider here is any changes to the physical network topology itself. As we mentioned in Section 2.4.2, changing the network topology generally belongs within the scope of the long-term planning (LTP) problem and involves major strategic factors in network planning, such as physical rights-of-way acquisition, installation of ducting, power and so on. Thus our two-part design model should only be applicable to medium-term (MTP), or short-term (STP) capacity investment problems, where under all future scenarios, the physical graph topology remains constant as given in the initial network.

Another comment related to this two-part design methodology is that demand uncertainty is classified in Level II (as discussed in Section 4.3.1), or is characterized by a set of plausible demand scenarios and each is associated with a probability measure. Under circumstances where the uncertainty goes beyond Level II, it is better to use more "descriptive" approaches (e.g., the framework suggested in Chapter 5) to evaluate the effect of uncertainty, but not incorporate the massive set of possible outcomes into a single planning model.

---

[22] By facilities-based providers, we refer to the ones that own or lease a substantial portion of the plant, property and equipment necessary to provide a broad range of integrated communications services. Level 3 Communications, Global Crossing and Qwest Communications are some examples.

Finally, while we acknowledge that there are many other different kinds of mesh-based survivability schemes in general (e.g., path restoration, meta-mesh, SBPP, *p*-cycles, etc.), the primary objective of this work is to propose and develop a basic framework using the span-restorable (SR) network as a vehicle for research, so that such a framework and associated principles can later be adapted to other survivability architectures as well.

6.4 The Combined Concept of Modularity and Economy of Scale

The incorporation of modularity and economy-of-scale (EoS) effects into optimal span-restorable capacity designs was first introduced by Doucette and Grover in [DGM99][DoG00] and later by Kennington and Lewis [KeL01] to the designs of path-restorable networks. In both studies, the researchers realized that available capacity increments of actual transmission systems are usually modular in nature, and in addition, the costs of increasing modular size follow some stair-step function versus capacity. For instance, typical module sizes in SONET may be OC-3, OC-12, OC-48, OC-192, etc. and an OC-192 will generally cost significantly less than four times the cost of an OC-48. The EoS captures the non-linear cost-capacity relationships in transmission capacity. These are the effects to be modeled in the following two-part capacity design formulations, and when network planners include these factors, the benefit of greater present expenditure on a large module might be warranted and produce somewhat forward-looking solution to reduce future recourse costs. This proposition might be especially valid when significant economy-of-scale effects are present. It is reasonable to expect that the combined effect of modularity and EoS may have an impact on reducing both present and future recourse costs.

6.5 Optimization Models for Span-Restorable Network Design under Uncertainty

We will now present the optimization model, including definition of the mathematical means through which we can capture the notion of future recourse to repair any shortcoming in the initial design in the face of future demand that is different from the nominal forecast. As mentioned, we work with span-restorable networks. In this regard, our starting point is to use the arc-path formulation introduced by Herzberg and Bye [HeB94] (presented in Section 4.2.2) for minimizing the total spare capacity cost of a fully span-restorable network, and then extend the model to create a *joint* design formulation. Recall from Section 4.2.2 that in the joint model, the routing of demands is simultaneously optimized with the placement of spare capacity such that the overall working capacity plus spare capacity is minimized in a survivable network.

6.5.1 *Two-Part Span-Restorable Design (TP-SR) without Modularity*

The key concept for the two-part span-restorable design is as follows. In the first part, a budget X is invested initially while the second part considers the cost of a corrective action $Y(k)$ if a future scenario k (modeled by a set of scenarios $k \in U$) occurs. In our problem, the present outlay X is the cost of an initial network design that is assured to serve and protect all demands of the defined *nominal forecast*, k_0. The expected recourse cost Y is the mathematical expectation of recourse costs over all future scenarios k that are possible and differ from k_0. Note that the nominal forecast can itself be arbitrarily certain – in many applications of this model – it can represent the *current actual* demand pattern. Under the assumption that for medium-term planning problems, the number of significantly different demand scenarios is typically in the order of tens (i.e., Level II uncertainty model), the original stochastic program can, in practice, be represented as an integer program of the deterministic equivalent form, for which standard solvers can be used. The two-part span-restorable capacity design (TP-SR) is as follows:

Sets:

- S Set of all spans in the network, indexed by j or i
- U Set of all possible future demand scenarios to be considered, index k
- D Set of all origin-destination (OD) pairs in a demand matrix, index r
- Q^r Set of pre-determined eligible working routes for OD pair r, index q
- P_i Set of pre-determined eligible restoration routes available upon the failure of span i, index p

Parameters:

- C_j Present cost of a unit capacity placed on span j
- R_j Recourse cost of placing an extra unit capacity on span j to cope with the unfolding of demand uncertainty. R_j can simply be a multiplicative value of C_j, or any other absolute value specific for each span j
- P_k Probability estimate for demand scenario k
- d_k^r Magnitude of the bi-directional (integer) demand on node pair r in scenario k
- $\zeta_j^{r,q}$ Equal to one if the q^{th} eligible route for demands between node pair r uses span j, zero otherwise
- $\delta_{i,j}^p$ Equal to one if the p^{th} eligible route for span i uses span j, zero otherwise

Variables:

- w_j Number of working capacity units on span j for the design

s_j Number of spare capacity units on span j for the design

$y_{j,k}$ Number of additional working capacity units that would have to be placed on span j in future to cope with scenario k

$z_{j,k}$ Number of additional spare capacity units required on span j under future demand scenario k

$g_k^{r,q}$ Working flow assigned on the q^{th} working route to serve OD pair r in scenario k

$f_{i,k}^p$ Restoration flow assigned on the p^{th} restoration route upon the failure of span i in scenario k

TP-SR: Minimize $$\sum_{j \in S} C_j \cdot (w_j + s_j) + \sum_{j \in S} \sum_{k=U} P_k \cdot R_j \cdot (y_{j,k} + z_{j,k}) \qquad (6.5.1)$$

Subject to:

$$\sum_{q \in Q^r} g_k^{r,q} = d_k^r \qquad \forall r \in D; \forall k \in U \qquad (6.5.2)$$

$$\sum_{r \in D} \sum_{q \in Q^r} \zeta_j^{r,q} \cdot g_k^{r,q} = w_j + y_{j,k} \qquad \forall j \in S; \forall k \in U \qquad (6.5.3)$$

$$\sum_{p \in P_i} f_{i,k}^p = w_i + y_{i,k} \qquad \forall i \in S; \forall k \in U \qquad (6.5.4)$$

$$s_j + z_{j,k} \geq \sum_{p \in P_i} \delta_{i,j}^p \cdot f_{i,k}^p \qquad \forall (i,j) \in S^2; i \neq j; \forall k \in U \qquad (6.5.5)$$

$$y_{j,k} = 0,\ z_{j,k} = 0 \qquad k = 0; \forall j \in S \qquad (6.5.6)$$

The objective is to minimize the total cost of the network design, i.e., the present cost denoted by the first term in (6.5.1) *plus* the expected value of the future costs to augment the design to serve each possible demand scenario $k \in U$. The parameter C_j is the present cost of a unit capacity on span j, and R_j is the recourse cost if extra working capacity $y_{j,k}$ and/or spare capacity $z_{j,k}$ must be added to span j in the future under scenario k.

In the general cost model, where recourse costs are specific to each span to reflect practical realities, such as dark fiber existing on some spans but not on others, the cost of leasing capacity on particular spans or routes from a third party carrier, and so on. For comparative study, we will use a common recourse cost factor for *all* spans, i.e., $R_j = \alpha * C_j$, and hereafter we refer α as the recourse cost factor.

In each scenario k, constraint (6.5.2) allocates the demand flows $g^{r,q}$ of OD pair r onto working routes q in Q^r, representing a set of pre-determined eligible routes for the demands. Constraint (6.5.3) determines the working capacity w_j required on each span to simultaneously serve the demand flows. $\zeta_j^{r,q}$ is an input parameter that is 1 if the q^{th} working route for OD pair r uses span j, or zero otherwise. For any scenario k where there is a mismatch between the level of demands and the initially installed working capacities, extra working capacities $y_{j,k}$ are added to serve the unexpected demands in the future design[23].

Constraints (6.5.4) and (6.5.5) correspond to the network survivability constraints based on span restoration. Note that other span-based restoration schemes (such as p-cycles) can be adapted to this formulation by employing a corresponding set of constraints that are specific to the particular restoration mechanism used[24]. Constraint (6.5.4) ensures that the total of all restoration flow $f_{i,k}^{p}$ assigned to the eligible routes in P_i when span i fails, satisfies the restoration requirement for that failure scenario (i.e., the total working capacity affected). Constraint (6.5.5) generates the required spare capacities s_j to support the largest of all simultaneously imposed restoration flows crossing each span under each failure scenario and in every demand scenario. If there is a shortage in spare capacity s_j on span j under possible scenario k, extra spare capacities $z_{j,k}$ would be added.

An important detail in this model is how we ensure that the nominal forecast *must* be satisfied. For all other scenarios, we consider only their cost of repair should they arise. This is done simply by imposing $y_{j,k} = 0$ and $z_{j,k} = 0$ for k_0 in constraint (6.5.6), which states that there can be no "extra" capacity of either type associated with ensuring the routability and restorability constraints above. This forces the design to contain adequate "present capacities," w_j *and* s_j, for the nominal scenario k_0. The corresponding constraints can, for all other scenarios, be satisfied by the admission of "non-zero" possible future additional capacities, $y_{j,k}$ and $z_{j,k}$. As a result of this effect, two other relevant types of design that can also be obtained by the same formulation:

Min-Expected Total Cost Design: If constraint (6.5.6) is deleted, we would obtain the network that represents the least expected (total) cost strategy over all possible futures. In this case, what

---

[23] Note that the "extra" working capacities $y_{j,k}$ (and later $z_{j,k}$ for spare capacities) take only zero or positive values. This means that no *removal* of initially installed capacity is ever anticipated. This does not imply, however, that the future demand scenarios only represent growth in demands. Under each future scenario here, some demands decrease while others increase. If, under a given future scenario, some initially placed capacity is unused, this would be accepted as an implication of what was an optimum overall strategy. On the other hand, if any present capacity is fully and efficiently re-used by the solver under *every* future scenario (before new recourse costs are added), such a capacity set simply implies a great built-in tendency not to have very much unused capacity in future scenarios.

[24] In Appendix C, we will illustrate how the two-part concept can be adapted to p-cycles capacity designs.

is built "today" is in effect the component of all possible future networks required. This is common enough to the range of future outcomes to be worth investing in at present, given the cost of capacity at present is less than in future (or with recourse cost factors >1). Conversely, if the recourse cost factor is less than 1, the optimal present network cost can in fact be zero, since it becomes obvious and economical to wait and build the network when uncertainty unfolds.

Most-Expensive Total Cost Design: If constraint (6.5.6) is asserted for *all* recourse capacity variables, i.e., $y_{j,k} = z_{j,k} = 0$ for all span j and all scenarios k, we would end up with the design that is guaranteed at initial construction to serve *all* defined future scenarios. In the language of Stochastic Programming [KaW94][BiL97], this brute-force kind of future-proofing solution is referred as the "fat solution": it serves all possible future scenarios by its basic design, but is the most expensive strategy in general. Under an unreasonably high recourse scenario (e.g., $R >$ 100), the optimal solver would produce a design that is close to the "fat solution," since the cost for possible recourse is so much higher than the cost of investing today.

In Section 6.6, we will make various comparisons between the main TP-SR design model and the two related extremes that can be easily derived from it simply by variants on Constraint (6.5.6). Unlike the traditional span-restorable design, whose objective is to minimize solely the initial total capacity, this two-part model allows us to minimize the present investment as well as the expected consequences and risk (characterized by R_j) of the present decision. It is important to address that an associated output from this model is not only full of details of the present network to build, but also each of the specific future *recourse actions* (through $y_{j,k}$ and $z_{j,k}$) that are required to cope for whatever demand scenario actually arises. Explicitly captured by this model, the coping or adaptation information not only tells us where to add capacities, how to route the unexpected (relative to nominal) working demand, and how to make updates to the restoration routing plans, but it may also suggest changes in the routing of one or more existing paths as part of the overall future adaptation or re-optimization plan.

6.5.2 Two-Part Span-Restorable Design with Modularity and Economy of Scale Effects (TP-MSR)

Based on the same two-part modeling framework, we now construct a more complete optimization model that combines modularity and economy-of-scale effects. This model (TP-MSR) requires some new notations, which are shown as follows. All previously defined sets, parameters and variables continue to apply.

Additional Set:

M Set of module capacities, indexed by m

Additional Parameters:

Z^m Number of capacity units for the m^{th} module size (e.g., 3, 12, 48, 192)

C_j^m Cost of a module of size m placed on span j and is used to reflect different degrees of economy-of-scale

R_j^m Recourse cost factor of a module of size m placed on span j relative to C_j^m

Additional Variables:

n_j^m Number of modules of type m placed on span j for the initial design

$e_{j,k}^m$ Number of extra modules of type m required on span j to cope with the uncertain demand scenario k

TP-MSR: Minimize $$\sum_{m\in M}\sum_{j\in S} C_j^m \cdot n_j^m + \sum_{m\in M}\sum_{j\in S}\sum_{k=U} P_k \cdot R_j^m \cdot e_{j,k}^m \qquad (6.5.7)$$

Subject to (6.5.2), (6.5.3), (6.5.4), (6.5.5) and

$$w_j + s_j \le \sum_{m\in M} Z^m \cdot n_j^m \qquad \forall j \in S \qquad (6.5.8)$$

$$y_{j,k} + z_{j,k} \le \sum_{m\in M} Z^m \cdot e_{j,k}^m \qquad \forall j \in S; \forall k \in U \qquad (6.5.9)$$

$$e_{j,k}^m = 0 \qquad k = 0; \forall j \in S; \forall k \in U \qquad (6.5.10)$$

The new objective function (6.5.7) minimizes the total of the cost of all *modules* initially placed plus the expected cost of extra capacity module placements in future. C_j^m is the cost of placing a single module m on span j at present, and R_j^m is the cost of placing new modules m on span j in the future as needed. Constraint (6.5.8) asserts that the capacity of the set of initially placed modules is adequate for the current demands and their protection. Constraint (6.5.9) relates the presently placed modular capacities to the unfulfilled requirements that are implied under each future outcome scenario, which collectively determine the expected recourse cost in the second part of the objective function. Constraint (6.5.10) plays the same role as (6.5.6), ensuring that the design is a fully feasible for the nominal demand forecast (or presently existing demand).

6.6 Experimental Design

6.6.1 Economy of Scale Model for Capacity

Let us now define a general model for module costs (i.e., C_j^m parameter) under various economies-of-scale assumptions. Given the cost of a minimum common-factor module, the cost of a larger module size (size2) is:

$$\text{For cost scheme } m \times nx\text{: } \text{Cost(size2)} = \text{Cost(size1)} \cdot n^{\frac{\log(\text{size2}/\text{size1})}{\log(m)}} \qquad (6.6.1)$$

where m and n characterize the economy of scale effect in that we obtain "m times capacity for n times the cost." This is denoted "mxnx" economny of scale. For example, the cost of 48-channel module under *4x2x* economy of scale is 120, provided that the cost of a 3-channel module is 30 (i.e., *size1* = 3; Cost(*size1*) = 30; *size2* = 48; $m = 4$, $n = 2$). Of course if we set any $m = n$, we will have a model where the cost is simply linear to the capacity. Table 6.1 lists the actual economy of scale cost-capacity progressions generated by this model and used in our following test cases.

Table 6.1. Cost of modules under various economy-of-scale scenarios.

Economy of Scale	Module Size 3	Module Size 12	Module Size 48	Module Size 192
2x2x	30	120	480	1920
3x2x	30	72	173	414
4x2x	30	60	120	240
6x2x	30	51	88	150

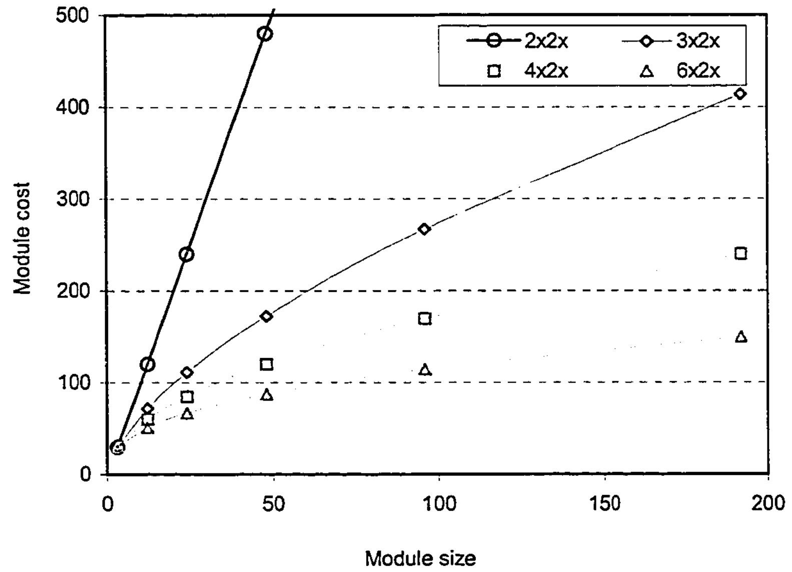


Figure 6.1. Differences in module cost among various economies-of-scale.

6.6.2 Test Networks and Nominal Demand Forecast

A well-documented pan European network, COST239 network [SGA02], is used to implement both non-modular and modular design formulations. This network has 11 nodes and 26 spans with an average nodal degree of 4.7. The topology is shown in Figure 6.2. The next step is to generate a nominal demand forecast and a set of plausible demand scenarios. For the nominal forecast, we chose to create it based on a gravity-based demand model (defined in Equation 5.6.1), and we repeat it here for convenience. We realize that although this demand model may not reflect the present real-world demand traffic, it does allow us and other researchers to reproduce the exact starting demand forecast or other repeatable demand scenarios for future comparative studies. The distances in Equation (6.6.2) refer to Euclidean distances between any two nodes (a,b) and the constant is simply a uniform scaling factor for adjusting the traffic to the desired volume level. Table 6.2 summarizes the properties of the network and nominal demand forecast.

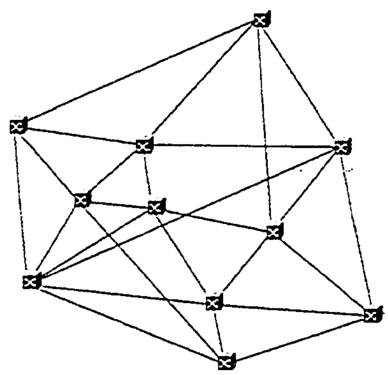


Figure 6.2. The COST239 network topology.

$$demand\,(a,b) = \text{int}\left[\frac{nodal\ degree_a\ x\ nodal\ degree_b}{distance_{a-b}} \bullet constant\right] \qquad (6.6.2)$$

Table 6.2. Topology and nominal forecast characteristics.

Nodes	Spans	Span distance [min, avg, max]	No. of OD pairs	Demand unit per pair	Total demand	Constant in (6.6.2)
11	26	[210, 579, 1310]	55	9.76	547	60

6.6.3 Alternate Futures for the Test Case

To reflect the alternative futures, a set of 20 future demand scenarios was generated, where one represents the "k_0" nominal forecast, and the other 19 demands patterns are generated by random variation around the values of the k_0 demand matrix and assigned a decreasing probability

$P(k)$ based on their total absolute value difference from the k_0 demand scenario, as shown in Figure 6.3. Note that although 20 scenarios were used for this particular study, one can always increase or reduce the number for different levels of uncertainty characterization. However, in order to solve a large-scale stochastic formulation (e.g., scenarios are in the order of thousands), stochastic sampling or decomposition techniques, as discussed in Section 4.3, might be required to break the problem down into manageable blocks.

Table 6.3 summaries the characteristics of the demand scenarios. For research purposes, we generate these future scenarios in a systematic way. In practice, however, network planners might substitute the actual "what if" scenarios that they are most interested in or concerned about, as the suite of scenarios given to the model. Note that these can be the same set of detailed what-if scenarios the planners may already typically develop for separate study with conventional single-forecast design tools.

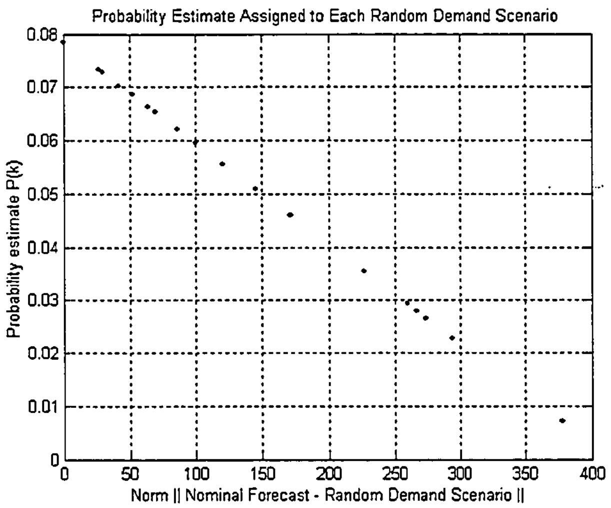


Figure 6.3. Probability assignment for input demand scenarios.

Table 6.3. Characteristics of the input demand scenarios.

Demand Scenario, k	Total Demand Volume, $\sum_{r\in D} d_k^r$	Relative Demand Volume, $\sum_{r\in D} d_k^r / \sum_{r\in D} d_0^r$	Pattern Forecast Accuracy, PFA	Assigned Probability, P_k
0 (nominal)	547	1.00	1.00	0.079
1	146	0.27	0.85	0.063
2	299	0.55	0.85	0.066
3	457	0.84	0.85	0.072
4	597	1.09	0.87	0.072
5	744	1.36	0.85	0.071
6	955	1.75	0.91	0.067
7	911	1.67	0.81	0.065
8	1107	2.02	0.86	0.061
9	1287	2.35	0.85	0.057
10	1358	2.48	0.81	0.051
11	1546	2.83	0.85	0.045
12	1571	2.87	0.85	0.043
13	1874	3.43	0.86	0.042
14	1878	3.43	0.84	0.039
15	2187	4.00	0.88	0.030
16	2088	3.82	0.83	0.029
17	2217	4.05	0.86	0.021
18	2367	4.33	0.86	0.019
19	2718	4.97	0.86	0.0072
Min	146	0.27	0.81	0.0072
Mean	1343	2.53	0.85	0.05
Max	2718	4.97	0.91	0.079

6.6.4 Eligible Routes for the Design Formulations

The final experimental aspect is the generation of eligible route sets (i.e., Q^r and P_i). While we can enumerate all distinct routes to form our eligible route sets, short-distance routes are often preferred to meet physical specifications such as optical signal path quality and restoration speed [Gro04]. The study by Herzberg and Bye [HeB94] shows that screening out the unnecessarily long routes helps speed up the computation process without losing true optimality. Hence for the following experiment; 5 shortest working routes (by distance) for each OD pair and 10 shortest restoration routes (also by distance) for each span are selected as the eligible route sets. These result in a complete set of 275 eligible working routes and 260 eligible restoration routes.

6.7 Results and Discussion

The two formulations were implemented in AMPL [FGK93] and solved with CPLEX 9.0 MIP Solver [ILO04] on a four-processor Ultrasparc at 450 MHz and 4 GB of RAM running Sun Solaris 8 OS. For the TP-SR formulation, all designs were obtained to a MIPGAP of 1% (guaranteed to be within 1% of the optimum) and within twenty minutes of run time. For the TP-MSR designs, run times were considerably longer due to the additional dimension of modularity *M*. The MIPGAP was therefore relaxed to 10%.

6.7.1 General Observations of Two-Part Capacity Planning Strategy

Table 6.4 shows the results of the TP-SR (non-modular) formulation and compares them to the conventional span-restorable design with four different recourse costs.

Table 6.4. Comparisons between conventional and two-part designs (cost in thousands).

Design	Conv.	TP-SR	Conv.	TP-SR	Conv.	TP-SR	Conv.	TP-SR
Recourse Cost Factor, α	1		2		3		5	
Initial Cost	532	533	532	942	532	1,308	532	1,527
Expected Future Cost	557	557	1,115	620	1,672	488	2,787	503
Total	1,089	1,090	1,647	1,562	2,204	1,796	3,319	2,030
Difference	0.09%		5.16%		18.51%		38.84%	

In the "conventional" approach, we consider a minimum-cost span-restorable mesh design based solely on the nominal forecast. The cost of this conventional design refers to the "initial cost." The initial cost for TP-SR designs is the cost to build the first part, which might include certain initially built-in added capacities to hedge against possible future costs of recourse. The "expected future cost" for both cases refers to the probability-weighted cost of adding needed capacity to adjusting the initial design to cope with future requirements, i.e.,

$$\sum_{j\in S}\sum_{k=U} P(k)\cdot R_j\cdot(y_{j,k}+z_{j,k}).$$

At low recourse cost (i.e., when $R_j = C_j$), the advantage of the two-part design is insignificant because it costs the same in the future to take recourse as it does to build it in now. However, as the recourse cost increases, the long-term benefit of building a more "future proof" network initially, and paying less in the future for recourse, becomes obvious. At a recourse cost factor of 3, the two-part design has an expected whole life cost that is approximately 19% lower than the strategy of building a currently optimal network to an assumed known forecast, and augmenting it as needed in the future. The cost benefit of the two-part design increases as the recourse cost assumption increases.

In Figure 6.4, we compare the two-part designs to conventional designs that attempt to have some future-proofing by considering demand matrices other than the nominal forecast. The "expected forecast" is the probability-weighted demand pattern calculated based on the 20 scenarios and the "maximum forecast" is where each OD pair takes the maximum demand of all the scenarios. We see that TP-SR approach always outperforms these pre-tuning forecast attempts with the conventional model. At low recourse cost, the "maximum forecast" design tends to over-build the capacity initially and fails to exploit the advantage of building in future. The "expected forecast" design also suffers from paying an expensive penalty in the high recourse region.

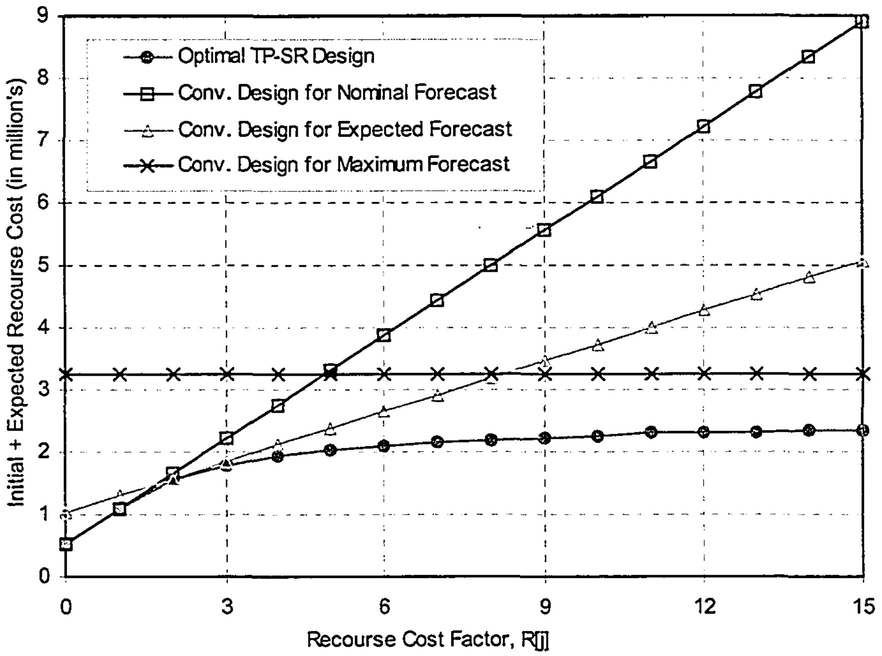


Figure 6.4. Cost-benefit of the (non-modular) future-aware designs over various conventional designs.

Thus, with no consideration of the recourse in advance and unconsciously making an investment plan targeted on single demand forecast, any single-period, snap-shot approaches can easily lead to a capacity plan that will suffer from either severe capacity surplus or deficiency.

While it is important to portray in general how the recourse factor affects the overall long-term cost, it is also meaningful to show the tradeoff between the long-term cost and the initial design cost under various recourse assumptions, as illustrated in Figure 6.5. As should be expected, at a recourse factor of one (or less), the optimal initial design is simply the one designed for the k_0 nominal forecast alone with conventional methods. If we increase the cost of the initial designs (i.e., the successive points to the right), we will end up over-building the capacity unnecessarily. This makes sense because under the low recourse assumption, we are encouraged to build only what is needed now, and wait for the future as there is so little penalty to add more later. As the recourse cost factor increases, however, we can optimize the present investment and come up with a capacity configuration that has the least expected repair cost to cope with future scenarios. For the highest recourse, the top curve indicates that the optimal initial design cost to invest is about $2.66 million, where the expected future cost is zero. In fact, this corresponds to an initial design that completely satisfies all of the possible scenarios without any future additions (i.e., the "fat solution" mentioned in Section 6.5.1).

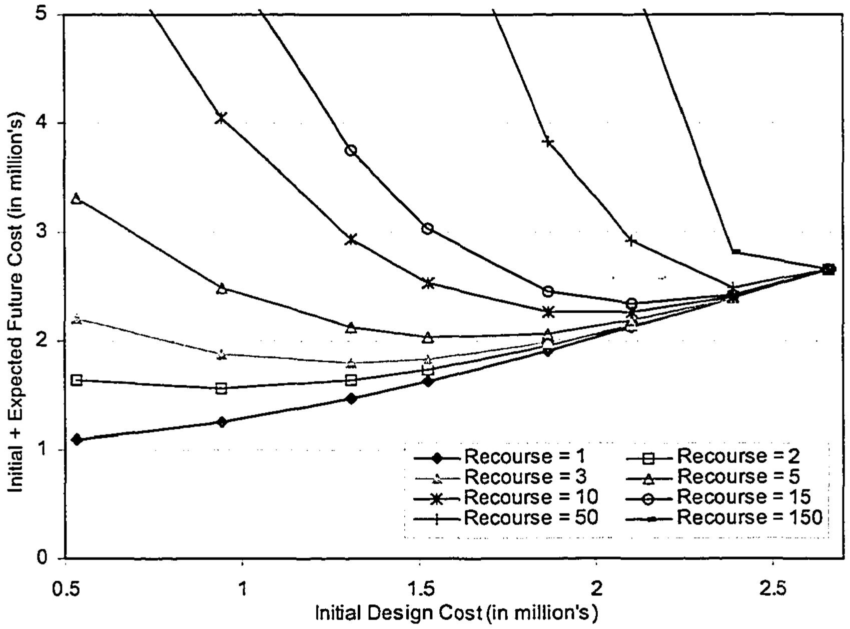


Figure 6.5. Total versus initial cost of future-aware designs at varying recourse cost factors.

6.7.2 *Effects of Modularity and Economy-of-Scale: Results with TP-MSR*

For tests with the modular capacity design formulation, we used the input demand sets described in Section 6.6. Four module sizes, namely Size-3, Size-12, Size-48 and Size-192, as well as three economies-of-scale (2x2x, 3x2x, 4x2x) were assumed. The associated module costs are listed in Table 6.1. Although in practical systems the absolute capacity values may differ from those used here, the total range of capacities represented and the number of such module types are quite characteristic of actual SONET OC-n line systems that are commercially available.

Tables 6.5 to 6.7 compare the results of the conventional (Conv.) and TP-MSR designs under different economy of scale assumptions. Similar to the finding of the non-modular designs in the previous section, the two-part modular model shows significantly lower expected total life cost than traditional designs. For a recourse cost factor less than one (i.e., $\alpha \leq 1$), the optimal designs are equivalent to the conventional designs that are strictly built for the nominal forecast. In cases where $\alpha = 3$ or $\alpha = 10$, the TP-MSR designs result in a total expected cost reduction of ~22% and ~63% (on average) compared to the conventional designs that are faced with the same range of possible futures. In particular, under the *3x2x* model, the cost reductions are the greatest.

Table 6.5. Comparisons between conventional and TP-MSR designs under the *2x2x* model.

Design	Conv.	TP-MSR	Conv.	TP-MSR
Recourse Cost Factor	3		10	
Initial Cost	5,407	13,740	5,407	18,953
Expected Future Cost	17,299	4,746	57,290	4,556
Total	22,706	18,486	62,697	23,509
Difference	18.59%		62.50%	

Table 6.6. Comparisons between conventional and TP-MSR designs under the *3x2x* model.

Design	Conv.	TP-MSR	Conv.	TP-MSR
Recourse Cost Factor	3		10	
Initial Cost	2,397	3,850	2,397	5,183
Expected Future Cost	4,640	1,008	15,526	603
Total	7,037	4,858	17,923	5,786
Difference	30.96%		67.72%	

Table 6.7. Comparisons between conventional and TP-MSR designs under the *4x2x* model.

Design	Conv.	TP-MSR	Conv.	TP-MSR
Recourse Cost Factor	3		10	
Initial Cost	1,559	2,459	1,559	3,068
Expected Future Cost	2,023	493	6,741	294
Total	3,582	2,952	8,300	3,362
Difference	17.59%		59.49%	

Figure 6.6 identifies a complete set of optimal designs. Each graph in the matrix corresponds to the optimal initial design for a unique recourse and economy-of-scale combination. The optimal designs are arranged with increasing recourse cost factors for each column and classified by different economies of scale in each row. As we move from the left to right column, we see that higher recourse costs generally encourage the building more expensive initial designs to reduce the expected penalty in the future. Moving from the top to the bottom

row, we also see how economy of scale generally favors the installation of large-size capacity modules. In the case where $\alpha \leq 1$, the largest size modules change from 179 *size-3* modules, to 23 *size-48* modules, to 12 *size-192* modules. For higher recourse cost factors, the benefit of deploying large size systems is even more obvious (i.e., the optimal size jumps from *size-3* to *size-192*). Probably the most interesting scenarios are where we have the strongest economy-of-scale and high recourse cost assumption, i.e., the 4x2x(α=3) and 4x2x(α=10) designs. In such cases, the optimal initial designs consist of only the largest modules.

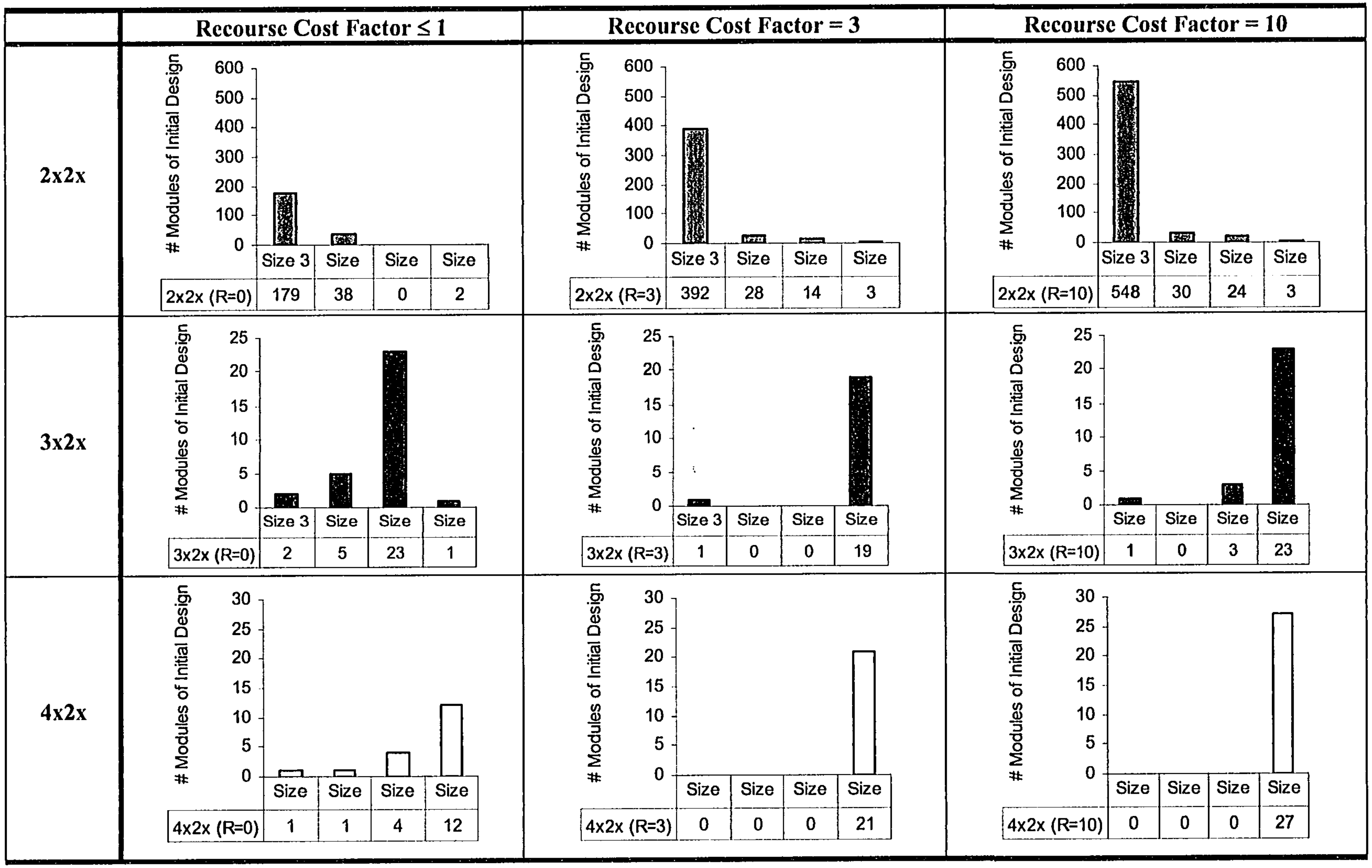


Figure 6.6. Initial design costs under various influences of recourse costs and economy-of-scales.

6.8 Summary

We have developed two integer program formulations for the design of non-modular and modular span-restorable networks under demand uncertainty. Stochastic programming was used as the mathematical framework to model the formulations that minimize the initial cost of network building and the expected value of future recourse actions to augment the design to serve the possible "what if" demand scenarios. One significant finding is that if the cost of building future capacity is greater than that of building it now, the notion of taking the recourse cost and future demand scenarios into a two-part capacity design becomes vital, since such a design could lead to huge long-term cost savings, compared to traditional designs that consider only a single nominal forecast. For the non-modular design under a recourse cost factor of 3, an approximately 19% cost reduction is observed. Under the 3x2x economy of scale model, the two-part modular design leads to 31% saving of the expected total life cost. Another interesting observation is that when modularity and moderate economy of scale (i.e., 3x2x or stronger) are considered, the most future-proof designs tend to deploy large modular systems rather than many small-size modules.

While we work with span-restorable networks in this study and assume a uniform recourse cost factor for all spans, the two-part formulations can be adapted to other survivable network architectures. The recourse costs can be set specifically for each span to reflect the cost of lighting up a new fiber system or leasing capacity from a third-party carrier or any other practical realities. In this regard, we will illustrate how the two-part framework can be adapted to a *p*-cycles design scheme in Appendix C, with the corresponding AMPL implementation shown in Appendix B.5.

Finally, some issues about recourse costs warrant specific comments. One general view of the future is that "capacity is always on an ever-decreasing cost curve" – so would recourse cost factors always be less than one? If this were so, then the optimum strategy is always just to build the minimum that is needed right now, and add anything else that is needed in the future. Given the economic hardships the telecom industry has recently endured, a common attitude is, perhaps understandably, to minimize costs now regardless of the future consequences.

However, from an operational and network management perspective, channels cannot be easily and cost-effectively added one at a time just when needed, and at ever decreasing cost [AiP94][ILO01][Mis04]. Clearly where actual installation of cables is involved, there is a very high recourse cost associated with pulling in more cable, or even digging up streets a second time. Similarly, simply adding a single wavelength system often involves labor-intensive costs, including performance monitoring, security management, configuration and provisioning costs.

Incremental growth in physical equipment to support new capacities can unconsciously exceed the space, power requirements, the maximum port counts on a cross-connect, etc. from an initially inadequate first installation and trigger a large recourse cost ($R >> 1$) to upgrade the infrastructure in the future. Thus, when assigning recourse costs in this type of foreword-looking model, it is important for planners to take all factors into account, and to consider all possible physical and "hidden" operational costs required by the existing capacity plan to adapt the future network. Even if the transmission equipment itself was given away by vendors, there are always significant operational and business costs associated with having to take corrective actions. This planning model provides a tool for planners to find the right balance between putting off some eventualities into the future, while building to accommodate others right now.

7 Max-Profit Demand Loading Strategy for Multi-QoP Survivable Mesh Networks

7.1 Introduction

This chapter presents a type of capacity management strategy, called Demand Loading, and summarizes the recent work published in [LeG05b]. Of particular interest, we develop an integer program model for the optimal selection of the demand pairs to maximize the net economic return, given a set of per-channel costs and a set available demands requiring different protection service classes and corresponding prices.

A unique concept that emerges in this chapter is that of "preferred demands," which the model identifies and that can guide service planning and market development efforts towards the most profitable mix of routes and demand types with which to load the network. Unlike the capacity planning problems, in which the objectives are to find the least-cost, minimum-capacity designs to strictly serve a given set of demands, the demand loading problem assumes that the capacity design is a given and operators have the freedom to choose which demands should be served. The issues of demand loading are therefore more tactical and operational in the sense that the objective is to determine how to utilize as-built capacity assets to maximize the profits from the available set of potential demands.

This work also attempts to gain insights into strategies for pricing and offering wavelength services of various protection classes over a span-restorable (SR) transport network. With span restoration, the cost of providing survivability to different paths is less than the cost of duplication, yet can be specific to each node-pair involved. This makes the problem of demand loading for maximum profit complex.

The service-oriented aspects of multiple survivability or quality of protection (multi-QoP) classes are also captured in the formulation. Specifically, three types of services – protected, unprotected and preemptible – are considered. Based on three case studies, we show the applicability of the formulation and illustrate how business planners and marketing groups can use it to compare strategies under different revenue, cost and demand assumptions. A novel strategy identified with the model is to combine preemptible service offerings with protected services. The overall profit can be the same as with protected-only services at higher prices.

7.2 Research Questions

Because the research questions are closely associated with the classification of multi-QoP service classes, we will explain the concept for the first time in this thesis and state the specific research questions at the end of this section.

7.2.1 Concept of Multi-QoP Service Classes in Wavelength Services

Wavelength-based transport services are expected to offer cost-effective and flexible solutions for many high-bandwidth applications, including storage area networking, data mirroring and grid computing applications. Unlike the era of dark fiber or private networks, it is believed by different equipment vendors and service providers that a dynamic wavelength-services solution should allow customers to lease only the bandwidth they need in a much shorter contract period (in the order of months) and effect an easier service provisioning process [NdF00][Gau03][Hun03][FPS03][MaG03].

As for the current wavelength service profile, many service providers have only offered a single type of wavelength service, the unprotected services, or at most two service types including protected services. Because new applications do not normally require the same level of survivability (for example, unprotected services may be sufficient for most Internet traffic, but fully protected services are usually needed for critical traffic connecting storage area networks), the ability to offer multiple protection classes can be one of the differentiators among competing network carriers.

In this study, our goal is to consider wavelength services with multiple quality-of-protection (multi-QoP) classes [GeS01][GrC02] – namely, protected, unprotected and preemptible classes – and to design a profit-maximizing service provisioning model. By protected services, we refer to a set of working paths or capacities that must be restored against any single span failure. In other words, these are the "guaranteed" failure-proof services. In a lower priority, unprotected services are working channels that do not receive any restoration efforts nor are they subject to any preemption for other failures. At the lowest priority class, preemptible services use spare capacity for transporting low-priority traffic, and if any of the protected services fails, it will seize the spare capacity to satisfy its own restoration requirement.

From the research literature, the optimal capacity design model of span restorable (SR) mesh network to support multi-QoP services was first proposed by Grover and Clouqueur in [GrC02]. The authors proposed an integer linear program formulation and showed that there was a surprisingly high potential to support preemptible services over the conventional spare capacity of a mesh network. In other words, that work revealed the possibility of designing a SR network where no truly idle protection capacity was needed. By properly routing both protected and preemptible services, the protection needs of the former class could be entirely met by preemption of the latter. Such findings motivate us in this study to go beyond the capacity planning problem of finding minimum-cost design to serve all multi-QoP demands, to a maximum net-revenue standpoint, or an optimal "preferred demand" standpoint.

In the "demand loading" context, we assume circumstances where operators have their already-deployed existing infrastructure with capacity limits. They have the latitude to choose which demands (node pairs and service types) they want to serve, from a set of potential demands. The cost to serve the selected demands should also be associated with the provisioning cost to serve each additional wavelength channel as well as the capacity resources required for service protection. Not all demands necessarily need to be served, nor in general would it be possible to serve all demands with the existing network from an operational perspective.

With the multi-QoP model formulated, we will try to address the following questions in three separate case studies in Section 7.5:

(1) How do the channel cost and the relative pricing of protected over unprotected service influence the preferred demand loading principles?

(2) How does the distance of multi-QoP services influence the preferred demand loading?

(3) What will be the potential benefit if network carriers introduce preemptible services into the service mix?

In the next section, we discuss the revenue and cost models used, followed by the demand loading formulation in Section 7.4. Section 7.5 provides the details of the experimental design, the three case studies, and discussion of the results. Concluding remarks are offered in Section 7.6.

7.3 Models for Cost and Revenue of Multi-QoP Services

An essential aspect to develop the demand loading formulation is the relative pricing policy (or the potential earnings) of the multi-QoP services. The difference in service distances obviously comes into play and influences what types of demands are best to try to serve. One model is a "flat rate" pricing scheme, where the revenues of all wavelength services are identical and insensitive to their distances. This is very much like the Internet today, but sheer distance independence may be hard to accept for some time when it comes to whole lightpath services (where the cost of the transponders used, and cost of optical amplifiers really do rise with distance [MaG03][CoW03][Dri04]). We will therefore consider a range of scenarios, where the cost and price of transport services depend in different ways on distance as well as service level.

Let us now define our meanings of cost and revenue used in the model. *C[j]* is the cost incurred when a new channel is provisioned on span *j*. The cost of an additional equipped channel on a span is the same regardless of what type of service path is used to support and also the same whether it is used as a working or protection ("spare") channel. For our experiments, we assume that the cost scales directly with the length of a span, which is a common convention used in many capacity design problems. The next definition is the revenue earned from an unprotected service path for demand pair r, R_o^r. This serves as a reference benchmark for the revenue models of the protected, R_+^r, and preemptible classes, R_-^r. The subscripts "+," "o" and "-" are used to indicate the protected, unprotected and preemptible classes, respectively. In addition, we define three different schemes for pricing each unprotected service on origin-destination (OD) pair r. This ranges from a distance-independent rate or "flat-rate" to "zone-rate" to a "linear" distance pricing assumption.

For flat-rate, the revenue for a given service class is identical for all demand pairs and is independent of the service distance. The zone-rate scheme assumes pricing of the demand pairs are classified in zones. For example, node pair r with service distance D that is within X_1 miles in radius, i.e., $D < X_1$ might cost P_1, and if $X_1 < D < X_2$, it might cost P_2 ($> P_1$), and so on. For linear-rate, the revenue of the demand pair is simply linearly proportional to its service distance. Note that these are models of revenue, not cost.

Next we must parameterize the ratio between protected and unprotected service revenues. We do this with $\alpha = R_+^r / R_o^r$. As a reasonable assumption, α is considered to be between 1 and 2 (see the pricing report by Drilling [Dri04]). In general, this could also be demand-pair specific. Finally, we define a parameter to use in varying the assumed price of preemptible to that of the unprotected services. This we denote by $\beta = R_-^r / R_o^r$. Obviously, the range for β is from near 0 to a maximum of one. Varying β lets us explore how the discount for preemptible service affects global net revenue. The effect is not obvious in advance because a lower β also lowers revenue for that service class. Under some circumstances, however, varying β in a multi-QoP mesh restorable network might suggest some alternative pricing strategies to reduce the revenue of protected services while keeping the total profit unchanged. Thus, α and β are the relative ratios that have direct impact on the optimal demand loading solutions. Note that these ratios apply per unit distance and per unit time to the respective service types they describe.

7.4 Economically Optimum Demand Loading Formulation

We now present the integer program for loading multi-QoP wavelength services onto span-restorable network with maximized revenue over costs. As a starting point, we adopt the constraint sets of the multi-QoP span-restorable capacity design model from [GrC02]. We then adapt and extend the model to have fixed span capacity limits, and set the objective to maximize the profit from the demand pairs and services chosen, but not to minimize cost to serve all demands. The max-profit multi-QoP demand loading (MP-QoP-DL) formulation is as follows:

Sets:

S	Set of all spans in the network, index i or j
C	Set of multi-QoP service classes, index c in the set {"+" for protected, "o" for unprotected, "- " for preemptible}
D_c	Set of all origin-destination (OD) pairs, index r
Q^r	Set of pre-determined eligible working routes for OD pair r, index q
P_i	Set of pre-determined eligible restoration routes available upon the failure of span i, index p

Parameters:

R_c^r	Revenue from serving OD pair r of service class c
C_j	Cost of provisioning or using a channel on span j
d_c^r	Number of lightpath requests of class c that may be served between OD pair r
T_j	The maximum number of channels that can be provisioned on span j
$\zeta_j^{r,q}$	Equal to one if the q^{th} eligible route for demands between node pair r uses span j, zero otherwise
$\delta_{i,j}^p$	Equal to one if the p^{th} eligible route for span i uses span j, zero otherwise

Variables:

x_c^r	Number of the lightpaths of class c that will be served on OD pair r. They are the "preferred demands."
w_j^c	Number of working channels used on span j for routing the demands of class c
s_j	Number of spare channels used on span j either for protecting the working links, or for routing the preemptible demands
$g_c^{r,q}$	Number of working paths assigned on the q^{th} working route to serve OD pair r of class c
f_i^p	Number of restoration paths assigned on the p^{th} eligible route for failure of span i

MP-QoP-DL:

$$\text{Maximize } \left\{ \left[\sum_{r \in D_- \cap D_o \cap D_-} \sum_{c \in \{+,o,-\}} (R_c^r \cdot x_c^r) \right] - \left[\sum_{j \in S} C_j \cdot (w_j^+ + w_j^o + w_j^- + s_j) \right] \right\} \tag{7.4.1}$$

Subject to:

$$\sum_{q \in Q^r} g_c^{r,q} = x_c^r \qquad \forall r \in D_c; \forall c \in \{+,o,-\} \tag{7.4.2}$$

$$\sum_{r \in D_c} \sum_{q \in Q^r} \zeta_j^{r,q} \cdot g_c^{r,q} = w_j^c \qquad \forall j \in S; \forall c \in \{+,o,-\} \tag{7.4.3}$$

$$\sum_{p \in P_i} f_i^p = w_i^+ \qquad \forall i \in S \tag{7.4.4}$$

$$w_j^- + s_j \geq \sum_{p \in P_i} \delta_{i,j}^p \cdot f_i^p \qquad \forall i \in S; \forall j \in S \mid i \neq j \tag{7.4.5}$$

$$\sum_{c \in \{+,o,-\}} w_j^c + s_j \leq T_j \qquad \forall j \in S \tag{7.4.6}$$

$$x_c^r \leq d_c^r \qquad \forall r \in D_c; \forall c \in \{+,o,-\} \tag{7.4.7}$$

The objective is to determine a subset, x_c^r, of the pool of available demands, d_c^r, to be served on each node pair in each service class such that the total revenue less total cost of the working and spare channels used (i.e., profit) is maximized. It is useful to identify the x_c^r solutions to this problem as "preferred demands" because if the latitude exists, then these are the subset of node pairs and service types that the network operator would most like to be asked to select and serve. It is important to realize that when there is a cost directly associated with equipping each incremental channel to provision a new path, the preferred demands are not always the full set of demands, nor a subset of demands that simply maximizes capacity utilization.

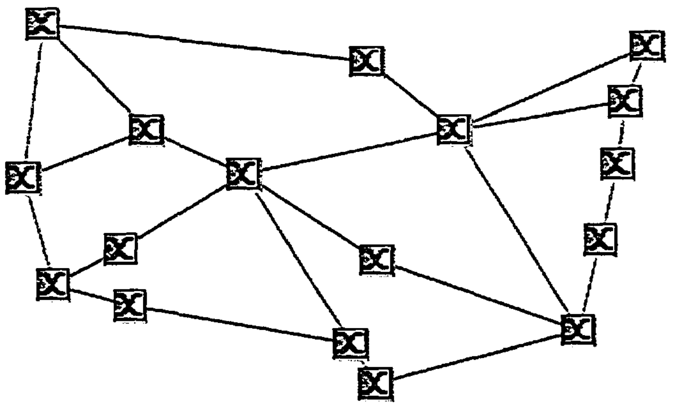
Constraints (7.4.2) and (7.4.3) determine the required working channels w_j^c on each span to simultaneously support the protected, unprotected and preemptible demands. The variables $g_c^{r,q}$ are the demand flows for each OD pair r and are specific to each service class. Constraints (7.4.4) and (7.4.5) correspond to the network survivability constraints based on span restoration. Constraint (7.4.4) ensures that the restoration flows f_i^p are always assigned for the recovery of the protected class traffic. With a preemptible service class considered, constraint (7.4.5) determines the spare capacities, s_j, required to support the restoration flows, recognizing that channels carrying the preemptible traffic on the same span are equivalent to spare channels from the standpoint of restoring protected class services. Constraint (7.4.6) ensures that the total working and spare capacities on each span in this loading plan are less than the available capacity T_j.

Finally, constraint (7.4.7) prevents the solution from routing more the maximum available demands in each class on each node pair, d_c^r.

Within this model, changing the values of the revenue for each demand, R_c^r, and the span-specific provisioning costs, C_j, can represent different sub-problems. If we do not consider the channel provisioning cost, i.e., setting C_j to zero, we then have a *max-revenue* demand loading model, where every demand served contributes to an increase in the revenue. This might apply to an existing network where it is already fully provisioned in terms of all its channels being turned up on each span as a result of some prior sunk cost investment. At the other extreme, if we set C_j high enough (relative to R_c^r), then at some point, the optimal loading solution would prefer not to serve any demand in order to keep the overall profit at zero rather than permit a negative value. The use of this model in the middle regimes can give insights about minimum revenue requirements to be profitable and identify the "preferred demands," x_c^r, as defined above. Various R_c^r assumptions may also greatly affect the preferred choice of potential demands to serve in the network. This can be used under any demand scenarios, any cost and any revenue assumptions to identify preferred (i.e., most theoretically profitable) city pairs to serve, and with what types of services. It is this intrinsic relationship of R_c^r and C_j to which we mainly address our experiments.

7.5 Experimental Design and Case Studies

The above formulation gives us a research framework within which C_j, R_c^r, T_j, α and β can be varied to represent a large number of possible or "what if" scenarios. In the following, we present three case studies that illustrate the potential use of the formulation and address the specific questions we stated at the end of Section 7.2. Now let us introduce the test topology and initial capacity assumption used for all case studies.



Span Lengths (in arbitrary units), L_j	
Mean	119.26
Total	2862.27
Std. Dev.	55.20
Min	37.22
Max	259.74

Service Distances, W^r	
Max	583.19
Min	37.22
Mean	304.64

Figure 7.1. 17-node, 24-span test network reported with span and service distances.

Figure 7.1 is a sample network topology used in this study. It is identical to the one we used in Section 5.6 (i.e., the *US* network [RBS01]). This network has 17 nodes and 24 spans with an average nodal degree of 2.82. Lengths of span j, L_j, and service distances (based on shortest path route), W^r, between node pairs r are also summarized on the side tables. Both L_j and W^r are used for the numerical values to determine the provisioning costs, C_j, and the revenues, R_c^r.

In all experiments, we have assumed that there may be a maximum of 128 lightwave channels on each span (i.e., $T_j = 128$ for all span j), and that each node is capable of performing wavelength conversion or has enough wavelength converters to make wavelength blocking negligible. We note that in real world situations, the maximum number of channels per span should be unevenly distributed over the network. In that case, one would simply adjust the parameters T_j to any capacity configuration. The fixed maximum plus the per-channel provisioning cost represents the situation in an optical network where the investment for basic DWDM infrastructure has been made (i.e., fiber, WDM mux, demux, optical amplifiers, common equipment) but each channel added has a direct provisioning cost (i.e., administrative charge per order, distance-related charge for channel transmitters, receivers, etc.).

In terms of the eligible route sets required by the demand loading formulation (i.e., the sets Q^r and P_i), we select the set of ten shortest working routes (by distance) as the eligible working route choices for each OD pair, and then set of ten successively-shortest distinct routes (also by distance) as the eligible route set for restoration flow assignment (of protected services only). This results in all problems having a total of 1360 eligible working routes and 240 eligible restoration routes. The formulation was implemented in AMPL and solved with CPLEX 9.0 MIP Solver on a four-processor Ultrasparc at 450 MHz and 4 GB of RAM running Sun Solaris 8 OS. All solutions were obtained to a MIPGAP of 0.01% (guaranteed to be within 0.01% of the optimum) and within three minutes of run time. We now present the individual case studies.

7.5.1 Case Study 1: Effect of Channel Cost and Protected-to-Unprotected Revenue Ratio on the Preferred Demands

In this study, only protected and unprotected classes are considered. Our interest is to see how the preferred demands change with the relative cost of a channel. This is studied by providing a set of five demands of each service class on each node pair. An unprotected service is by definition assumed to provide revenue of $R_o^r = W^r$, which corresponds to the service distance, while a protected service earns twice this ($\alpha = 2$) on any node pair. We then vary the cost of a channel on each span to correspond to the different cost factors $C \in \{0, 0.2, 0.4, ..., 1.4\}$ multiplying the length of the span, i.e., $C_j = C * L_j$. This creates conditions varying from one extreme (at $C = 0$) where the max-revenue solution tends to fill with mostly protected demands and unprotected ones to the other extreme (at $C = 1.4$) where the network reaches a "cutoff" state that serves no demands, as shown in Figure 7.2.

We repeated the experiment with $\alpha = 1.6$ and $\alpha = 1.2$, and plotted the results in Figures 7.3 and 7.4. Similar results with three distinct loading regions were observed. Using Figure 7.3 as an example, when the provisioning cost is within the first regions (where $C = 0.2$ to 0.8), optimal demands of both types are served. As the cost keeps increasing and reaches the second region ($C = 1$), only the service class with sustainable profit (in this case protected services that can share protection capacity well) could be served. At $C = 1.2$ and up, the max-profit solution reaches the cutoff region where none of the demands can generate enough revenue to cover the provisioning cost.

Comparing Figures 7.2, 7.3, and 7.4, we see how α generally influences the preferred demand types. For a high α value, protected services are preferred, while unprotected services are relatively more profitable at a low α. Finally we plotted the net utilization of total capacity against the provisioning cost at $\alpha = 1.2$ in Figure 7.5. As might be expected, profit corresponds to high network utilization only when channel cost is much lower than unit revenue. As channel cost becomes more significant, simply managing for high network utilization correlates less and less with the net economic return.

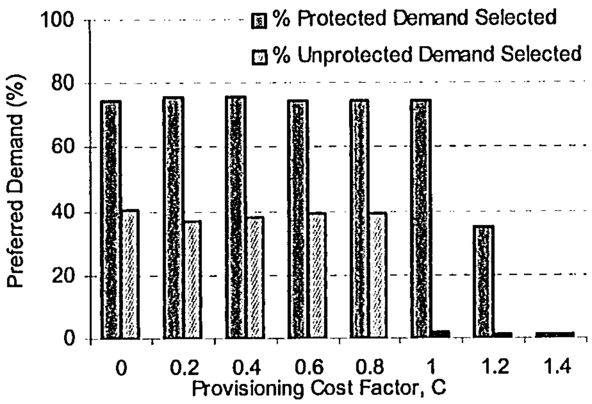


Figure 7.2. Percent of total available demand in each service class selected in the maximum-profit solution as a function of channel cost ($\alpha = 2$).

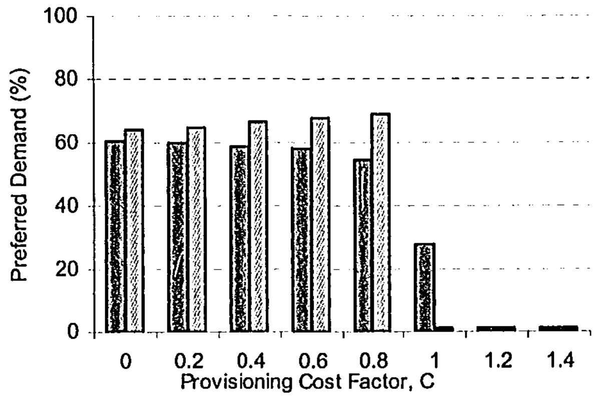


Figure 7.3. Percent of total available demand in each service class selected in the maximum-profit solution as a function of channel cost ($\alpha = 1.6$).

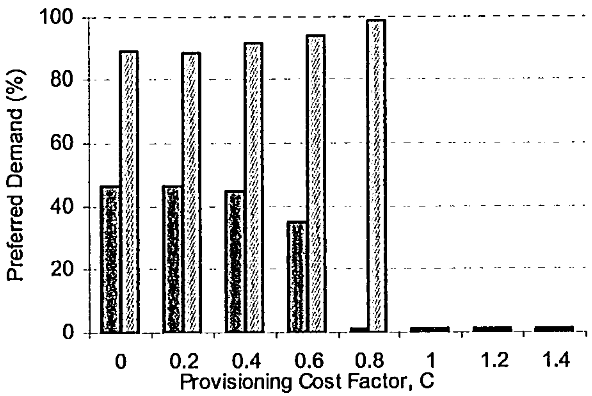


Figure 7.4. Percent of total available demand in each service class selected in the maximum-profit solution as a function of channel cost ($\alpha = 1.2$).

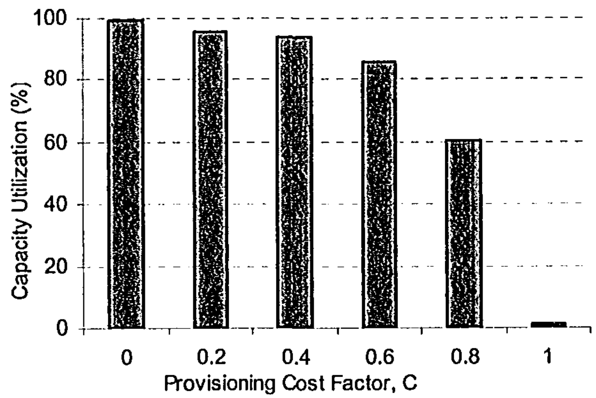


Figure 7.5. Total network capacity utilization by preferred demands as channel cost rises ($\alpha = 1.2$).

7.5.2 *Case Study 2: Effect of Distance-Sensitive Pricing on Preferred Demands*

This case study examines how different distance-related pricing strategies affect max-profit loading decisions. As in the previous section, only protected and unprotected services are considered. Again, five protected and five unprotected services are assumed to be possible for each node pair and the provisioning cost of each span is fixed at $C_j = C * L_j$, where C is set to 0.4. The changing parameter is the revenue of the unprotected services, R_o^r, and thus also of the protected service $R_+^r = \alpha * R_o^r$. α is set to 1.6 in these results. The particular selection of $C = 0.4$ and $\alpha = 1.6$ allows us a cross check against results we had in Figure 7.3. Three different pricing schemes, namely, Flat-rate, Zone-rate and Linear-rate, are modeled. The following table summarizes their parameters:

Table 7.1. Three distance-related revenue assumptions.

Distance Model	**Unprotected service revenue for node pair r, R_o^r**
Flat-rate	$R_o^r = \text{Avg}[W^r] = 304.64$
Zone-rate	$R_o^r\ \{r \mid W^r \le 100\} = 100$; $R_o^r\ \{r \mid 100 < W^r \le 250\} = 200$; $R_o^r\ \{r \mid 250 < W^r \le 500\} = 250$; $R_o^r\ \{r \mid W^r > 500\} = 280$
Linear-rate	$R_o^r = W^r$

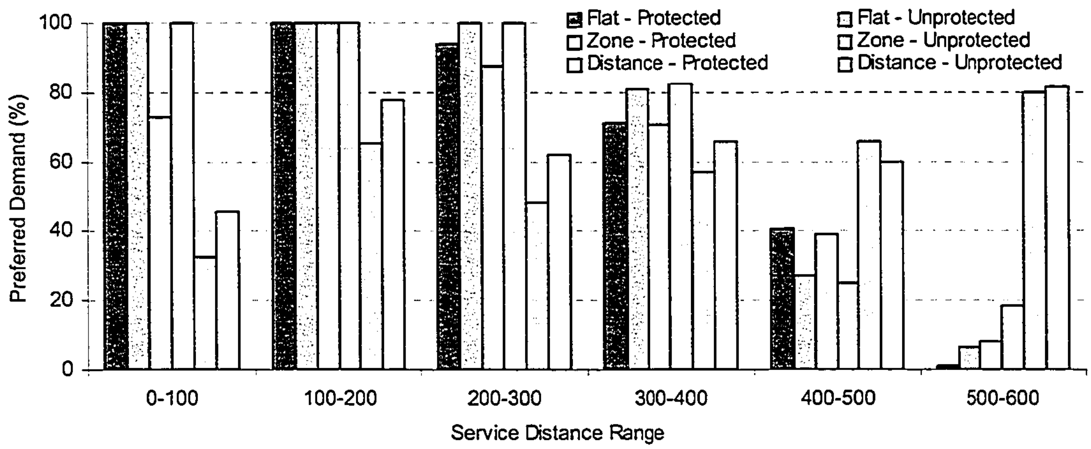


Figure 7.6. Locality of preferred demand pairs under different distance-related revenue assumptions.

Figure 7.6 shows the breakdown of the preferred demands into six different distance regions. To understand the plot, node-pairs are grouped by their distances. Each pair of bars shows the fraction of these demands that are selected for node pairs in that distance range, under the three different revenue-distance models. It is important to understand that the preferred demands in each distance range are being selected *in the presence of* the simultaneously offered demands on node pairs of all other distances as well. They are not evaluated in isolation.

The results show that under flat-rate pricing, almost all demands of both types are preferred on node pairs up to 300 in distance. Capacity is best used to serve those relatively local demands, and the drop-off in selected demands at longer distances thereafter is almost linear. At distance 500, almost no node-pairs are selected because for each one such long path, *several* shorter reach services could have been provided for. And under flat rate, since each service served provides the same revenue, the number of node pairs served is most important.

Zone-based pricing behaves similarly, except that at the shortest distance unprotected services are not fully served. This is probably due to the smallest zone earning of 100 for distances up to 100, not the mean value of distances as in the other zones, while costs are still distance-proportional. In other words, instead of using some capacity here as spare capacity on short protected services, it seems more profitable to use that capacity for working or protection purposes on service paths in the next-longer two zones. This may not be a general effect, but rather only an outcome of the specific zone-price model.

Finally, under distance-direct pricing we see the solver being extremely selective and careful across all distances. At no distance were *all* offered demands preferred. It makes sense that under linear-earnings with distance, the selected demands should be spread over all distances as they are because costs here are also linear; the two are basically in balance at all scales. This means that preferred demands of either type can be found across all distances. Also, at all scales, roughly the same numbers of protected and unprotected services were preferred, but in no case are all demands (or should be) picked up if one desires to achieve the maximum potential profit. Strategically, the solver is using capacity in very specific ways to make detailed choices of how to incur channel cost, how to make better trade-offs between spending directly for channels for an unprotected service here, and for protection capacity to enable two or three protected services there (which earn more), and so on.

7.5.3 Case Study 3: Effects of Preemptible Service on Preferred Demands

Most, if not all, service providers have included protected and unprotected services in their transport services portfolios. To our knowledge, preemptible services, which use protection (or spare) channels and get discarded when protected services fail, have not been as widely introduced in a mesh network context[25]. Here we analyze how preemptible services affect the preferred demands and impact on overall profit.

All three service classes are now considered and for the available demands, we have five possible demands to serve per node pair in each service class. The provisioning cost of each span, as before, is $C_j = 0.4 * L_j$ and the revenue of the unprotected services is $R_o^r = W^r$, which corresponds to the service distance. The parameters of relevance in this study are the revenue ratios, α and β, as defined in Section 7.3.

Figures 7.7 to 7.9 show the preferred demands in each class as we increase α from 1 to 2 for every β. From these figures, we see that increasing α (from left to right) generally increases our preference to serve more protected services and preemptible services, while it decreases the unprotected demands. Increasing β (from bottom to top), on the other hand, suggests that we should serve more protected *and* preemptible services, as a synergistic pair of services. What seems to be less intuitive is that when we increase β from 0, where we provide no incentive for serving any preemptible services, to β of 0.2, there is an abrupt change in preferred demands. We prefer to serve less protected services while significantly serving more preemptible services, as shown in Figure 7.9. An increase in α (for non-zero β) also has a positive influence on serving more preemptible services. The interpretation is that as α increases, the max-revenue solution tends to serve more protected services. That requires more spare capacity and hence gives room and incentive to provision additional preemptible services because this effectively provides the spare capacity for the former.

Figure 7.10 shows an interesting aspect of the optimal profit strategies for all α and β combinations. Assuming, from the curves a value for discussion of, say, total profit of a quarter million, we have *more than one* mix of preferred demands to meet this target. We can either set α to 1.9 (on β=0) or introduce preemptible services with $\beta \cong 0.4$ while reducing α to 1.7. In other words, even a pricing decrease in the protected services can achieve the same profit goal, if we introduce preemptible services and charge accordingly (in this example, 40% of the unprotected services.) This may seem surprising, but it is an effect explained by the fact that with $\alpha \sim 2$ and β=0, we earn more for protected services but we must bear the cost of conventional explicit spare channels for protection of those services. Evidently, we can earn just as much profit by lowering protection service costs and admitting an even more-discounted preemptible

[25] In ring-based network, this might correspond to using the "extra traffic" feature of ring ADMs. In a mesh network, the dual use of protection capacity in this way is much more flexible and general than with rings [Gro04].

service class, because the working channels of the latter effectively subsidize the provisioning of spare channels for the former.

7.6 Summary

We have provided a maximum-profit demand loading model for wavelength-service networks with multiple service classes and tested it in three case studies. This model can be used to gain insight and conduct research about many scenarios, strategies and "what ifs." From an operational standpoint, its value lies in helping to identify and understand the "preferred demands" for a given network, cost, and revenue situation. Although a network operator cannot literally pick and chose individual demands as precisely as the maximum-profit solution suggests, studies with the optimization model can reveal strategies and insights about what *would be* most profitable, and ultimately guide and influence sales and marketing development.

An interesting case illustrating one possible use of the model is that we revealed a strategy option in which a mix of preemptible and protected service offerings can earn as much profit as an offering of only protected services. If an operator's business environment provided few customers willing to pay nearly twice the price for protection, then it would be useful to know that a mixed environment of preemptible and protected services at lower prices can achieve the same profit target. Combined with operations support "back office" systems, the maximum-profit model given might even be adapted as an on-line, powerful inventory-based, decision-support tool for service providers to guide their business planning in today's competitive environment.

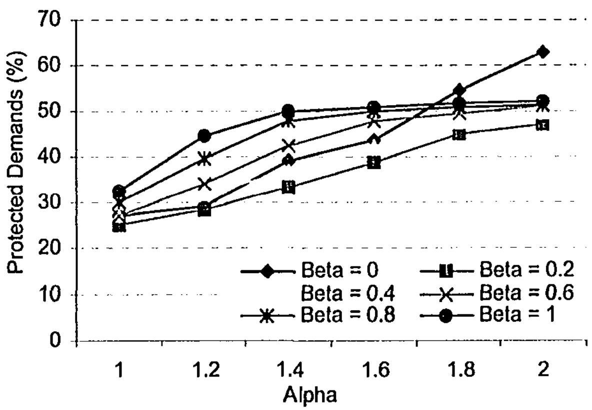


Figure 7.7. Percentage of *protected* demand served under various α and β values.

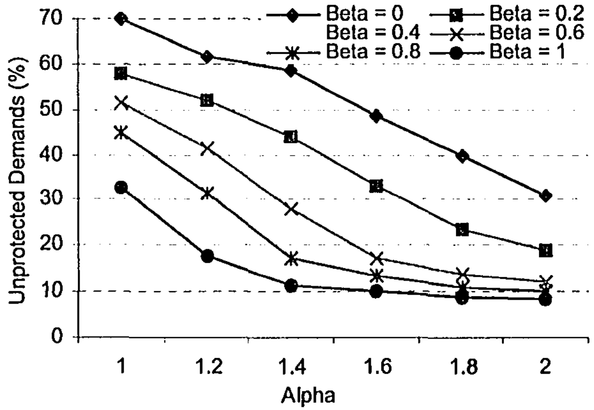


Figure 7.8. Percentage of *unprotected* demand served under various α and β values.

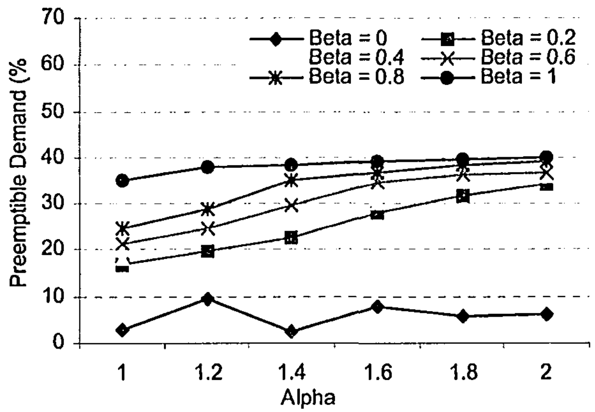


Figure 7.9. Percentage of *preemptible* demand served under various α and β values.

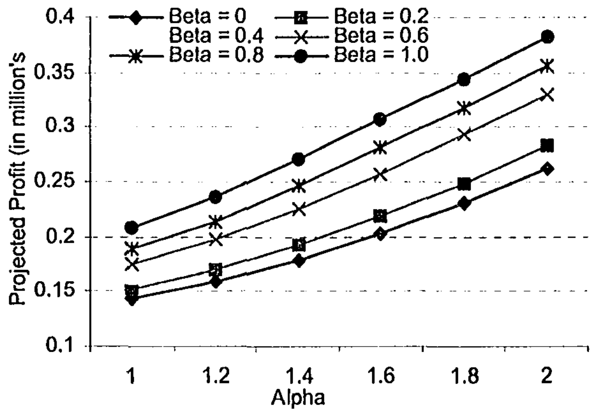


Figure 7.10. Potential profit generated by introducing preemptible services into service offerings.

8 Capacity Re-optimization of Mesh-based Survivable Networks

8.1 Introduction

This chapter summarizes the collaborative work conducted at Osaka University, by Dr. Arakawa and Professor Murata, investigating re-optimization issues for span restorable mesh networks. Through this summer program, co-sponsored by the Japan Society for the Promotion of Science (JSPS)[26] and NSERC, we developed several re-optimization strategies that enhance the network's potential to carry random incremental traffic. Full results were published in [LAM05].

Recall from Section 4.4.2 that we have classified some important related work on reconfiguration and re-optimization problems. The work in this chapter falls under the "type C" problem category, where a current topology is given. The objective is to determine a new logical topology or capacity configuration that has greater readiness to adapt to uncertain traffic. As we pointed out earlier, an important aspect of this work is to exploit the possible use of protection or spare capacity to achieve the objective of re-optimization. This option is particularly attractive to operators because idle protection capacity assignments in survivable networks can be re-arranged as often as one wishes, without affecting any working paths and always remaining in a restorable state.

We realize that one should never rearrange existing service paths (or active working channels) during the re-optimization process. However, for a complete analysis, we try not to make such an assumption and allow adjustment of in-service working paths. We do so because the synergetic rearrangement of both working and protection channels should theoretically give the best capacity performance. Indeed, some companies have reported that it is now possible to a "swap" or "bridge-and-roll" lightpath from one route to another (in a single administrative, single-vendor environment) without noticeable impact on services [Mer04]. Thus, we consider re-optimization *with* and *without* permission to re-arrange in-service paths. It depends on the technological choice for the network management system (NMS), or it is up to the operators who quantify the cost-benefit of each possible operational strategy.

Another unique aspect of this work is that, without the precise knowledge of the future traffic, we could not re-optimize *directly* for enhanced measures of routing efficiency or reduced

[26] Website URL is provided here for academic researchers who might be interested in this program, http://www.jsps.go.jp/english/e-summer/.

spare capacity in the existing networks. Rather, we try to maximize the traffic "carrying potential" of the network so that it *indirectly* enhances the blocking performance when faced with subsequent random-arrival growth in demand. We achieve this by the following model.

8.2 Models of Re-optimization on Span-Restorable Mesh Networks

We consider span restorable networks and develop four re-optimization strategies based on integer linear programs (ILP). Our goal here is to take any existing span restorable network configuration (comprised of demands already in service, spare channels pre-planned for span failure protection, and a remaining set of equipped but unused channels) and perform an offline re-optimization to create a configuration that not only serves and protects all existing demands, but also has a better readiness to serve continuing unpredictable growth in demands. There is one master ILP model within which the choice of objective function, and the latitude to re-arrange existing demands, provides the four different re-optimization strategies. We now provide a full explanation of the following formulation:

Sets:

S Set of all spans in the network, indexed by j or i

D Set of all origin-destination (OD) pairs in a demand matrix, index r

Q^r Set of pre-determined eligible working routes for OD pair r, index q

P_i Set of pre-determined eligible restoration routes available upon the failure of span i, index p

Parameters:

d^r *Existing* bi-directional demand on node pair r

T_j Total as-built capacity for span j from any given design

$\zeta_j^{r,q}$ Equal to one if the q^{th} eligible route for demands between node pair r uses span j, zero otherwise

$\delta_{i,j}^p$ Equal to one if the p^{th} eligible route for span i uses span j, and zero otherwise

a^r Optional parameter for setting priority among different OD pair r

Variables:

x^r *Projected* bi-directional demand that *could* be served on node pair r

λ Largest possible number of demand units that could be served uniformly on *all* OD pair

w_j Number of working capacity units on span j to support existing demand set

$\hat{w}_j$ Number of *idle* working capacity units allocated on span j to support future demands

s_j Number of spare capacity units on span j to support existing demand set

$\hat{s}_j$ Number of *idle* spare capacity units allocated on span j to support future demands

$g^{r,q}$ Working flow assigned on the q^{th} working route to serve OD pair r in existing demand set (Note: $g^{r,q}$ becomes an input parameter if rearranging working paths is not allowed)

$\hat{g}^{r,q}$ Potential working flow to be assigned on the q^{th} working route to serve OD pair r in future demand set

f_i^p Restoration flow assigned on the p^{th} restoration route upon the failure of span i in existing demand set

$\hat{f}_i^p$ Potential restoration flow to be assigned on the p^{th} restoration route upon the failure of span i in future demand set

Max-Fair: Maximize λ (8.2.1)

Max-Vol: Maximize $\sum_{r\in D} a^r \cdot x^r$ (8.2.2)

Subject to:

$$\sum_{q\in Q^r} g^{r,q} = d^r \qquad \forall r \in D \qquad (8.2.3)$$

$$\sum_{r\in D}\sum_{q\in Q^r} \zeta_j^{r,q} \cdot g^{r,q} = w_j \qquad \forall j \in S \qquad (8.2.4)$$

$$\sum_{p\in P_i} f_i^p = w_i \qquad \forall i \in S \qquad (8.2.5)$$

$$s_j \geq \sum_{p\in P_i} \delta_{i,j}^p \cdot f_i^p \qquad \forall (i,j) \in S^2; i \neq j \qquad (8.2.6)$$

$$\sum_{q\in Q^r} \hat{g}^{r,q} = x^r \qquad \forall r \in D \qquad (8.2.7)$$

$$\sum_{r\in D}\sum_{q\in Q^r} \zeta_j^{r,q} \cdot \hat{g}^{r,q} = \hat{w}_j \qquad \forall j \in S \qquad (8.2.8)$$

$$\sum_{p\in P_i} \hat{f}_i^p = \hat{w}_i \qquad \forall i \in S \qquad (8.2.9)$$

$$\hat{s}_j \geq \sum_{p \in P_i} \delta_{i,j}^p \cdot \hat{f}_i^p \qquad \forall (i,j) \in S^2; i \neq j \qquad (8.2.10)$$

$$s_j + w_j + \hat{s}_j + \hat{w}_j \leq T_j \qquad \forall j \in S \qquad (8.2.11)$$

$$x^r \geq \lambda \qquad \forall r \in D \qquad (8.2.12)$$

The ILP model basically has two sets of constraints. Constraints (8.2.3) to (8.2.6) represent demand routing and restoration flow assignment plans to serve existing demands, d^r, given the total capacity on each span is T_j. Constraints (8.2.7) to (8.2.10) characterize how a set of *projected* demand variables, x^r, would be routed and protected in capacity that is not used by d^r. It is important to emphasize that these "projected demands" are not *known* future demands. They only represent *potential* extra demands on each node pair that *could* be served. The hypothesis here is that maximizing these properties should lead to the enhanced readiness for future demand. We will test this hypothesis through simulations in the next section.

The two objective functions (8.2.1) and (8.2.2) express different goals for maximizing the future demand carrying potential. Coupled with constraint (8.2.12), *Max-Fair* re-optimizes existing routes and protection plans so that the largest minimum number λ of new paths is possible on all demand pairs. For example, the new network state with $\lambda = 3$ will support *at least* three new paths (with restorability guarantee) on every OD pair. In contrast, *Max-Vol* re-optimizes the capacity so that the bulk total volume of potential new paths is highest, with no consideration of fairness between node pairs. By using a weighting parameter a^r in *Max-Vol*, we have the option to assign preferences to the most desirable node pairs and for which to create growth readiness. Thus, if there exist some historical data on which preferable OD pairs are available, the explicit parameterization of a^r would be valuable. As in the case of *Max-Fair*, we can exclude less important node pairs from constraint (8.2.12) if desired.

Two sets of variables (with and without arrow tops) distinguish between the current and projected demands and their corresponding routing and protection plans. Constraint (8.2.3) indicates the demand flows $g^{r,q}$ onto working paths q to support the current demands d^r of each demand pair r. Constraint (8.2.4) generates the working capacity w_j required on each span to simultaneously serve the demand flows. (8.2.5) and (8.2.6) are standard survivability constraints for span restoration. (8.2.5) ensures that there are enough restoration flows f_i^p assigned to eligible restoration routes P_i when span i fails, and (8.2.6) generates the required spare capacities s_j to support each single span failure.

Readers might notice that the constraints (8.2.3) to (8.2.6) are identical to those in the basic spare capacity design of span restorable network. However, in the context of re-optimization without rearrangement of in-service paths, the working paths $g^{r,q}$ are no longer variables but become fixed input parameters to the problem. In this case, only the restoration paths f_i^p and assigned spare channel quantities s_j are the variables in (8.2.5) and (8.2.6), and are altered during the re-optimization process. Similarly, constraints (8.2.7) to (8.2.10) deal with the demand routing and protection flow assignments sub-problem for the projected demands. Note that the variables ultimately of interest to the solution are *not* potential demand variables themselves (i.e., x^r), but the changes made to current spare channel and working route assignments, s_j, and f_i^p, to maximize the potential if only protection rearrangement is considered, or variables w_j, s_j, $g^{r,q}$, f_i^p if complete rearrangement is allowed. The changes in these variables are the substance of the re-optimization actions to implement on the network itself.

8.3 Experimental Design and Results

The combination of the two objective functions (denoted by *Max-Fair* and *Max-Vol*) with and without in-service path rearrangement (denoted by *Complete* and *SpareOnly*) yields four distinct strategies. To evaluate the robustness of each one, we first create an initial capacity and routing configuration. The initial configuration (specifically d^r, $g^{r,q}$ and T_j) is then passed to the ILP for re-optimization. In the final phase, we generate random incremental requests to evaluate the blocking performance after the re-optimization changes are put into effect. By incremental traffic, we mean demands whose connection (e.g., lightpath) requests arrive sequentially from some random OD nodes. Once a lightpath is established for each connection, the lightpath remains in the network indefinitely. Readers therefore should not confuse these traffic characteristics with those of the "dynamic" traffic (as discussed in Section 3.2.1) in which lightpaths and capacity resources are released after some finite amount of time.

To create the initial capacity configuration, we use the basic minimum-cost span restorable mesh design with an average demand of 4.58 lightpaths per node-pair. The 11-node, 26-span COST239 network (also shown in Figure 6.2 and Appendix A.4) is used for the results. We assume that each span on the network has a total as-built capacity of 40 channels and each node is capable of performing wavelength conversion or has enough wavelength converters to make wavelength blocking negligible. The initial capacity scenario results in nearly half of the total capacity being used (493 out of 1040 channels in total). The distribution of used capacity over the total is illustrated in Figure 8.1. The capacity utilizations after the four re-optimization procedures are shown in Figures 8.2 to Figure 8.5.

Carefully analyzing these figures, we see how each re-optimization strategy re-distributes the existing capacities. For the *SpareOnly* strategies, the distributions of the capacity seem to retain the characteristics and shape of the initial design. In fact, simple correlation calculation measures in Table 8.1 reflect that the *Max-Vol-SpareOnly* design has the least amount of capacity re-distributions. In contrast, both *Max-Fair-Complete* and *Max-Vol-Complete* strategies tend to re-distribute the existing capacities in such a way that they are more evenly spread over all spans, even if more total capacity might be used.

Table 8.1. Capacity utilization before and after applying each re-optimization strategy.

Re-optimization Strategy	Not re-optimized	Max-Fair-SpareOnly	Max-Fair-Complete	Max-Vol-SpareOnly	Max-Vol-Complete
Total fixed capacity	1040	1040	1040	1040	1040
Total capacity used after re-optimization[27]	493	540	506	498	489
Total unused capacity after re-optimization	547	500	534	542	551
Correlation measure of re-optimized to initial design	1.00	0.82	0.70	0.94	0.72

Following each re-optimization method, we randomly generate 200 incremental requests as the future demands and load them onto each capacity configuration. The *Max-Vol* cases used d^r = 1 for all node pairs (assuming no prior knowledge of the OD pairs is given). In the incremental traffic model, each request is for one lightpath and its origin-destination nodes are selected uniformly. Such incrementally-arriving demands are then routed by solving an algorithm of finding the route that uses the minimum total (working plus spare) channels for protection.

[27] Readers might question why the *Max-Vol-Complete* strategy could produce a configuration that has less overall capacity than the initial design, supporting the identical set of demands. It is so because the initial design is optimized based on minimizing capacity "cost," whereas in our re-optimization formulations, the capacity cost is no longer a parameter to the model. In fact, we confirmed that the initial design still produces the lowest cost possible design of 242405 cost units, wheareas the *Max-Vol-Complete* design cost 264035 units, which is approximately 9% more costly than the initial design.

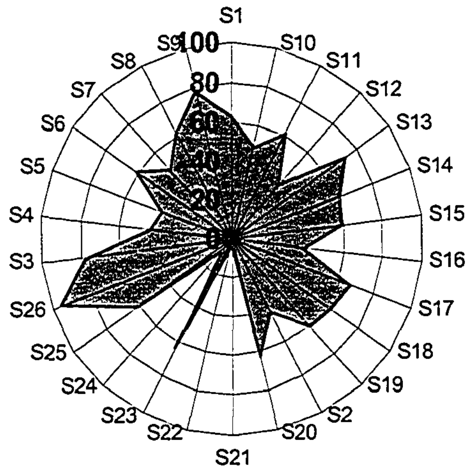


Figure 8.1. Used-to-total capacity ratio for each span, *S*, in the initial design (in percentages).

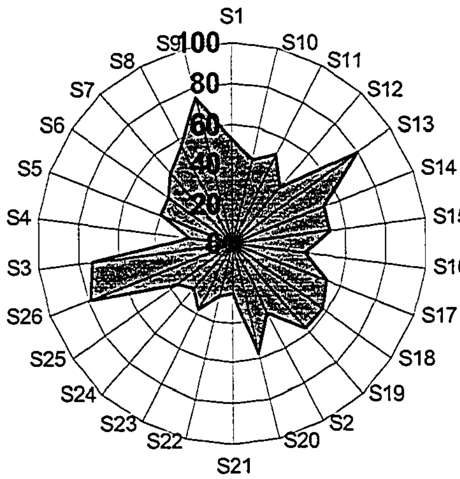


Figure 8.2. Used-to-total capacity ratio for *Max-Fair-SpareOnly* design.

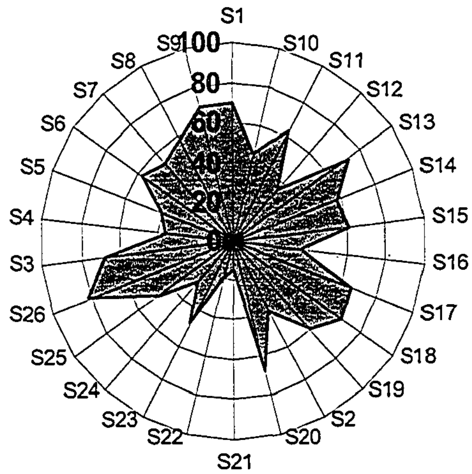


Figure 8.3. Used-to-total capacity ratio for *Max-Vol-SpareOnly* design.

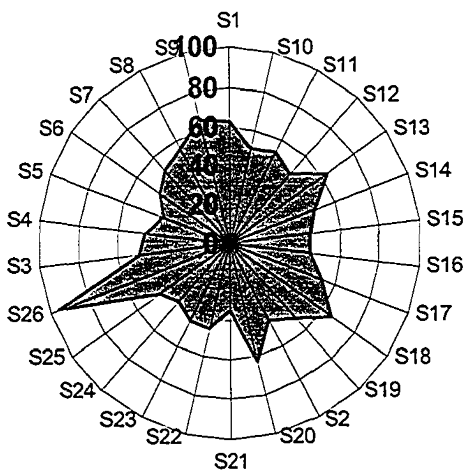


Figure 8.4. Used-to-total capacity ratio for *Max-Fair-Complete* design.

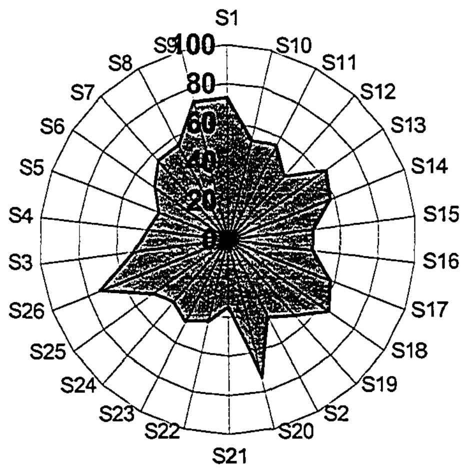


Figure 8.5. Used-to-total capacity ratio for *Max-Vol-Complete* design.

Five eligible working routes are considered and ten eligible restoration routes are represented for span restoration in solving each of the incremental minimum-cost protected routing problems. If the working path cannot be set up after trying all possible eligible routes, or if it cannot be fully protected within existing remaining capacity, then the request is blocked.

Table 8.2 shows the main results, in terms of the blocking performances of each strategy in coping with the continued random incremental growth, as well as the number of rearrangements that have taken place. Blocking is estimated from 12 trials of 200 demand-arrival experiments. It is interesting that even if *only* rearrangement of protection is allowed, the blocking improvements are still significant. Also, even though *Max-Fair-SpareOnly* and *Max-Vol-SpareOnly* have less total unused capacity (as indicated in Table 8.1), their blocking performances are still better than the non-optimized case. As expected, if re-arranging working paths is permitted, further improvement in blocking can be achieved by the *Complete* schemes. The *Max-Vol-Complete* gives the best overall blocking reduction, provided that the underlying network management systems are capable of re-routing nearly 15% of the OD pairs without significant impact.

Table 8.2. Blocking improvement from four re-optimization strategies.

Re-optimization Strategy	Not re-optimized	Max-Fair-SpareOnly	Max-Fair-Complete	Max-Vol-SpareOnly	Max-Vol-Complete
Average blocking (%) in random demand arrival test	54.33	40.75	21.08	43.17	19.58
Standard dev. of blocking (%)	6.12	6.78	6.80	6.13	6.69
% improvement in blocking	benchmark	25.00	61.20	20.55	63.96
Number of node pairs experiencing working path rearrangement	0/55	0/55	13/55	0/55	8/55
Number of restoration flow assignments changed during re-optimization	0/114	53/114	64/114	46/114	48/114

8.4 Summary

We have presented an integer linear program formulation to characterize four re-optimization strategies for span-restorable mesh networks, and we showed that the potential improvement in blocking from each strategy to face random arrival of subsequent demands. Such blocking improvement might imply that one can serve demand growth while deferring unnecessary capital investment for transport capacity. Overall, the best strategy tends to re-optimize for maximum "volume" of potential future paths, with working paths rearrangement allowed. Re-optimization of only restoration flows and spare channels still gave a 20% reduction in blocking in the test cases presented here.

While this study might provide the numerical justification of *why* re-optimization should be warranted and *what* capacity configuration could maximize the readiness to adapt to uncertain demand growth, these questions only represent a part of the overall issue. For example, how to obtain an accurate view of existing network elements or capacity assets poses a difficult and real practical challenge on service providers, especially when the task is to get an updated, consolidated capacity view in multi-vendor, multi-administrative network environment. To optimize, one must have an accurate, central view on inventory of capacities as a prerequisite. *Readers should note that this is also a key assumption of this chapter for us to measure the* "stranded" capacities, to manage and finally to optimize them accordingly.

9 Concluding Remarks

9.1 Best Strategic Tool for Demand Uncertainty?

Given the new research tools presented in this thesis, a question might now be: What is the best strategy under demand uncertainty? An answer to this question is: it depends how much we know about the future. There is no *single* best, or "one-size-fits-all," solution to this problem.

Recall that in Section 4.3.1, we have presented the conceptual framework by Courtney, Kirkland and Viguerie and classified the notion of uncertainty into Level I (A Clear-Enough Future), Level II (Alternative Futures), Level III (A Range of Futures) and Level IV (True Ambiguity). Our optimal selection on the type of strategies would depend on the level of demand information available at the time of the decision-making.

Under circumstances where it is possible to identify clear trends on demand traffic and a single forecast is precise enough to capture the future, the use of traditional planning approaches that optimize the routing and transport capacity assignment for a single target planning view is most appropriate. All capacity design methods in Section 4.2 provide possible design tools. In today's highly competitive, multi-service environment, obtaining a single demand forecast might seem impossible. Indeed, in the past when the telecommunication industry was a monopoly, when telephony was the dominant traffic and exhibited constant growth in transport network, demand information was inherently knowable and precisely predictable.

At Level II, we assume that uncertainty can be described as a set of discrete, plausible demand scenarios. If additional information is available, one can assign a probability measure to each of the outcomes. Based on such uncertainty assumptions, our stochastic programming (SP) based models in Chapter 6 can provide a more suitable planning strategy, compared to traditional ones where they might produce, at best, the same SP-based capacity design solution (when the recourse factor is less than or equal to 1), and, at worst, very expensive designs (when the recourse factor is above 1). These supporting results were fully discussed and illustrated in Figure 6.4.

When no distinct scenario can be identified, the strategies proposed in Chapter 5 can be used to evaluate the robustness of possible designs to the range of outcomes. The SP-based capacity methods are not recommended and might produce misleading solutions when Level III uncertainty is assumed. This is so because SP-based methods work best for suggesting optimal solutions based on a set of distinct scenarios. If we are forced to incorporate a full "range" of demand scenarios into a SP-based formulation, we will complicate the complexity of the problem, which leads to a design solution with only inconclusive insights (as discussed in Section 4.3.1.) On the other hand, evaluative approaches based on the PFA-Servability framework (and the ones discussed in Section 4.3.3) enable decision makers to more effectively identify and compare robust designs that use traditional planning approaches. Another recommendation is that if capacity planning decisions can be postponed and it is possible to wait for more information on the demand until some discrete scenarios are revealed, Level II strategies could still be employed.

The demand loading and re-optimization strategies presented in Chapters 7 and 8 are the most suitable strategies for Level IV uncertainty, where there is simply no basis to forecast the future. From these strategies, we are no longer bounded by the precision of the demand forecast (or the set of scenarios) as a crucial input parameter to the formulations, but rather we can apply them to whatever demand scenario might arise. In other words, the optimality from the demand loading or re-optimization solutions is always valid, regardless of the ambiguity of the demand forecast. If we happen to have a demand scenario that exceeds the available capacity provided by the existing network, the demand loading formulation can be used to select the most profitable demands to provision, specific to which node-pair and in which multi-QoP service class. For cases where the network has surplus capacity to support the already in-service demands, re-optimization strategy can be used to turn "stranded" working and spare capacities into productive ones for adapting the unforeseen future.

To conclude, we have presented a portfolio of strategic tools and ideas of how demand uncertainty can be handled in the planning and management of mesh survivable transport networks. We have also acknowledged the limitation and underlying implications of each strategy, so that network planners, researchers and decision makers can judiciously pick and choose the right model to be applied.

9.2 Summary of Publications

The following journal, magazine and conference papers have been published or have been accepted for publication during the course of my doctoral program (listed in chronological order).

1. **D. Leung**, W. D. Grover, "Maximum-Profit Model for Study of Multi-QoP Wavelength Service Offerings in Survivable Mesh Networks," in Proceedings of the Optical Fiber Communication Conference & Exposition and the National Fiber Optic Engineers Conference (OFC/NFOEC), Anaheim Convention Center, Anaheim, California, March 6-11, 2005,

2. **D. Leung**, S. Arakawa, M. Murata, W. D. Grover, "Re-optimization Strategies to Maximize Traffic-Carrying Readiness in WDM Survivable Mesh Networks," in Proceedings of the Optical Fiber Communication Conference & Exposition and the National Fiber Optic Engineers Conference (OFC/NFOEC), Anaheim, California, March 6-11, 2005.
3. **D. Leung**, W. D. Grover, "Capacity Planning of Survivable Mesh-based Transport Networks under Demand Uncertainty," Accepted to Journal of Photonic Network Communications, February 14, 2005.
4. **D. Leung**, W. D. Grover, "Capacity Design of *p*-Cycle Networks in Face of Demand Forecast Uncertainty," in Proceedings of the 9th OptoElectronics and Communications Conference / 3rd International Conference on Optical Internet (OECC/COIN 2004), Pacifico Yokohama, Kanagawa, Japan, July 12-16, 2004.
5. **D. Leung**, W. D. Grover, "Restorable Mesh Network Design under Demand Uncertainty: Toward 'Future-proofed' Transport Investments," in Proceedings of the Optical Fiber Communication Conference (OFC 2004), Los Angeles, California, February 22-27, 2004.
6. **D. Leung**, W. D. Grover, "Comparative Ability of Span Restorable and Path Protected Network Designs to withstand Uncertainty in the Demand Forecast," in Proceedings of the 18th National Fiber Optic Engineers Conference (NFOEC 2002), Dallas, Texas, Sept. 15-19, 2002.
7. W. D. Grover, J. Doucette, M. Clouqueur, **D. Leung**, D. Stamatelakis, "New Options and Insights for Survivable Transport Networks," IEEE Communications Magazine, vol. 40, no. 1, January 2002.
8. M. Clouqueur, W. D. Grover, **D. Leung**, O. Shai, "Mining the Rings: Strategies for Ring-to-Mesh Evolution," in Proceedings of the Third International Workshop on Design of Reliable Communication Networks (DRCN), Budapest, Hungary, October 2001.
9. W. D. Grover, M. Clouqueur, **D. Leung**, "Evolution of a Telecommunications Network from Ring to Mesh Structure," U.S. Patent No. 60,301,120, June 28, 2001, CDN Patent No. 2,392,123, June 28, 2002.

Some comments must be given on my contributions to the published works 7, 8 and 9 above, related to the idea of "Ring Mining." The idea of Ring Mining, originated from Dr. Wayne Grover, suggests a potential strategy for network operators who operate ring-based fibre-optic networks. Without physically adding new capacity, operators could increase the network capacity of an existing ring-based fibre-optic network through an operational transformation from a ring-based to a mesh-based topology. Somewhat similar to the re-optimization strategies, the ring mining strategy might allow operators to serve future growth by efficiently utilizing their existing infrastructure, and hence, potentially postpone major capital additions for a significant period of time.

Results from 7, 8 and 9 are based on a project that I conducted in a graduate course entitled "Survivable Networks" in Fall 2000. In that project, I undertook the preliminary investigation of this idea and then developed two optimization models showing that this strategy could be beneficial. In several test cases, we found that the capacity of an existing ring-based network, when reclaimed and used by a mesh-based architecture, could support a demand up to 60% higher than the original demand served by the ring-based network. Because I had to complete three other courses in the subsequent winter term, this project was transferred to Matthieu Clouqueur, a past research engineer at TRLabs, and Ofer Shai, a past co-op student with TRLabs. Thus I only participated in the initial implementation and paper revision process, but not in the final writing of the joint publications.

9.3 Problems for Future Research

9.3.1 Long-Term Capacity Investment under Demand Uncertainty

Recall that in Table 2.3 we illustrated the differences between the long-term, medium-term and short-term planning problems. While studies in this thesis are strictly within the scope of medium- and short-term capacity planning contexts, the effects of demand uncertainty on a long-term planning problem could be investigated. In the long-term context, both capacity and network topology (in particular, network spans) can be altered along with the demand evolution. If an optimization-based approach is taken, the publication by Grover and Doucette [GrD01b] and other references documented in the Master's thesis by Ezema [Eze03] might provide a good starting point for tackling this problem.

If our question of interest is not to decide where and how much capacity needs to be allocated, but more strategically, to determine *whether* and *when* new capacity should be invested, an analytical technique called Real Options might be an appropriate approach for evaluating the time-discounted value of the capacity investment under uncertainty [DiP94][AlN99][All02].

In much of the real options literature (including [DiP94][AlN99][All02]), the real options approach is compared to the discounted cash flow (DCF) analysis. In a basic DCF analysis, a net present value (NPV) is calculated based on summing a series of discounted cash flows (e.g., revenues and costs). From engineering economics, we learned that if NPV is greater than zero (or some threshold values), it implies that investment should be made. On the other hand, a negative

NPV would suggest the investment should not be made at present (or in the future). One of the new dimensions of real options (over traditional DCF analysis) is that this approach considers the possibility of delay and makes the investment decision at a later time. Such a delay might have some value referred as "the option value of waiting," and might exploit the flexibility or option to defer, abandon, contract, switch and expand the investment. Another noticeable difference of the two approaches is the characterization of uncertainty in real options is more sophisticated than that in traditional DCF analysis. Dixit and Pindyck [DiP94] have given an excellent treatment to real options. Its applications to telecommunications can also be found in [AlN99][All02].

9.3.2 Heuristics for Solving TP-SR, TP-MSR Formulations

Like any other integer program formulations, the stochastic programming-based formulations in Chapter 6 have their own limitations in terms of computational complexity. For many real-world network topologies that have tens of nodes and spans, heuristic or decomposition techniques are warranted to find approximate, sub-optimal solutions.

As mentioned in Section 4.3.2, decomposition techniques, such as Benders decomposition and Lagrangian Relaxation, might be used to break down large problems into more manageable sub-problems. Sampling methods, such as the stochastic quasi-gradient method, importance sampling and stochastic decomposition, might be useful options.

In terms of heuristics, one obvious approach to tackling the two-part formulation is to first solve the joint capacity placement (JCP) problem for each demand scenario, and then combine all capacity designs with recourse-aware factors. Another possibility is to use the set of demand scenarios as input and generate a master demand matrix. From this, we then solve the joint capacity placement problem. Note that the reduction in computation time with heuristic approaches will also affect the capability of completely characterizing the variables in the original model.

9.3.3 Re-configuration Policies for Transport Capacity Management

In Section 4.4.2, we described various types of re-configuration problems. An interesting extension to our study of re-optimization might be to look into the questions of *when* and *how often* re-optimization processes should be triggered. Event-triggered, periodic-based, threshold-based reconfiguration policies (as discussed at the end of Section 4.4.2) might be possibilities well worth investigating. Recall that in our study, only incremental demand arrivals are considered. Dynamic demands, with traffic arrivals and departures, can be studied within this problem scope.

Bibliography

[ADB97] Asian Development Bank, *Guidelines for the Economic Analysis of Telecommunications Projects*, Economics and Development Resource Center, September 1997.

[ACM81] G. R. Ash, R. H. Cardwell, R. P. Murray, "Design and optimization of networks with dynamic routing," Bell System Technical Journal, vol. 60, no. 8, 1981, pp. 1787-1820.

[AHK80] A. Ali, R. Helgason, J. Kennington, H. Lall, "Computational Comparison among Three Multicommodity Network Flow Algorithms," Operations Research, Vol. 28, No. 4, July 1980, pp. 995-1000.

[AiP94] S. Aidarous, T. Plevyak (eds), *Telecommunications Network Management into the 21st Century: Techniques, Standards, Technologies, and Applications*, The Institution of Electrical and Electronics Engineers (IEEE Press), New York NY, 1994.

[AKQ00a] V. Anand, T. Katarki, C. Qiao, "Profitable connection assignment in all optical WDM networks," Proceedings of Optical Networks Workshop, Dallas, TX, Feb 2000.

[AKQ00b] V. Anand, T. Katarki, C. Qiao, "Profitable connection assignment for incremental traffic in all-optical WDM networks," Academia/Industry Working Conference on Research Challenges, April 27-29, 2000, pp. 355 – 360.

[All02] J. Alleman, "A New View of Telecommunications Economics," Telecommunications Policy, vol. 26, 2002, pp. 87-92.

[AlN99] J. Alleman, E. Noam, *The New Investment Theory of Real Options and its Implication for Telecommunications Economics*, Kluwer Academic Publishers, Massachusetts, 1999.

[ATI01] The Alliance for Telecommunications Industry Solutions (ATIS) Committee T1A1, ATIS Telecom Glossary 2000, T1.523-2001, February 28, 2001, Online: http://www.atis.org/tg2k, [date accessed: January 3, 2005]

[ATT04] AT&T Enterprise Business, "Achieving Resilience – Best Practices in Business Continuity," August 11, 2004.

[BaB97] S. Baroni, P. Bayvel, "Wavelength requirements in arbitrarily connected wavelength-routed optical networks," Journal of Lightwave Technology, Volume 15, Issue 2, Feb. 1997, pp. 242 – 251.

[BaK00] D. Backman, U. Killat, "Planning of Survivable ATM Networks Based on the Virtual Path Concept," 2nd International Workshop on the Design of Reliable Communication Networks (DRCN 2000), Munich, Germany, April 2000.

[BaM96] D. Banerjee, B. Mukherjee, "A practical approach for routing and wavelength assignment in large wavelength-routed optical networks," IEEE JSAC, vol. 14, pp.903-908, June 1996.

[BaM97] D. Banerjee, B. Mukherjee, "Wavelength-routed optical networks: linear formulation, resource budgeting tradeoffs, and a reconfiguration study," IEEE

Sixteenth Annual Joint Conference of the Computer and Communications (INFOCOM '97), volume 1, April 7-11, 1997, pp. 269 – 276.

[BaR99] I. Baldine, G. N. Rouskas, "Dynamic reconfiguration policies for WDM networks," IEEE Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '99), Volume 1, March 21-25, 1999, pp. 313 – 320.

[BCM97] D. Bai, T. Carpenter, J. Mulvey, "Making a case for robust optimization models," Management Science, vol.43, no.7, July 1997, pp.895-907.

[BeW93] P. Bernt, M. Weiss, *International Telecommunications,* Sams Publishing, 1993.

[BiL97] J. R. Birge, F. Louveaux, *Introduction to Stochastic Programming*, Springer-Verlag, Now York, 1997.

[BKO03] G. Birkan, J. Kennington, E. Olinick, A. Ortynski, G. Spiride, "Making a case for using integer programming to design DWDM networks," Optical Networks Magazine, vol. 4, no. 6, November 2003, pp. 107-120.

[BLR05] E. Bouillet, J.-F. Labourdette, R. Ramamurthy, S. Chaudhuri, "Lightpath Re-Optimization in Mesh Optical Networks," IEEE/ACM Transactions on Networking, Volume 13, Issue 2, April 2005, pp. 437 – 447.

[BML02] E. Bouillet, P. Mishra, J.-F. Labourdette, K. Perlove, S. French, "Lightpath Re-optimization in Mesh Optical Networks," Proc. 7th European Conference on Networks & Optical Communications (NOC), Darmstadt, Germany, June 2002.

[BMO84] A. B. Borison, P. A. Morris, S. S. Oren, "A State-Of-The-World Decomposition Approach to Dynamics and Uncertainty in Electric Utility Generation Expansion Planning," Operations Research, Vol. 32, No. 5, Sep 1984, pp. 1052-1068.

[BMS02] A. Balakrishnan, T. L. Magnanti, J. S. Sokol, Y. Wang, "Spare-Capacity Assignment For Line Restoration Using a Single-Facility Type," INFORMS Operations Research, vol. 50, no. 4, July – August 2002, pp. 617 –635.

[CCF98] T. W. Chung, J. Coulter, J. Fitchett, S. Mokbel, B. Arnaud, "Architectural and Engineering Issues for Building an Optical Internet," draft, CANARIE, September 1998, Online: http://www.canet2.net [date accessed: July 30, 2003]

[CGK92] I. Chlamtac, A. Ganz, and G. Karmi, "Lightpath Communications: An Approach to High Bandwidth Optical WANs," IEEE Transactions on Communications, vol.40, no.7, July 1992, pp. 1171-1182.

[Che88] J. Cheney, "The application of optimisation methods to the design of large scale telecommunication networks," IEE Colloquium on Large-Scale and Hierarchical Systems, Mar 1988, pp. 2/1 - 2/2.

[ChG95] S. G. Chang, B. Gavish, "Lower Bounding Procedures for Multiperiod Telecommunications Network Expansion Problems," Special Issue on Telecommunications Systems: Modeling, Analysis and Design, Operations Research, Vol. 43, No. 1, Jan., 1995, pp. 43-57.

[CHS98] T. Carpenter, D. Heyman, I. Saniee, "Studies of random demands on network costs," Telecommunication Systems, vol. 10, no. 3-4, September 1998, pp.409-421.

[ChV98] J. Chamberlain, D. Vokey, "Metallic Armored vs. All Dielectric Fiber Optic Cable, the Pros and Cons," Proceedings of National Fiber Optic Engineers Conference (NFOEC), 1998.

[Cis03] Cisco Systems, "Cisco ONS 15327 Reference Manual: Product and Documentation, Release 4.0," March 2003.

[ClG02] M. Clouqueur, W. D. Grover, "Mesh-restorable Networks with Complete Dual-failure Restorability and with Selectively Enhanced Dual-failure Restorability Properties," Proceedings of the SPIE Optical Networking and Communications Conference (OptiComm), Boston, July 29-Aug. 2, 2002, paper 4874-1, pp.1-12.

[CKV97] H. Courtney, J. Kirkland, P. Viguerie, "Strategy Under Uncertainty," Harvard Business Review, vol. 75, no. 6, November 1997, pp. 67-79.

[CoW03] C. Courcoubetis, R. Weber, *Pricing Communication Networks: Economics, Technology and Modelling*, Wiley Press, 2003.

[DaK95] H. T. Daugherty, W. J. Klein, "U.S. Network Reliability Issues and Major Outage Performance," Proceedings of IEEE Symposium on Computers and Communications, June 27-29, 1995, pp. 114 – 119.

[Dan55] G. B. Dantzig, "Linear Programming under Uncertainty," Management Science, vol. 1, no. 3/4, April – July 1955, pp. 197-206.

[DaW60] G. B. Dantzig, P. Wolfe, "Decomposition Principle for Linear Programs," Operations Research, Vol. 8, No. 1, Jan 1960, pp. 101-111.

[DGM99] J. Doucette, W. D. Grover, R. Martens, "Modularity and Economy-of-Scale Effects in the Optimal Design of Mesh-Restorable Networks," IEEE Canadian Conference Electrical & Comp. Engineering (CCECE'99), Edmonton, May 9-12, 1999, vol.1, pp. 226-231.

[DiP94] A. K. Dixit, R. S. Pindyck, *Investment under Uncertainty*, Princeton University Press, New Jersey, 1994.

[DiP95] A. K. Dixit, R. S. Pindyck, "The Options Approach to Capital Investments," Harvard Business Review, no. 73, 1995, pp. 105-115.

[DML94] R. D. Doverspike, J. A. Morgan, W. Leland, "Network design sensitivity studies for use of digital cross-connect systems in survivable network architectures," IEEE Journal on Selected Areas in Communications, Volume 12, Issue 1, Jan. 1994 pp. 69 – 78.

[DoG00] J. Doucette, W. D. Grover, "Influence of Modularity and Economy-of-Scale Effects on Design of Mesh-Restorable DWDM Networks," IEEE Journal on Selected Areas in Communications (JSAC), vol. 18, no. 10, October 2000, pp. 1912-1923.

[DoG01] J. Doucette, W. D. Grover, "Comparison of Mesh Protection and Restoration Schemes and the Dependency on Graph Connectivity," Proceedings 3rd International Workshop on the Design of Reliable Communication Networks (DRCN 2001), Budapest, Hungary, October 2001, pp. 121-128.

[DoG02] J. Doucette, W. D. Grover, "Capacity Design Studies of Span-Restorable Mesh Networks with Shared-Risk Link Group (SRLG) Effects," Proceedings of the SPIE Optical Networking and Communications Conference (OptiComm), Boston, July 29-Aug. 2, 2002, paper 4874-3, pp.25-38.

[Dov91] R. Doverspike, "A multi-layered model for survivability in intra-LATA transport networks," IEEE Global Telecommunications Conference (GLOBECOM), vol.3, Dec 2-5, 1991, pp. 2025 – 2031.

[DoW94] R. Doverspike and B. Wilson, "Comparison of capacity efficiency of DCS network restoration routing techniques," Journal of Network and System Management, vol. 2, no. 2, 1994, pp. 95-123.

[Dri04] E. Drilling, "SBC Multi-Service Optical Network Ring Service," Southwestern Bell Telephone, Little Rock, Arkansas, May 10, 2004. Online: http://www.sbc.com/Large-Files/RIMS/Arkansas/Digital_Link/ar-dl-16.pdf, [date accessed: July 20, 2004]

[DSB01] S. Datta, S. Sengupta, S. Biswas, S. Datta, "Efficient Channel Reservation for Backup Paths in Optical Mesh Networks," IEEE Globecom 2001, San Antonio, TX, November 2001.

[DuL92] A. Dutta, J. Lim, "A Multiperiod Capacity Planning Model for Backbone Computer Communication Networks," Operations Research, Vol. 40, No. 4, July 1992, pp. 689-705.

[Dup02] J. Dupaová, "Applications of stochastic programming: Achievements and questions," European Journal of Operational Research, Volume 140, Issue 2, July 16, 2002, pp. 281-290.

[Dut94] A. Dutta, "Capacity planning of private networks using DCS under multibusy-hour traffic," IEEE Transactions on Communications, Volume 42, Issue 7, July 1994 pp. 2371 – 2374.

[Dwi03] A. Dwivedi, "Efficiency, Utilization and Evolution of Current Networks," Proceedings of the 19th National Fiber Optic Engineers Conference (NFOEC 2003), Orlando, FL, Sept. 7-11, 2003, pp. 731-739.

[DwW00] A. Dwivedi, R. Wagner, "Traffic model for USA long-distance optical network," Proc. of Optical Fiber Communication Conference (OFC), Baltimore, March 2000, vol. 1, TuK1-1, pp.156-158.

[EFL99] R. J. Ellison, D. A. Fisher, R. C. Linger, H. F. Lipson, T. Longstaff, N. R. Mead, "Survivable Network Systems: An Emerging Discipline," Technical Report CMU/SEI-97-TR-013 and ESC-TR-97-013, Carnegie Mellon University, Software Engineering Institute, Nov 1997, Rev. May 1999.

[EUR00a] EURESCOM, Project P709, "Planning of Optical Network," Deliverable 3, Optical Network Planning, Volume 1 of 9: Main Report, March 2000.

[EUR00b] EURESCOM, Project P709, "Planning of Optical Network," Deliverable 3, Optical Network Planning, Volume 7 of 9: Main Report, March 2000.

[Eva03] S. Evans (editor), *Telecommunications Network Modelling, Planning and Design*, Institution of Electrical Engineers, London, 2003.

[Eze03] C. A. Ezema, *Topology Design of Mesh-restorable Networks*, M.Sc. Thesis, University of Alberta, Spring 2003.

[FGK93] R. Fourer, D. Gay, B. Kernighan, AMPL: A Modeling Language for Mathematical Programming. Fraser Publishing Company, Danvers, MA, 1993.

[FiK02] R. Fildes, V. Kumar, "Telecommunications demand forecasting - a review," International Journal of Forecasting, Vol. 18 (4), 2002, pp. 489-522.

[FOE89] T. Flanagan, S. Oxner, D. Elkaim, "Principles and Technologies for Planning Survivability – A Metropolitan Case Study," IEEE Global Telecommunications Conference (GLOBECOM), vol.2, Nov. 27-30, 1989, pp. 813 – 820.

[FPS03] S. French, D. Pendarakis, D. Saha, "Mining Services from the Optical Layer," Proceedings of the 19th National Fiber Optic Engineers Conference (NFOEC 2003), Orlando, FL, Sept. 7-11, 2003, pp.1058-1067.

[GAD01a] N. Geary, A. Antonopoulos, E. Drakopoulos, J. O'Reilly, "Analysis of Optimisation Issues in Multi-Period DWDM Network Planning," Proceedings of the IEEE Conference on Computer Communications (INFOCOM), Anchorage, Alaska, April 2001, pp. 152-158.

[GAD01b] N. Geary, A. Antonopoulos, E. Drakopoulos, J. O'Reilly, J. Mitchell, "A Framework for Optical Network Planning under Traffic Uncertainty," 3rd International Workshop on the Design of Reliable Communications Networks (DRCN 2001), Budapest, Hungary, 2001.

[Gai95] A. Gaivoronski, "Stochastic Programming Approach to the Network Planning under Uncertainty," Optimization in Industry 3, John Wiley & Sons, 1995, pp.145-163.

[GAM03] N. Geary, A. Antonopoulos, J. Mitchell, "Network and business modeling under traffic forecast uncertainty: a case study," 4th International Workshop on the Design of Reliable Communications Networks (DRCN 2003), Banff, Alberta, Canada, Oct 19-22, 2001, pp. 304-310.

[GaR99] D. T. Gardner, J. S. Rogers, "Planning Electric Power Systems under Demand Uncertainty with Different Technology Lead Times," Management Science, vol. 45, no. 10, Oct 1999, pp. 1289-1306.

[Gau03] R. Gaudet, "Advances in Wavelength Services," Proceedings of the 19th National Fiber Optic Engineers Conference (NFOEC 2003), Orlando, FL, Sept. 7-11, 2003, pp.1505-1510.

[GBV91] W. D. Grover, T. D. Bilodeau, B. D. Venables, "Near optimal spare capacity planning in a mesh restorable network," IEEE GLOBECOM, Dec 2-5, 1991, pp. 2007 – 2012.

[GDC02] W.D. Grover, J. Doucette, M. Clouqueur, D. Leung, D. Stamatelakis, "New Options and Insights for Survivable Transport Networks," IEEE Communications Magazine, vol.40, no.1, January 2002, pp. 34-41.

[Gea03] N. Geary, *Optical network planning: process, analysis and optimization*, Ph.D. Thesis, Communications Engineering Doctorate Centre, Department of Electronic and Electrical Engineering, University College London, May 2003.

[GeM03] A. Gencata, B. Mukherjee, "Virtual-Topology Adaptation for WDM Mesh Networks Under Dynamic Traffic," IEEE/ACM Transactions on Networking, Vol.11, No.2, April 2003, pp.236-247.

[GeS01] O. Gerstel, G. Sasaki, "Quality of Protection (QoP): A Quantitative Unifying Paradigm to Protection Service Grades," Proceedings of the SPIE Optical Networking and Communications Conference (OptiComm), Aug. 2001.

[GIZ99] W.D. Grover, R.R. Iraschko, Y. Zheng, "Comparative Methods and Issues in Design of Mesh-Restorable STM and ATM Networks," Telecommunication Network Planning, B. Sanso, P. Soriano (editors), Kluwer Academic Publishers, 1999, pp. 169-200.

[Glo05] Global Crossing Limited, Interactive Network Maps, Online: http://www.globalcrossing.com, [date accessed: February 2, 2005]

[GoB04] W. Golab, R. Boutaba, "Policy-Driven Automated Reconfiguration for Performance Management in WDM Optical Networks, IEEE Communications Magazine," Special Issue on Management of Optical Networks, January 2004.

[GrC02] W. D. Grover, M. Clouqueur, "Span-Restorable Mesh Network Design to Support Multiple Quality of Protection (QoP) Service Classes," Proceeding of 1st International Conference on Optical Communications and Networks (ICOCN02), Singapore, Nov. 11-14, 2002, pp.321-323.

[GrD01a] W. D. Grover and J. Doucette, "Increasing the Efficiency of Span-restorable Mesh Networks on Low-connectivity Graphs," 3rd International Workshop on the Design of Reliable Communications Networks (DRCN), Budapest, Hungary, Oct. 2001, pp. 99–106.

[GrD01b] W. D. Grover, J. Doucette, "Topology Optimization of Survivable Mesh-Based Transport Networks," Annals of Operations Research, 2001, vol. 106, pp. 79 – 125.

[Gre01] P. Green, "Progress in Optical Network," IEEE Communications Magazine, vol.39, no.1, Jan 2001, pp. 54-61.

[Gro03] W. D. Grover, Mesh-based Survivable Networks: Options and Strategies for Optical, MPLS, SONET and ATM Networking, Prentice-Hall PTR, 2003.

[Gro04] W. D. Grover, *Mesh-based Survivable Networks: Options and Strategies for Optical, MPLS, SONET and ATM Networking*, Prentice Hall, 2004.

[GVB91] W. D. Grover, T. D. Bilodeau, B. D. Venables, "Near Optimal Synthesis of a Mesh Restorable Network," Proceeding of IEEE Global Telecommunications Conference (GLOBECOM), vol.3, Dec. 1991, pp. 2007-2012.

[GVS90] W. D. Grover, B. D. Venables, J. H. Sandham, A. F. Mine, "Performance studies of a selfhealing network protocol in Telecom Canada long haul networks," Proceeding of IEEE Global Telecommunications Conference (GLOBECOM), Dec. 2-5, 1990, pp. 452 – 458.

[HBB03] S. J. Hood, S. Bermon, F. Barahona, "Capacity Planning Under Demand Uncertainty for Semiconductor Manufacturing," IEEE Trans. On Semiconductor Manufacturing, vol. 16, no. 2, May 2003, pp. 273-280.

[HBU95] M. Herzberg, S. J. Bye, A. Utano, "The hop-limit approach for spare-capacity assignment in survivable networks," IEEE/ACM Transactions on Networking, vol. 3, no. 6, Dec. 1995, pp. 775 – 784.

[HeB94] M. Herzberg, S. Bye, "An Optimal Spare Capacity Assignment Model for Survivable Network with Hop Limits," IEEE GLOBECOM 1994, San Francisco, USA, pp. 1601-1606.

[HeH95] A. Herschtal, M. Herzberg, "Dynamic capacity allocation and optimal rearrangement for reconfigurable networks," IEEE Global Telecommunications Conference (GLOBECOM '95), Volume 2, Nov. 13-17, 1995, pp. 946 – 951.

[HiS91] J. L. Higle, S. Sen, "Stochastic decomposition: An algorithm for stage linear programs with recourse," Mathematical Operations Research, vol. 16, 1991, pp. 650-669.

[HLB95] M. Hopkins, G. Louth, H. Bailey, R. Yellon, A. Ajibulu, M. Niva, "A multi-faceted approach to forecasting broadband demand and traffic," IEEE Communications Magazine, vol. 33, no. 2, Feb. 1995, pp. 36 – 42.

[HoM01] P. H. Ho, H. T. Mouftah, "SLSP: a new path protection scheme for the optical Internet," Optical Fiber Communication Conference and Exhibit (OFC), vol. 2, TuO1, 2001, pp. TuO1-1 - TuO1-3.

[HoM02] P. H. Ho, H. T. Mouftah, "A framework for service-guaranteed shared protection in WDM mesh networks," IEEE Communications Magazine, vol. 40, no. 2, Feb. 2002, pp. 97 – 103.

[Hun03] J. Hunt, "Managed Wavelength Services: A Reality Check," Proceedings of the 19th National Fiber Optic Engineers Conference (NFOEC 2003), Orlando, FL, Sept. 7-11, 2003, pp.1499-1504.

[IAM03] S. Ishida, S. Arakawa, M. Murata, "Reconfiguration of Logical Topologies with Minimum Traffic Disruptions in Reliable WDM-based Mesh Networks," Photonic Network Communications, June 2003, pp. 265-277.

[IEC04] International Engineering Consortium, "Optical Switches: Making Optical Networks a Brilliant Reality," Web ProForums, 2002.

[IFM02] T. Islam, D.G. Fiebig, N. Meade, "Modelling multinational telecommunications demand with limited data," International Journal of Forecasting, Vol. 18 (4), 2002, pp. 605-624.

[ILO01] ILOG, "Network and Service Management: Smart Components for Next-Generation OSS," ILOG Datasheets, September 2001. Available Online, http://www.ilog.com/industries/communications/datasheets/ [date accessed: June 22, 2004]

[ILO04] ILOG corporate website, "ILOG Optimization Product - ILOG CPLEX," Online: http://www.ilog.com/products/cplex/ [date accessed: June 4, 2004]

[IMG96] R.R. Iraschko, M.H. MacGregor, W. D. Grover, "Optimal capacity placement for path restoration in mesh survivable networks," IEEE International Conference on Communications (ICC), vol. 3, June 23-27, 1996, pp. 1568 – 1574.

[IMG98] R. R. Iraschko, M. H. MacGregor, W. D. Grover, "Optimal capacity placement for path restoration in STM or ATM mesh-survivable networks," IEEE/ACM Transactions on Networking, vol. 6, no. 3, June 1998, pp. 325 – 336.

[Ira96] R. R. Iraschko, *Path-Restorable Networks*, Ph.D. Thesis, University of Alberta, Fall 1996.

[ITU02] ITU-T Recommendation E.360.6, "Series E: Overall Network Operation, Telephone Service, Service Operation and Human Factors: QoS Routing and Related Traffic Engineering Methods – Capacity Management Methods," May 2002.

[ITU05] International Telecommunication Union (ITU), Study Group 15, Optical Technologies, Online: http://www.itu.int/ITU-T/studygroups/com15/otn, [date accessed: February 17, 2005]

[JaB02] R. Jain, N. Bisht, "Current Issues and Trends in Optical Networking," Business Briefing: Global Optical Communications, World Markets Research Centre, 2002.

[KaM02] O. Kabranov, D. Makrakis, "Performance evaluation of optimal wavelength allocation and flow assignment for optical networks using profit maximization under demands uncertainty," IEEE Canadian Conference on Electrical and Computer Engineering (CCECE 2002), Volume 3, May 12-15, 2002, pp. 1461 – 1466.

[KaW94] P. Kall, S. Wallace, *Stochastic Programming*, John Wiley & Sons, Chichester, New York, 1994.

[KCK91] T. Kikuno, C. Chen, K. Kawashima, Y. Kakuda, "Spare channel assignment for restoration in fault-tolerant loop network," Proceedings of Fault Tolerant Systems, Pacific Rim International Symposium, Sept. 26-27, 1991, pp. 108 – 113.

[KDP95] K. R. Krishnan, R. D. Doverspike, C. D. Pack, "Improved survivability with multi-layer dynamic routing," IEEE Communications Magazine, Volume 33, Issue 7, July 1995, pp. 62 – 68.

[KeL01] J. Kennington and M. Lewis, "The Path Restoration Version of the Spare Capacity Allocation Problem with Modularity Restrictions: Models, Algorithms, and an Empirical Analysis," INFORMS Journal On Computing, vol. 13, no. 3, Summer 2001, pp. 181-190.

[KKD98] L.G. Kazovsky, G. Khoe, M. Deventer, "Future Telecommunication Networks: Major Trend Projections," IEEE Communications Magazine, vol. 36, no. 11, November 1998, pp.122-127.

[KLO01] J. Kennington, K. Lewis, E. Olinick, A. Ortynski, G. Spiride, "Robust solutions for the WDM routing and provisioning problem: models and algorithms," Technical Report, 01-EMIS-03, EMIS Dept., School of Engineering, SMU, Dallas, TX, July 2001.

[KMC01] O. Kabranov, D. Makrakis, C. Charalambous, D. Ionescu, G. Bochman, "Optimal wavelength allocation and flow assignment for optical networks for profit maximization," International Conferences on Info-tech (ICII 2001), Beijing, Volume 2, Oct. 29 - Nov. 1, 2001, pp. 182 – 187.

[KoL01] Murali Kodialam, T. V. Lakshman, "Dynamic routing of locally restorable bandwidth guaranteed tunnels using aggregated link usage information," IEEE INFOCOM on Computer Communications, no. 1, April 2001, pp. 376-385.

[KOL03] J. Kennington, E. Olinick, K. Lewis, A. Ortynski, G. Spiride, "Robust Solutions for the DWDM Routing and Provisioning Problem: Models and Algorithms," Optical Networks Magazine, vol. 4, no. 2, March 2003, pp. 74-84.

[Ku95] A. Ku, *Modelling Uncertainty in Electricity Capacity Planning*, Ph.D. Thesis, London Business School, University of London, February 1995.

[LAM05] D. Leung, S. Arakawa, M. Murata, W. D. Grover, "Re-optimization Strategies to Maximize Traffic-Carrying Readiness in WDM Survivable Mesh Networks," Proceedings of Optical Fiber Communication Conference & Exposition and the National Fiber Optic Engineers Conference (OFC/NFOEC), Anaheim, California, March 6-11, 2005.

[Lar94] K. G. Laretto, "Sprint Network Survivability," IEEE Military Communications Conference (MILCOM'94), Fort Monmouth, NJ, Oct 2-5, 1994.

[Lee86] L. Lee, *An Introduction to Telecommunications Network Traffic Engineering*, Alta Telecom International Ltd., Edmonton, Alberta, Canada, 1986.

[LeG02] D. Leung, W. D. Grover, "Comparative Ability of Span Restorable and Path Protected Network Designs to withstand Uncertainty in the Demand Forecast," Proceedings of the 18th National Fiber Optic Engineers Conference (NFOEC 2002), Dallas, Texas, Sept. 15-19, 2002.

[LeG04a] D. Leung, W. D. Grover, "Restorable Mesh Network Design under Demand Uncertainty: Toward 'Future-proofed' Transport Investments," Proceedings of Optical Fiber Communication Conference (OFC 2004), Los Angeles, California, February 22-27, 2004.

[LeG04b] D. Leung, W. D. Grover, "Capacity Design of *p*-Cycle Networks in Face of Demand Forecast Uncertainty," Proceedings of the 9th OptoElectronics and Communications Conference / 3rd International Conference on Optical Internet (OECC/COIN 2004), Pacifico Yokohama, Kanagawa, Japan, July 12-16, 2004.

[LeG05a] D. Leung, W. D. Grover, "Capacity Planning of Survivable Mesh-based Transport Networks under Demand Uncertainty," Accepted to Journal of Photonic Network Communications, February 14, 2005.

[LeG05b] D. Leung, W. D. Grover, "Maximum Revenue Loading for Mesh-based Survivable Networks: A Way of Coping with Demand Uncertainty," Proceedings of Optical Fiber Communication Conference & Exposition and the National Fiber Optic Engineers Conference (OFC/NFOEC), Anaheim, California, March 6-11, 2005.

[Lev02] Level (3) Communications Annual Report, 2002.

[Lev03] S. Levine, "Why the back office is big business," America's Network, vol. 107, issue 16, Nov 1, 2003, pp. 22-27.

[LHA94] J.-F. P. Labourdette, G. W. Hart, and A. S. Acampora, "Branch-Exchange Sequences for Reconfiguration of Lightwave Networks," IEEE Transactions on Communications, vol. 42, no. 10, , Oct. 1994, pp. 2822-2832.

[Lit79] S. C. Littlechild, *Elements of Telecommunications Economics*, Institution of Electrical Engineers, London, 1979.

[LiT01] Y. Liu, D. Tipper, "Successive survivable routing for node failures," IEEE Global Telecommunications Conference (GLOBECOM), vol. 4, Nov. 25-29, 2001, pp. 2093 – 2097.

[LMS89] D. N. Lee, K. T. Medhi, J. Strand, R. Cox and S. Chen, "Solving Large Telecommunications Network Loading Problems," AT&T Technical Journal , Vol. 68, No. 3, May / June 1989, pp. 48-56.

[LOV99] A. Lisser, A. Ouorou, J.-Ph. Vial, J. Gondzio, "Capacity planning under uncertain demand in telecommunication networks," Logilab Technical Report 99.13, Department of Management Studies, University of Geneva, Switzerland, October 1999.

[LTS01] Y. Liu, D. Tipper, P. Siripongwutikorn, "Approximating optimal spare capacity allocation by successive survivable routing," Proceedings of IEEE INFOCOM, vol. 2, April 22-26, 2001, pp. 699 – 708.

[LuD89] J. F. Luby, M. A. Dziatkiewicz, "Considerations and Concerns for Survivability of the Chicago MSA Fiber Optic Networks," IEEE GLOBECOM, Nov. 27-30, 1989, vol.2, pp. 808 – 812.

[Lus82] H. Luss, "Operations Research and Capacity Expansion Problems: A Survey," Operations Research, Vol. 30, No. 5, Sep., 1982, pp. 907-947.

[MaG03] P. MacLeod, A. Godin, "Managed Wavelength Services: A Service Provider's Business Case," Proceedings of the 19th National Fiber Optic Engineers Conference (NFOEC 2003), Orlando, FL, Sept. 7-11, 2003, pp.283-292.

[MaS98] A. Marin, J. Salmeron, "Electric capacity expansion under uncertain demand: decomposition approaches," IEEE Transactions on Power Systems, Volume 13, Issue 2, May 1998, pp. 333 – 339.

[Mau99] T. A. Maufer, *IP Fundamentals*, Prentice-Hall, Upper Saddle River NJ, 1999.

[Mau02a] C. Mauz, "Considering Variations of the Traffic Pattern in the Capacity Dimensioning Process for Transport Networks," Business Briefing: Global Optical Communications, June 2002, pp. 1-4.

[Mau02b] C. Mauz, "Dimensioning of a Transport Network Being Robust to Changes of the Traffic Pattern," Proceedings of the Forth International Conference on Transparent Optical Networks (ICTON'02), Warsaw, Poland, April 2002.

[MaZ94] S. A. Malcolm, S. A. Zenios, "Robust Optimization for Power Systems Capacity Expansion under Uncertainty," The Journal of the Operational Research Society, Vol. 45, No. 9, Sept. 1994, pp. 1040-1049.

[MCF03] J. Moteff, C. Copeland, J. Fischer, "Critical Infrastructures: What Makes an Infrastructure Critical?" Report from Congressional Research Service, January 2003.

[MCK92] K. Miyazaki, T. Chujo, H. Komine, T. Ogura, "Spare Capacity Assignment for Multiple-Link Failures," Proceedings of International Workshop on Advanced Communications and Applications for High Speed Networks, March 16 – 19, 1992, pp. 191 – 197.

[Med95] D. Medhi, "Multi-hour, multi-traffic class network design for virtual path-based dynamically reconfigurable wide-area ATM networks," IEEE/ACM Trans. On Networking, Vol. 3, No. 6, December 1995, pp. 809-818.

[MeK95] D. Medhi, R. Khurana, "Optimization and Performance of Network Restoration Schemes for Wide-Area Teletraffic Networks," Journal of Network and Systems Management, Vol. 3, No. 3, September 1995, pp. 265-294.

[MeL97] D. Medhi and C.-T. Lu, "Dimensioning and Computational Results for Wide-Area Broadband Networks with Two-level Dynamic Routing," IEICE Trans. on Communications, Vol. E80-B, No. 2, 1997, pp. 273-281.

[Mel04] S. Melle, "A Digital Optical Approach," Telecommunications Magazine, October 2004.

[Mer04] Meriton Networks, "8600 NMS: Simplifying management of optical networks," whitepaper, Online, http://www.meriton.com. [date accessed: June 10, 2004]

[MeS93] D. Medhi and S. Sankarappan, "Impact of a Transmission Facility Link Failure on Dynamic Call Routing Circuit-Switched Networks under Various Circuit Layout Policies," Journal of Network and Systems Management, Vol. 1, 1993, pp. 143-169.

[MeT98] D. Medhi and D. Tipper, "Some Approaches to Solving a Multi-Hour Broadband Network Capacity Design Problem with Single-Path Routing," Telecommunication Systems, Vol. 13, No. 2, 2000, pp. 269-291.

[MGR97] M.H. MacGregor, W.D. Grover, K. Ryhorchuk, "Optimal Spare Capacity Preconfiguration for Faster Restoration of Mesh Networks," Journal of Network and Systems Management, vol. 5, no. 2, June 1997, pp. 159-171.

[MiL80] D. Minoli, E. Lipper, "Cost Implications for Survivability of Terrestrial Networks Under Malicious Failure," IEEE Transactions on Communications, vol. 28, no. 9, Sept. 1980, pp. 1668 – 1674.

[Mis04] K. Misra, *OSS for Telecom Networks: An Introduction to Network Management,* Springer-Verlag, London, 2004.

[MSC02] G. Madden, S. J. Savage, G. Coble-Neal, "Forecasting United States–Asia international message telephone service," International Journal of Forecasting, Vol. 18 (4), 2002, pp. 523-543.

[MuK95] K. Murakami, H. S. Kim, "Joint optimization of capacity and flow assignment for self-healing ATM networks," IEEE International Conference on Communications (ICC), vol. 1, June 18-22, 1995, pp. 216 – 220.

[Muk97] B. Mukherjee, *Optical Communication Networking,* McGraw-Hill, 1997.

[Muk00] B. Mukherjee, "WDM Optical Communication Networks: Progress and Challenges," IEEE Journal on Selected Areas in Communications (JSAC), vol.18, no.10, October 2000.

[Mul96] J. M. Mulvey, "Solving robust optimization models in finance," IEEE/IAFE 1996 Conference on Computational Intelligence for Financial Engineering, March 24-26, 1996, pp. 1 – 13.

[MuR95] J. M. Mulvey, A. Ruszczynski, "A New Scenario Decomposition Method for Large-Scale Stochastic Optimization," Operations Research, Vol. 43, No. 3, May 1995, pp. 477 - 490.

[MVZ95] J. M. Mulvey, R. J. Vanderbei, S. A. Zenios, "Robust optimization of large-scale systems," Operations Research, vol. 43, no. 2, March 1995, pp. 264-281.

[NaN04] M. P. Narayanan, V. K. Nanda, *Finance for Strategic Decision Making: What Non-Financial Managers Need to Know,* Jossey-Bass, San Francisco, 2004.

[NdF00] M. Ndesandjo, G. Friesen, "Managed Wavelength Services, the New Wave Bandwidth Options, Commercial Applications & Values," Proceedings of the 16th National Fiber Optic Engineers Conference (NFOEC 2000), Denver, Colorado, Aug. 27-31, 2000.

[Nor96] Nortel Networks, Introduction to SONET Networking, Whitepaper, October 30, 1996.

[NRC85] National Research Council Network Working Group, "Transport Protocols for Department of Defense Data Networks," Report to the Department of Defense and the National Bureau of Standards Washington, D.C., February 1985.

[NRI93] Network Reliability & Interoperability Council (NRIC), "Fiber Optic Cable Dig-Ups: Causes and Cures, Network Reliability: A Report to the Nation," Technical Report, June 1993.

[NRI04] Network Reliability & Interoperability Council (NRIC) Focus Group 1C, "Analysis of Effectiveness of Best Practices Aimed at E911 and Public Safety," Technical Report, September 23, 2004.

[NSE04] Natural Sciences and Engineering Research Council of Canada (NSERC), March 2004. Online: http://www.nserc.ca/news/2004, [date accessed: January 3, 2005]

[Oke04] S. O'Keefe, "Back Office Makeover," Telecommunications Americas, vol. 38, issue 13, Dec 2004, pp. 20-22. [OuT95] I. Ouveysi and Y. Tham, "Network design for multi-hour traffic profile," Proceedings of Australian Telecommunication Networks and Applications Conference 1995, Sydney, Australia, 1995, pp. 461-466.

[OSW98] I. Ouveysi, F. Safael, A. Wirth, "A Dimensioning and dynamic reconfiguration of hierarchical multi-service crossconnect platforms for multihour traffic profiles," IEEE International Conference on Communications (ICC 98), Volume 1, June 7-11, 1998, pp. 249 – 252.

[Oxf02] J. Black, *A Dictionary of Economics*, Oxford University Press, 2002, Online: http://www.oxfordreference.com, [date accessed: September 5, 2004]

[OMZ02] C. Ou, B. Mukherjee, H. Zang, "Sub-path protection for scalability and fast recovery in WDM mesh networks," Optical Fiber Communication Conference and Exhibit (OFC), March 17-22, 2002, pp. 495 – 496.

[PiD99] M. Pickavet, P. Demeester, "Long-Term Planning of WDM Networks: A Comparison between Single-Period and Multi-Period Techniques," Photonic Network Communications, vol. 1, no. 4, August 1999, pp. 331-346.

[PiM04] M. Pióro, D. Medhi, Routing, Flow and Capacity Design in Communication and Computer Networks, Morgan Kaufmann, July 2004.

[PKR91] D. Paraskevopoulos, E. Karakitsos, R. Rustem, "Robust capacity planning under uncertainty," Management Science, vol. 37, no. 7, July 1991, pp. 787–800.

[Ran92] K. Rana, "A decomposition technique for mixed integer programming problems," Computers & Operations Research, Volume 19, Issue 6, August 1992, pp. 505-519.

[RaR00] B. Ramamurthy, A. Ramakrishnan, "Virtual topology reconfiguration of wavelength-routed optical WDM networks," IEEE GLOBECOM 2000, vol. 2, Dec. 2000, pp. 1269 – 1275.

[RaS95] R. Ramaswami, K. N. Sivarajan, "Routing and wavelength assignment in all-optical networks," IEEE/ACM Transactions on Networking, Volume 3, Issue 5, Oct. 1995, pp. 489 – 500.

[RaS02] R. Ramaswami, K. N. Sivarajan, *Optical Networks: A Practical Perspective*, 2nd Edition, Morgan Kaufmann, San Francisco CA, 2002.

[RBS01] R. Ramamurthy, Z. Bogdanowicz, S. Samieian, D. Saha, B. Rajagopalan, S. Sengupta, S. Chaudhuri, K. Bala, "Capacity Performance of Dynamic Provisioning in Optical Networks," Journal of Lightwave Technology, vol. 19, no. 1, January 2001.

[Rea04] Reach Global Services Limited, *Global Reach Interactive Maps*, Online: http://www.reach.com, [date accessed: December 4, 2004]

[RiA02] M. Riis, K. A. Andersen, "Capacitated Network Design with Uncertain Demand," INFORMS Journal on Computing, Volume 14, Issue 3, pp. 247 – 260.

[Ros02] E. Rosenberg, "Capacity Requirements for Node and Arc Survivable Networks," Telecommunication Systems, vol. 20, no. 1 – 2, May 2002, pp. 107 – 131.

[RSC01] R. Ramamurthy, S. Sengupta, S. Chaudhuri, "Comparison of Centralized and Distributed Provisioning of Lightpaths in Optical Networks," Optical Fiber Communications Conference (OFC), Anaheim, CA, March 2001.

[Sat96] K. Sato, *Advances in Transport Network Technologies: Photonic Networks, ATM, and SDH*, Artech House, Norwood MA, 1996.

[SAG05] T. Santoso, S. Ahmed, M. Goetschalckx, A. Shapiro, "A stochastic programming approach for supply chain network design under uncertainty," European Journal of Operational Research, Volume 167, Issue 1, 16 November 2005, pp. 96-115.

[SDC94] S. Sen, R. Doverspike, S. Cosares, "Network planning with random demand," Telecommunication Systems, vol. 3 (1994), 1994, pp.11-30.

[SeG80] J. K. Sengupta, S. K. Gupta, "Optimal bus scheduling and fleet selection: A programming approach," Computers & Operations Research, volume 7, issue 4, 1980, pp. 225-237.

[SGA02] D. A. Schupke, C. G. Gruber, A. Autenrieth, "Optimal Configuration of *p*-Cycles in WDM Networks," IEEE International Conference on Communications (ICC 2002), New York City, NY, April 28 - May 2, 2002, pp. 2761-2765.

[Sha96] R. Sharma, *Broadband Optoelectronic Switching and Signal Processing*, Ph.D. Thesis, University of Alberta, Fall 1996.

[Sha00] J. K. Shaw, *Strategic Management in Telecommunications*, Artech House, 2000.

[ShG03] G. Shen, W.D. Grover, "Capacity Requirements for Network Recovery from Node Failure with Dynamic Path Restoration," Proc. Optical Fiber Communications Conference (OFC 2003), Atlanta, March 24-27 2003, pp.775-777.

[ShG04] G. Shen, W. D. Grover, "Segment-based approaches to survivable translucent network design under various ultra-long-haul system reach capabilities," OSA Journal of Optical Networking, vol. 3, no.1, January 2004, pp.1-24.

[SKH02] J. P. Sterbenz, R. Krishnan, R. R. Hain, A. W. Jackson, D. Levin, R. Ramanathan, J. Zao, "Survivable Mobile Wireless Networks: Issues, Challenges, and Research Directions," ACM Workshop on Wireless Security (WiSe02), September 28, 2002, Atlanta, Georgia, USA.

[SNH90] H. Sakauchi, Y. Nishimura, S. Hasegawa, "A self-healing network with an economical spare-channel assignment," Proceeding of IEEE Global Telecommunications Conference (Globecom), Dec. 2-5, 1990, pp. 438 – 443.

[SOH92] H. Sakauchi, Y. Okanoue, S. Hasegawa, "Spare-channel design schemes for self-healing networks," IEICE Transaction on Communication, vol. E75-B, no. 7, July 1992, pp. 624-633.

[Sor00] Sorrento Networks, "Metropolitan Optical Networks: Overview and Requirements," Whitepaper, June 2000.

[SpB98] J. Späth and S. Bodamer, "Routing of Dynamic Poisson and Non-Poisson Traffic in WDM Networks with Limited Wavelength Conversion," European Conf. on Optical Communications, vol. 1, Sept. 1998, pp. 359-360.

[SrS00] M. Sridharan, A. K. Somani, "Revenue Maximization in Survivable WDM Networks," Proceedings of the SPIE Optical Networking and Communications Conference (OptiComm), Dallas, Texas, October 22-26, 2000, pp. 291-302.

[SSA97] S. Subramaniam, A. K. Somani, M. Azizoglu, R. A. Barry, "A performance model for wavelength conversion with non-Poisson traffic," IEEE Sixteenth Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '97), volume 2, April 7-11, 1997, pp. 499 – 506.

[SSM84] H. D. Sherali, A. L. Soyster, F. H. Murphy, S. Sen, "Intertemporal Allocation of Capital Costs in Electric Utility Capacity Expansion Planning under Uncertainty," Management Science, vol. 30, no. 1, January 1984, pp. 1-19.

[SSS02] A. K. Somani, M. Sridharan, and R. Srinivasan, "Dynamic routing in survivable WDM networks," Proceedings of 40th Annual Allerton Conference on Communication, Control, and Computing, October 2002.

[StB99] T. E. Stern, K. Bala, *Multiwavelength Optical Networks: A Layered Approach*, Addison-Wesley, May 1999.

[StM95] K. Stordahl, E. Murphy, "Forecasting long-term demand for services in the residential market," IEEE Communications Magazine, Volume 33, Issue 2, Feb. 1995, pp. 44 – 49.

[SwM79] D. J. Sweeney, R. A. Murphy, "A Method of Decomposition for Integer Programs," Operations Research, Vol. 27, No. 6, Nov 1979, pp. 1128-1141.

[Thr00] Three-Sixty (360) Networks Annual Report, 2000.

[T1A01] T1A1.2 Working Group on Network Survivability Performance, Technical Report on Enhanced Network Survivability Performance, January 2001.

[TCK90] E. I. Tsai, B. A. Coan, M. Kerner, M. P. Vecchi, "A comparison of strategies for survivable network design: reconfigurable and conventional approaches," IEEE Global Telecommunications Conference (GLOBECOM '90), Dec. 2-5, 1990, vol.1, pp. 49 – 55.

[VCP03] S. Verbrugge, D. Colle, M. Pickavet, P. Demeester, "Common planning practices for network dimensioning under traffic uncertainty," 4th International Workshop on the Design of Reliable Communication Networks (DRCN 2003), Banff, Alberta, Canada, October 19-22, 2003, pp. 317-324.

[VeP02] Presentation by A. J. Vernon, J. D. Portier, "Protection of Optical Channels in All-Optical Networks," 18th Annual National Fiber Optic Engineers Conference (NFOEC 2002), Dallas, TX, September 2002, pp. 1695 – 1706.

[VGM93] B. D. Venables, W. D. Grover, M. H. MacGregor, "Two strategies for spare capacity placement in mesh restorable networks," IEEE International Conference on Communications (ICC), vol. 1, May 23-26, 1993, pp. 267 – 271.

[VHT02] S. Verbrugge, A. Hallez, G. De Tré, J. Verstraete, M. Pickavet, R. De Caluwe, P. Demeester, "Modelling of uncertain demands in optical network planning," Proceedings of the 7th European Conference on Networks & Optical Communications (NOC 2002), June 18-21, Darmstadt, Germany, 2002, pp. 29-36.

[VVS95] J. Veerasamy, S. Venkatesan, J. C. Shah, "Spare capacity assignment in telecom networks using path restoration," Proceedings of the Third International Workshop on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems, Jan. 18-20, 1995, pp. 370 – 374.

[WCB91] T.H. Wu, R.H. Cardwell, M. Boyden, "A Multi-Period Design Model for Survivable Network Architecture Selection for SONET Interoffice Networks," IEEE Transactions on Reliability, vol. 40, no. 4, October 1991, pp. 417-427.

[WCW88] T. H. Wu, R. H. Cardwell, W. E. Woodall, "Decreasing survivable fiber network cost using optical switches," IEEE Global Telecommunications Conference (GLOBECOM), vol.1, Nov. 28 - Dec. 1, 1988, pp. 93 – 97.

[WGT95] O. J. Wasem, A. M. Gross, G. A. Tlapa, "Forecasting broadband demand between geographic areas," IEEE Communications Magazine, vol. 33, no. 2, Feb. 1995, pp. 50 – 57.

[WIL63] T. Williams, "The Design of Survivable Communications Networks," IEEE Transactions on Communications, vol. 11, no. 2., June 1963, pp. 230 – 241.

[Yag73] B. Yaged, "Minimum cost routing for dynamic network models," Networks, Vol. 3, 1973, pp. 193-224.

[Yam95] J. Yamada, "A spare capacity design method for restorable networks," IEEE Global Telecommunications Conference (GLOBECOM), vol. 2, Nov. 13-17, 1995, pp. 931 – 935.

[YRL99] J. Yates, M. Rumsewicz, J. Lacey, "Wavelength converters in dynamically-reconfigurable WDM networks," IEEE Communication Surveys, Second Quarter 1999, pp. 2-15.

[ZaM01] H. Zang and B. Mukherjee, "Connection Management for Survivable Wavelength-Routed WDM Mesh Networks," Optical Networks Magazine, July/August 2001, pp. 17-28.

[Zad74] N. Zadeh, "On building minimum cost communication networks over time," Networks, Vol. 4, 1974, pp. 19-34.

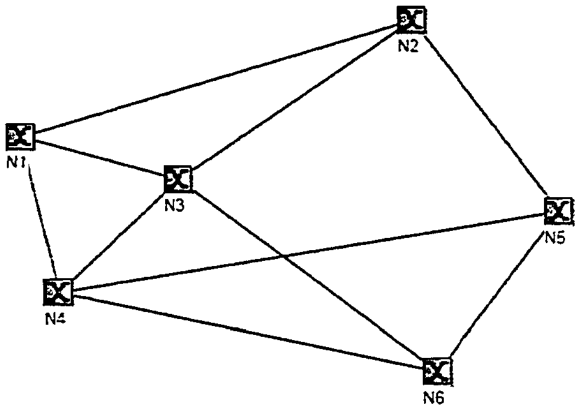
[ZJM00] H. Zang, J. P. Jue, B. Mukherjee, "A Review of Routing and Wavelength Assignment Approaches for Wavelength-routed Optical WDM Networks," Optical Network Magazine, vol. 1, Jan. 2000, pp. 47-60.

[ZJS01] H. Zang, J. P. Jue, L. Sahasrabuddhe, R. Ramamurthy, B. Mukherjee, "Dynamic lightpath establishment in wavelength routed WDM networks," IEEE Communications Magazine, vol. 39, issue 9, Sept. 2001, pp. 100 – 108.

Appendix A: Detailed Descriptions of Test Networks

We show all test topologies – namely, Metro, Germany, US and COST 239 networks – used in this thesis. Precise descriptions of each are also presented.

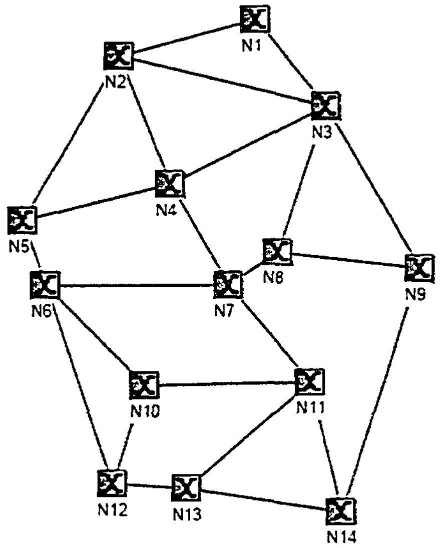
A.1 Metro Networks (6 nodes, 10 spans)



Node	X-coord.	Y-coord.	Nodal Size
N1	87	161	3
N2	400	70	3
N3	213	195	4
N4	116	283	4
N5	517	221	3
N6	420	346	3

Span	Origin	Destination	Length
S1	N1	N2	212
S2	N1	N3	121
S3	N2	N3	147
S4	N1	N4	125
S5	N3	N4	147
S6	N2	N5	226
S7	N3	N6	191
S8	N4	N6	237
S9	N5	N6	161
S10	N4	N5	346

A.2 Germany Networks (14 nodes, 24 spans)

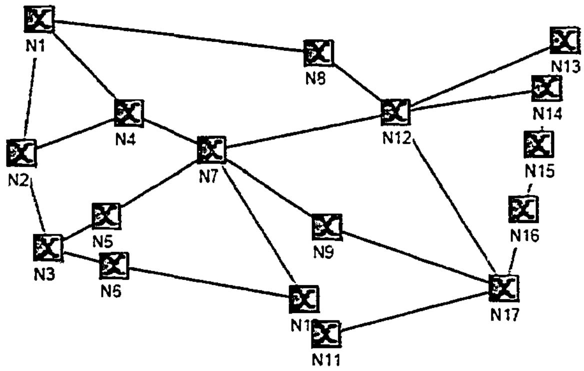


Node	X-coord.	Y-coord.	Nodal Size
N1	363	53	2
N2	261	80	4
N3	416	116	5
N4	299	174	4
N5	187	201	3
N6	205	248	4
N7	343	249	4
N8	379	224	3
N9	485	236	3
N10	279	322	3
N11	403	319	4
N12	255	394	3
N13	311	397	3
N14	427	414	3

Span	Origin	Destination	Length
S1	N1	N2	106
S2	N1	N3	82
S3	N2	N3	159
S4	N3	N4	131
S5	N2	N4	101
S6	N2	N5	142
S7	N4	N5	115
S8	N3	N9	138
S9	N3	N8	114
S10	N8	N9	107
S11	N8	N7	44

S12	N4	N7	87
S13	N5	N6	50
S14	N6	N7	138
S15	N6	N12	154
S16	N6	N10	105
S17	N10	N12	76
S18	N7	N11	92
S19	N10	N11	124
S20	N12	N13	56
S21	N11	N13	121
S22	N11	N14	98
S23	N13	N14	117
S24	N9	N14	187

A.3 US Networks (17 nodes, 24 spans)

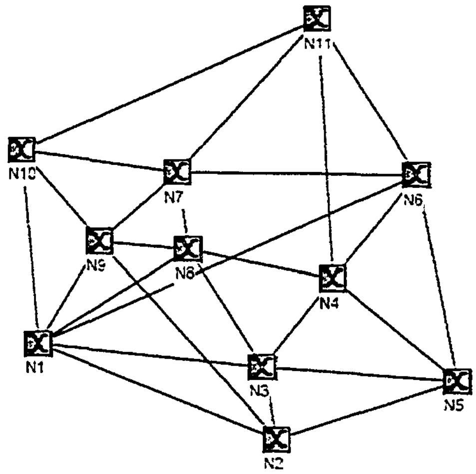


Node	X-coord.	Y-coord.	Nodal Size
N1	112	151	3
N2	95	273	3
N3	119	359	3
N4	195	236	3
N5	173	330	2
N6	181	375	2
N7	271	270	5
N8	370	181	2
N9	378	339	2
N10	357	405	3
N11	376	437	2
N12	441	235	5
N13	597	169	2
N14	580	213	3
N15	573	263	2
N16	560	323	2
N17	542	394	4

Span	Origin	Destination	Length
S1	N1	N2	123
S2	N1	N4	119
S3	N2	N4	107
S4	N2	N3	89
S5	N3	N5	61
S6	N3	N6	64
S7	N5	N7	115
S8	N4	N7	83
S9	N1	N8	260
S10	N8	N12	89
S11	N6	N10	179
S12	N10	N11	37
S13	N7	N10	160
S14	N7	N9	127

S15	N11	N17	171
S16	N9	N17	173
S17	N7	N12	174
S18	N12	N17	188
S19	N12	N13	169
S20	N12	N14	141
S21	N13	N14	47
S22	N14	N15	50
S23	N15	N16	61
S24	N16	N17	73

A.4 COST 239 Networks (11 nodes, 26 spans)



Node	X-coord.	Y-coord.	Nodal Size
N1	140	281	6
N2	315	350	4
N3	304	298	5
N4	356	235	5
N5	447	308	4
N6	417	161	5
N7	242	159	5
N8	250	214	5
N9	185	208	5
N10	128	143	4
N11	344	50	4

Span	Origin	Destination	Length
S1	N1	N2	820
S2	N1	N3	600
S3	N1	N6	1090
S4	N1	N8	400
S5	N1	N9	300
S6	N1	N10	450
S7	N2	N3	320
S8	N2	N5	820
S9	N2	N9	930
S10	N3	N4	565
S11	N3	N5	730
S12	N3	N8	350
S13	N4	N11	740
S14	N4	N5	320
S15	N4	N6	340
S16	N4	N8	730
S17	N5	N6	660
S18	N6	N11	390

S19	N6	N7	660
S20	N7	N11	760
S21	N7	N8	390
S22	N7	N9	210
S23	N7	N10	550
S24	N8	N9	220
S25	N9	N10	390
S26	N10	N11	1310

Appendix B: AMPL Formulations

B.1 Maximum Servability for Span Restorable Network (SR-MS Formulation)

```
# ---------------------------------------------------------------------
# Maximum Servability for Span Restorable Network Formulation
#
# November 30, 2001 by Dion Leung
# Copyright (C) 2001 TRLabs, Inc. All Rights Reserved.
# Model File:  SR-MS.mod
# ---------------------------------------------------------------------
# Given a capacitated capacity design and test demand scenario, this
# formulation (minimizes) maximizes the (un-)servability of a span-
# restorable network.
# ---------------------------------------------------------------------

# ---------------------------------------------------------------------
# SET
# ---------------------------------------------------------------------

# set of all spans:
set SPANS;

# set of all demands:
set DEMANDS;

# set of eligible working routes for each demand pair r:
set WORK_ROUTES{r in DEMANDS};

# set of eligible restoration routes for each span i:
set REST_ROUTES{i in SPANS};

# ---------------------------------------------------------------------
# PARAMETERS
# ---------------------------------------------------------------------

# number of existing capacity units on span j:
param Totalcap{j in SPANS};

# the test demand pattern:
param DemUnits{r in DEMANDS};

# equal to 1 if qth eligible working route for demand pair r crosses
span j, 0 otherwise:
param Zeta{j in SPANS, r in DEMANDS, q in WORK_ROUTES[r]} default 0;

# equal to 1 if pth restoration route for failure of span i uses span j
and 0 otherwise:
param Delta{i in SPANS, j in SPANS, p in REST_ROUTES[i]} default 0;

# ---------------------------------------------------------------------
# VARIABLES
# ---------------------------------------------------------------------
```

```
# total number of un-served demands for OD pair r
var unserved{r in DEMANDS} >= 0, <= 100000 integer;

# working flow required by qth working route for demand between node
pair r:
var workflow{r in DEMANDS, q in WORK_ROUTES[r], k in SCENARIOS} >=0,
<=200000 integer;

# restoration flow through pth restoration route for failure of span i:
var restflow{i in SPANS, p in REST_ROUTES[i], k in SCENARIOS} >=0,
<=200000 integer;

# number of spare links placed on span j:
var spare{j in SPANS} >=0, <=300000 integer;

# number of working links placed on span j:
var work{j in SPANS} >=0, <=300000 integer;

# -----------------------------------------------------------------------
# OBJECTIVE AND CONSTRAINTS
# -----------------------------------------------------------------------

minimize UNSERVABILITY:
     sum{r in DEMANDS} unserved[r];

subject to ROUTABILITY{r in DEMANDS}:
     sum{q in WORK_ROUTES[r]} workflow[r,q] = DemUnits[r] - unserved[r];

subject to WORKING_CAPS{j in SPANS}:
     sum{r in DEMANDS, q in WORK_ROUTES[r]} Zeta[j,r,q] * workflow[r,q]
     <= work[j];

subject to RESTORABILITY{i in SPANS}:
     sum{p in REST_ROUTES[i]} restflow[i,p] = work[i];

subject to SPARE_CAPS{i in SPANS, j in SPANS: i <> j}:
     spare[j] >= sum{p in REST_ROUTES[i]} Delta[i,j,p] * restflow[i,p];

subject to LIMITED_TOTAL_CAPS{i in SPANS}:
     work[i] + spare[i] <= Totcap[i];

subject to UNSERVED_UPPERBOUND{r in DEMANDS}:
     DemUnits[r] >= unserved[r];
```

B.2 Maximum Servability for SBPP Network (SBPP-MS Formulation)

```
# ----------------------------------------------------------------------
# Maximum Servability for Shared Backup Path Protected (SBPP) Network
# Formulation
#
# November 30, 2001 by Dion Leung
# Copyright (C) 2001 TRLabs, Inc. All Rights Reserved.
# Model File:  SBPP-MS.mod
# ----------------------------------------------------------------------
# Given a capacitated capacity design and test demand scenario, this
# formulation (minimizes) maximizes the (un-)servability of a SBPP
# network.
# ----------------------------------------------------------------------
# Additional Notes:
# (1) Each OD demand unit is routed over a single shortest working
#     route (from WORKING_ROUTES pathsets).
# (2) When there is any span failure, the affected OD pairs are
#     restored over their corresponding span-disjointed backup paths.
# ----------------------------------------------------------------------

# ----------------------------------------------------------------------
# SET
# ----------------------------------------------------------------------

# set of all spans:
set SPANS;

# set of all demands:
set DEMANDS;

# set of eligible span-disjoint backup routes for each demand pair r:
set BACKUP_ROUTES{r in DEMANDS};

# ----------------------------------------------------------------------
# PARAMETERS
# ----------------------------------------------------------------------

# number of existing capacity units on span j:
param Totalcap{j in SPANS};

# the test demand pattern:
param DemUnits{r in DEMANDS};

# equal to 1 if b-th backup path of the OD pair r uses span j, 0
otherwise
param Delta{r in DEMANDS, j in SPANS, b in BACKUP_ROUTES[r]} default 0;

# equal to 1 if the primary path of the OD pair r uses span i, 0
otherwise
param Beta{r in DEMANDS, i in SPANS} default 0;

# ----------------------------------------------------------------------
# VARIABLES
# ----------------------------------------------------------------------
```

```
# total number of un-served demands for OD pair r
var unserved{r in DEMANDS} >= 0, <= 100000 integer;

# equal to 1 if the backup path for demand r uses the b-th route, 0
otherwise
var backup_flow {r in DEMANDS, b in BACKUP_ROUTES[r]} >=0, <=1 integer;

# number of spare links placed on span j:
var spare{j in SPANS} >=0, <=300000 integer;

# number of working links placed on span j:
var work{j in SPANS} >=0, <=300000 integer;

# a fractional number which indicates the portion of demand units of OD
pair r that is being served and protected by the bth backup route
var restore_fraction {r in DEMANDS, b in BACKUP_ROUTES[r]} >=0, <=1;

# -------------------------------------------------------------------
# OBJECTIVE AND CONSTRAINTS
# -------------------------------------------------------------------

minimize UNSERVABILITY:
     sum{r in DEMANDS} unserved[r];

subject to WORKING_CAPS {j in SPANS}:
     sum{r in DEMANDS} Beta[r,j] * (DemUnits[r]-unserved[r]) <= work[j];

subject to SINGLE_BACKUP_ONLY {r in DEMANDS}:
     sum{b in BACKUP_ROUTES[r]} backup_flow[r,b] = 1;

subject to SPARE_CAPS {i in SPANS, j in SPANS: i <> j}:
     sum {r in DEMANDS, b in BACKUP_ROUTES[r]}
     Delta[r,j,b]* Beta[r,i] * DemUnits[r] * restore_fraction[r,b] <=
     spare[j];

subject to RESTORABILITY {r in DEMANDS, b in BACKUP_ROUTES[r]}:
     restore_fraction[r,b] >=
     (backup_flow[r,b] - unserved[r] /DemUnits[r]);

subject to LIMITED_TOTAL_CAPS {i in SPANS}:
     work[i] + spare [i] <= Total_caps[i];

subject to UNSERVED_UPPERBOUND {r in DEMANDS}:
     unserved[r] <= DemUnits[r];
```

B.3 Two-Part Span Restorable Capacity Design (TP-SR Formulation)

```
# -------------------------------------------------------------------
# Two-Part Span Restorable Capacity Design Formulation (based on
# Stochastic Programming framework)
#
# July 21, 2003 by Dion Leung
# Copyright (C) 2003 TRLabs, Inc. All Rights Reserved.
# Model File:  TP-SR.mod
# -------------------------------------------------------------------
# Given a set of discrete forecast scenarios, this formulation
# minimizes the initial design cost plus the expected recourse
# costs to cope with the demands.
# -------------------------------------------------------------------
# Additional Notes:
# (1) Nominal demand scenario is denoted as "01".
# (2) RecourseCost[j] should always be relative to the present Cost[j].
# -------------------------------------------------------------------

# -------------------------------------------------------------------
# SET
# -------------------------------------------------------------------

# set of all spans
set SPANS;

# set of all restoration paths for each span failure:
set REST_ROUTES{i in SPANS};

# set of all demand pairs or node pairs:
set DEMANDS;

# set of all working routes for each demand pair r:
set WORK_ROUTES{r in DEMANDS};

# set of possible forecast scenarios:
set SCENARIOS;

# -------------------------------------------------------------------
# PARAMETERS
# -------------------------------------------------------------------

# cost of span j:
param Cost{j in SPANS};

# recourse cost for each span j, relative to Cost[j]:
param RecourseCost{j in SPANS};

# number of demand units between OD pair r in forecast scenario k:
param DemUnits{k in SCENARIOS, r in DEMANDS};

# equal 1 if pth restoration route for failure of span i uses span j, 0
otherwise:
param Delta{i in SPANS, j in SPANS, p in REST_ROUTES[i]} default 0;
```

```
# equal 1 if qth working route for demand between node pair r uses span
j, 0 otherwise:
param Zeta{j in SPANS, r in DEMANDS, q in WORK_ROUTES[r]} default 0;

# probability estimate of each demand scenario k:
param Prob{k in SCENARIOS};

# ---------------------------------------------------------------------
# VARIABLES
# ---------------------------------------------------------------------

# number of working links placed on span j in the present design:
var work{j in SPANS} >=0, <=300000 integer;

# number of spare links placed on span j in the present design:
var spare{j in SPANS} >=0, <=300000 integer;

# working capacity required by qth working route for demand between
node pair r:
var workflow{r in DEMANDS, q in WORK_ROUTES[r], k in SCENARIOS} >=0,
<=200000 integer;

# restoration flow through pth restoration route for failure of span i:
var restflow{i in SPANS, p in REST_ROUTES[i], k in SCENARIOS} >=0,
<=200000 integer;

# additional working capacity needed to support demand scenario k:
var extrawork{j in SPANS, k in SCENARIOS} >=0, <=300000 integer;

# additional spare capacity needed to support demand scenario k:
var extraspare{j in SPANS, k in SCENARIOS} >=0, <=300000 integer;

# total capacity of span j in the present design:
var totalcap{j in SPANS};

# total cost of the present or initial capacity design:
var initialcost;

# total expected recourse cost required to cope with all scenarios:
var totalrecourse;

# total initial plus future expected cost (objective value):
var totalcost;

# ---------------------------------------------------------------------
# OBJECTIVE AND CONSTRAINTS
# ---------------------------------------------------------------------

minimize INITIALplusEXPECTED_FUTURE_COST:
     sum{j in SPANS} Cost[j] * ( work[j] + spare[j] ) +
     sum{j in SPANS, k in SCENARIOS} Prob[k] * RecourseCost[j] *
     ( extrawork[j,k] + extraspare[j,k] );

subject to ROUTABILITY {r in DEMANDS, k in SCENARIOS}:
     sum{q in WORK_ROUTES[r]} workflow[r,q,k] = DemUnits[k,r];

subject to WORKING_CAPS {j in SPANS, k in SCENARIOS}:
```

```
     work[j] + extrawork[j,k] >= sum{r in DEMANDS, q in WORK_ROUTES[r]}
     Zeta[j,r,q] * workflow[r,q,k];

subject to RESTORABILITY {i in SPANS, k in SCENARIOS}:
     sum{p in REST_ROUTES[i]} restflow[i,p,k] =
     work[i] + extrawork[i,k];

subject to SPARE_CAPS {k in SCENARIOS, i in SPANS, j in SPANS: i <> j}:
     spare[j] + extraspare[j,k] >=
     sum{p in REST_ROUTES[i]} Delta[i,j,p] * restflow[i,p,k];

subject to SERVE_NOMINAL_SPARE {j in SPANS, k in SCENARIOS}:
     extraspare[j,"O1"] = 0;

subject to SERVE_NOMINAL_WORK {j in SPANS, k in SCENARIOS}:
     extrawork[j,"O1"] = 0;

subject to EVALUATE_TOTALCAP {j in SPANS}:
     totalcap[j] = work[j] + spare[j];

subject to EVALUATE_INITIAL_COST:
     initialcost = sum{j in SPANS} Cost[j] * ( work[j] + spare[j] );

subject to EVALUATE_TOTALRECOURSE:
     totalrecourse = sum{j in SPANS, k in SCENARIOS} Prob[k] *
     RecourseCost[j] * ( extrawork[j,k] + extraspare[j,k] );

subject to EVALUATE_OVERALL_COST:
     totalcost = sum{j in SPANS} Cost[j] * ( work[j] + spare[j] ) +
     sum{j in SPANS, k in SCENARIOS} Prob[k] * RecourseCost[j] *
     ( extrawork[j,k] + extraspare[j,k] );
```

B.4 Two-Part Span Restorable Modular Capacity Design (TP-MSR Formulation)

```
# ----------------------------------------------------------------------
# Two-Part Span Restorable MODULAR Capacity Design Formulation (based
# on Stochastic Programming framework)
#
# July 21, 2003 by Dion Leung
# Copyright (C) 2003 TRLabs, Inc. All Rights Reserved.
# Model File:  TP-MSR.mod
# ----------------------------------------------------------------------
# Given a set of discrete forecast scenarios, this formulation
# minimizes the initial design cost plus the expected recourse
# costs to cope with the demands.  Modularity and economy-of-scale
# effects are considered.
# ----------------------------------------------------------------------
# Additional Notes:
# (1) Nominal demand scenario is denoted as "O1".
# (2) RecourseCost[m,j] should always be relative to the present
#     modular Cost[m,j].
# ----------------------------------------------------------------------

# ----------------------------------------------------------------------
# SET
# ----------------------------------------------------------------------

# set of all spans
set SPANS;

# set of all restoration paths for each span failure:
set REST_ROUTES{i in SPANS};

# set of all demand pairs or node pairs:
set DEMANDS;

# set of all working routes for each demand pair r:
set WORK_ROUTES{r in DEMANDS};

# set of possible forecast scenarios:
set SCENARIOS;

# set of capacity modules:
set MODULES;

# ----------------------------------------------------------------------
# PARAMETERS
# ----------------------------------------------------------------------

# cost of each module type m on span j:
param Cost{m in MODULES, j in SPANS};

# recourse cost of each module m on span j:
param RecourseCost{m in MODULES, j in SPANS};

# maximum capacity supported by each module type m:
param ModType{m in MODULES};
```

```
# number of demand units between OD pair r in forecast scenario k:
param DemUnits{k in SCENARIOS, r in DEMANDS};

# equal 1 if pth restoration route for failure of span i uses span j, 0
otherwise:
param Delta{i in SPANS, j in SPANS, p in REST_ROUTES[i]} default 0;

# equal 1 if qth working route for demand between node pair r uses span
j, 0 otherwise:
param Zeta{j in SPANS, r in DEMANDS, q in WORK_ROUTES[r]} default 0;

# probability estimate of each demand scenario k:
param Prob{k in SCENARIOS};

# --------------------------------------------------------------------
# VARIABLES
# --------------------------------------------------------------------

# number of working links placed on span j in the present design:
var work{j in SPANS} >=0, <=300000 integer;

# number of spare links placed on span j in the present design:
var spare{j in SPANS} >=0, <=300000 integer;

# working capacity required by qth working route for demand between
node pair r:
var workflow{r in DEMANDS, q in WORK_ROUTES[r], k in SCENARIOS} >=0,
<=200000 integer;

# restoration flow through pth restoration route for failure of span i:
var restflow{i in SPANS, p in REST_ROUTES[i], k in SCENARIOS} >=0,
<=200000 integer;

# number of modules of type m placed on span j in the present design:
var modules{j in SPANS, m in MODULES} >=0, <=10000 integer;

# additional working capacity needed to support demand scenario k:
var extrawork{j in SPANS, k in SCENARIOS} >=0, <=300000 integer;

# additional spare capacity needed to support demand scenario k:
var extraspare{j in SPANS, k in SCENARIOS} >=0, <=300000 integer;

# additional modules of type m placed on span j needed to support
demand scenario k:
var extramodules{k in SCENARIOS, j in SPANS, m in MODULES} >=0, <=10000
integer;

# total cost of the present (initial) design:
var initialcost;

# total expected recourse cost to cope with all scenarios:
var totalrecourse;

# total initial plus future expected cost (objective value):
var totalcost;
```

```
# ----------------------------------------------------------------------
# OBJECTIVE AND CONSTRAINTS
# ----------------------------------------------------------------------

minimize INITIALplusEXPECTED_FUTURE_COST:
     sum{m in MODULES, j in SPANS} Cost[m,j] * modules[j,m] +
     sum{m in MODULES, j in SPANS, k in SCENARIOS} Prob[k] *
     RecourseCost[m,j] * extramodules[k,j,m];

subject to ROUTABILITY {r in DEMANDS, k in SCENARIOS}:
     sum{q in WORK_ROUTES[r]} workflow[r,q,k] = DemUnits[k,r];

subject to WORKING_CAPS {j in SPANS, k in SCENARIOS}:
     work[j] + extrawork[j,k] >= sum{r in DEMANDS, q in WORK_ROUTES[r]}
     Zeta[j,r,q] * workflow[r,q,k];

subject to RESTORABILITY {i in SPANS, k in SCENARIOS}:
     sum{p in REST_ROUTES[i]} restflow[i,p,k] =
     work[i] + extrawork[i,k];

subject to SPARE_CAPS {k in SCENARIOS, i in SPANS, j in SPANS: i <> j}:
     spare[j] + extraspare[j,k] >=
     sum{p in REST_ROUTES[i]} Delta[i,j,p] * restflow[i,p,k];

subject to MODULARITY{j in SPANS}:
     spare[j] + work[j] <= sum{m in MODULES} ModType[m] * modules[j,m];

subject to MODULARITY_FOR_EXTRACAP{j in SPANS, k in SCENARIOS}:
     extraspare[j,k] + extrawork[j,k] <=
     sum{m in MODULES} ModType[m] * extramodules[k,j,m];

subject to SERVE_NOMINAL_CONSTRAINT {k in SCENARIOS, j in SPANS, m in
MODULES}:
     extramodules["O1",j,m] = 0;

subject to EVALUATE_INITIALCOST:
     initialcost =
     sum{m in MODULES, j in SPANS} Cost[m,j] * modules[j,m];

subject to EVALUATE_TOTALRECOURSE:
     totalrecourse = sum{m in MODULES, j in SPANS, k in SCENARIOS}
     Prob[k] * RecourseCost[m,j] * extramodules[k,j,m];

subject to EVALUATE_TOTAL_COST:
     totalcost = sum{m in MODULES, j in SPANS} Cost[m,j] * modules[j,m]+
     sum{m in MODULES, j in SPANS, k in SCENARIOS} Prob[k] *
     RecourseCost[m,j] * extramodules[k,j,m];
```

B.5 Two-Part *p*-Cycles Capacity Design (TP-PC Formulation)

```
# -------------------------------------------------------------------
# Two-Part p-Cycles Capacity Design Formulation (based on Stochastic
# Programming framework)
#
# May 19, 2004 by Dion Leung
# Copyright (C) 2004 TRLabs, Inc. All Rights Reserved.
# Model File:  TP-PC.mod
# -------------------------------------------------------------------
# Given a set of discrete forecast scenarios, this formulation
# minimizes the initial design cost plus the expected recourse
# costs to cope with the demands.
# -------------------------------------------------------------------
# Additional Notes:
# (1) Nominal demand scenario is denoted as "O1".
# (2) RecourseCost[j] should always be relative to the present Cost[j].
# -------------------------------------------------------------------

# -------------------------------------------------------------------
# SET
# -------------------------------------------------------------------

# set of all spans
set SPANS;

# set of eligible cycles for protection:
set PCYCLES;

# set of all demand pairs or node pairs:
set DEMANDS;

# set of all working routes for each demand pair r:
set WORK_ROUTES{r in DEMANDS};

# set of possible forecast scenarios:
set SCENARIOS;

# -------------------------------------------------------------------
# PARAMETERS
# -------------------------------------------------------------------

# cost of span j:
param Cost{j in SPANS};

# recourse cost for each span j, relative to Cost[j]:
param RecourseCost{j in SPANS};

# number of demand units between OD pair r in forecast scenario k:
param DemUnits{k in SCENARIOS, r in DEMANDS};

# equal 1 if qth working route for demand between node pair r uses span
j, 0 otherwise:
param Zeta{j in SPANS, r in DEMANDS, q in WORK_ROUTES[r]} default 0;
```

```
# equal 2 if the failed span i is a straddling span, 1 if it is an on-
cycle span, 0 if it has no relationship to the eligible p-cycles:
param Xpi{p in PCYCLES, i in SPANS} default 0;

# equal 1 if p-cycle p uses span j, 0 otherwise. i.e., if Xpi[p,j] = 1,
then p-cycle p crosses span j:
param pCrossesj{p in PCYCLES, j in SPANS} = if Xpi[p,j] = 1 then 1 else
0;

# probability estimate of each demand scenario k:
param Prob{k in SCENARIOS};

# ---------------------------------------------------------------------
# VARIABLES
# ---------------------------------------------------------------------

# number of working links placed on span j in the present design:
var work{j in SPANS} >=0, <=300000 integer;

# number of spare links placed on span j in the present design:
var spare{j in SPANS} >=0, <=300000 integer;

# working capacity required by qth working route for demand between
node pair r:
var workflow{r in DEMANDS, q in WORK_ROUTES[r], k in SCENARIOS} >=0,
<=200000 integer;

# copies of cycle p used for protecting demands in scenario k:
var p_cycle_usage{p in PCYCLES, k in SCENARIOS} >=0 integer, <=100000;

# additional working capacity needed to support demand scenario k:
var extrawork{j in SPANS, k in SCENARIOS} >=0, <=300000 integer;

# additional spare capacity needed to support demand scenario k:
var extraspare{j in SPANS, k in SCENARIOS} >=0, <=300000 integer;

# total capacity of span j in the present design:
var totalcap{j in SPANS};

# total cost of the present or initial capacity design:
var initialcost;

# total expected recourse cost required to cope with all scenarios:
var totalrecourse;

# total initial plus future expected cost (objective value):
var totalcost;

# ---------------------------------------------------------------------
# OBJECTIVE AND CONSTRAINTS
# ---------------------------------------------------------------------

minimize INITIALplusEXPECTED_FUTURE_COST:
     sum{j in SPANS} Cost[j] * ( work[j] + spare[j] ) +
     sum{j in SPANS, k in SCENARIOS} Prob[k] * RecourseCost[j] *
     ( extrawork[j,k] + extraspare[j,k] );
```

```
subject to ROUTABILITY {r in DEMANDS, k in SCENARIOS}:
     sum{q in WORK_ROUTES[r]} workflow[r,q,k] = DemUnits[k,r];

subject to WORKING_CAPS {j in SPANS, k in SCENARIOS}:
     work[j] + extrawork[j,k] >= sum{r in DEMANDS, q in WORK_ROUTES[r]}
     Zeta[j,r,q] * workflow[r,q,k];

subject to RESTORABILITY {i in SPANS, k in SCENARIOS}:
     sum{p in PCYCLES} Xpi[p,i] * p_cycle_usage[p,k] >=
     work[i] + extrawork[i,k];

subject to SPARE_CAPS {k in SCENARIOS, i in SPANS, j in SPANS: i <> j}:
     spare[j] + extraspare[j,k] >=
     sum{p in PCYCLES} pCrossesj[p,j] * p_cycle_usage[p,k];

subject to SERVE_NOMINAL_SPARE {j in SPANS, k in SCENARIOS}:
     extraspare[j,"O1"] = 0;

subject to SERVE_NOMINAL_WORK {j in SPANS, k in SCENARIOS}:
     extrawork[j,"O1"] = 0;

subject to EVALUATE_TOTALCAP {j in SPANS}:
     totalcap[j] = work[j] + spare[j];

subject to EVALUATE_INITIAL_COST:
     initialcost = sum{j in SPANS} Cost[j] * ( work[j] + spare[j] );

subject to EVALUATE_TOTALRECOURSE:
     totalrecourse = sum{j in SPANS, k in SCENARIOS} Prob[k] *
     RecourseCost[j] * ( extrawork[j,k] + extraspare[j,k] );

subject to EVALUATE_OVERALL_COST:
     totalcost = sum{j in SPANS} Cost[j] * ( work[j] + spare[j] ) +
     sum{j in SPANS, k in SCENARIOS} Prob[k] * RecourseCost[j] *
     ( extrawork[j,k] + extraspare[j,k] );
```

B.6 Maximum-Profit Multi-QoP Demand Loading (MP-QoP-DL Formulation)

```
# ------------------------------------------------------------------------
# Maximum Profit demand loading over span-restorable mesh with multi-
# QoP demand services
#
# October 1, 2003 by Dion Leung
# Copyright (C) 2003 TRLabs, Inc. All Rights Reserved.
# Model File:  MP-QoP-DL.mod
# ------------------------------------------------------------------------
# Given a capacitated network, cost and revenue models, this
# formulation selects the specific demands such that they generate the
# greatest net profit(or greatest revenue if cost is ignored).
# ------------------------------------------------------------------------
# Additional Notes:
# Multi-QoP are considered.
# "protected" services - both 100% routed and restorable
# "unprotected" services - only routed but not protected
# "preemptible" services - use spares for demands, and preempt for
# restoring Protected Services
# ------------------------------------------------------------------------

# ------------------------------------------------------------------------
# SET
# ------------------------------------------------------------------------

# set of all spans:
set SPANS;

# set of all demands:
set DEMANDS;

# set of adjacent demands:
set ADJ_DEMANDS within DEMANDS;

# set of non-adjacent demands:
set NON_ADJ_DEMANDS := DEMANDS diff ADJ_DEMANDS;

# set of multi-QoP service classes:
set CLASSES;

# set of eligible working routes for each demand pair r:
set WORK_ROUTES{r in DEMANDS};

# set of eligible restoration routes for each span i:
set REST_ROUTES{i in SPANS};

# ------------------------------------------------------------------------
# PARAMETERS
# ------------------------------------------------------------------------

# the relative revenue for a particular demand pair r of class c:
param Revenue{r in DEMANDS, c in CLASSES};

# the cost of provisioning a channel on span r:
param Cost{i in SPANS};
```

```
# the shortest path distance between demand pair r:
param Distance{r in DEMANDS};

# equal to 1 if r is an adjacent demand pair, 0 otherwise:
param Adjacency{r in DEMANDS} default 0;

# the distance of each span r:
param SpanLength{i in SPANS};

# the desired demand requests to be served:
param DemandList{r in DEMANDS, c in CLASSES};

# number of existing capacity units on span j:
param Totalcap{j in SPANS};

# equal to 1 if qth eligible working route for demand pair r crosses
span j, 0 otherwise:
param Zeta{j in SPANS, r in DEMANDS, q in WORK_ROUTES[r]} default 0;

# equal to 1 if pth restoration route for failure of span i uses span j
and 0 otherwise:
param Delta{i in SPANS, j in SPANS, p in REST_ROUTES[i]} default 0;

# ----------------------------------------------------------------------
# VARIABLES
# ----------------------------------------------------------------------

# number of demand units for demand pair r:
var demandselected{r in DEMANDS, c in CLASSES} >= 0, <= 10000000
integer;

# working flow on route q for demand service r of class c:
var workflow{r in DEMANDS, c in CLASSES, q in WORK_ROUTES[r]} >= 0,
<=100000 integer;

# flow on route p for restoration of span i:
var restflow{i in SPANS, p in REST_ROUTES[i]} >=0, <=2500000 integer;

# number of working capacity units used by a service class c on span j:
var work{j in SPANS, c in CLASSES} >=0, <=250000 integer;

# number of total working capacity units on span j:
var totalwork{j in SPANS} >=0, <=250000 integer;

# number of spare capacity units on span j:
var spare{j in SPANS} >=0, <=250000 integer;

# the operational cost for provisioning the demands:
var opex >= 0;
```

```
# ----------------------------------------------------------------------
# OBJECTIVE AND CONSTRAINTS
# ----------------------------------------------------------------------

maximize PROFIT:
     sum{r in DEMANDS, c in CLASSES} Revenue[r,c] * demandselected[r,c]
     - opex;

subject to WORKING_FLOW{r in DEMANDS, c in CLASSES}:
     sum{q in WORK_ROUTES[r]} workflow[r,c,q] = demandselected[r,c];

subject to WORKING_CAPS{j in SPANS, c in CLASSES}:
     work[j,c] = sum{r in DEMANDS, q in WORK_ROUTES[r]} Zeta[j,r,q] *
     workflow[r,c,q];

subject to TOTAL_WORKING_CAPS{j in SPANS}:
     totalwork[j] = sum{c in CLASSES} work[j,c];

subject to RESTORATION_FLOW_FOR_PROTECTED{i in SPANS}:
     sum{p in REST_ROUTES[i]} restflow[i,p] = work[i,"protected"];

subject to USE_SPARE_CAPS_WITH_PREEMPT{i in SPANS, j in SPANS: i<>j}:
     spare[j] + work[j,"preemptible"] >= sum{p in REST_ROUTES[i]}
     Delta[i,j,p] * restflow[i,p];

subject to TOTAL_CAPS{j in SPANS}:
     totalwork[j] + spare[j] <= Totalcap[j];

subject to DEMAND_UPPERBOUND{r in DEMANDS, c in CLASSES}:
     demandselected[r,c] <= DemandList[r,c];

subject to TOTAL_OPERATING_COST:
     sum{j in SPANS} Cost[j]*(totalwork[j] + spare[j]) <= opex;
```

B.7 Maximum-Fairness Re-optimization Model (Max-Fair Formulation)

```
# -------------------------------------------------------------------
# Max-Fair Re-optimization Formulation for Span-restorable Networks
#
# July 28, 2004 by Dion Leung (at Osaka University, Osaka, Japan)
# Copyright (C) 2004 TRLabs, Inc. All Rights Reserved.
# Model File:  Max-Fair.mod
# -------------------------------------------------------------------
# Given a capacitated span-restorable network and an existing capacity
# configuration (comprised of demands already in service, spare
# channels pre-planned for span failure protection, and a remaining set
# of equipped but unused channels), this formulation re-configures
# existing (working and/or restoration) routes such that the new
# configuration has a better ability to serve incremental future
# demands (tested in a separate step).
# -------------------------------------------------------------------

# -------------------------------------------------------------------
# SET
# -------------------------------------------------------------------

# set of all spans
set SPANS;

# set of all demand pairs or node pairs:
set DEMANDS;

# set of all working routes for each demand pair r:
set WORK_ROUTES{r in DEMANDS};

# set of all restoration paths for each span failure:
set REST_ROUTES{i in SPANS};

# -------------------------------------------------------------------
# PARAMETERS
# -------------------------------------------------------------------

# total as-built capacity units on span j:
param Totalcap{j in SPANS} default 40;

# existing or in-service demand on node pair r:
param DemUnits{r in DEMANDS};

# equal to 1 if q-th eligible working route for demand pair r crosses
span j, 0 otherwise:
param Zeta{j in SPANS, r in DEMANDS, q in WORK_ROUTES[r]} default 0;

# equal to 1 if p-th restoration route for failure of span i uses span
j and 0 otherwise:
param Delta{i in SPANS, j in SPANS, p in REST_ROUTES[i]} default 0;
```

```
# ---------------------------------------------------------------------
# VARIABLES FOR EXISTING DEMANDS
# ---------------------------------------------------------------------

# working flow assigned on the q-th working route to serve OD pair r in
# existing demands (Note: these become input parameters if
# re-arrangement of working paths is not allowed:
var existing_workflow{r in DEMANDS, q in WORK_ROUTES[r]} >=0, <=10000
integer;

# restoration flow assigned on the p-th restoration route upon the
failure of span i in existing demands:
var existing_restflow{i in SPANS, p in REST_ROUTES[i]} >=0, <=2500000
integer;

# number of working capacity units on span j to support existing
demands:
var existing_work{j in SPANS} >=0, <=250000 integer;

# number of spare capacity units on span j to support existing demands:
var existing_spare{j in SPANS} >=0, <=2500000 integer;

# ---------------------------------------------------------------------
# VARIABLES FOR PROJECTED DEMANDS
# ---------------------------------------------------------------------

# projected demands that could be served on node pair r:
var future_demand{r in DEMANDS} >=0, <=10000 integer;

# working flow to be assigned on the q-th working route to serve OD
pair r for projected demands:
var future_workflow{r in DEMANDS, q in WORK_ROUTES[r]} >=0, <=10000
integer;

# restoration flow to be assigned on the p-th restoration route upon
the failure of span i for projected demands:
var future_restflow{i in SPANS, p in REST_ROUTES[i]} >=0, <=2500000
integer;

# number of idle working capacity units allocated on span j to support
projected demands:
var future_work{j in SPANS} >=0, <=250000 integer;

# number of idle spare capacity units allocated on span j to support
projected demands:
var future_spare{j in SPANS} >=0, <=2500000 integer;

# largest possible number of demand units could be served uniformly on
# all OD pair:
var lambda >=0, <=1000 integer;
```

```
# ------------------------------------------------------------------
# OBJECTIVE AND CONSTRAINTS
# ------------------------------------------------------------------

maximize MAX_FAIRNESS:  lambda;

subject to Demand_Routing_for_Existing_Demands {r in DEMANDS}:
     sum{q in WORK_ROUTES[r]} existing_workflow[r,q] = DemUnits[r];

subject to Demand_Routing_for_Future_Demands {r in DEMANDS}:
     sum{q in WORK_ROUTES[r]} future_workflow[r,q] = future_demand[r];

subject to Working_Capacity_for_Existing_Demands {j in SPANS}:
     sum{r in DEMANDS, q in WORK_ROUTES[r]} Zeta[j,r,q] *
     existing_workflow[r,q] <= existing_work[j];

subject to Working_Capacity_for_Future_Demands {j in SPANS}:
     sum{r in DEMANDS, q in WORK_ROUTES[r]} Zeta[j,r,q] *
     future_workflow[r,q] <= future_work[j];

subject to Restoration_for_Existing_Demands {i in SPANS}:
     sum{p in REST_ROUTES[i]} existing_restflow[i,p] = existing_work[i];

subject to Restoration_for_Future_Demands {i in SPANS}:
     sum{p in REST_ROUTES[i]} future_restflow[i,p] = future_work[i];

subject to Spare_Capacity_for_Existing_Demands {i in SPANS, j in SPANS:
i <> j}:
     existing_spare[j] >= sum{p in REST_ROUTES[i]} Delta[i,j,p] *
     existing_restflow[i,p];

subject to Spare_Capacity_for_Future_Demands {i in SPANS, j in SPANS: i
<> j}:
     future_spare[j] >= sum{p in REST_ROUTES[i]} Delta[i,j,p] *
     future_restflow[i,p];

subject to Total_Capacity {i in SPANS}:
     existing_work[i] + future_work[i] + existing_spare[i] +
     future_spare[i] <= Totalcap[i];

subject to Demand_Fairness {r in DEMANDS}:
     future_demand[r] >= lambda;
```

B.8 Maximum-Volume Re-optimization Model (Max-Vol Formulation)

```
# ----------------------------------------------------------------------
# Max-Volume Re-optimization Formulation for Span-Restorable Networks
#
# July 28, 2004 by Dion Leung (at Osaka University, Osaka, Japan)
# Copyright (C) 2004 TRLabs, Inc. All Rights Reserved.
# Model File:  Max-Vol.mod
# ----------------------------------------------------------------------
# Given a capacitated span-restorable network and an existing capacity
# configuration (comprised of demands already in service, spare
# channels pre-planned for span failure protection, and a remaining set
# of equipped but unused channels), this formulation re-configures
# existing (working and/or restoration) routes such that the new
# configuration has a better ability to serve incremental future
# demands (tested in a separate step).
# ----------------------------------------------------------------------

# ----------------------------------------------------------------------
# SET
# ----------------------------------------------------------------------

# set of all spans
set SPANS;

# set of all demand pairs or node pairs:
set DEMANDS;

# set of all working routes for each demand pair r:
set WORK_ROUTES{r in DEMANDS};

# set of all restoration paths for each span failure:
set REST_ROUTES{i in SPANS};

# ----------------------------------------------------------------------
# PARAMETERS
# ----------------------------------------------------------------------

# total as-built capacity units on span j:
param Totalcap{j in SPANS} default 40;

# existing or in-service demand on node pair r:
param DemUnits{r in DEMANDS};

# equal to 1 if q-th eligible working route for demand pair r crosses
span j, 0 otherwise:
param Zeta{j in SPANS, r in DEMANDS, q in WORK_ROUTES[r]} default 0;

# equal to 1 if p-th restoration route for failure of span i uses span
j and 0 otherwise:
param Delta{i in SPANS, j in SPANS, p in REST_ROUTES[i]} default 0;

# weight factor for prioritizing OD pairs r [optional]:
param DemandPriority{r in DEMANDS} default 1;
```

```
# ---------------------------------------------------------------------
# VARIABLES FOR EXISTING DEMANDS
# ---------------------------------------------------------------------

# working flow assigned on the q-th working route to serve OD pair r in
# existing demands (Note: these become input parameters if
# re-arrangement of working paths is not allowed:
var existing_workflow{r in DEMANDS, q in WORK_ROUTES[r]} >=0, <=10000
integer;

# restoration flow assigned on the p-th restoration route upon the
failure of span i in existing demands:
var existing_restflow{i in SPANS, p in REST_ROUTES[i]} >=0, <=2500000
integer;

# number of working capacity units on span j to support existing
demands:
var existing_work{j in SPANS} >=0, <=250000 integer;

# number of spare capacity units on span j to support existing demands:
var existing_spare{j in SPANS} >=0, <=2500000 integer;

# ---------------------------------------------------------------------
# VARIABLES FOR PROJECTED DEMANDS
# ---------------------------------------------------------------------

# projected demands that could be served on node pair r:
var future_demand{r in DEMANDS} >=0, <=10000 integer;

# working flow to be assigned on the q-th working route to serve OD
pair r for projected demands:
var future_workflow{r in DEMANDS, q in WORK_ROUTES[r]} >=0, <=10000
integer;

# restoration flow to be assigned on the p-th restoration route upon
the failure of span i for projected demands:
var future_restflow{i in SPANS, p in REST_ROUTES[i]} >=0, <=2500000
integer;

# number of idle working capacity units allocated on span j to support
projected demands:
var future_work{j in SPANS} >=0, <=250000 integer;

# number of idle spare capacity units allocated on span j to support
projected demands:
var future_spare{j in SPANS} >=0, <=2500000 integer;
```

```
# ----------------------------------------------------------------------
# OBJECTIVE AND CONSTRAINTS
# ----------------------------------------------------------------------

maximize MAX_VOLUME:
     sum{r in DEMANDS} DemandPriority[r] * future_demand[r];

subject to Demand_Routing_for_Existing_Demands {r in DEMANDS}:
     sum{q in WORK_ROUTES[r]} existing_workflow[r,q] = DemUnits[r];

subject to Demand_Routing_for_Future_Demands {r in DEMANDS}:
     sum{q in WORK_ROUTES[r]} future_workflow[r,q] = future_demand[r];

subject to Working_Capacity_for_Existing_Demands {j in SPANS}:
     sum{r in DEMANDS, q in WORK_ROUTES[r]} Zeta[j,r,q] *
     existing_workflow[r,q] <= existing_work[j];

subject to Working_Capacity_for_Future_Demands {j in SPANS}:
     sum{r in DEMANDS, q in WORK_ROUTES[r]} Zeta[j,r,q] *
     future_workflow[r,q] <= future_work[j];

subject to Restoration_for_Existing_Demands {i in SPANS}:
     sum{p in REST_ROUTES[i]} existing_restflow[i,p] = existing_work[i];

subject to Restoration_for_Future_Demands {i in SPANS}:
     sum{p in REST_ROUTES[i]} future_restflow[i,p] = future_work[i];

subject to Spare_Capacity_for_Existing_Demands {i in SPANS, j in SPANS:
i <> j}:
     existing_spare[j] >= sum{p in REST_ROUTES[i]} Delta[i,j,p] *
     existing_restflow[i,p];

subject to Spare_Capacity_for_Future_Demands {i in SPANS, j in SPANS: i
<> j}:
     future_spare[j] >= sum{p in REST_ROUTES[i]} Delta[i,j,p] *
     future_restflow[i,p];

subject to Total_Capacity {i in SPANS}:
     existing_work[i] + future_work[i] + existing_spare[i] +
     future_spare[i] <= Totalcap[i];
```

Appendix C: Two-Part p-Cycles (TP-PC) Capacity Design

In Section 6.5, we presented the two-part span restorable (TP-SR) capacity design formulation. The same two-part concept can be applied to p-cycles capacity design, which we now demonstrate. For the complete set of results and explanation of the optimal p-cycles design model, readers can refer to [LeG04b]. Of particular note is that the set $\boldsymbol{P}$, parameters $\boldsymbol{x_{p,j}}$ and $\boldsymbol{\delta_{p,j}}$, and the variable $\boldsymbol{n_{p,k}}$ are replacing P_i, $\delta_{i,j}^{p}$ and $f_{i,k}^{p}$ in TP-PC formulation from Section 6.5.1. The constraint sets (C.4) and (C.5) are substituted for (6.5.4) and (6.5.5) to reflect the restorability restriction by p-cycles.

Sets:

- S Set of all spans in the network, indexed by j or i
- U Set of all possible future demand scenarios to be considered, index k
- D Set of all origin-destination (OD) pairs in a demand matrix, index r
- Q^r Set of pre-determined eligible working routes for OD pair r, index q
- $\boldsymbol{P}$ Set of pre-determined eligible cycles available upon the failure of span i, index p

Parameters:

- C_j Present cost of a unit capacity placed on span j
- R_j Recourse cost of placing an extra unit capacity on span j to cope with the unfolding of demand uncertainty. R_j can simply be a multiplicative value of C_j, or any other absolute value specific for each span j
- P_k Probability estimate for demand scenario k
- d_k^r Magnitude of the bi-directional (integer) demand on node pair r in scenario k
- $\zeta_j^{r,q}$ Equal to one if the q^{th} eligible route for demands between node pair r uses span j, zero otherwise
- $x_{p,j}$ Equal to one if failed span i is part of the cycle p; equal to two if span i straddles cycle p; equal to zero if there is no relationship between span i and cycle p.
- $\delta_{p,j}$ Equal to one if if cycle p passes over span j; zero otherwise

Variables:

- w_j Number of working capacity units on span j for the design
- s_j Number of spare capacity units on span j for the design
- $y_{j,k}$ Number of additional working capacity units that would have to be placed on span j in future to cope with scenario k

$z_{j,k}$ Number of additional spare capacity units required on span j under future demand scenario k

$g_k^{r,q}$ Working flow assigned on the q^{th} working route to serve OD pair r in scenario k

$n_{p,k}$ Number of copies of cycle p used in the p-cycle design in scenario k

TP-PC: Minimize $$\sum_{j\in S} C_j \cdot (w_j + s_j) + \sum_{j\in S}\sum_{k=U} P_k \cdot R_j \cdot (y_{j,k} + z_{j,k}) \quad \text{(C.1)}$$

Subject to:

$$\sum_{q\in Q^r} g_k^{r,q} = d_k^r \qquad \forall r \in D; \forall k \in U \qquad \text{(C.2)}$$

$$\sum_{r\in D}\sum_{q\in Q^r} \zeta_j^{r,q} \cdot g_k^{r,q} = w_j + y_{j,k} \qquad \forall j \in S; \forall k \in U \qquad \text{(C.3)}$$

$$w_i + y_{i,k} \le \sum_{p\in P} x_{p,i} \cdot n_{p,k} \qquad \forall i \in S; \forall k \in U \qquad \text{(C.4)}$$

$$s_j + z_{j,k} \ge \sum_{p\in P} \delta_{p,j} \cdot n_{p,k} \qquad \forall j \in S; \forall k \in U \qquad \text{(C.5)}$$

$$y_{j,k}, z_{j,k} = 0 \qquad k = 0; \forall j \in S \qquad \text{(C.6)}$$

The objective function (C.1) plus constraint sets (C.2), (C.3) and (C.6) are identical to those explained in the TP-SR formulation. The only differences in this model are the restorability asserting and spare capacity generating constraints, which are now based on the p-cycle network design. $x_{p,j}$ and $\delta_{p,j}$ are pre-determined input parameters for each cycle in set P indicating the number of protection relationships cycle p provides to span j. $x_{p,j}$ can be {0, 1 or 2}; $x_{p,j}$ = 1 if span j is part of the cycle p; $x_{p,j}$ = 2 if span j straddles cycle p; $x_{p,j}$ = 0 if there is no relationship between span j and cycle p. Like (6.5.4), constraint (C.4) ensures that the sum of all cycles $n_{p,k}$, provides sufficient pre-connected spare capacity for protecting every span failure i; similarly, constraint (C.5) determines the spare capacities s_j necessary to support a protecting cycle set. $\delta_{p,j}$ takes a value of one if the cycle p uses the span j; it is zero otherwise.

Appendix D: Examples of Outage Index Calculation

The outage index is defined as a measure of the impact of the failure on customer experiences. Essentially the higher the index, the greater the impact on the customer. The index values are always positive, and small outage would have an index near zero. Note that the formulas, especially the weights from the index calculation, are based on empirical data available from or estimated from actual outage reports. It is important to note that these calculations only consider outages in Public Switched Telephone Networks (PSTN). Today's data services, such as the Internet and enterprise's data traffic, had not been taken into the account, and that might warrant some future research and investigations.

Example 1: Dedicated (Local Switch) Partial Services Outage Example with Same Service Outage Durations. From [T1A01].

Report Data:

Start Time: .... 3:00 am, August 10, 1995 (Thursday)

Number of Lines Affected:........................... 42,291

Types of Services Affected:.......................... IntraLATA Interoffice, InterLATA Interoffice, 911

Duration of Outage: .... 34 minutes (same for all services affected)

Number of Blocked Calls: 9,000

Outage Category: .... Local Switch (Adjunct Processor Failure).

Outage Index Calculation:

Method Used: .... Lines

Time Factor Used: .... 0.1 (IntraLATA Interoffice, InterLATA services),
1.0 (911 Service).

Table D.1. Example 1: Outage Index Calculation. From [T1A01]

Method Used	Service	Weights			Product
		Service (W_S)	Duration (W_D)	Magnitude (W_M)	W_s W_D W_M
Lines	911	3	1.05	0.477	1.50
Lines	IntraLATA Intraoffice	N/A	N/A	N/A	N/A
Lines	IntraLATA Interoffice	2	1.05	0.00477	0.0100
Lines	InterLATA Interoffice	2	1.05	0.00477	0.0100
Block Calls	All services except 911	N/A	N/A	N/A	N/A
Outage Index = Sum of Products = 1.52					

Example 2: Diversified (Facilities, Fiber) Outage Durations. From [T1A01].

Report Data:

Start Time: .... 4:30 pm, August 9, 1995 (Wednesday)

Number of Lines Affected:........................... Not Reported

Types of Services Affected:.......................... InterLATA Interoffice

Duration of Outage .... 10 hours and 47 minutes (same for all services affected)

Number of Blocked Calls: 102,144

Outage Category: .... Facilities (Fiber Cable).

Outage Index Calculation:

Method Used: .... Blocked Calls

Time Factor Used: .... N/A.

Table D.2. Example 2: Outage Index Calculation. From [T1A01].

Method Used	Service	Weights			Product
		Service (W_S)	Duration (W_D)	Magnitude (W_M)	$W_s\ W_D\ W_M$
Lines	911	N/A	N/A	N/A	N/A
Lines	IntraLATA Intraoffice	N/A	N/A	N/A	N/A
Lines	IntraLATA Interoffice	N/A	N/A	N/A	N/A
Lines	InterLATA Interoffice	N/A	N/A	N/A	N/A
Block Calls	All services except 911	2	2.32	0.309	1.44
Outage Index = Sum of Products = 1.44					