FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTOCTORALES

uOttawa

L'Université canadienne
Canada's university

FACULTY OF GRADUATE AND POSDOCTORAL STUDIES

Yiming Zhang
AUTEUR DE LA THÈSE / AUTHOR OF THESIS

Ph.D. (Electrical Engineering)
GRADE / DEGREE

School of Information Technology and Engineering)
FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

Virtual Topology Optimization in Wavelength Routed Optical Networks
TITRE DE LA THÈSE / TITLE OF THESIS

Oliver Yang
DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

EXAMINATEURS (EXAMINATRICES) DE LA THÈSE / THESIS EXAMINERS

Mourad Elhadef

Amiya Nayak

Changcheng Huang

Brigitte Jaumard (absent-on phone)

Gary W. Slater
Le Doyen de la Faculté des études supérieures et postdoctorales / Dean of the Faculty of Graduate and Postdoctoral Studies

VIRTUAL TOPOLOGY OPTIMIZATION
IN
WAVELENGTH ROUTED OPTICAL NETWORKS

by

Yiming Zhang

A thesis submitted to
the Faculty of Graduate and Postgraduate Studies,
University of Ottawa
in partial fulfillment of the requirements
for the degree of Ph.D.

August 2006

Ottawa-Carleton Institute for Electrical and Computer Engineering
School of Information Technology and Engineering
University of Ottawa
Ottawa, Ontario, Canada

Your file Votre référence
ISBN: 978-0-494-18615-2
Our file Notre référence
ISBN: 978-0-494-18615-2

Abstract

In static and semi-dynamic routing, the traffic requirements in the Wavelength Division Multiplexing (WDM) network are given as a set of point-to-point connections of different granularities between pairs of access stations. Finding the exact optima for most of the network optimization problems is generally impossible with today's computation facility even for the medium size networks. Therefore, finding a sub-optimal solution within a reasonable computation time is the only choice, while knowing the proximity of the sub-optimal solution to the exact-optimum will be an additional advantage.

We provide in this thesis an efficient optimization framework to static and semi-dynamic connection routing problems, which contains new integer linear programming (ILP) formulations and a solution methodology employing Lagrangean Relaxation and Subgradient Method (LRSM). We are able to provide a good bound at the same time as obtaining a sub-optimal solution, with reasonable time complexity. We prove the efficiency and the easy adaptation of our framework by studying the optimization problems for various WDM networks with different models, different constraints and different objective functions in generalized network topologies. We use it to investigate the static Routing and Wavelength Assignment (RWA) problem, semi-dynamic RWA problem, as well as the traffic grooming problem. We also study a wide range of influences, such as the number of resources and the network design parameters, on the network behaviours under different network structures and assumptions. On the other hand, we provide new formulations for the network optimization considering various aspects of the network operation and design concerns. Compared with many other methodologies proposed in some previous research, our framework demonstrates a high efficiency in terms of both computation results and time complexity.

Acknowledgement

I am sincerely grateful to my supervisor Dr. Oliver W. W. Yang for his unfailing knowledgeable guidance and helpful suggestions on my work.

I want to give my great thanks to Dr. Jing Wu for the selflessly sharing of his knowledge and his valuable comments for my work.

My research has been partially supported in part by a grant from CITO, and in part by the National Science and Engineering Research Council (NSERC) and industrial partners, through the Agile All-Photonic Networks (AAPN) Research Network.

I want to dedicate this thesis to my beloved grandmother and grandfather who passed away during my study towards my Ph.D. degree.

Table of Content

Title i
Abstract ii
Acknowledgement iii
Table of Content iv
List of Figures vi
List of Tables viii
Table of Notations ix
Table of Acronyms xii
Chapter One Introduction 1
1.1 Overview 1
1.1.1 Review of Routing and Wavelength Assignment (RWA) Algorithms 2
1.1.2 Review of Network Grooming Optimization 8
1.2 Mathematical Background 12
1.2.1 Lagrangean Relaxation 12
1.2.2 Subgradient Method 14
1.2.3 Other approaches 15
1.3 Motivations 17
1.4 Objectives 21
1.5 Approaches and Methodologies 21
1.6 Contributions 24
1.7 Organizations 25
1.8 Publications 25
Chapter Two Network Operations, Models and Assumptions 27
2.1 Wavelength Routed Network 27
2.2 Network Operation 28
2.2.1 Operation of Wavelength Division Multiplexing 28
2.2.2 Operation of Wavelength Switching 29
2.2.3 Operation of Network Reconfiguration 29
2.3 Component Operation 30
2.3.1 Operation of Wavelength Add/Drop Multiplexer 30
2.3.2 Operation of Wavelength Converter 31
2.3.3 Operation of Electronic Switching 32
2.4 Network Models 33
2.4.1 Mapping Between Physical and Logical Topologies 33
2.4.2 Wavelength Converter Models 34
2.4.3 Electronic and Photonic Switching Model 36
2.5 General Assumptions 37
Chapter Three Static RWA Problems in All-Optical WRNs 38
3.1 Minimum-Cost Problem 38
3.1.1 Problem Formulation 38
3.1.2 Mathematical Solution Procedure 40
3.1.3 Performance Evaluation 47
3.2 Maximum Traffic Problem 49
3.2.1 Problem Formulation 50

3.2.2 Mathematical Solution Procedure ... 52
3.2.3 Computation Analysis ... 56
3.2.4 Performance Evaluation ... 56
3.3 Concluding Remarks ... 67
Chapter Four Integrated Design of Traffic Grooming on WRN Structure ... 69
4.1 Design and Optimization Objective ... 69
4.2 Assumptions for the Traffic Grooming Network ... 71
4.3 Problem Formulation ... 71
4.4 Mathematical Solution Procedure ... 74
4.4.1 Overview of the Solution Framework ... 74
4.4.2 The LR Solution ... 75
4.4.3 Solving the Sub-problems ... 76
4.4.4 Solving the Dual Problem ... 79
4.4.5 Constructing a Feasible RWA and Routing Scheme ... 79
4.4.6 Computation Complexity ... 81
4.4.7 Overall Structure of the Algorithm ... 82
4.5 Performance Evaluation ... 83
4.6 Concluding Remarks ... 91
Chapter Five Semi-dynamic Reconfiguration Problem in WRN Network ... 92
5.1 Problem Formulation ... 92
5.1.1 Pre-formulation Processing ... 93
5.1.2 Objective and Constraints ... 94
5.2 Solution Methodology ... 98
5.2.1 The LR Solution Procedure ... 99
5.2.2 RWSS and CGSS ... 100
5.3 The LRSM Framework ... 100
5.3.1 Solving RWSS and CGSS ... 101
5.3.2 Updating Lagrange Multipliers ... 103
5.3.3 Constructing a Feasible RWA Scheme ... 103
5.3.4 Computation Complexity ... 106
5.4 Performance Evaluation ... 107
5.4.1 Effect of Different Factors in The Semi-dynamic Reconfiguration ... 108
5.4.2 Fairness Study ... 111
5.5 Concluding Remarks ... 113
Chapter Six Design Guideline and Recommendations ... 114
6.1 Range of Applicability ... 114
6.2 Selection of Rejection Penalties ... 114
6.3 Stopping Criteria ... 115
6.4 Multiplier Reuse ... 116
6.5 Recommendation for Employing LRSM Framework ... 117
Chapter Seven Conclusion ... 119
7.1 Future work ... 120
References ... 121
Appendix A Algorithms ... 129

List of Figures

Figure 2.1. All-Optical WDM Network Consisting Of Routing Nodes Interconnected By Point-To-Point Fibers ... 27
Figure 2.2. The Wavelength Division Multiplexing ... 28
Figure 2.3. A $WN \times WN$ Crossconnect without wavelength Conversion ... 31
Figure 2.4. An Integrated Traffic Grooming Node ... 32
Figure 2.5. The description of the two-layer topology ... 33
Figure 3.1. The schematic depiction of the algorithm ... 46
Figure 3.2. Randomly generated 11-node network map ... 47
Figure 3.3. Connection demand matrix ... 48
Figure 3.4. Overall cost v.s. W and F ... 48
Figure 3.5. Overall cost v.s. W and F ... 48
Figure 3.6. NSFNET with 14 Nodes and 21 Edges ... 57
Figure 3.7. Traffic Demand Matrix for NSFNET ... 57
Figure 3.8. Computation Time Comparison of Our Algorithm and LP-relaxation ... 59
Figure 3.9. Number of rejected connection demands vs. W and F (v=1) ... 60
Figure 3.10. Number of rejected connection demands vs. W and F (v=2) ... 61
Figure 3.11. Number of rejected connection demands vs. W and F (v=3) ... 61
Figure 3.12. Relation between S and D (v=3) ... 63
Figure 3.13. S vs. Number of rejected connection demands (v=3) ... 64
Figure 3.14. S vs. Number of rejected connection demands (v=3) ... 64
Figure 3.15. S vs. Number of rejected connection demands (v=3) ... 65
Figure 3.16. Network With 22 Nodes And 35 Edges ... 65
Figure 3.17. Traffic Demand Matrix for the 22-node Network ... 66
Figure 4.1. The schematic depiction of the overall algorithm ... 82
Figure 4.2. 14-node NSFNET topology ... 83
Figure 4.3. J and q versus T_i/R_i ... 85
Figure 4.4. J and q versus W ... 85
Figure 4.5. J and q versus t_i/r_i ... 86
Figure 4.6. J and q versus d_{ijc} ... 87
Figure 4.7. Average hop number of semi-lightpaths versus d_{ijc} ... 87
Figure 4.8. Average hop number of traffic versus V_{pqz} ... 88
Figure 4.9. Percentage of OC-3 traffic rejected versus P_{pqz} ... 88
Figure 4.10. A 6-node network ... 90
Figure 5.1. The illustration graph of the rejection penalty for (s, d) ... 95
Figure 5.2. Schematic depiction of the overall algorithm ... 101
Figure 5.3. The flowchart of the heuristic algorithm to obtain a feasible solution ... 104
Figure 5.4. NSFNET with 14 Nodes and 21 Edges ... 107
Figure 5.5. Traffic Demand Matrix for NSFNET ... 107
Figure 5.6. Simulated Traffic Demand Matrix for NSFNET at Previous Session ... 108
Figure 5.7. The number of rejected connection demands and the number of reconfigured connections ... 109
Figure 5.8. The comparison of different number of wavelength converters ... 109
Figure 5.9. Influence of different conversion degree (v) ... 110
Figure 5.10. The congestion (γ) and the number of rejected connection demands ... 110

Figure 5.11. The congestion penalty (G) v.s. the final results of q and J 111
Figure 5.12. The rejection penalty stepsize (S) v.s. the number of rejected connection demands 112
Figure 5.13. The rejection penalty stepsize(S) v.s. number of totally disconnected pairs (D) 112
Figure 5.14. The rejection penalty stepsize (S) v.s. the final results of q and J 113

List of Tables

Table 3.1. Number of Connections Rejected in Results of [SwSi02] ... 58
Table 3.2. Number of Connections Rejected by Our Algorithm ... 58
Table 3.3. Number of rejected connection demands vs. W and F ... 62
Table 3.4. Number of Connections Rejected Obtained for the 22-node Network(F=4) ... 66
Table 4.1. Traffic demand matrices ... 84
Table 4.2. Influence of F_i on J ... 89
Table 4.3. Traffic matrices for the 6-node network (Equivalent to 988 OC-1 bandwidth in total). 90
Table 4.4. The comparison between the results in [ZhMu02] and our results ... 90

Table of Notations

		Section of the first appearance:
a_{ij}	the weight of the edge (i, j) in the wavelength graph;	Section 3.2.1
c_j	the cost of using a wavelength converter at node j;	Section 3.2.1
d_{ijc}	the cost of using WC (Wavelength Channel) w_{ijc};	Section 2.1
e_{ij}	the physical fiber between node i and node j; $e_{ij} \in \mathcal{E}$;	Section 3.1.1
f	the number of 'dummy' demands, which is equal to $u(X'_{sd}-N_{sd})$;	Section 5.1.1
i,j	the nodes at the end of a fiber;	Section 3.1.1
n_{ij}	the number of WCs (Wavelength Channel) on the fiber e_{ij};	Section 3.1.1
p,q	endpoint nodes of a traffic;	Section 4.3
r_j	the cost of a receiver at node j;	Section 3.1.1
s,d	the source and destination nodes of a semi-lightpath;	Section 3.1.1
s_{sdn}	the nth semi-lightpath between node pair (s, d);	Section 3.1.1
t_j	the cost of a transmitter at node j;	Section 3.1.1
t'_{sdn}	the nth existing connection (semi-lightpath) between (s, d) from the previous session;	Section 5.1.1
u_i	the label associated with vertex i in the wavelength graph;	Section 3.2.1
u_{ij}	the label associated with wavelength j of vertex i in the wavelength graph;	Section 3.2.1
v	the degree of wavelength conversion;	Section 3.2.1
w_{ijc}	the cth WC on the fiber e_{ij};	Section 3.1.1
x_{pqz}	the zth traffic demand between node pair (p,q);	Section 4.3
A	the set of admission status of all the semi-lightpath matrix, i.e., $\{\alpha_{sdn}\}$;	Section 3.2.1
A_{sd}	the variable set $\{\alpha_{sdn}\}_{sd}$, the set of admission statuses of all the connection demands of (s, d);	Section 5.1.2
B	the variable set $\{\beta_{sdn}\}$;	Section 5.1.2
D	the number of source-destination pairs that have traffic demands, but are not assigned any semi-lightpath;	Section 3.2.1
D_{sdn}	the total routing dual cost of the s_{sdn};	Section 3.2.1
$\mathcal{E}$	the set representing all fiber links in the network;	Section 3.1.1
$\mathcal{E}_v$	the set of all the unidirectional semi-lightpaths that could be set up;	Section 4.3
E_{sdn}	the total network resource cost of s_{sdn};	Section 3.1.1
E_{pqz}	the total electronic routing cost of the traffic x_{pqz};	Section 4.3
F_i	the number of full-range wavelength converters at node i;	Section 2.1
F_{ic}	the number of wavelength converters that convert a signal from wavelength c to other wavelengths on node i;	Section 3.2.1
G	the penalty coefficient of the load on the links;	Section 5.1.1
G_{pqz}	the bandwidth of the traffic x_{pqz};	Section 2.1
H_{sd}	the integer equals $\max(N_{sd}, X'_{sd})$;	Section 5.1.1
$I_i(c)$	the set of wavelengths, to which the traffic from wavelength c can be converted on node i;	Section 3.2.1
M	the semi-lightpath set, i.e., $\{s_{sdn}\}$;	Section 3.2.1

N	the number of nodes in the network;	Section 3.1.1
N_{sd}	the maximum number of semi-lightpaths between node pair (s, d);	Section 2.1
P_{sdn}	the penalty for rejecting s_{sdn};	Section 3.2.1
P_{pqz}	the penalty coefficient for rejecting the traffic demand x_{pqz};	Section 4.3
$P_{sd}(k)$	the penalty coefficient for rejecting k connection demands of (s, d), which can be represented by $\sum_{h=1}^{k} P_{sdh}$. $P_{sd}(0) = 0$;	Section 5.1.1
Q	the penalty coefficient for reconfiguring one existing connection;	Section 5.1.2
Q_{sdn}	the dual routing cost for the semi-lightpath s_{sdn};	Section 4.3
R	the set of rejection statuses of all the semi-lightpath demands, i.e., $\{r_{sdn}\}$;	Section 3.2.1
R_i	the number of receivers on node i;	Section 2.1
R_{pqz}	the dual routing cost of the traffic x_{pqz};	Section 4.3
S	the penalty step size for fairness consideration;	Section 3.2.1
T	the overall number of source-destination pairs that have the traffic demands;	Section 3.2.1
T_i	the number of transmitters at node i;	Section 2.1
$\mathcal{V}$	the set representing all the nodes in the network;	Section 3.1.1
$\mathcal{V}_v$	the set of all the electronic switches at nodes;	Section 4.3
$(\mathcal{V},\mathcal{E})$	bi-directional graph representing the physical WDM network,;	Section 3.1.1
$(\mathcal{V}_v,\mathcal{E}_v)$	a unidirectional graph representing the virtual WDM network topology;	Section 4.3
V_{pqz}	the coefficient for the electronic routing cost of the traffic x_{pqz};	Section 4.3
W	the number of wavelengths used in the network;	Section 3.2.1
$\boldsymbol{W}_{ij}$	the WC set on the fiber e_{ij}. We generally assume all fibers use the same number of wavelengths and denoted as W;	Section 3.1.1
WE_{jhi}	the weight of the edge in the wavelength graph from column j *to* column h in the ith row (equals ∞ if there is not arc);	Section 3.2.1
X'_{sd}	the number of existing connections between (s, d) from the previous session;	Section 5.1.1
Y	the set of use of semi-lightpaths by all traffic, i.e., $\{\upsilon_{sdn}^{pqz}\}$;	Section 4.3
Y_{pqz}	the set $\{\upsilon_{sdn}^{pqz}\}_{pqz}$, representing the routing of the traffic x_{pqz};	Section 4.3
Z_{pq}	the maximum number of traffic demands between node pair (p, q). Note that Z_{pq} not necessarily equals Z_{qp}, i.e., traffic demands may be asymmetric;	Section 4.3
$*\alpha_{sdn}$	the admission status of the semi-lightpath s_{sdn}, which equals to one, if the semi-lightpath s_{sdn} is set up; zero otherwise;	Section 3.2.1
β_{sdn}	a 0-1 integer variable indicating the reconfiguration status of t'_{sdn}, equal to one, if t'_{sdn} is reconfigured; zero otherwise;	Section 5.1.2
γ	a variable representing the maximum load (congestion) on all e_{ij} ($0 \leq \gamma \leq 1$);	Section 5.1.2
γ_{pqz}	the admission status of the traffic x_{pqz}, which equals to zero, if the traffic x_{pqz} is rejected; one otherwise;	Section 4.3
$*\delta_{ijc}^{sdn}$	a 0-1 integer variable, representing the use of wavelength channel w_{ijc} by semi-lightpath s_{sdn};	Section 3.1.1
λ_i	wavelength i multiplexed on the same fiber	Section 2.2

$^*\phi_{j,ab}^{sdn}$	a 0-1 integer variable, representing the use of wavelength converters (used on limited-range wavelength converters);	Section 3.2.1
ϕ_j^{sdn}	a 0-1 integer variable, representing the use of wavelength converters (used on full-range wavelength converters);	Section 3.1.1
υ_{sdn}^{pqz}	the use of the semi-lightpath s_{sdn} by the traffic x_{pqz}, which equals to one, if the traffic x_{pqz} goes through s_{sdn}; zero otherwise;	Section 4.3
Γ	the set of admission status of all traffic, i.e., $\{\gamma_{pqz}\}$;	Section 4.3
$^*\Delta_{sdn}$	the variable set $\{\delta_{ijc}^{sdn}\}_{sdn}$, representing the wavelength assignment for s_{sdn};	Section 3.1.1
Δ	the variable set $\{\delta_{ijc}^{sdn}\}=\{\Delta_{sdn}\}$;	Section 3.1.1
$^*\Phi_{sdn}$	the variable set $\{\phi_{j,ab}^{sdn} / \phi_j^{sdn}\}_{sdn}$, representing the wavelength converter assignment for s_{sdn};	Section 3.1.1
$^*\Phi$	the variable set $\{\phi_{j,ab}^{sdn} / \phi_j^{sdn}\}=\{\Phi_{sdn}\}$;	Section 3.2.1

Note: For all the variables marked with '*', we denote the same variables/sets for the connections that are from the previous RWA session with an apostrophe. For example, the notation Δ'_{sdn} represents the set Δ_{sdn} of the existing RWA of t'_{sdn} from the previous session. Please see Chapter Six for the details.

Table of Acronyms

Notations	Explanations	Section of the first appearance:
CGSS	Congestion Subproblem Set	Section 5.2.2
DQDB	Distributed Queue Dual Bus	Section 1.1
ETRS	Electronic Traffic Routing Sub-problems	Section 4.4.3
FDDI	Fiber Distributed Data Interface	Section 1.1
FWC	Full-range Wavelength Converter	Section 3.1.1
ILP	Integer Linear Programming	Section 1.1.1
LAN	Local Area Network	Section 1.1
LR	Lagrangean Relaxation	Section 1.1.1
LRSM	Lagrangean Relaxation and Subgradient Method	Section 1.1
LP	Linear Programming	Section 1.1.1
MILP	Mixed Integer Linear Programming	Section 1.1.1
MCSLP	Modified Minimum Cost Semi-lightpath	Section 3.2.1
OLT	Optical Line Terminal	Section 2.3.1
OXC	Optical Cross-Connect	Section 2.2.2
RMCSLP	Revised Minimum Cost Semi-lightpath	Section 4.4.3
RSPA	Revised Shortest Path Algorithm	Section 4.4.3
RWA	Routing and Wavelength Assignment	Section 1.1.1
RWAS	RWA Sub-problems	Section 4.4.3
RWSS	RWA Subproblem Set	Section 5.2.2
SONET	Synchronous Optical Network	Section 1.1.1
SPAWG	Shortest Path Algorithm for Wavelength Graph	Section 3.1.1
TDM	Time Division Multiplexing	Section 1.1
WAN	Wide Area Network	Section 1.1
WC	Wavelength Channel	Section 1.1.1
WDM	Wavelength Division Multiplexing	Section 1.1.1
WRN	Wavelength Routed Network	Section 1.1.1

Chapter One
Introduction

1.1 Overview

The data communication networks can be categorized into three generations, based on the technologies used in the physical-layer [Gree91]. The first-generation networks, such as ARPANET, Ethernet, and Token Ring, which appeared before the emergence of fiber optic technology, use only the electronic technology. In the second-generation networks, optical techniques are mainly used for transmission, and the traditional electronic equipment is used for the switching. The typical examples are the Distributed Queue Dual Bus (DQDB) linear network and the Fiber Distributed Data Interface (FDDI) ring network designed for Local/Metropolitan Area Network (LAN/MAN). Due to the bottleneck of the electronic multiplexing/de-multiplexing and switching, the transmission speed of optical fibers is limited by the electronic processing speed [Mukh92a]. The third generation networks are the all-optical networks, i.e., once a packet enters the all-optical network, it remains in the optical domain without any electronic processing until it arrives at the destination. All-optical networks using optical switching and Wavelength Division Multiplexing (WDM) have emerged to be promising candidates for the next generation of transport networks as they can economically satisfy the growing demand for bandwidth. WDM technology allows the transmission of signals of different wavelengths on the same fiber link, thus much greater bandwidth on the fiber can be achieved by the wavelength multiplexing, even if the transmission speed of the transceivers for individual signal remains the same. On the other hand, the equipment for the ultra high-speed WDM network is very expensive, which demands the economic network planning and the efficient use of the network bandwidth. The dominating WDM networks are the Wavelength Routed Networks (WRN), which can efficiently switch the optical signals of different wavelengths to the intended destinations. Routing and Wavelength Assignment (RWA) is extremely important in these kinds of networks to better utilize the expensive network resources. However, a WRN can only provide a whole wavelength as the smallest bandwidth unit, while sub-wavelength connections are more suitable for non-backbone routers to support lower-bandwidth channels. *Traffic Grooming* is thus introduced to take sub-wavelength traffic and to sort it into the most efficient arrangement possible in WRN.

1.1.1 Review of Routing and Wavelength Assignment (RWA) Algorithms

In this section, we will provide a general review of the existing RWA algorithms for different scenarios.

1.1.1.1 Static RWA Algorithms

RWA problems can be generally classified into static RWA, semi-dynamic RWA and dynamic RWA problems according to different traffic models and assumptions in the WDM network. In static RWA problems, all the traffic demands are known in advance and do not change over time. The problem is then to route the traffic demands in a global fashion with respect to a certain objective function such as minimizing the network resources and maximizing the number of connections. Static RWA problem is often referred to as optical *logical/virtual topology* design problems. Unlike the static RWA problems, all the traffic demands in semi-dynamic RWA problems are known in advance for every session and do not change within one session. However, the number of lightpaths (see Section 2.2.2 for the definition) might change from one session to the next due to the increase/decrease of its capacity demand from the clients of the network services. The problem now has to take into account the evolution of the traffic load in the network for the consecutive sessions. In dynamic RWA problems at the extreme, the arrival and holding times of demands can be quite random albeit there might be a pattern of changes. Thus, whenever a connection request arrives, an algorithm must be executed in real time to determine whether it is feasible to accommodate the request, and if so, to perform routing and wavelength assignment. If the request cannot be accepted because of the lack of resources, it is blocked [Rous01]. The real time nature, the randomness of connection requests and the insufficient time to obtain the global information make dynamic RWA problem more difficult to optimize, and therefore, straightforward heuristic algorithms are generally used to improve the probabilistic characteristics in the network. Simple standards such as blocking ratio and average hop-number are commonly used in the evaluation of dynamic RWA solutions.

In access networks, the traffic is generally of low capacity and very volatile, opposite to the traffic in the backbone networks. The traffic patterns in the access networks are transient and the global optimization is unlikely because during the time of performing the optimization and communicating the information, the traffic pattern has already changed. In the wide-area backbone optical networks, the traffic is much more stable, and the traffic requirements are generally given as a set of point-to-point connections of different granularities between pairs of access stations (nodes). The set of point-to-point connections also defines a *virtual topology*, which is a logical topology imbedded in the physical network. By setting up a point-to-point connection, two nodes that are physically apart can be virtually neighbours. The virtual topology formed by these connections provides somewhat independence from the physical topology, although constrained by the physical topology, and different virtual topologies could be implemented given the same physical topology. The connections between the nodes (thus the virtual topology) usually are reconfigured only after remaining in place for a relatively long period of time. The optimization problems under this situation are called static or semi-dynamic optimization problems. Virtual topology design is an important issue in the network planning. Considering the high price of the devices, a good virtual topology can result in big saving and high performance of the network.

The RWA optimization problem, which is essential to the efficiency of the all-optical WRNs, has attracted lots of attentions in the previous research [Gree96, Mukh97, RaSi01, StBa99]. In most of the solutions, RWA problem is partitioned into two subproblems: (1) routing and (2) wavelength assignment, and each subproblem is solved independently. The authors in [BaMu96] proposed practical approximation algorithms to solve the RWA problem for large networks, and graph-coloring algorithms are used to allocate wavelengths to the lightpaths once the lightpaths are routed. However, the two subproblems are actually closely related, thus the brutal separation in the solution procedure results in the degradation of the performance of the algorithm.

There are also some other research trying to solve the RWA problem as a whole [ChGa92, Mukh97, StBa99], without solving it in two subproblems, while cutting off the relationship between the two problems. These researchers generally formulate RWA optimization problems as Mixed Integer Linear Programming (MILP) problems, and try to solve the problems with mathematical tools. Most of these problems are NP-complete or NP-hard, e.g., [ChGa92, GaJo79, ZaJu00]. To achieve different goals, various solution techniques have been proposed, which can be classified into two main groups [Bert99], i.e., sub-optimal methods and exact methods. Sub-optimal methods in many cases are acceptable and require a limited computational effort. The exact methods are much more computationally intensive and generally do not scale well with the network size, and only small-size RWA optimization problems can be solved exactly. However, because the exact methods are able to obtain the absolute optimal solution, they play a fundamental role either as direct planning tools or as benchmarks to validate and test the heuristic/sub-optimal methods [ToMa02]. For larger size networks, it is normally impossible to obtain the exact optimum solution, due to the high computation complexity. The only alternative is to generate an acceptable sub-optimal solution within a reasonable computation time. The most commonly used methods are heuristic methods. Compared with other non-heuristic methods, which are mostly classic mathematical optimizations, the heuristic methods are more intuitive, easier to understand, generally simpler to implement, and having fewer requirements on the formulation of the optimization problems. Therefore, the heuristic approaches are the most commonly used in the optical network optimizations. From different angles and starting points, various objectives and constraints have been proposed to formulate those problems, and many kinds of heuristic algorithms have been proposed to solve those problems approximately. Examples are the minimum-hop heuristic [BaBa96], Monte Carlo approach [Boui98] and many others [ChBa95, ChGa93, ZhAc95]. Most of the heuristic methods usually generate satisfactory schedules in a reasonable computation time, but they generally cannot demonstrate their effectiveness in providing an estimate of how close they are from the optimal solution. In the non-heuristic approaches, many classic mathematical tools developed for complex MILP / ILP optimization problems have been used to obtain optimal or sub-optimal solutions to RWA optimization problems. These approaches include relaxation of integer constraints [BaMu00, KrSi01, SwSi2], Lagrangean relaxation [SaLu02, YeLi01a, ZhYa03a], search space dimensionality (SSD) reduction [RaSi96, ZaVu03], branch-and-bound [KuPu03], etc.

Many types of RWA problems with different objectives and constraints have been proposed. Among them, the minimum cost routing problem in the network has always been an interesting topic in the previous research. The routing of a lightpath with the minimum cost has been studied in [ChFa96]. An algorithm is provided in [ChWa02] to minimize the cost of a multicast tree in an all-optical WDM network of an arbitrary topology, while the multicast nodes are pre-assigned. However, both papers do not consider the collisions produced by the routing of simultaneous connections in a network with limited resources. Optical level configuration has been studied jointly with the electronic level in [BaCh96] to minimize the cost. There is another typical RWA problem, which is to maximize connections for a given traffic session with a fixed set of wavelengths, i.e., the Max-RWA problem [KrSi01]. In the new generation of wavelength routing networks such as Ca*net4 [Cana03], it is extremely important for network service providers to maximize the number of connections to satisfy the demand of the customers, and therefore to maximize the profit for the existing network configuration. The Max-RWA problem is therefore subject to considerable interest and study, e.g., [KrSi01, SaLu02, SwSi02]. Both [KrSi01] and [SwSi02] used a branch-and-bound LP-relaxation to relax some integer constraints in order to obtain a lower bound for the solution. Then a heuristic algorithm is used to obtain a feasible RWA scheme based on the solution to the relaxed problem. However, the relaxed ILP problem is very computationally expensive to solve even for networks of very small sizes, and therefore not suitable for large networks. Also, the manner in which the connection demands are accepted might not be fair for different node pairs. For example, while all connections between some node pairs are accepted, the connections between some other node pairs might be rejected altogether. Partial wavelength conversion considered in [SwSi02] with the assumption that the wavelength conversion has an unlimited capacity has also made the formulation less practical. In general, LP-relaxation has the shortcoming that all the computation has to be done again for any small change in the network configuration or in the traffic demand. In [RaSi95], Ramaswami and Sivarajan used a Linear Programming (LP) program to determine the upper bound on the amount of carried traffic that any RWA algorithm could be able to accommodated. LRSM (Lagrangean relaxation and subgradient method) has been employed to solve the wavelength assignment and routing problem in all-optical network [YeLi01a], which uses the idea of relaxing the wavelength capacity constraint, and minimizes the cost of the usage of wavelengths on links in the optical network with no wavelength conversion. However, in the practical networks, the cost of the wavelength converters sometimes dominates over the cost of wavelengths on links, and the cost of using wavelengths on links might be only a small part of the overall cost. An extensive survey on the RWA problem can be found in [ZaJu00].

1.1.1.2 Semi-dynamic RWA Algorithms

In WRNs, lightpaths may have to be quickly reconfigured in reaction to traffic pattern changes, network failures or new network resource installations. Compared with the pure static Routing and Wavelength Assignment (RWA), the lightpath reconfiguration features a strong correlation between the current and the new projected RWA configurations. Taking such correlation into account, the WDM network reconfiguration has been studied in different scenarios, e.g., survivable rerouting [BoLa05], pure RWA reconfiguration [ZhZh02], single-hop broadcast networks [BaRo01], virtual topology reconfiguration in ring networks [MoEr03], dynamic virtual topology adaptation in response to measured load imbalance [GeMu03], heuristics based virtual topology migration [LeZh03, RiSa02, SrMu01], and virtual topology design to minimize average packet hop distance [BaRo01], etc.

To optimize the WDM network reconfiguration, it is critical to model the correlation between the existing and the new RWA configurations. A two-step approach was proposed in [BaMu00], where the first step is to find a new RWA configuration that best matches a new traffic pattern, and the second step is to design a new RWA configuration that requires the least changes from the current RWA configuration and at the same time matches the new traffic pattern. A variation of this two-step approach was presented in [GeMu03], where the second step was changed to choose the closest configuration that requires the least number of lightpaths. Also a multi-objective evolutionary algorithm was also proposed to obtain the best performance [PrSt04].

For all the two-step approaches and their variations described above, a major challenge is that the best performance sometimes cannot be defined properly (e.g., [BaMu00]), or can rarely be achieved in different RWA configurations. Therefore, the reconfiguration options are very much limited when the first priority is set to the best performance (e.g., in minimal blocking) in an independent static RWA (i.e., ignoring the correlation). Consequently, many potential reconfiguration options that require much less changes (at a small performance cost) are excluded just because they cannot achieve the questionable objective of the best performance, which is set up at the onset of an independent static RWA problem. Moreover, within a two-step framework, it is difficult to study the tradeoffs between the performance of a new RWA configuration and the changes from the existing RWA configuration.

To overcome these problems of the two-step approaches, a heuristic solution is provided in [SaLu05], optimizing the two steps at the same time. However, the re-routing of the existing connections is not considered, thus the trade-offs between the re-routing and the rejection of the connection demands cannot be studied. Also, the objective function does not consider other design objectives such as congestion and fairness. Finally, since the paths have to be pre-selected, the path-selection is an optimization problem by itself, which makes their path-based implementation much harder. Note that one-step approach was also attempted in [PrSt04] by using a multi-objective function, but their best performance (minimal blocking) is still used as the dominant objective, while the other objectives are not really integrated in the optimization process.

An optimization algorithm was developed for a broadcast WDM network and compared to the performance bound derived from a greedy algorithm [BaRo01]. To quantify the benefits of network reconfiguration, the optimization objective was defined as maximizing the reconfiguration gain, which is the difference between a performance reward and the reconfiguration cost. Similarly, the reconfiguration gain was used in a wavelength-routed WDM network [LeZh03], which solves the traffic predication based on-line reconfiguration problem. However, the performance of the network reconfiguration schemes in a wavelength-routed WDM network cannot be determined, because no performance bound can be obtained for such networks yet. As a different approach, the performance rewards and the reconfiguration costs were modeled as conflicting objectives for wavelength-routed WDM networks [ZhZh02]. However, the applicability of such model is limited, because the associated stochastic searching based algorithm cannot guarantee the convergence and the optimality of the result is unknown. Also, there are some obvious flaws in the formulation, making it questionable whether a meaningful mathematical solution can be obtained.

The impact of limited number and conversion range of wavelength converters has not been fully studied in the context of the WDM network reconfiguration. The previous studies on WDM network reconfiguration assume either full wavelength conversion at all nodes [BoLa05, PrSt04, GeMu03, RiSa02, ZhZh02, BaMu00], or no wavelength conversion at all [MoEr03, SrSi01]. Although in the static RWA, the wavelength conversion only demonstrates a marginal contribution [ZhYa04, SwSi02], the real value of the wavelength conversion has not been demonstrated in the virtual topology optimization. The classic static RWA problem can be considered as a special case of the general WDM network reconfiguration problem without considering the existing network configurations. In [SaLu05], no wavelength conversion capability is considered. This assumption makes the formulation simpler, but the contribution of wavelength conversion is neglected.

As for the methodologies to optimize the Wavelength Routed Networks (WRN), the majority of previous studies use either heuristics [GeMu03, SrSi01, BaMu00, SaLu05], or the branch-and-bound based linear programming relaxation that is provided in the commercial tools such as CPLEX [RiSa02], or other methods such as a stochastic searching based genetic algorithm [ZhZh02]. It has been realized that from the experiments with small-scale networks, no general conclusions can be made about the optimality of the sub-optimal solution for large networks [RiSa02]. Obviously, no performance bound is available for the previous approaches for large networks. Only in special cases, performance bounds have been derived, e.g., for a network with limited tunability [Labo97], a broadcast network [BaRo01]. Meanwhile, the Lagrangean Relaxation and Subgradient Method (LRSM) has demonstrated its potential in solving large-scale optimization problems such as static RWA [SaLu04, ZhYa04, LeYu04], virtual path planning and packet routing in a layered network [SuSo03], two-layer routing problems [HoHe98, PaMa03]. Despite the advances made by the LR and subgradient methods in the static RWA problems [SaLu04, ZhYa04, LeYu04], the reconfigurable network in the real world cannot directly adopt the same approach, because the static RWA assumes that all the existing lightpaths are freely rerouted in the reconfiguration.

1.1.2 Review of Network Grooming Optimization

In Wavelength Division Multiplexing (WDM) networks, the huge capacity of wavelength channels is generally much larger than the bandwidth requirement of individual traffic streams from network users. Traffic grooming techniques aggregate low-bandwidth traffic streams onto high-bandwidth wavelength channels. The design of wavelength-routed Wavelength Division Multiplexing (WDM) networks generally employs a two-step approach [RaSi01]. In the first step, a virtual topology is designed for traffic that is directly carried over lightpaths [DuRo00, LeMe00] and each traffic flow is routed onto the designed virtual topology. Depending on the network architecture, such traffic may be Time Division Multiplexing (TDM) traffic in Synchronous Optical Network (SONET) over WDM networks, or IP traffic in IP-over-WDM networks, or other traffic. A lightpath is a dedicated optical connection using non-overlapping frequency or Wavelength Channels (WCs). The payload information is carried transparently from the source to the destination of a lightpath. The electronic processing of payload information only happens at the source and destination of a lightpath. In the second step, point-to-point virtual links obtained from the first step are routed as lightpaths onto physical fibers and each lightpath is assigned to WCs. If wavelength converters are used, a lightpath may change wavelength along its route. The second step is the RWA (Routing and Wavelength Assignment).

Such two-step approaches were effective for early-deployed WDM networks, because at that time RWA was independent of the virtual topology design. Although two-step approaches reduce the complexity of design and optimization, they do not yield the best network resource utilization.

Integrated design approaches are required to improve the network resource utilization. This is of particular interest to the new generation WDM networks, which consist of network elements (nodes) capable of manipulating traffic in the electronic domain. Such new nodes integrate switching and add-drop multiplexing in the optical domain with traffic de-multiplexing, multiplexing and grooming in the electronic domain [ZhZa03, ZhZh03]. The traffic in the electronic domain can be TDM, IP or other traffic carried directly over lightpaths, depending on the network architecture. In this thesis, our discussion on the grooming network is based on the SONET-over-WDM architecture. So the de-multiplexing, multiplexing and grooming in the electronic domain are with respect to TDM time slots. The terms of traffic flow, traffic demand or simply traffic all refer to TDM traffic. Grooming means the aggregation of low bandwidth connections onto high bandwidth lightpaths. The proposed concept and solution method are generally applicable to many network architectures, e.g., SONET-over-WDM [ZhMu02] and IP-over-WDM [MaPi03].

As an integrated design approach, traffic grooming in WDM networks has received considerable attention recently [Cink03, DuRo02, MoLi01, OuZh03, ZhZa03, ZhZh03]. Traffic grooming consists of four sub-problems, which are not necessarily independent: 1) determining the virtual topology of lightpaths; 2) routing low-speed traffic stream onto the virtual topology; 3) routing the lightpaths over fiber networks; and 4) performing wavelength assignment to the lightpaths. Traffic grooming is studied for two scenarios, i.e., static network planning and dynamic network operation. Static traffic patterns are used in network planning, which means all traffic requests are given in advance. The design and optimization method proposed in this thesis is for network planning. In network operational stage, traffic requests arrive one after another. The best effort is made to provide service to as many traffic requests as possible with given network resources [OkSh03, SrSo04, XiQi03, XiQi03, ZhZh03]. Capacity fairness issue is studied [ThSo01]. Reconfiguring wavelength assignment may be used incrementally for new traffic pattern [ZhRa03]. Most previous studies were conducted based on the assumption that all traffic requests should be accepted and tried to directly or indirectly minimize the network cost

[GeRa00, Hu02a, ZhQi00]. Simply accepting all traffic requests is not a practical approach, because the traffic might be more than the network capacity. Network cost optimization needs to be related to the revenue generated by accepting traffic demands. Although the optimization objective used in [ZhMu02] modelled the revenue as the weighted network throughput, the cost for accepting traffic demands is not considered. Prathombutr *et al.* [PrSt03] use multi-objectives in the traffic grooming optimization. Maximizing traffic throughput and minimizing number of transceivers or lightpaths are modelled as two competing objectives. However, the relation to the network profit is not revealed.

The majority of research for traffic grooming in WRN networks is based on the ring topology because of the relation to SONET networks and the simplicity of the ring topology [ChMo00, ChRo04, DuRo02, GeRa00, SoZh04, ZhQi00]. Different traffic models, network architectures and switching capability have been extensively studied. A number of heuristic algorithms were proposed and theoretical bounds on performance were discovered for ring networks [BeCo03, BiWa03, MoLi01]. Theoretical bounds were also discovered for other regular topologies such as star or tree topology [DuHu03, DuHu03a]. However, emerging SONET-over-WDM networks are increasingly deployed in the mesh topology due to unprecedented efficiency of the mesh topology [CoSa01]. Traffic grooming on the mesh topology is very challenging and only a few recent publications reported such research [AlEl03, BiWa03, GeMu03, WeHe03, ZhZa03, ZhHa03, ZhMu02, ZhMu03, ZhZa03a, ZhZh03]. The traffic grooming problem in a mesh network is NP-complete, where the number of variables and equations increases drastically with the number of nodes and links in a network. Thus, the traffic grooming problem requires dramatic amounts of time to solve for large networks. For practical scale networks, the traffic grooming problem formulated as the Integer Linear Programming (ILP) cannot be directly solved by ILP solver package software such as CPLEX [Hu02]. Even for a small example network of 6 nodes and 8 links, CPLEX cannot compute the optimal solution due to high computational complexity and the result presented in [ZhMu02] is actually a sub-optimal solution by terminating the optimization process after a period of computational time.

A Lagrangean-based heuristic algorithm was proposed in [PaMa03] for traffic grooming in WDM networks. However, the physical topology was not considered, so the constraints from the physical configuration were neglected. Essentially, the problem is simplified to a virtual topology design problem which, as discussed early on, usually does not give the best network utilization when separately solved with the RWA problem.

Some researches (e.g. [ZhZh04]) have used a multi-objective approach. The problem with this kind of approaches is the obscurity in performance definition which makes the comparison between two routing schemes difficult. For example, a penalty based objective function is used to incorporate all the objective functions and then used as the performance measure to evaluate the network routing scheme [ZhZh04]. This creates a dilemma (conflicting causal effect) that the optimization procedure is not optimizing the problem according to the measure defined. One solution to this dilemma is to use the penalty-based evaluation measure as the objective function.

Two types of optimization problems have been studied for traffic grooming under static traffic patterns. The first optimization type is to minimize network cost [ChRo04, DuRo02]. In minimizing network cost, a given static traffic matrix is completely satisfied and the optimization objective is to use minimal network resources to provide services to all traffic demands. Fundamentally, minimizing network cost is to allocate network resources to match a given set of traffic demands. Various cost functions are adopted, e.g., [ChRo04, DuRo02t]. For example, the total number of lightpaths is used as a cost function [DuRo02t], which represents the total cost of transmitters and receivers, as well as wavelength switching fabric. Another cost function is the total capacity of SONET switches [DuRo02t], which represents the cost of electronic multiplexing, de-multiplexing and switching, and indirectly represents the cost of transmitters and receivers. Similarly, the maximum number of lightpaths terminating/originating at a node serves as another cost function, which reflects the cost of transmitters and receivers, as well as the cost of SONET switches.

The second optimization type is to maximize network throughput [ZhMu02]. In maximizing network throughput, a given network has such limited resources that not all given traffic demands can be accepted, and therefore the optimization objective is to maximize the total traffic that is carried by the network. Essentially, to maximize network throughput is to prioritize traffic demands based on how efficient the traffic demands can be provisioned and only the traffic demands that most efficiently use network resources are accepted. Maximizing the total network throughput implicitly evaluates the efficiency of traffic flow provisioning. A variation of maximizing network throughput is network revenue optimization, where the network revenue is modelled as a weighted network throughput [ZhMu02] or based on service differentiation for lightpath protection [SrSo01]. Since for a given network, the total network cost is fixed, maximal network revenue will lead to maximal network profit. However, it is difficult to study the impact of network cost factors to the overall network profit, because the network cost factors are not explicitly modelled. In maximizing network throughput, it is still unknown whether providing service to a certain traffic demand is profitable or not. The reason is that an arbitrarily selected network configuration will unlikely match a given traffic matrix and such miss-matching has two negative impacts. In some cases, non-profitable traffic demands are accepted because the given network resources are abundant. It is also possible that potential profitable traffic demands cannot be accepted because the given network resources are less than necessary. By adding small amounts of network resources with relatively low cost, the potential profitable traffic can be actually accepted to generate profit. Unfortunately, the network throughput optimization does not provide clues for improving network configurations, i.e., no guide is provided for adding network resources.

1.2 Mathematical Background

The network optimization problem is generally formulated as an ILP problem. In this section, we will introduce the Lagrangean relaxation technique, the subgradient method and many other mathematical approaches that are extensively used in solving ILP. The basic mathematical background for optimization theories can be found in [Bert99].

1.2.1 Lagrangean Relaxation

Optimization problems with convex, but not necessarily differentiable, objective functions are called *nondifferentiable optimization* problems. Due to the wide applications in various fields, these problems received extensive attention. Lagrangean duality is one of the important areas of application of *nondifferentiable optimization* techniques. The concept of *Lagrangean duality* was first used by Falk [Falk67] for nonlinear programming by applying *Lagrangean relaxation* (LR). Geoffrion then accomplished some important work to this subject in his seminal work on LR [Geof70, Geof70a]. Held and Karp also used the *Lagrangean duality* concept [HeKa70] to solve the traveling salesman problem, which is a discrete optimization problem. The same methodology was also employed to solve the routing and capacity assignment problem in the packet switching networks with various constraints [GaNe89, YeFr01].

In most of network routing and capacity allocation problems, the integer constraint results in the NP-complete problem complexity. Various famous integer programming problems, such as scheduling problem, traveling salesman problem, have been successfully solved by Lagrangean relaxation method [Fish81]. In these applications, by using the LR technique, the original problem can usually be decomposed after being relaxed into a few independent subproblems (with respect to the decision variables). The independent subproblem is generally easier to solve than the original problem because of integrality of its decision variables. Hence, the exponential increase in the computation complexity for the integer programming problem would be effectively eliminated by applying LR method, which makes it suitable for solving integer programming problem. Besides decomposition, Lagrangean relaxation method generates a legitimate lower bound to verify the solution quality of the upper bound calculated by the proposed algorithms. The main topics that arise in LR methods are the dualization strategies, the solution processes, and the primal recovery techniques. More specifically, due to the existence of duality gap, dualization schemes are affected by the quality of the lower bound obtained by solving the *Lagrangean duality* problems. The general background, *Lagrangean duality* theories, of the LR technique can be found in [Bert99]. As compared to linear programming relaxation, the bounds provided by Lagrangean relaxation are better in many instances [LeeS99]. In addition to these integer programming problems, Gavish solved a series of network planning and capacity management problems by using the LR method [GaNe89, GaHa83, GaNe92]. In these works, Gavish showed that LR method is not only effective and efficient in solving integer programming problem, but also in solving non-convex programming problems.

Recently, there has been a lot of attention on optimization methodologies to solve the optical network optimization problems. The methods include the branch-and-bound method [KuPu03], LP (Linear Programming) [WiVe03], and LP-relaxation [KrSi01]. Branch-and-bound and LP-relaxation (not LP) are used to solve the integer problems by enumerating partially the routing and wavelength assignment schemes. Because the number of possible routing schemes generally grows exponentially as the problem size increases, these methods become very computational intensive for even small-sized networks.

1.2.2 Subgradient Method

When applying Lagrangean relaxation to integer programming problems, methods such as the subgradient, bundle, and cutting plane [Bert99] are often used to maximize the dual function since the dual function is polyhedral concave and nondifferentiable. Subgradient method, originated by Shor [Shor64] and further explored by Polyak [Poly69], is the most widely used method to optimize the nondifferentiable *Lagrangean dual*, where the subgradient direction is obtained after all the subproblems are minimized, and the multipliers are updated along this subgradient direction. Some convergence conditions of the famous subgradient method are provided in [Poly69] and [Ermo66]. Bundle method can provide better directions than the classical subgradient method. However, to obtain a better direction may require minimizing all the subproblems many times [HiLe93]. The cutting plane method often tends to converge slowly, even for problems where the Lagrangean dual function is polyhedral.

The subgradient method is a simple algorithm for minimizing a nondifferentiable function, which may be viewed as an extension of the gradient and gradient projection methods [Bert99]. The method is very much like the ordinary gradient method for differentiable functions, using subgradients as directions to reduce the distance to the optimum. However, unlike the gradient, which is always in the steepest ascent direction, subgradient is not a "descent direction", which means the new iterate may not decrease the function value. However, under certain conditions, such as when the stepsize is small enough, the distance of the current iterate to the optimal solution set can usually be reduced, so that the solution converges to the optimum. On the other hand, the subgradient method uses stepsizes that are fixed before the computation of the subgradient, instead of a precise or approximate line search as in the gradient method. The subgradient method is far slower than Newton's method, but is much simpler and can be applied to a far wider variety of problems. By combining the subgradient method with primal or dual decomposition techniques, it is sometimes possible to develop a simple distributed algorithm for a problem [Bert99].

As classical textbook knowledge, the advantages of the subgradient method are the efficient computation, the intuitively clear idea, the straight-forward implementation, and proved robustness and effectiveness [Bert99]. However, it is known that subgradient methods are quite promising for solving the dual problem but, in general, the optimal solution of the subproblem is not feasible to the primal problem. [Sho85] and [Bert99] provide detailed proofs and references on the subgradient method and primal or dual decomposition.

1.2.3 Other approaches

Besides the Lagrangean relaxation method, there are many other mathematical approaches or tools used in the optical network optimizations, which include LP [WiVe03], LP-relaxation [KrSi01], and branch-and-bound [KuPu03], etc. Linear Programming (LP) is the mostly commonly used form of constrained optimization. The objective of an LP problem is to maximize or minimize a linear function of the decision variables. Each constraint must be a linear equation or a linear inequality, and a restriction of sign is associated with each variable. The variables are assumed to have real-valued, i.e., they can take on fractional values [WiVe03].

There are basically two methods used to solve LP problems [Bert99, Chin04]. One is the simplex method, which is a basic methodology to solve LP problems. The other method is the interior-point technique, which is newer and can be used for extremely large problems. Simplex method proceeds along the exterior of a feasible region, moving from vertex to vertex, while the interior-point method moves along the interior of the feasible region, and not visiting vertex at all until the final solution [Bert99].

Many optimization problems, in addition to the usual equality and inequality constraints, involve integer constraints. An Integer Linear Programming (ILP) problem is an LP in which part or all of the variables are required to be non-negative integers. If all variables are required to be integers, it is called a pure Linear Integer Programming (pure ILP) problem. If some variables are required to be integers and others are allowed to be either integers or non-integers, it is called mixed integer programming problem (MILP). An ILP in which all the variables must be 0 or 1 is called a 0-1 ILP.

ILP is a combinatorial optimization problem, which is much more difficult to solve than LP. There is a vast temptation to just solve the problem by standard LP, and then to round any non-integer variable values to the closest integer value. However, it can be easily proved that we cannot use real-valued solution methods to optimize integer-valued models, because they might result in some infeasible solutions that are far from the optimum or are not feasible. Another way to get the feasible optimum for the ILP is to enumerate all possible solutions and then choose the best solution. The original/basic branch-and-bound [Bert99] approach is to enumerate all possible solutions to obtain the optimum for the ILP. The method is based on the observation that the enumeration of integer solutions has a tree structure [Bert99, Chin04]. The solution based on enumeration in the tree structure can be very time-consuming in general and can therefore solve problems of a very small size. This method, however, is in principle capable of yielding the exact optimal solution [Bert99] if run to completion. Three important techniques used in the branch-and-bound approach are branching, bounding and pruning. Branching is to select a node for further growth and create the next generation of children of that node. Bounding is to estimate the bound on the best value obtained by growing a node. Pruning is to cut off and permanently discard nodes when it can be shown that the node and its descendents will never be feasible or optimal. Pruning prevents the search tree from growing too large.

The concept of LP-relaxation of an ILP problem plays a key role in the solution of ILPs. It is one of the most widely used methods to solve the complicated ILP optical network optimization problems [BaMu00, KrSi01, SwSi02] and can be used together with many other techniques, such as branch-and-bound or randomized rounding. The LP obtained by omitting all integer constraints on some decision variables is called the *LP-relaxation* of the ILP. Typically in applying LP-relaxation, the partially relaxed problem is first solved by using classic branch-and-bound (to solve the integer variables) and LP (to solve the relaxed variables). Then a heuristic algorithm is launched to obtain a feasible solution rounding the relaxed linear variables into integers. Any ILP may be viewed as the LP-relaxation plus additional constraints (the constraints that state which variables must be integers). The feasible solution space for any ILP problem is contained in the feasible solution space for the corresponding LP-relaxed problem. Therefore, the optimal solution of the LP-relaxed problem poses a lower bound to the minimization ILP problem (upper bound to the maximization ILP problem). If we solve the LP-relaxation of an ILP and obtain a solution, in which all required integer variables are integers, then the optimal solution to the LP-relaxation is also the optimal solution to the ILP [WiVe03]. LP-rel.axation technique can reduce the computation complexity to solve the ILP problem by reducing the searching space, so the size of the problems solved by LP-relaxation (to obtain a sub-optimum solution) can be larger than those solved by ILP (to obtain the exact optimum) [KrSi01]. However, LP-relaxation cannot eliminate the enumeration of the values for integer variables, and it has generally an exponential computation complexity. Therefore, the optimization algorithm still does not scale well as the network size increases.

1.3 Motivations

In static and semi-dynamic routing, the traffic requirements in the WDM network are given as a set of point-to-point connections of different granularities between pairs of access stations. It is extremely important for network service providers to use the network resource with respect to the given traffic demand in an optimized way, and therefore the profit for the existing network configuration can be maximized. These connections quite often remain in place for a relatively long period of time; so it is worthwhile to optimize the way, in which the recourses in the network are allocated to the traffic demand with respect to a given performance measure, even if the optimization may require a considerable amount of computation.

As discussed in previous sections, to maximize the profit for the existing network configuration, it is very important for network service providers to optimize the allocation of network resources to satisfy the demand of the customers. With the same configuration (thus the cost), the profit from an optimized scheme to utilize the network resources differentiates drastically from that from an un-optimized scheme. The goal of maximizing profit is the ultimate target of the network service provider, and this goal fuels the study on the optimization methods for the network actions, such as the optimal assignment of the wavelengths and the wavelength converters.

Constrained optimization problems are generally used to describe the design target and the resource constraints of the network. Ideally, for a given objective function and given constraints, the exact optimum is obtained, and the network should be configured as what is provided in the best optimized scheme. However, the computation facilities are normally limited compared with the bandwidth and QoS demand from the users. The computation complexity to obtain the optimal solution is extremely high, and for most of the cases, impractical for large networks. The only option is to obtain a sub-optimum solution in a reasonable computation time. Various mathematical optimization and heuristic methodologies are therefore used to obtain the sub-optimum solutions. Most of the heuristic methods do not provide the bounds to the problem being optimized. Some other methodologies, such as branch-and-bound and LP-relaxation, do not scale well with the network size (thus the size of the optimization problem). It is important to provide an optimization framework that has good characteristics such as, good scalability with the network size, providing tight bounds, fast (reconfiguration) computation, extensive theoretical background (for such as the convergence condition and bounds) and easy implementation, etc. On the other hand, lots of important problems, such as the fairness issues and the network reconfigurations, are not well studied, due to the lack of mathematical tools.

The objective functions of the optimization vary largely from case to case, depending on the network design goals and the network operation goals. In most of the cases, the network optimizations either optimize the network operations in a given network configuration, or optimize the network planning with respect to a given traffic. There are only seldom papers considering both situations (e.g., [AlAy99]). The most commonly used network optimization objective functions are those with the direct physical meanings, which include the number of wavelengths used, the average hop numbers, the total number of add-drop multiplexers, and the total number of connections routed, etc. In most of the cases, these objective functions provide good help in the network planning, but do not have clear relations with the network revenues. Only some researches minimizing the cost of some network equipment can provide the direct ideas about the cost of the network planning. There are also many abstract objective functions ([OzBe03, StBa99, YeLi01a, ZhYa03]) that do not have direct physical meanings. These objective functions are generally penalty-based, such as the rejection penalty, and the summation of several penalty items. The penalties can be understood as a virtual indirect 'cost' for the network investor or operator, which not necessarily result in the direct money saving. However, the 'virtual' cost will eventually results in less revenue or higher cost through some indirect way. To have a better understanding of the virtual cost, let us consider a network with limited resources charging the network users by the bandwidth they use. Assume that the maximum bandwidth that can be allocated to the network users is B, and N network users all require bandwidth B. Then we can compare the following two bandwidth allocation schemes: 1) all bandwidth is allocated to a single user, and 2) every user is granted with B/N bandwidth. Generally, the second bandwidth allocation scheme is a better scheme, because it is fairer. The unfairness created by the first scheme might eventually cause the network operator losing money in the competitive market, due to the grudge from the network users. Therefore, it is very reasonable to estimate a virtual cost for the unfairness created by the routing scheme, and optimize it in a way that will eventually results in a fairer solution.

In many cases, the capacity of wavelength channel is larger than bandwidth requirement of individual traffic streams from network users. This will result in the inefficient usage of the network bandwidth. For example, if the bandwidth requirement between two network nodes is very small compared with the bandwidth provided by the lightpaths at the optical domain, to set up the connection in an all-optical WDM network, a whole lightpath has to be employed to transmit the low-capacity traffic of the connection, and a large proportion of the bandwidth in the lightpath is wasted. To have a better utilization of the network capacity, the electronic switching technique has to be employed on the top of the wavelength routing WDM network, so that many low-speed traffic streams can be bundled and routed together to use a higher percentage of the lightpath capacity, i.e., traffic grooming. Among the various types of traffic grooming technologies, SONET-over-WDM is the most widely implemented technology. As discussed before, the majority of research for traffic grooming in networks is based on the ring topology because of the relation to SONET networks and the simplicity of the ring topology. However, emerging traffic grooming networks are increasingly deployed in the mesh topology due to unprecedented efficiency of the mesh topology. Traffic grooming on the mesh topology is very challenging and only a few recent publications reported such research. Although heuristic algorithms were proposed and the efficiency of the heuristics was demonstrated for a practical scale network (e.g. [ZhMu02]), it is still an open issue as to how close the results obtained from the heuristics are to the optimal results, i.e., what is the theoretical bound of the performance metrics. No theoretical performance bound is available yet. The research on the mesh topology is to be extended.

The traffic grooming ILP problem is NP-complete, and the number of variables and constraints increases drastically as the number of nodes and links in a network increase. Thus, the traffic grooming problem requires dramatic amounts of time to solve for large networks using traditional tools, such as CPLEX. On the other hand, heuristic methods generally can solve problems in a relatively short time, but their performances vary from case to case, while their results cannot be evaluated by demonstrating how far the results are from the optima. The optimization method, instead of simple heuristic algorithms, that can solve large networks with better insured performance and acceptable computation complexity is urgently needed.

Finally, in the traffic grooming network, maximizing the weighted network throughput does not necessarily mean the real profit is maximized, because the profit is the excess of revenue over cost. Accepting a traffic stream that can only generate less revenue than the cost of using the network resource is not acceptable to the network service provider. Multi-objective approaches (e.g. [PrSt03]), considering competing objectives, cannot reveal the relation of the revenue and cost to the network profit. Traditional single-objective approaches (e.g. [ZhMu02]), on the other hand does not consider both revenue and cost. So unlike the common single-objective approach, the potential profit generated by accommodating traffic demands can be optimized by integrating both cost and revenue in the same objective function. A good formulation with a reasonable penalty function with parameters that can be easily adjusted is essential for the network operators to find a good scheme to ensure the maximum profit (sometimes in the sense of a long term) from the using of the network resources. However, the selection of the particular penalty values is generally a subjective matter and in close relation with the specific consideration of the network operations, and is thus outside the scope of this thesis.

In WRNs, lightpaths may have to be quickly reconfigured in reaction to traffic pattern changes, network failures or new network resource installations. Pure static network optimization problem assumes that for every reconfiguration, all the existing connections in the network are free to be reconfigured, which is very impractical when used in the real network operation, because the reconfiguration in the optical layer will disrupt the upper layer traffic and eventually results in poor quality of service. This leads to the semi-dynamic network optimization problem, which considers the correlation between the existing network configuration and the projected network configuration. There has not been any satisfying work that can provide a good reconfiguration optimization tool and at the same time, provide a good performance measure. It is of great interest to develop some mathematical tool to facilitate this purpose and to provide a proper formulation to precisely describe the reconfiguration problem. Also, previous research (e.g. [SaLu05]) does not consider many useful design objectives such as congestion and fairness. Our study is motivated to tackle this challenge and provide a solution to study tradeoffs between the reconfiguration and the rejection/congestion. To relate the current session and the previous session in the optimization, two rules shall be adopted to ensure the stability of the network, i.e., between any node pair, 1) the capacity allocated has to be more (or equal to) the existing capacity, if there is more (or the same) capacity demand in the new session. 2) the capacity allocated has to be equal to the new demand, if there is less capacity demand in the new session. In the objective function, we want to optimize the demand rejection, the reconfiguration, and the network congestion in the network, as well as the fairness concern.

1.4 Objectives

Our general objective of the network optimization is the ultimate goal of most network investors, which is to maximize the revenue, and to minimize the cost. At the same time, we want to study various network behaviors in the network operations. Specifically, we would like to:

1) study the WDM network optimization problems and provide formulations that can be mathematically solved.
2) set up a solution framework (to our formulations) that is scalable to large networks, while providing a method to evaluate the quality of any sub-optimal but acceptable results. This solution framework should have a low computation complexity, a high efficiency and good adaptability to different kinds of network constraints and design objectives.
3) explore area not studied before due to the lack of mathematical tools, including the fairness issue of accepting demands in the all-optical networks, the maximizing-profit problem in the traffic-grooming network, and the performance measure in dynamic reconfiguration problems.
4) quantitatively demonstrate the advantages of having wavelength conversion in WDM networks for simplifying the network operations, administration and management, which have been speculated in the literature (e.g. [KaAy98]).
5) study the influence of design and network parameters on the network behavior to provide guidelines for the network planning and operations; study the trade-offs between different network parameters

1.5 Approaches and Methodologies

To accomplish our objectives, we would like to focus our study on the static and the semi-dynamic RWA, as well as traffic grooming optimization problems in a backbone transport WRN network, rather than the access networks. For simplicity, we shall refer to both static RWA and semi-dynamic RWA problems as RWA optimization problems in the rest of the thesis.

Same as most of the studies on network optimization, we shall propose new integer linear programming (ILP) formulations to facilitate our mathematical solution procedure. We shall also provide a general optimization algorithm to solve the proposed problems. Generally, to obtain the exact distance from the obtained solution to the optimum is very computational expensive, because it requires the exact solution of the problem being optimized. Therefore, we would like to use performance measures that can efficiently estimate the sub-optimality and the computation complexity of our optimization algorithms. The computation complexity can normally be obtained through mathematical analysis and verified through calculating the running time on the computer. The computation complexity of an algorithm can generally be classified into three classes, i.e., NP-hard, NP-complete, and polynomial [DuHu03a, PaRa88]. The algorithms with NP-hard or NP-complete computation complexities do not scale well as the problem size increases, while the algorithms with polynomial computation complexity generally scales much better with the problem size. We shall provide algorithms with polynomial computation complexity against the existing NP-complete/NP-hard algorithms, while providing better results.

We shall assemble various solution approaches into a framework that would accommodate various objective functions and constraints, so that it can be applied to diverse network models and network optimization problems. For example, by appropriately relaxing some of the constraints in the formulation using Lagrange multipliers, a separable dual problem may be obtained. We want to investigate how the solution to the dual problem decomposes into a finite number of subproblems that can be solved with polynomial complexity, even though the complexity to obtain the exact optimum of the original problem is generally NP-complete or NP-hard.

We shall apply our framework to various network architectures and assumptions. We want to show the excellent performance and good adaptability of our solution framework. We will also compare our methodology with other existing work to show the big improvement in terms of both computation complexity and results in generalized network topologies.

In the virtual topology design for an all-optical WRN network, we shall study two network models with different objective functions. The influences of the number of the network resources and different wavelength converter structures will also be studied. Unlike the path-based formulation (e.g. [SaLu02]), we shall use a link-based formulation, because the pre-selection of paths (in the path-based formulation) itself is a complicated optimization problem, which makes the implementation very difficult. We would like to simplify the implementation of the algorithm and demonstrate high efficiency of our framework when compared to other newest research using LP-relaxation [KrSi01, SwSi02].

We would also like to study the fairness issue, which to the best of our knowledge has not been done before on similar problems. We shall provide the formulation and solution to fairness problem within our LR and subgradient framework. The definition of the fairness is very network-operation-dependant. So we shall provide in this thesis a simple definition, which is how evenly the traffic acceptance is distributed among the source-destination pairs that have traffic demands ((s, d)) (the details of the definition will be provided in Section 3.2.1).

We will also explore the reconfiguration problems in the WRN networks, using our framework. We would like to consider a penalty-based objective function that would consider various factors during the network reconfiguration. Although WDM network reconfiguration is fundamentally reactive traffic engineering, we want to achieve better long-term performance by integrating the proactive traffic engineering (load-balancing improvement) into the reactive traffic engineering. Specifically, we would like to use the objective function to spread the traffic into the network in a balanced manner in a proactive approach, so that critical network resources can be preserved for future traffic demands.

For the networks with small-capacity connection demands, we shall propose a network model using traffic grooming technology. We would like to take into account many factors, such as the optical routing cost, the electronic routing cost, and the rejection penalty, integrating various practical constraints in our formulation. We want to simplify our solution by decomposing the dual problem into RWA subproblems and the traffic routing subproblems. We would also like to study the influence to the optimization result using many parameters.

We shall include the revenue penalty for rejecting a traffic demand in our objective function, in order to incorporate the revenue factor into the network cost function. We shall assign the amount of revenue penalty to be the same as the amount of revenue the traffic demand generates, if it is accepted, so that the actual optimization objective becomes minimizing the summation of the revenue penalty and the total network cost (, which is equal to minimizing profit loss or maximizing profit). Unlike any of the previously used indirect cost functions, we shall use a direct measurement of the total cost, similar to the monetary value in the presence of revenue penalty. The details of the formulation will be presented later in the thesis.

We shall use C++ as our software development platform, because it has much higher code execution efficiency than Java, whose cross-platform compatibility does not benefit our study. When compared with C, the Object Oriented structure of C++ provides a very good modeling system of the real world entities. C++ also facilitates the reuse of the developed classes, which makes the continuous code development much more efficient. Other tools such as OPNET or CPLEX have also been considered, but they cannot be used to control and verify the correctness of every step of our optimization algorithm. Since we are developing our own optimization software, we would like to implement consistency check at difference stages, such as verifying the constraint satisfaction and comparing with the existing results.

1.6 Contributions

The following are the contributions of this thesis:

1) A new optimization framework called the Lagrangean Relaxation and Subgradient Method (LRSM) for optical network optimization problems. It has an excellent adaptability and a high efficiency, which can at the same time provide a tight bound and a good (near-optimal) solution. This is the first time such systematical approach is introduced to optical network optimization problems.
2) A new *link-based* formulation for the RWA problem, considering simultaneously the fairness issue and limited wavelength conversion constraints in our formulation, which to the best of our knowledge, has not been undertaken yet.
3) A generalized optimization formulation for traffic grooming mesh networks, which considers both the cost of using network resources and the revenue generated from accepting traffic streams.
4) A bound on the result for the traffic grooming mesh network optimization problem.
5) A new formulation and solution to study the WDM network reconfiguration problem, which can provide a good performance measure while generating excellent results with high efficiency. A quantitative study on the performance gain of using limited wavelength converters in the WDM network dynamic reconfiguration problem has been carried out.
6) Discovery of the wavelength converter's real value (precisely, the wavelength converter has marginal contribution in the static scenario) of in the virtual network optimization problem in the semi-dynamic scenario.
7) The trade-off study on the semi-dynamic reconfiguration and rejection/congestion/ fairness.

8) Demonstration of the influence of various network and traffic parameters on the network behaviours, as well as the superiority of our solution to other newest research through our performance evaluation of some 'real' networks.

To the best of our knowledge, Contribution 2, 4 and 5 have not been achieved before.

1.7 Organizations

In Chapter 2, the network operations and network models used in the rest of the thesis are presented. We study the WRN with full-range and limited wavelength conversion in Chapter 3, while providing the basic LR and subgradient framework. We first study the problem of minimizing the overall routing cost. The WRN model with limited wavelength conversion degree is then discussed. A formulation to maximize the routed traffic addressing the fairness concern is also provided. The results and computation complexity are compared with some other existing research. Chapter 4 covers the more generalized network models with traffic grooming capability. The objective in this chapter is to maximize the network operation profit, taking into consideration both the network cost and revenue. The overall problem is decomposed into two subproblem sets, which are solved independently. In Chapter 5, the semi-dynamic reconfiguration problems are studied, which considers the existing network configuration at the reconfiguration. The formulation incorporates the accepted traffic, the congestion and the reconfiguration, while inheriting the fairness consideration from Chapter 3. Chapter 6 provides some design guidelines and recommendations for the practical implementation of our framework. In Chapter 7, we conclude the work of this thesis.

1.8 Publications

The following is a list of publications related to this research work:

[1] Yiming Zhang, Oliver Yang, Jing Wu, and Michel Savoie, "Connection Routing Optimization in A Semi-Dynamic Reconfigurable Network", submitted to *IEEE Journal On Selected Areas In Communications*

[2] Yiming Zhang, Oliver Yang, Jing Wu, and Michel Savoie, "Lightpath Reconfiguration in a Semi-Dynamic Reconfiguration Network", to be presented on *IEEE Globecom'06*, San Francisco, USA, 27 Nov. – 1 Dec., 2006

[3] Yiming Zhang, Jing Wu, Oliver Yang, and Michel Savoie, "Network Profit Optimization in Traffic-Grooming Networks", submitted to *Journal Of High Speed Networks*

[4] Yiming Zhang and Oliver Yang, “A Proposal for a Semi-dynamically Reconfigurable Optical Network Optimization”, *Proceed. of SPIE Photonic North*, 12 -14 September, 2005, Toronto, Canada

[5] Yiming Zhang, Jing Wu, Oliver Yang, and Michel Savoie, “A Lagrangian-Relaxation Based Network Profit Optimization for Mesh SONET-over-WDM Networks”, *Journal of Photonic Network Communications*, pp. 155-178, September, Vol. 10, 2005

[6] Yiming Zhang, Oliver Yang and Haomei Liu, "A Lagrangean Relaxation And Subgradient Approach For The Routing And Wavelength Assignment Problem In WDM Networks", Nov. 2004, *IEEE Journal On Selected Areas In Communications*, pp. 1752- 1765, Vol. 22, Issue 9

[7] Yiming Zhang, Jing Wu, Oliver W.W. Yang and Michel Savoie, “Lagrangian-Relaxation Based Mesh Traffic Grooming For Network Profit Optimization”, pp. 406-410, *IEEE Proceed. of Workshop on High Performance Switching and Routing*, Hong Kong, P. R. China, 12-14 May, 2005

[8] Yiming Zhang and Oliver Yang, "The optimization issues in an agile all-photonic backbone network", pp. 1233-1244, *Proc. SPIE APOC (Asia-Pacific Optical Communications)*, 7-12 Nov., 2004, Beijing, China

[9] Yiming Zhang, Oliver Yang and Haomei Liu, "A Lagrangean Relaxation Approach To The Maximizing-Number-Of-Connection Problem In WDM Networks", pp. 23 – 28, *Proceed. of IEEE Workshop on High Performance Switching and Routing*, Turin, Italy, June 24-27, 2003

[10] Yiming Zhang and Oliver Yang, "An Effective Approach To The Connection Routing Problem Of All-Optical Wavelength Routing DWDM Networks With Wavelength Conversion Capability", pp. 1370 - 1374 vol.2, *Proceed. of IEEE ICC* (*International Conference on Communications*), Alaska, USA, May 11–15, 2003

[11] Yiming Zhang, Oliver Yang and Haomei Liu, "An Algorithm for the Routing and Wavelength Assignment Problem in WDM Networks”, pp. 935 - 938 vol.2, *Proc. IEEE CCECE* (*Canadian Conference on Electronic and Computer Engineering*), 2003, Montreal, Canada

[13] Yiming Zhang and Oliver Yang, "A Lagrangean And Subgradient Method For The Opti misation Of Wavelength Routing Network Design", *Proceed. 21st Biennial Symposium on Communications*, Queen's University, Kingston, Ontario, Canada, June 2 - 5, 2002

Chapter Two

Network Operations, Models and Assumptions

In this chapter, we will provide the detailed description about the operations of the network we study, as well as the network models and assumptions that will be used in the mathematical formulation in this thesis.

2.1 Wavelength Routed Network

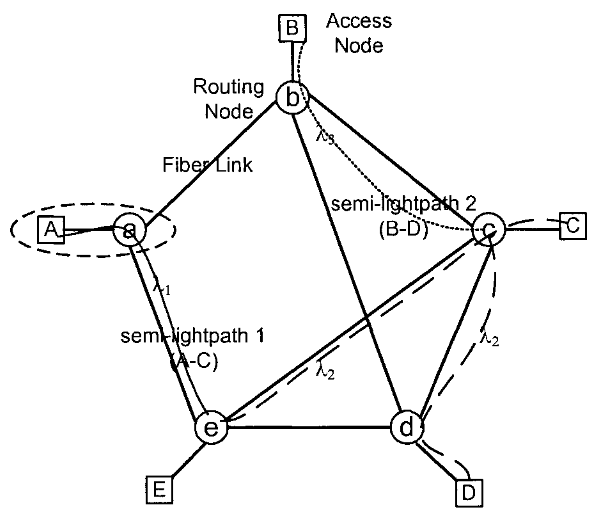


Figure 2.1. All-Optical WDM Network Consisting Of Routing Nodes Interconnected By Point-To-Point Fibers

We consider a generalized mesh Wavelength Routed Network (WRN) using WDM technology with N nodes, L links, and W wavelengths (See Figure 2.1). Our WRNs are capable of static or semi-dynamic *wavelength selectivity* at the network nodes and allow many simultaneous connections through optical multiplexing and multi-access techniques. *Wavelength selectivity* means the optical signals of different wavelengths in the same fiber link can be switched to different destinations on a network node (with optical cross-connect). With this feature, the network nodes in our network allow spectrum reuse and hence can improve network performance through the use of appropriate connection control algorithms. Our networks are capable of supporting general physical topologies with/without wavelength conversion capability.

Comparing with the traditional networks with only electronic switching, the WRN network using WDM technology has a simpler architecture, lower delay and higher throughput, which facilitate the user access to the huge bandwidth provided by the WDM technology with cheap price and high agility. These good attributes makes the WRN network an ideal candidate for the backbone transport networks.

For each node i in the WRN, there are T_i transmitters, R_i receivers, and F_i wavelength converters. The distance/cost of a fiber link between two nodes i and j is d_{ij}. The bandwidth of one Wavelength Channel (WC) is C, and the granularity of the traffic demands is G_{pqz} (the zth traffic demand between nodes p and q). The traffic demand matrix in the network is given as $\{N_{sd}\}$. Other parameters are to be specified as needed in other chapters.

2.2 Network Operation

The following are a few fundamental but important network operations in our RWA networks.

2.2.1 Operation of Wavelength Division Multiplexing

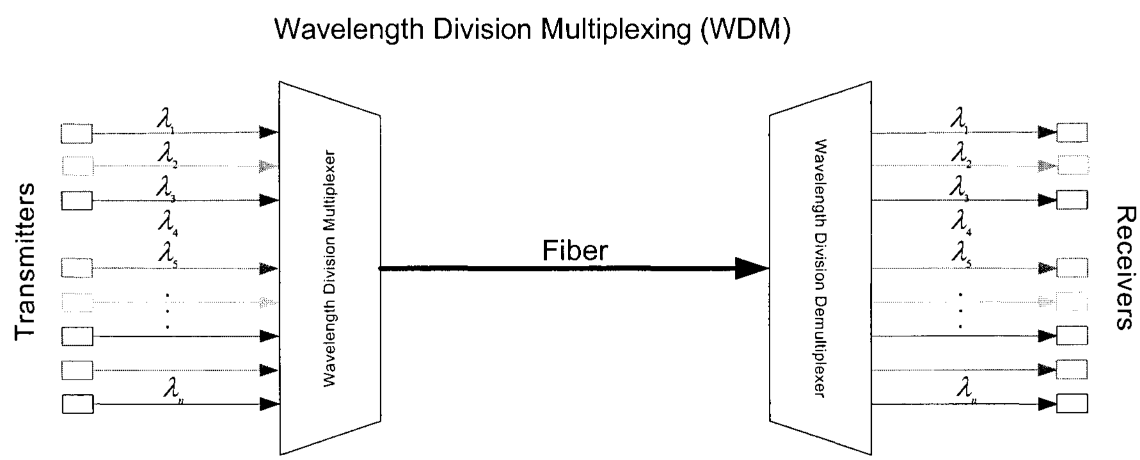


Figure 2.2. The Wavelength Division Multiplexing

WDM technology divides the huge optical bandwidth available in fiber into lower-capacity wavelengths. The wavelengths in fiber can be considered as separate non-interfering channels. A graphic description of WDM technology is shown in Figure 2.2. The multiple wavelengths (λ_i) from the transmitters are multiplexed and sent over the fiber to the destination. The wavelengths are then de-multiplexed and recovered to the original signals. The optical signals in different frequencies are transmitted separately and multiplexed in to the same fiber without interfering each other. At the de-multiplexer, the optical signals can be de-multiplexed, which means the optical signals of different frequencies are separated. A wavelength channel (WC) on a fiber, which allows the transmission of a signal of the corresponding frequency, is considered to be a logical virtual channel between the network nodes that the fiber connects, and this logical channel is only available to the signals of the corresponding frequency. In our WDM network model, we do not consider the physical nonlinear effects such as the signal degradation and signal interferences. In the current practical WDM networks we limit the network configurations as specified by the International Telecommunications Union (ITU) [BrPo98].

2.2.2 Operation of Wavelength Switching

Our network nodes in WDM networks are generally equipped with optical cross-connects (OXCs) devices, which can optically switch a signal on a given wavelength from any input port to certain output ports on different wavelength ports. Based on this approach, two nodes in our network are connected logically through optical channels (lightpaths) defined to be a concatenated sequence of wavelength channels of the same wavelength (color) [ChGa93]. Further flexibility can be achieved if wavelength converters are allowed at the nodes so that several lightpaths of different wavelengths (colors) are chained together to form a *semi-lightpath* (denoted by s_{sdn}) [ChFa96] (See Figure 2.1). We shall consider in this thesis a WDM network with wavelength converters so that one connection demand between any two nodes (and in particular between a source node s, and a destination node d) is equivalent to one semi-lightpath. Note that there could be multiple connection demands between any source-destination pair, and therefore there could be more than one semi-lightpath being set up between them. The individual wavelength channel (WC) provides huge capacity (to OC-192/10Gb/s and beyond), and the semi-lightpath, formed by chaining together many WCs, connects two nodes that are physically apart with the bandwidth equal to a whole WC, thus makes the two nodes in our network virtual neighbours.

2.2.3 Operation of Network Reconfiguration

The WDM network reconfiguration is an off-line optimization process, where traffic demands are processed in batches. For every batch of traffic demands, a RWA reconfiguration session is conducted. Then the network does not establish or remove any lightpaths for a period of time until the next reconfiguration session. New traffic demands are accumulated and handled in the next reconfiguration session. The accepted traffic demands in the current session may be reduced in the next reconfiguration session because of a traffic pattern change. In every reconfiguration session, traffic demands are selected based on the overall network resource availability, the demand situation and the optimization objective. To provision for the accepted/selected traffic demands, some established lightpaths may have to be rearranged. Further details can be found in Chapter Five.

2.3 Component Operation

In this section, we introduce the operations of individual network components that will be used in our RWA network models.

2.3.1 Operation of Wavelength Add/Drop Multiplexer

Each network node has a Wavelength Add/Drop Multiplexer (WADM) that can modify the signals of different wavelengths on fibers, i.e., it adds/drops a small number of wavelengths and let pass most wavelengths with no processing. The signal of a certain wavelength is *dropped*, when it is terminated optically in the WADM. On the other hand, a signal of a certain wavelength is *added*, if this signal is injected to the fiber at this node. The interface between our optical system and our electronic processing system (i.e., converting the dropped wavelengths to electronic signals and converting the electronic signals to optical signals for transmission) is called an Optical Line Terminal (OLT), which contains the transceivers and the WADM equipments.

Our Optical Cross-Connect (OXC), sometimes called a wavelength cross-connect, routes each signal of a certain wavelength from an input port to one or more selected output ports, independent of the other wavelengths. The typical generalized $N \times N$ OXC of W wavelengths is made of W independent switches. All switches are preceded by a wavelength demultiplexer and ended by a wavelength multiplexer, as shown in Figure 2.3. If some of the wavelengths are terminated/ originated at the node that is connected to our OXC, i.e., those wavelengths are to be dropped/added at the node, our WADM has to be used together with the OXC.

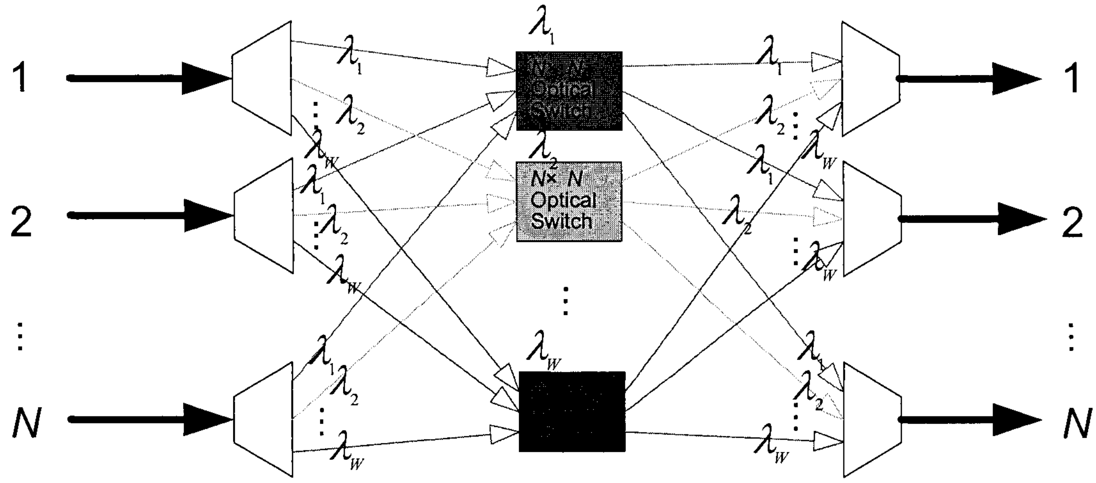


Figure 2.3. A $WN \times WN$ Crossconnect without wavelength Conversion

2.3.2 Operation of Wavelength Converter

We consider the networks with Wavelength Converters, which is to convert the signal of a certain wavelength to a different wavelength, installed in the network nodes. Wavelength conversion allows a clear optical channel to be carried on different wavelengths on different physical links. A wavelength converter can be implemented either all-optically, or with O-E-O conversion. However, in the modeling of the wavelength converter, there is no differentiation between the physical implementations. We do not consider the physical nonlinear effects or optical loss of the wavelength converter, which means that the wavelength converter is considered with the function of converting the signal to another wavelength color without any loss of interference. Depending on the physical characteristics of the wavelength converters, there can be different models (thus constraints) for wavelength converters.

Basically, there can be three categories: full wavelength conversion, limited wavelength conversion, and fixed wavelength conversion [Dutt01]. Full wavelength conversion means that an input wavelength can be converted to any other output wavelength. Limited wavelength conversion implies that an input wavelength can be converted to any wavelength in a specific set, which generally does not contain all the wavelengths. If the converter can only convert a wavelength to exactly one another wavelength, which is a special case of the limited wavelength conversion, it is called fixed wavelength conversion. The use of the converters (full or limited conversion) installed on the network nodes can influence the virtual topology design. The wavelength conversion can fully or partially remove the wavelength continuity constraint in the routing of the lightpaths, which generally results in larger solution space and therefore better solution for the virtual topology optimization. Thus, wavelength use is generally more efficient. However, the installation of converters results also in the extra higher cost, as well as the complexity of the optimization problem. The extra cost might be reduced by using limited conversion instead of full conversion. Our network model supports both full wavelength conversion and partial wavelength conversion, which are formulated as constraints in our mathematical description.

2.3.3 Operation of Electronic Switching

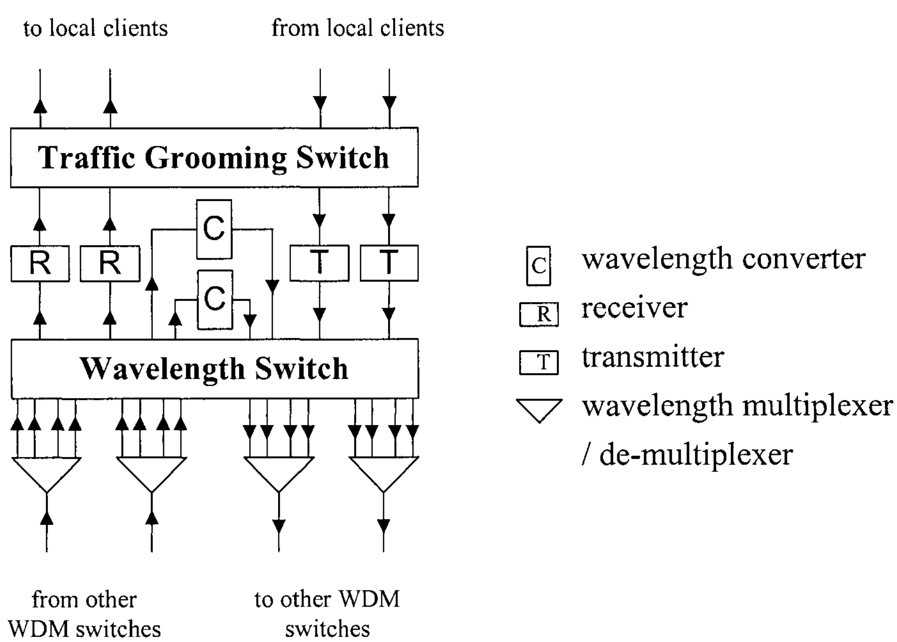


Figure 2.4. An Integrated Traffic Grooming Node

Figure 2.4 depicts a node capable of integrated traffic grooming. In the optical domain, WCs in the input fibers are de-multiplexed and then switched to either receivers for optical-to-electronic conversion, or wavelength converters, or output fibers. All the transmitters and receivers are tuneable to any wavelength used in the network. If there is any wavelength converter (some nodes may not have any such wavelength converter at all), it operates at the full wavelength range, i.e., it can convert any input wavelength to another output wavelength.

The integrated design approach optimizing the electronic layer and the optical layer together is adopted in this thesis. The transmitters, receivers and wavelength converters are arranged in a share-per-node structure, because this structure requires the least number of such components to achieve the same performance [XiLe99]. The traffic grooming switch is to add/drop local traffic and pass the traffic streams to the intended wavelength channel. The integrated design approaches are to optimize the electronic domain and the optical domain together. This is of particular interest to the new generation of WDM networks, which consist of switches capable of manipulating traffic in both the electronic domain and optical domain. Such new switch operations integrate the optical domain switching (i.e., WC, waveband and fiber switching), optical add-drop multiplexing, and traffic de-multiplexing, multiplexing and grooming in the electronic domain [ZhZh03]. The traffic in the electronic domain can be SONET, IP or other traffic carried directly over semi-lightpaths, depending on the network architecture.

2.4 Network Models

Below are the major models we rely on in the networks studied in this thesis. The optimization problems we solve can be compared with other similar research works and can be applied to any network topologies. In our network models, the number of network resources is formulated as constraints, instead of a part in the objective function.

2.4.1 Mapping Between Physical and Logical Topologies

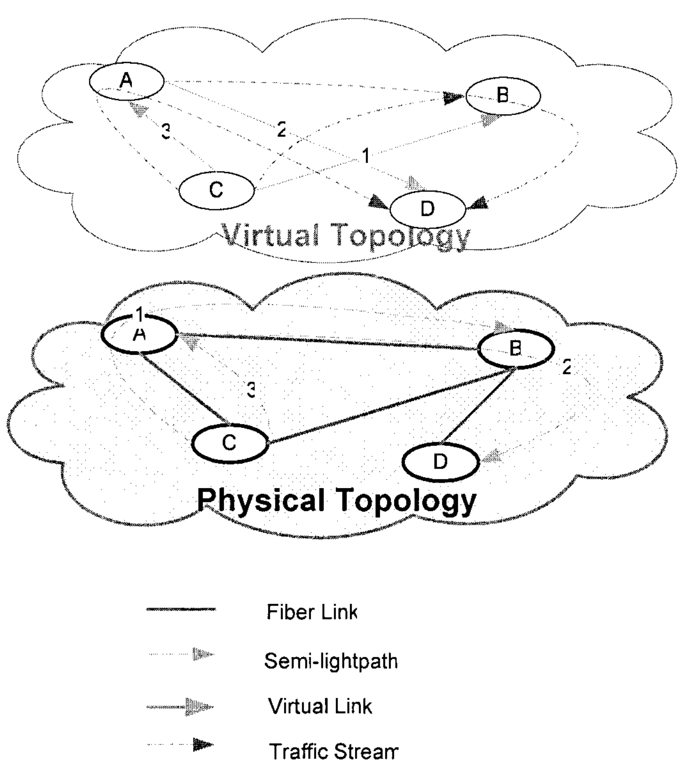


Figure 2.5. The description of the two-layer topology

The semi-lightpaths that are set up between the network nodes form a virtual topology that is relatively independent of the physical topology, and the traffic switching on the electronic layer does not have direct relation with the physical layer. The electronic layer only sees the virtual links formed by the semi-lightpaths, and the union of the network nodes that are the terminals of the virtual links. We can thus separate the formulation of the electronic switching network into a two subproblems, i.e., the optical layer problem and the electronic layer problem. Note that the two subproblems are closely related, which means obtaining the optimal solutions for each subproblem independently does not result in the optimal solution for the overall problem.

Figure 2.5 gives a graphical description of the virtual topology and the physical topology in the grooming network. We can see that the virtual links in the virtual topology corresponds to the semi-lightpath routed through the physical topology, e.g., semi-lightpath 1 from node C to node B forms the virtual link 1 in the virtual topology. The traffic streams are routed in the virtual topology formed by the virtual links, and the traffic streams cannot see the actual fiber link connections in the physical topology

2.4.2 Wavelength Converter Models

We consider both the full-range wavelength converters and the limited range limited number of wavelength converters. In our wavelength conversion models, the number of wavelength converters on each node is limited, and all our wavelength converters have share-per-node architecture [LeLi93], which means the number of simultaneous wavelength conversion on each node can not exceed the number of wavelength converters installed on that node. The converters are modeled as constraints in the formulation for the optimization procedure.

Full-range Wavelength Converter Model:

The full-range wavelength converters are modeled as converter capacity constraints and wavelength conversion constraints:

Converter capacity constraints:

$$\sum_{(s,d)} \sum_{0<n\leq N_{sd}} \phi_j^{sdn} \leq F_j, \forall j, \tag{2.1}$$

where ϕ_j^{sdn} is a 0-1 integer variable, representing the use of wavelength converters on node j by s_{sdn}(used on full-range wavelength converters), F_j is the number of converters at node j, and N_{sd} is the maximum number of semi-lightpaths between node pair (s, d); Constraints (2.1) confine that the number of converters used in one node cannot exceed the number of available converters.

Wavelength conversion constraints:

$$\phi_j^{sdn} = \begin{cases} 1, & \text{if } \exists m, k \in \mathcal{V} \text{ and } b \neq a, \delta_{mja}^{sdn} = \delta_{jkb}^{sdn} = 1, \\ 0, & \text{otherwise.} \end{cases} \qquad (2.2)$$

where δ_{ijc}^{sdn} is a 0-1 integer variable, representing the use of wavelength channel w_{ijc} by semi-lightpath s_{sdn}. Constraints (2.2) stipulate that a converter with on an intermediate node j is used only when different wavelengths are assigned to s_{sdn} for the incoming and outgoing signals at node j. Note that (m, j) and (j, k) are the incoming and outgoing fiber links.

<u>*Limited-range Wavelength Converter Model:*</u>

In the networks with limited range and limited number of wavelength converters, without loss of generality, we assign the same index c to all the wavelength channels of the same wavelength. The wavelength converters have a limited wavelength conversion degree v, i.e., the traffic from a wavelength c can be only converted to a set of wavelengths $I_i(c) = \{c, c+1, \ldots, (c+(v\text{-}1)) \bmod W\}$, where W is the number of wavelengths supported by the fiber. The limited wavelength converters are modeled as converter capacity constraints, wavelength conversion constraints and limited wavelength conversion constraints. The wavelength conversion constraints are similar to Constraints (2.2). The constraints are listed as follows:

Limited wavelength conversion degree constraints:

$$\delta_{ijc}^{sdn} = \sum_{y \in \mathcal{V}} \sum_{w_{jyx} \in I_j(c) \cap W_{jy}} \delta_{jyx}^{sdn} \;\; \forall (s,d), 0 < n \leq N_{sd}, 0 < c \leq n_{ij};\; \forall (i,j), j \neq d, i \neq s \qquad (2.3)$$

These constraints stipulate that a wavelength can only be converted to a certain set of wavelengths ($I_j(c)$) that are allowed by the wavelength converters and physical links.

Converter capacity constraints:

$$\sum_{(s,d)} \sum_{0 < n \leq N_{sd}} \sum_{j \in \mathcal{V}} \sum_{0 \leq a < n_{ij}} \phi_{i,ca}^{sdn} \leq F_{ic} \quad \forall i, 0 \leq c < W \qquad (2.4)$$

These constraints mean that the number of occupied converters with an index c in node i cannot be more than the number of the converters available in this node.

Wavelength conversion constraints:

$$\phi_{j,ab}^{sdn} = \begin{cases} 1, & \text{if } \exists m, k \in \mathcal{V} \text{ and } b \neq a, \delta_{mja}^{sdn} = \delta_{jkb}^{sdn} = 1, \\ 0, & \text{otherwise.} \end{cases} \qquad (2.5)$$

where $\phi_{j,ab}^{sdn}$ is a 0-1 integer variable representing the use of wavelength converters by s_{sdn} at node

j to convert a signal from wavelength a to wavelength b. The value '1' means in-use and '0' not-in-use. They stipulate that a converter with index a (corresponding to wavelength channel a) on an intermediate node j is used only when different wavelengths are assigned to s_{sdn} for the incoming and outgoing signals at node j.

2.4.3 Electronic and Photonic Switching Model

The electronic switching is modeled as a layer of sub-wavelength traffic switching in the virtual topology formed by the semi-lightpaths set-up in the optical layer. The model of the traffic switching contains the virtual link capacity constraints and the traffic stream continuity constraints:

Traffic flow continuity constraint:

$$\sum_{d}\sum_{0<n\le N_{sd}} \upsilon_{sdn}^{pqz} - \sum_{d}\sum_{0<n\le N_{ds}} \upsilon_{dsn}^{pqz} = \begin{cases} \gamma_{pqz} & \text{if } s = p, \\ -\gamma_{pqz} & \text{if } s = q, \quad \forall (p,q), 0 < z \le Z_{pq}, \\ 0 & \text{otherwise.} \end{cases} \tag{2.5}$$

where υ_{sdn}^{pqz} represents the use of the semi-lightpath s_{sdn} by the traffic stream x_{pqz}, which equals to one, if the traffic x_{pqz} goes through s_{sdn}; zero otherwise; γ_{pqz} is the admission status of the traffic x_{pqz}, which equals to zero, if the traffic x_{pqz} is rejected; one otherwise. Traffic flow continuity means if x_{pqz} is admitted, x_{pqz} has to be continuous from source to destination. Thus, the balance of entering and exiting traffic flow at all intermediate nodes does not change. Only at the source and destination nodes, the traffic stream adds the number of entering and exiting traffic streams by γ_{pqz} (zero or one), respectively.

Semi-lightpath capacity constraint:

$$\sum_{(p,q)}\sum_{0<z\le Z_{sd}} \upsilon_{sdn}^{pqz} G_{pqz} \le C\alpha_{sdn}, \ \forall (s,d), \tag{2.6}$$

where G_{pqz} is the bandwidth of the traffic x_{pqz}, C is the bandwidth of a semi-lightpath, and α_{sdn} is the admission status of the semi-lightpath s_{sdn}, which equals to one, if the semi-lightpath s_{sdn} is set up; zero otherwise. Constraints (2.6) mean the total traffic being routed through a semi-lightpath should be less than the capacity of the semi-lightpath.

2.5 General Assumptions

Unless specified otherwise, the following assumptions pertain to the remainder of this thesis, especially in the problem formulation.

1. The connection demands are static (to semi-dynamic). This is due to its long-term nature, thus all the connection demands are scheduled at a time in a centralized manner.
2. All the connection matrices and node connection matrices are assumed to be known and fixed over time once scheduled, as in other static RWA problems [GhDi00]. This is because these matrices can usually be established from a given physical topology and the knowledge of the long-term network traffic requirement.
3. The smallest unit of connection demands of the source-destination pairs is one semi-lightpath in the pure WRN network without electronic switching, and all semi-lightpaths have the same capacity and the wavelengths on the same fiber span assume the same cost. One source-destination demands can have multiple (≥ 1) connection demands to consider more general network traffics.
4. Propagation delay is much larger than queuing delay in the network.
5. The share-per-node tunable transceiver structure is used in the OXC's, because this structure requires the least number of FWC to achieve the same performance [XiLe99].
6. The wavelength channels (denoted as set $\boldsymbol{W}_{ij}$) are bi-directional, and the lights of the same wavelength can travel in both directions simultaneously, i.e., $\boldsymbol{W}_{ij}= \boldsymbol{W}_{ji}$. One wavelength channel allows at most one semi-lightpath being routed through in each direction. Obviously, if there is no link between nodes i and j, $\boldsymbol{W}_{ij}=\boldsymbol{\Phi}$.

Chapter Three
Static RWA Problems in All-Optical WRNs

We start our study on the static Routing and Wavelength Assignment (RWA) problems for Wavelength Routed Networks (WRNs) with wavelength conversion capability of a given physical resource configuration and topology. We will provide the basic LRSM (Lagrangean Relaxation and Subgradient Method) framework, which will be used throughout the rest of the thesis.

We first study the minimum-cost problem in the WRNs with full-range wavelength conversion capability. We then investigate the generalized MAX-RWA problem in a network with limited wavelength conversion capability. The study in this chapter is focused on the configuration of the optical layer only.

3.1 Minimum-Cost Problem

In this section, we study the WRNs with full-range wavelength conversion capability, i.e., the full-range wavelength converter (FWC) is used on every network node. At the same time, we will provide the basic LRSM framework. In addition to the general assumptions introduced in Section 2.4, we assume in this section that all traffic demands have to be accommodated.

3.1.1 Problem Formulation

We use the overall cost of the routing scheme to formulate our objective function, which is comprised of the use of wavelengths on fibers and the cost of using wavelength converters. Practical constraints, such as transceiver constraints and converter constraints need to be considered.

Let C_{sdn} be the capacity of s_{sdn} (the nth connection demand of (s, d) pair), N_{sd} be the number of connection demands between (s, d), and E_{sdn} be the cost of s_{sdn}, which is defined as $\sum_{(i,j)} \sum_{0<c\le n_{ij}} d_{ijc}\delta_{ijc}^{sdn} + \sum_j c_j \phi_j^{sdn}$ (see the details of the description later). In this thesis, only the costs of wavelength converters and wavelength channels are modelled. We can then formulate our

objective function as follows,

Objective Function:

$$\min_{\Delta,\Phi} \quad \{\sum_{(s,d)} \sum_{0<n\le N_{sd}} C_{sdn} E_{sdn}\} \tag{3.1}$$

The summation of $C_{sdn}E_{sdn}$ is the total traffic-cost product of all (s, d)'s. Since the capacity is generally the same for all wavelength channels and a semi-lightpath connection has the same capacity as one wavelength channel, we consider C_{sdn} as a constant, and the objective can be rewritten as:

$$\min_{\Delta,\Phi} \quad \{J \equiv \sum_{(s,d)} \sum_{0<n\le N_{sd}} E_{sdn} = \sum_{(s,d)} \sum_{0<n\le N_{sd}} (\sum_{(i,j)} \sum_{0<c\le n_{ij}} d_{ijc}\delta_{ijc}^{sdn} + \sum_{j} c_j \phi_j^{sdn})\} \tag{3.2}$$

The variable δ_{ijc}^{sdn} equals 1 if s_{sdn} goes through wavelength channel w_{ijc} on link (i, j), and zero otherwise. The notation w_{ijc} represents the cth WC on the fiber e_{ij}, and n_{ij} represents the number of WCs (Wavelength Channel) on the fiber e_{ij}. We allow the fiber links to be assigned with different number of wavelengths to consider more general network configurations. Without losing generality, the WCs on all fibers with the same index c are assumed using the same wavelength, and this wavelength is referred to as the *wavelength c*. Similarly, $\phi_j^{sdn}=1$ means the use of one wavelength converter on node j by s_{sdn}, and $\phi_j^{sdn}=0$ otherwise. The notations Φ and Δ represent variable sets $\{\phi_j^{sdn}\}$ and $\{\delta_{ijc}^{sdn}\}$, respectively. When the network topology is given, the network design problem is dependent on several parameters: the number of source-destination pairs, the number of connection demands (semi-lightpaths) of each source-destination pair, and the number of resources (which include transceivers, wavelength converters and wavelengths on links). Therefore, we have the following constraints.

a) Transceiver capacity constraints:

Since the switching nodes have the share-per-node transceiver structure, the constraints of transceiver can be formulated as follows,

$$\sum_{j} \sum_{0<c\le n_{ij}} \sum_{(s,d)} \sum_{0<n\le N_{sd}} \delta_{ijc}^{sdn} \le T_i \ , \qquad \forall i \in \mathcal{V} \tag{3.3}$$

$$\sum_{i} \sum_{0<c\le n_{ij}} \sum_{(s,d)} \sum_{0<n\le N_{sd}} \delta_{ijc}^{sdn} \le R_j \ , \qquad \forall j \in \mathcal{V} \tag{3.4}$$

Constraints (3.3) and (3.4) mean the total number of transmitters and receivers used on a node cannot exceed the corresponding number of transmitters and receivers installed on this node. Generally, we would assume that $T_i = R_i$ for simplicity, although it is not a strict requirement.

b) Semi-lightpath continuity constraints:

Semi-lightpath continuity means every semi-lightpath has to be continuous from source to destination, and for every source-destination pair, the semi-lightpaths assigned to them terminate at the two end nodes. So, we have

$$\sum_{j}\sum_{0<c\le n_{ij}}\delta_{ijc}^{sdn}-\sum_{j}\sum_{0<c\le n_{ij}}\delta_{jic}^{sdn}=\begin{cases}1 & if\ i=s,\\ -1 & if\ i=d,\\ 0 & otherwise,\end{cases}\qquad \forall(s,d),0<n\le N_{sd} \tag{3.5}$$

$$\text{and}\qquad \sum_{j}\sum_{0<c\le n_{ij}}\sum_{0<n\le N_{sd}}\delta_{sjc}^{sdn}=\sum_{i}\sum_{0<c\le n_{ij}}\sum_{0<n\le N_{sd}}\delta_{idc}^{sdn}=N_{sd}\,. \tag{3.6}$$

c) Wavelength channel capacity constraints:

$$\sum_{(s,d)}\sum_{0<n\le N_{sd}}\delta_{ijc}^{sdn}\le 1\qquad \forall(i,j),0<c\le n_{ij} \tag{3.7}$$

These constraints mean every wavelength channel on each fiber allows only one semi-lightpath being routed through in one direction.

d) Converter capacity constraints:

$$\sum_{(s,d)}\sum_{0<n\le N_{sd}}\phi_{j}^{sdn}\le F_{j},\ \forall j\in\mathcal{V} \tag{3.8}$$

Constraints (3.8) confine that the number of converters used in one node cannot exceed the number of available converters at that node.

e) Wavelength conversion constraints:

$$\phi_{j}^{sdn}=\begin{cases}1, & \text{if } \exists m,k\in\mathcal{V} \text{ and } b\ne a,\delta_{mja}^{sdn}=\delta_{jkb}^{sdn}=1,\\ 0, & \text{otherwise.}\end{cases} \tag{3.9}$$

These constraints are due to the semi-lightpath continuity constraints in (3.5). They stipulate that a converter with on an intermediate node j is used only when different wavelengths are assigned to s_{sdn} for the incoming and outgoing signals at node j.

3.1.2 Mathematical Solution Procedure

Our formulation can first be relaxed by the LR technique. Then it can be decomposed into subproblems, which in turn can be solved by the SPAWG (Shortest Path Algorithm for Wavelength Graph) proposed in [ChFa96]. The dual problem generated from LR will be solved by subgradient method.

3.1.2.1 The LR Solution

We can relax the wavelength channel capacity constraints (3.7), converter capacity constraints (3.8), and transceiver capacity constraints (3.3) and (3.4), by using Lagrange multipliers ξ_{ijc}, λ_i, η_i, θ_j, respectively. This leads to the following Lagrangean dual problem, denoted as DP.

$$\max_{\xi,\lambda,\eta,\theta\geq 0} q = \min_{\Delta,\Phi}\{\sum_{(s,d)}\sum_{0<n\leq N_{sd}} E_{sdn} + \sum_{(i,j)}\sum_{0<c\leq n_{ij}} \xi_{ijc}(\sum_{(s,d)}\sum_{0<n\leq N_{sd}}\delta_{ijc}^{sdn} - 1) + \sum_j \lambda_j(\sum_{(s,d)}\sum_{0<n\leq N_{sd}}\phi_j^{sdn} - F_j)$$
$$+\sum_i \eta_i(\sum_j\sum_{0<c\leq n_{ij}}\sum_{(s,d)}\sum_{0<n\leq N_{sd}}\delta_{ijc}^{sdn} - T_i) + \sum_j \theta_j(\sum_i\sum_{0<c\leq n_{ij}}\sum_{(s,d)}\sum_{0<n\leq N_{sd}}\delta_{ijc}^{sdn} - R_j)\}, \tag{3.10}$$

subject to the constraints (3.5) and (3.9), where ξ, λ, η, θ are, respectively, the vectors of Lagrange multipliers $\{\xi_{ijc}\}$, $\{\lambda_i\}$, $\{\eta_i\}$, $\{\theta_j\}$.

After regrouping the relevant terms, the dual function leads to the following problem:

$$\min_{\Delta,\Phi}\{\sum_{(s,d)}\sum_{0<n\leq N_{sd}}[E_{sdn} + \sum_{(i,j)}\sum_{0<c\leq n_{ij}}\xi_{ijc}\delta_{ijc}^{sdn} + \sum_{(i,j)}\sum_{0<c\leq n_{ij}}\eta_i\delta_{ijc}^{sdn} + \sum_{(i,j)}\sum_{0<c\leq n_{ij}}\theta_j\delta_{ijc}^{sdn} + \sum_j\lambda_j\phi_j^{sdn}]$$
$$-\sum_{(i,j)}\sum_{0<c\leq n_{ij}}\xi_{ijc} - \sum_i \eta_i T_i - \sum_j \theta_i R_j - \sum_j \lambda_j F_j\}.$$

Because the last four terms are independent of the decision variables, the problem can be simplified as:

$$\min_{\Delta,\Phi}\{\sum_{(s,d)}\sum_{0<n\leq N_{sd}}[E_{sdn} + \sum_{(i,j)}\sum_{0<c\leq n_{ij}}\xi_{ijc}\delta_{ijc}^{sdn} + \sum_{(i,j)}\sum_{0<c\leq n_{ij}}\eta_i\delta_{ijc}^{sdn} + \sum_{(i,j)}\sum_{0<c\leq n_{ij}}\theta_j\delta_{ijc}^{sdn} + \sum_j\lambda_j\phi_j^{sdn}]\}, \tag{3.11}$$

which is referred to as RP.

The RP can also be rewritten as follows:

$$\sum_{(s,d)}\sum_{0<n\leq N_{sd}}\min_{\Delta_{sdn}}\{E_{sdn} + \sum_{(i,j)}\sum_{0<c\leq n_{ij}}\xi_{ijc}\delta_{ijc}^{sdn} + \sum_{(i,j)}\sum_{0<c\leq n_{ij}}\eta_i\delta_{ijc}^{sdn} + \sum_{(i,j)}\sum_{0<c\leq n_{ij}}\theta_j\delta_{ijc}^{sdn} + \sum_j\lambda_j\phi_j^{sdn}\}, \tag{3.12}$$

which can be decomposed into $\sum_{(s,d)}\sum_{0<n\leq N_{sd}} 1$ subproblems, each of which corresponds to exactly one s_{sdn}. Each semi-lightpath-level subproblem (denoted as SP_{sdn}), corresponding to s_{sdn} (the nth semi-lightpath of source-destination pair (s, d)), is defined as follows:

$$\min_{\Delta_{sdn}}\{E_{sdn} + \sum_{(i,j)}\sum_{0<c\leq n_{ij}}\xi_{ijc}\delta_{ijc}^{sdn} + \sum_{(i,j)}\sum_{0<c\leq n_{ij}}\eta_i\delta_{ijc}^{sdn} + \sum_{(i,j)}\sum_{0<c\leq n_{ij}}\theta_j\delta_{ijc}^{sdn} + \sum_j\lambda_j\phi_j^{sdn}\}, \tag{3.13}$$

subject to constraints (3.5) and (3.9).

We can rewrite equation (3.13) as

$$\min_{\Delta_{sdn}}\{\sum_{(i,j)}\sum_{0<c\leq n_{ij}}\delta_{ijc}^{sdn}(d_{ijc} + \xi_{ijc} + \eta_i + \theta_j) + \sum_j (c_j + \lambda_j)\phi_j^{sdn}\}. \tag{3.14}$$

In (3.14), the subproblem has the same form as the optimization problem that has been solved by

SPAWG [ChFa96]. So we can modify the SPAWG to solve the subproblems (3.13) (see Appendix A.3), assuming the cost of using wavelength w_{ijc} is $d_{ijc} + \xi_{ijc} + \eta_i + \theta_j$, and the cost of using one wavelength converter on node j is $c_j + \lambda_j$.

3.1.2.2 Solving the Dual Problem

In studying the network behaviors, we often need to isolate one network resource from others to have a better understanding on the influence of the single resource. We therefore sometimes need to ignore some of the constraints from certain resources. To reach this goal, we need to modify the subgradient method to update more than one set of multipliers separately, so that when we update the multipliers for the constraints corresponding to a network resource, we can ignore the other multipliers simply by not updating them, while ensuring the convergence of the modified subgradient method. For example, if we only want to study the impact from the number of wavelengths on the network, we can simply ignore the transceiver constraints and do not update the corresponding Lagrange multipliers. This improves the efficiency of the algorithm and benefits the implementation in the future study on network behaviors.

Suppose the dual problem includes two sets of Lagrange multipliers, ξ and η. The subgradient method generates a sequence of dual feasible points according to the iteration

$$\begin{pmatrix} \xi^{(k+1)} \\ \eta^{(k+1)} \end{pmatrix} = \begin{bmatrix} \xi^{(k)} + r^{(k)} g(\xi^{(k)}) \\ \eta^{(k)} + s^{(k)} g(\eta^{(k)}) \end{bmatrix}^{+},$$

where $\xi^{(k)}$, $\eta^{(k)}$ denote the vectors ξ, η obtained at the kth iteration; the function $g(\cdot^{(k)})$ denotes the subgradient; the notation $[\cdot]^{+}$ denotes projection on the closed convex constraint set; the notation $\|\cdot\|$ is the Euclidean norm, and $r^{(k)}, s^{(k)}$ are positive scalar stepsizes.

Theorem 1: If $\xi^{(h)}$ and $\eta^{(h)}$ are not optimal, then for every dual optimal solution ξ^{*} and η^{*}, we have

$$\left\| \begin{pmatrix} \xi^{(k+1)} \\ \eta^{(k+1)} \end{pmatrix} - \begin{pmatrix} \xi^{*} \\ \eta^{*} \end{pmatrix} \right\| < \left\| \begin{pmatrix} \xi^{(k)} \\ \eta^{(k)} \end{pmatrix} - \begin{pmatrix} \xi^{*} \\ \eta^{*} \end{pmatrix} \right\|,$$

for all stepsizes $r^{(k)}, s^{(k)}$ such that

$$0 < r^{(k)} < \frac{2(q(\xi^{*}, \eta^{(k)}) - q(\xi^{(k)}, \eta^{(k)}))}{\|g(\xi^{(k)})\|^{2}}$$
$$0 < s^{(k)} < \frac{2(q(\xi^{(k)}, \eta^{*}) - q(\xi^{(k)}, \eta^{(k)}))}{\|g(\eta^{(k)})\|^{2}}.$$

Proof:

By using the subgradient inequalities:

$$(\xi^* - \xi^{(k)})' g(\xi^{(k)}) \ge q(\mu^{(k)}, \xi^*) - q(\mu^{(k)}, \xi^{(k)})$$
$$(\eta^* - \eta^{(k)})' g(\eta^{(k)}) \ge q(\mu^*, \xi^{(k)}) - q(\mu^{(k)}, \xi^{(k)}),$$

we can derive the following

$$\left\| \begin{pmatrix} \xi^{(k+1)} \\ \eta^{(k+1)} \end{pmatrix} - \begin{pmatrix} \xi^* \\ \eta^* \end{pmatrix} \right\|^2 = \left\| \begin{pmatrix} \xi^{(k)} \\ \eta^{(k)} \end{pmatrix} + \begin{pmatrix} r^{(k)} g(\xi^{(k)}) \\ s^{(k)} g(\eta^{(k)}) \end{pmatrix} - \begin{pmatrix} \xi^* \\ \eta^* \end{pmatrix} \right\|^2$$

$$= \left\| \begin{pmatrix} \xi^{(k)} \\ \eta^{(k)} \end{pmatrix} - \begin{pmatrix} \xi^* \\ \eta^* \end{pmatrix} \right\|^2 - 2 \begin{pmatrix} \xi^* - \xi^{(k)} \\ \eta^* - \eta^{(k)} \end{pmatrix}' \begin{pmatrix} r^{(k)} g(\xi^{(k)}) \\ s^{(k)} g(\eta^{(k)}) \end{pmatrix} + \left\| \begin{matrix} r^{(k)} g(\xi^{(k)}) \\ s^{(k)} g(\eta^{(k)}) \end{matrix} \right\|^2$$

$$= \left\| \begin{pmatrix} \xi^{(k)} \\ \eta^{(k)} \end{pmatrix} - \begin{pmatrix} \xi^* \\ \eta^* \end{pmatrix} \right\|^2 - 2 r^{(k)} \left(\xi^* - \xi^{(k)} \right)' g(\xi^{(k)})$$
$$- 2 s^{(k)} \left(\eta^* - \eta^{(k)} \right)' g(\eta^{(k)}) + \left\| r^{(k)} g(\xi^{(k)}) \right\|^2 + \left\| s^{(k)} g(\eta^{(k)}) \right\|^2$$

$$\le \left\| \begin{pmatrix} \xi^{(k)} \\ \eta^{(k)} \end{pmatrix} - \begin{pmatrix} \xi^* \\ \eta^* \end{pmatrix} \right\|^2 - \frac{\phi^{(k)} (2 - \phi^{(k)}) (q(\mu^{(k)}, \xi^*) - q(\mu^{(k)}, \xi^{(k)}))^2}{\left\| g(\xi^{(k)}) \right\|^2}$$
$$- \frac{\varphi^{(k)} (2 - \varphi^{(k)}) (q(\mu^*, \xi^{(k)}) - q(\mu^{(k)}, \xi^{(k)}))^2}{\left\| g(\eta^{(k)}) \right\|^2}$$

$$\le \left\| \begin{pmatrix} \xi^{(k)} \\ \eta^{(k)} \end{pmatrix} - \begin{pmatrix} \xi^* \\ \eta^* \end{pmatrix} \right\|^2 - \frac{\phi^{(k)} (2 - \phi^{(k)}) (q(\mu^*, \xi^{(k)}) - q(\mu^{(k)}, \xi^{(k)}))^2}{\left\| g(\xi^{(k)}) \right\|^2}$$
$$- \frac{\varphi^{(k)} (2 - \varphi^{(k)}) (q(\mu^{(k)}, \xi^*) - q(\mu^{(k)}, \xi^{(k)}))^2}{\left\| g(\eta^{(k)}) \right\|^2}$$

$$< \left\| \begin{pmatrix} \xi^{(k)} \\ \eta^{(k)} \end{pmatrix} - \begin{pmatrix} \xi^* \\ \eta^* \end{pmatrix} \right\|^2,$$

where

$$\phi^{(k)} = \frac{s^{(k)} \left\| g(\xi^{(k)}) \right\|^2}{q(\mu^*, \xi^{(k)}) - q(\mu^{(k)}, \xi^{(k)})}$$

$$\varphi^{(k)} = \frac{r^{(k)} \left\| g(\eta^{(k)}) \right\|^2}{q(\mu^{(k)}, \xi^*) - q(\mu^{(k)}, \xi^{(k)})}.$$

Since the projection is nonexpansive [Bert99], we have

$$\left\| \left[\begin{pmatrix} \xi^{(k)} \\ \eta^{(k)} \end{pmatrix} + \begin{pmatrix} r^{(k)} g(\xi^{(k)}) \\ s^{(k)} g(\eta^{(k)}) \end{pmatrix} \right]^+ - \begin{pmatrix} \xi^* \\ \eta^* \end{pmatrix} \right\| \le \left\| \begin{pmatrix} \xi^{(k)} \\ \eta^{(k)} \end{pmatrix} + \begin{pmatrix} r^{(k)} g(\xi^{(k)}) \\ s^{(k)} g(\eta^{(k)}) \end{pmatrix} - \begin{pmatrix} \xi^* \\ \eta^* \end{pmatrix} \right\|.$$

So eventually, we can have

$$\left\| \begin{pmatrix} \xi^{(k+1)} \\ \eta^{(k+1)} \end{pmatrix} - \begin{pmatrix} \xi^{*} \\ \eta^{*} \end{pmatrix} \right\| < \left\| \begin{pmatrix} \xi^{(k)} \\ \eta^{(k)} \end{pmatrix} - \begin{pmatrix} \xi^{*} \\ \eta^{*} \end{pmatrix} \right\|.$$

Q.E.D.

The proof of more than 2 sets of Lagrange multipliers is similar but more complicated.

According to the theorem, the vectors ξ, λ, η, θ can be updated using the following formula with guaranteed convergence:

$$\xi^{(h+1)} = \xi^{(h)} + \alpha^{(h)} g(\xi^{(h)}) \tag{3.15}$$

$$\lambda^{(h+1)} = \lambda^{(h)} + \beta^{(h)} g(\lambda^{(h)}) \tag{3.16}$$

$$\eta^{(h+1)} = \eta^{(h)} + \gamma^{(h)} g(\eta^{(h)}) \tag{3.17}$$

$$\theta^{(h+1)} = \theta^{(h)} + \rho^{(h)} g(\theta^{(h)}), \tag{3.18}$$

where $\xi^{(h)}$, $\lambda^{(h)}$, $\eta^{(h)}$, and $\theta^{(h)}$ denote the value of vectors ξ, λ, η, and θ obtained at the hth iteration, respectively. The notations $\alpha^{(h)}$, $\beta^{(h)}$, $\gamma^{(h)}$, and $\rho^{(h)}$ denote the step size at the hth iteration. The notations $g(\xi)$, $g(\lambda)$, $g(\eta)$, and $g(\theta)$ are the subgradients of q with respect to ξ, λ, η, and θ, respectively. The vectors $g(\xi)$, $g(\lambda)$, $g(\eta)$, $g(\theta)$ are composed of $g_{ijc}(\xi)$, $g_j(\lambda)$, $g_i(\eta)$, $g_j(\theta)$, respectively where

$$g_{ijc}(\xi) = \sum_{(s,d)} \sum_{0<n\le N_{sd}} \Delta^{sdn}_{w_{ijc}} - 1,$$

$$g_j(\lambda) = \sum_{(s,d)} \sum_{0<n\le N_{sd}} \delta^{sdn}_j - F_j,$$

$$g_i(\eta) = \sum_j \sum_{(s,d)} \sum_{0<n\le N_{sd}} \Delta^{sdn}_{w_{ijc}} - T_i,$$

$$g_j(\theta) = \sum_i \sum_{(s,d)} \sum_{0<n\le N_{sd}} \Delta^{sdn}_{w_{ijc}} - R_j. \tag{3.19}$$

The step sizes are given by

$$\alpha^{(h)} = \mu \times \frac{q^U - q^{(h)}}{g^T(\xi^{(h)}) g(\xi^{(h)})},$$

$$\beta^{(h)} = \nu \times \frac{q^U - q^{(h)}}{g^T(\lambda^{(h)}) g(\lambda^{(h)})},$$

$$\gamma^{(h)} = \omega \times \frac{q^U - q^{(h)}}{g^T(\eta^{(h)}) g(\eta^{(h)})},$$

$$\rho^{(h)} = o \times \frac{q^U - q^{(h)}}{g^T(\theta^{(h)}) g(\theta^{(h)})}, \tag{3.20}$$

where q^U is an estimate of the optimal solution of (3.10), and $q^{(h)}$ is the value of q at the hth iteration. Generally, the best value of the objective function J of the feasible routings obtained is used to be q^U. The parameters μ, v, ω, o and q^U are changed adaptively as the algorithm converges. The rate of the change can be determined by experimental experiences. It has also been shown in the operation research field that the adaptive change of the parameters with the progress of the subgradient algorithm will speed up the convergence of the algorithm. Specifically, if the value of $q^{(h)}$ remains roughly the same for several iterations, the parameter values are decreased, and if the value of $q^{(h)}$ keeps increasing for several iterations, the parameter values are increased [HoLu93].

3.1.2.3 Constructing a Feasible Routing Scheme

Because some of the constraints are relaxed by the Lagrange multipliers, the solution to the dual problem is generally associated with an infeasible routing scheme, i.e., some wavelength channel capacity constraints (3.7), converter capacity constraints (3.8), and transceiver capacity constraints (3.3) and (3.4) might be violated at some links and some nodes. Note that other constraints will not be violated because of the way subproblems (3.13) are solved. To construct a feasible routing scheme, a heuristic algorithm has to be employed to decide which semi-lightpaths should be rerouted and which path the rerouted semi-lightpaths should take.

First we set the arc lengths of the wavelengths that have violated constraints (wavelength channel or transceiver capacity constraints) to infinity, and employ Esau-Williams [Haye84] algorithm to recalculate the costs of source-destination pairs. Then a heuristic shown in Appendix A.4, assuming the wavelength conversion degree $v=W$ (full wavelength conversion), based on the estimated cost change in J (the original cost function (3.2)) is used to determine which semi-lightpaths should take the 'critical resources'.

To decide the sequence to route the collided semi-lightpaths, let E'_{sdn} be the cost obtained in the recalculation, and let E_{sdn} be the cost obtained in the solution of the dual problem for the nth semi-lightpath of (s, d). The estimated cost change of the nth semi-lightpath of (s, d) can be defined as

$$X_{sdn} = E'_{sdn} - E_{sdn}.$$

If there is no feasible solution for rerouting the semi-lightpath (s, d), X_{sdn} can be set to a very big positive number.

The semi-lightpath with the highest X_{sdn} is deployed first with the route derived from dual solution. If there is capacity violation in the route, a new search for a feasible route will be launched until all demands are satisfied. The further details are included in Appendix A.4.

3.1.2.4 Evaluation of the Feasible Solution

In order to gauge the effectiveness of the optimization, we need to know how close our solution is to the optimum value, which generally cannot be obtained except by exhaustive search in this case. However, the LR approach can provide an upper bound on the distance of the final result obtained to the optimal solution [Bert99]. The value of the objective function J of any feasible routing scheme obtained is an upper bound on the optimal objective J^*, because J cannot be lower than optimum. The value of the dual function q^*, on the other hand, is a lower bound on J^*. The difference $(J^* - q^*)$ is known as duality gap [Bert99]. Moreover, $(J - q^*)$ provided an upper bound to the duality gap, and the approximate relative duality gap, defined as $(J - q^*)/q^*$, is a measure of the sub-optimality of the feasible routing scheme with respect to the optimal scheme. In the rest of the thesis, this evaluation method for the quality of the feasible solution will be used.

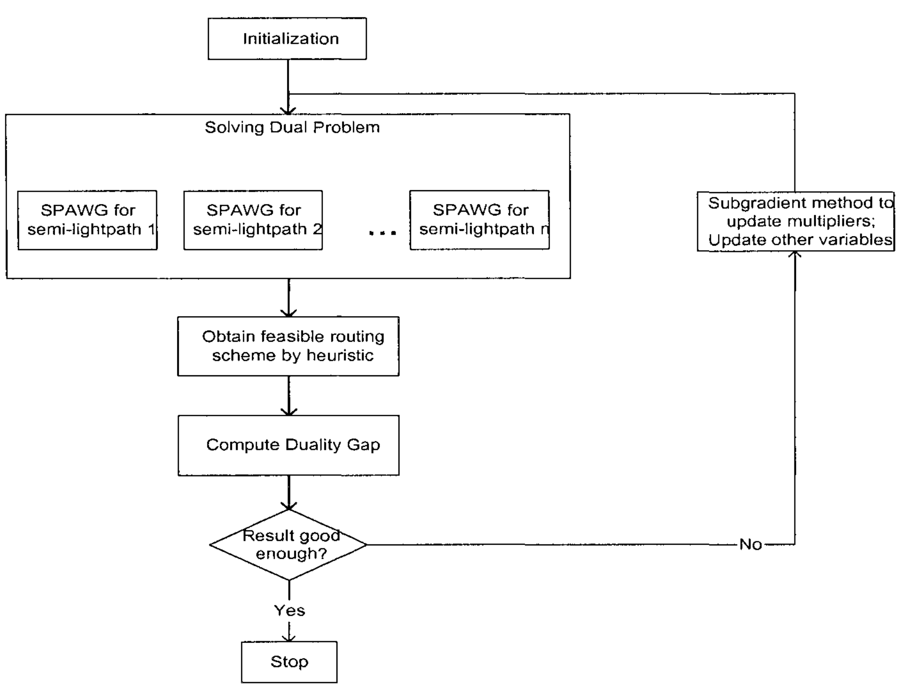


Figure 3.1. The schematic depiction of the algorithm

3.1.2.5 Overall Structure of the Algorithm

Figure 3.1 depicts the overall structure of the algorithm. Solving the subproblems can be thought of as the low level problems and the updating of the multipliers can be thought of as the high level problem. There are various criteria that can be used to terminate the algorithm, such as the duality gap, the number of iterations and the objective value. In the current implementation, we simply use the number of iterations as the stopping criterion (1000 iterations in this case). The selection of other stopping criteria will be introduced in detail in Chapter Six.

3.1.3 Performance Evaluation

In this section, we will show some computation examples for some networks with a limited number of full-range wavelength converters. More generalized network models and more network samples will be studied in the following chapters. To simplify the comparison, we used the same number of wavelengths (W) and converters (F) on all fibers and nodes. The algorithm is developed in Visual C++ 6.0 environment with C++ standard class architecture. The network resources, such as the wavelength channels, the wavelength converters, the fiber links, and the nodes, are all packaged as C++ classes with functions describing the behaviours of the corresponding physical device, and variables recording the device' status and statistics of the network data. The optimization algorithms, such as the heuristic algorithm and the Lagrange multiplier updating structure, as well as other network behaviours are also packaged in the C++ classes, with interfaces that can be easily invoked and inherited. All the computation testing is done on a 1.6GHz Intel Pentium M® laptop with 256MB RAM and Windows XP® operating system. We will use the same computation environment all through the thesis.

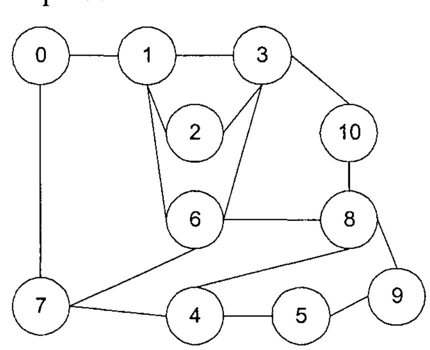


Figure 3.2. A Randomly Generated 11-Node Network

To test the performance of the algorithms and study the characteristics of the solutions, we generated a random 11-node network, as shown in Figure 3.2. The traffic matrix, with 123

0	0	0	0	0	0	1	0	1	1	2	1	0
2	0	2	1	0	1	1	1	1	0	2	2	1
0	2	0	1	1	2	0	1	2	1	2	1	1
0	1	2	0	0	0	2	1	2	2	1	1	1
2	2	0	1	0	2	0	0	0	0	1	0	0
0	2	0	1	1	0	2	0	1	0	2	0	1
0	0	2	2	0	0	0	1	0	1	2	0	1
1	0	0	0	2	0	1	0	0	0	0	2	1
1	0	1	0	1	0	1	0	0	0	0	1	1
2	2	2	0	2	1	2	0	1	0	0	0	0
0	1	0	1	0	1	0	2	0	1	0	1	0
0	2	0	1	1	2	1	1	0	0	1	0	1
1	0	0	1	0	2	1	0	0	1	0	1	0

Figure 3.3. Connection demand matrix

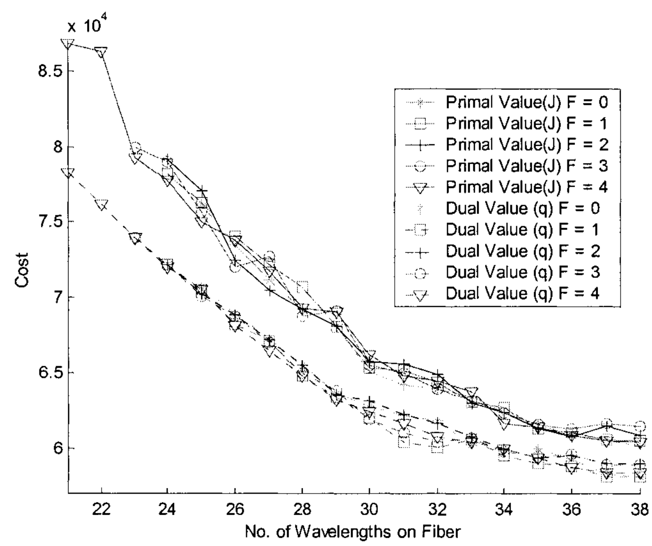
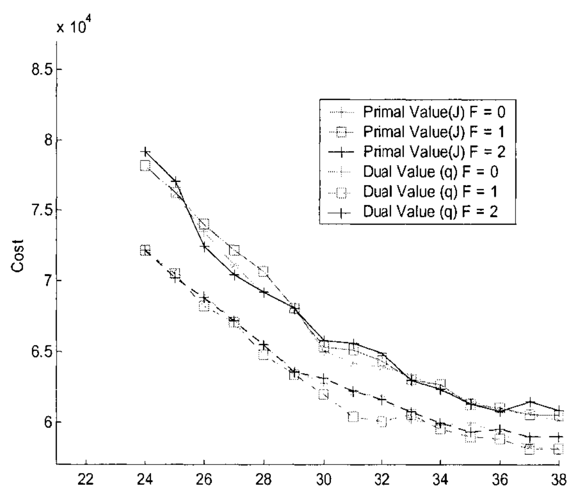
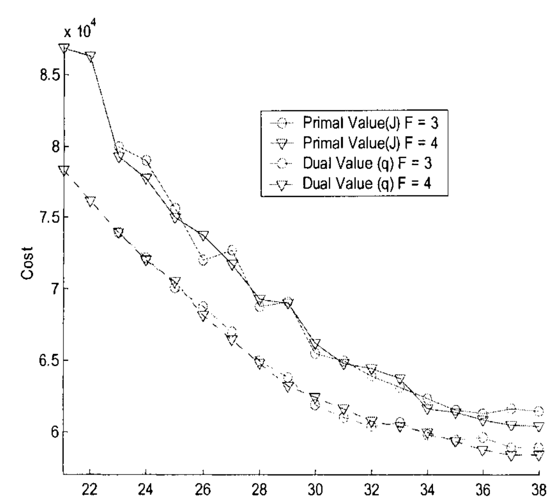


Figure 3.4. Overall cost v.s. *W* and *F*



(a) For $F = 0$, 1 and 2



(b) For $F = 3$ and 4

Figure 3.5. Overall cost v.s. *W* and *F*

connection demands in total, of the network is shown in Figure 3.3. The values of d_{ijc} are randomly generated integers ranging from 100 to 800, and all c_j's are assumed to be 800.

In the computation, we assign the costs of the wavelengths in the same fiber to slightly different values for the preference of using some wavelengths to others. The computation result is shown in Figure 3.4, and the cases with the number of converters ranging from 0 to 4 have been studied. For the readers to have a better vision of the curves, the graph in Figure 3.4 is separated into two shown in Figure 3.5. We can see that most of the final costs, i.e. the primal values, generated are within 5% of the lower bounds, i.e. the dual values. Most of the results are obtained within 30 seconds. By running the algorithm longer, better bounds and primal values can be generated. However, for the cases close to infeasibility, longer time is required for the algorithm to search for the feasible wavelength assignment and routing scheme. The detailed complexity analysis of the algorithm will be performed in Section 3.2.

We can see from Figure 3.4 that more wavelength converters on the nodes do not result in lower overall routing cost, because the formidable cost of using the converters. However, more converters result in lower minimum number of wavelengths required on the fiber to get a feasible solution. The lowest numbers of wavelengths required for the cases with converter numbers (F) equal to 0, 1, 2, 3, and 4 are 25, 24, 24, 23, and 21, respectively. Note that to the left of the curves, no feasible results could be obtained, because that there are not enough resources to accommodate all the traffic demands. Therefore, installing the converters can reduce the necessary number of wavelengths required to obtain a feasible routing scheme. Also, we noticed in the experiments that generally converters on nodes with higher nodal degree are more frequently used.

Installing more wavelengths on fibers reduces the overall cost largely when the number of wavelengths is low, but when the number of wavelengths reaches 34, more wavelengths hardly results in any reduction in the overall routing cost.

3.2 Maximum Traffic Problem

In this section, we will study networks with more complicated (limited) wavelength conversion model. Unlike Section 3.1, the objective here is to maximize the accepted connections. A novel formulation (considering the fairness of the traffic demand acceptance) and the solution methodology for a generalized MAX-RWA problem (see Section 1.1.1) will be proposed. The simplifying assumption that is used in Section 3.1 that all traffic should be accepted is not used in this section. Our formulation allows the trade-off between the fairness and the total number of rejected connections. In this section, we will also provide a complete analysis on the complexity of our algorithm.

3.2.1 Problem Formulation

In formulating the network design problem in this section, we assume that the wavelength converters have a limited (but the same) conversion degree v and a limited conversion capacity F_{ic}. The detailed description of the wavelength conversion model has been described in Section 2.4.2. Without losing generality, we assign the index c to all the wavelength converters that can convert wavelength c to other wavelengths. The number of converters with the same index in one node is limited. This will be translated into Constraint (3.26) later to make our formulation more practical (although harder).

We shall re-formulate the MAX-RWA problem as a minimizing-rejection-penalty problem to address the fairness issue (described in Section 1.5). Let s_{sdn} be the nth semi-lightpath demand of source-destination pair (s, d). Let α_{sdn} be a 0-1 integer variable, representing the rejection status of s_{sdn}. It has a value of 0, if s_{sdn} is rejected, and 1 if admitted. Let P_{sdn} be the penalty for rejecting s_{sdn}. Let A, Δ and Φ represent sets $\{\alpha_{sdn}\}$, $\{\delta_{ijc}^{sdn}\}$ and $\{\Phi_{sdn}\}$, respectively. We use the overall rejection penalty of the traffic demands to formulate our objective function as follows,

Objective Function:

$$\min_{A,\Delta,\Phi} \quad \{J\}, \qquad \text{with } J \equiv \sum_{(s,d)} \sum_{0<n\le N_{sd}} P_{sdn}(1-\alpha_{sdn}), \tag{3.21}$$

where N_{sd} represents the number of semi-lightpath demands of source-destination pair (s, d), and the summation of $P_{sdn}(1\text{-}\alpha_{sdn})$ is the total rejection penalty for the overall traffic demand in the network. To take the fairness into account, for the same source-destination pair (s, d), P_{sdn} is a decreasing function of index n, so that the rejection of the semi-lightpath demand from all source-destination pairs can be balanced in the optimization procedure.

Constraints:

a) Semi-lightpath flow continuity constraints:

Semi-lightpath continuity means that if a semi-lightpath is admitted, the semi-lightpath has to be continuous along the path of every source-destination pair. Since the assigned semi-lightpaths terminate at the two end nodes, we have

$$\sum_{j\in\mathcal{V}}\sum_{0<c\le n_{ij}}\delta_{ijc}^{sdn}-\sum_{j\in\mathcal{V}}\sum_{0<c\le n_{ij}}\delta_{jic}^{sdn}=\begin{cases}\alpha_{sdn} & \text{if } i=s,\\ -\alpha_{sdn} & \text{if } i=d,\quad \forall(s,d), 0<n\le N_{sd},\\ 0 \text{ otherwise,}\end{cases} \tag{3.22}$$

where δ_{ijc}^{sdn} is a 0-1 integer variable, which equals to one if wavelength channel w_{ijc} is used by s_{sdn}, and zero otherwise; $\mathcal{V}$ is node set. Note that δ_{ijc}^{sdn} equals 0, if $\alpha_{sdn}=0$. If s_{sdn} is accepted, (i.e. $\alpha_{sdn}=1$) at the source node ($i=s$), there is one unit of flow going out of this node, thus equation (3.22) equals 1. At the destination node ($i=d$), there is a flow of 1 coming into this node, thus equation (3.22) equals –1. Finally, at the intermediate nodes, equation (3.22) equals 0 due to the conservation of flows. If the s_{sdn} is rejected (i.e., $\alpha_{sdn}=0$), equation (3.22) equals 0 at any node.

b) Wavelength conversion constraints:

$$\phi_{j,ab}^{sdn}=\begin{cases}1, & \text{if } \exists m,k\in\mathcal{V} \text{ and } b\ne a, \delta_{mja}^{sdn}=\delta_{jkb}^{sdn}=1,\\ 0, & \text{otherwise.}\end{cases} \tag{3.23}$$

where $\phi_{j,ab}^{sdn}$ is a 0-1 integer variable representing the use of wavelength converters by s_{sdn} at node j to convert a signal from wavelength a to wavelength b. The value '1' means in-use and '0' not-in-use. These constraints are due to the semi-lightpath flow continuity constraints in (3.22). They stipulate that a converter with index a on an intermediate node j is used only when different wavelengths are assigned to s_{sdn} for the incoming and outgoing signals at node j.

c) Wavelength channel capacity constraints:

$$\sum_{(s,d)}\sum_{0<n\le N_{sd}}\delta_{ijc}^{sdn}\le 1 \qquad \forall(i,j),\ 0<c\le n_{ij} \tag{3.24}$$

These constraints mean every wavelength channel on a fiber allows only one semi-lightpath routed in the same direction, as stipulated in Assumption 5 in Section 2.4.

d) Limited wavelength conversion degree constraints:

$$\delta_{ijc}^{sdn}=\sum_{y\in\mathcal{V}}\sum_{w_{jyx}\in I_j(c)\cap W_{jy}}\delta_{jyx}^{sdn}\ \forall(s,d),\ 0<n\le N_{sd},\ 0<c\le n_{ij};\ \forall(i,j), j\ne d, i\ne s \tag{3.25}$$

These constraints stipulate that a wavelength can only be converted to a certain set of wavelengths that are allowed by the wavelength converters and physical links.

e) Converter capacity constraints:

$$\sum_{(s,d)}\sum_{0<n\leq N_{sd}}\sum_{j\in V}\sum_{0\leq a<n_{ij}}\phi_{i,ca}^{sdn}\leq F_{ic}\qquad\forall i,0\leq c<W \tag{3.26}$$

These constraints mean that the number of occupied converters with an index c in node i cannot be more than the number of the converters available in this node.

Special Cases

Note that the Max-RWA problem [KrSi01, SwSi02] is merely a special case of our formulation above, when all P_{sdn}'s are assigned the same value. In this case, our objective function becomes the minimization of the number of connections rejected, i.e.,

$$\min_{\mathrm{A},\Delta,\Phi}\ \{\sum_{(s,d)}\sum_{0<n\leq N_{sd}}(1-\alpha_{sdn})\}\,. \tag{3.27}$$

When the traffic demand matrix is given, equation (3.27) is equivalent to the objective function of maximizing the number of connections, which means the objective function does not consider the fairness of the solution and the only consideration is the overall rejected number of connection demands.

To measure the fairness of the scheduling result, we introduce a performance measure *disconnection ratio* defined to be D (the number of source-destination pairs that have traffic demands, but are not assigned any semi-lightpath) to T (the overall number of source-destination pairs that have the traffic demands). In essence, this measure gives us an idea on the percentage of the source-destination pairs (with traffic demands) that are totally disconnected.

3.2.2 Mathematical Solution Procedure

For a network with L links, N nodes, W wavelengths and Z semi-lightpath demands, the problem defined in Section 3.2.1 contains $|A|+|\Delta|+|\Phi| = L\times W\times Z+N\times W+Z$ variables, where $|\cdot|$ denotes the number of elements in the set. The problem is usually a very complex formulation with hundreds of thousands of integer variables. For example, a small network with L=21, N=14, W=20 and Z=268, has 113,108 variables in total. Finding the exact optimum for a problem of this size is hardly possible for the computing facilities available today. Therefore, finding a sub-optimal solution within a reasonable computation time is the only choice, while knowing the proximity of the sub-optimal solution to the real-optimum will be an additional advantage.

3.2.2.1 The LR Solution

We first use the Lagrange multipliers ξ_{ijc}, λ_{ic} to relax respectively the wavelength channel capacity constraints (3.24) and converter capacity constraints (3.26). This leads to the following Lagrangean dual problem, which we shall denote as DP (the Dual Problem):

$$\max_{\xi,\lambda\geq 0} q = \min_{\mathrm{A},\Delta,\Phi}\{\sum_{(s,d)}\sum_{0<n\leq N_{sd}} P_{sdn}(1-\alpha_{sdn}) + \sum_{(i,j)}\sum_{0<c\leq n_{ij}} \xi_{ijc}(\sum_{(s,d)}\sum_{0<n\leq N_{sd}}\delta_{ijc}^{sdn}-1)$$
$$+\sum_{i\in\mathcal{V}}\sum_{0\leq c<W}\lambda_{ic}(\sum_{(s,d)}\sum_{0<n\leq N_{sd}}\sum_{j\in\mathcal{V}}\sum_{0\leq a<n_{ij}}\phi_{i,ca}^{sdn}-F_{ic})\}, \tag{3.28}$$

subject to the constraints (3.22), (3.23) and (3.25), where ξ, λ are respectively the vectors of Lagrange multipliers $\{\xi_{ijc}\}$, $\{\lambda_{ic}\}$.

After regrouping the relevant terms, the dual function leads to the following problem:

$$\min_{\mathrm{A},\Delta,\Phi}\{\sum_{(s,d)}\sum_{0<n\leq N_{sd}}[P_{sdn}(1-\alpha_{sdn}) + \sum_{(i,j)}\sum_{0<c\leq n_{ij}}\xi_{ijc}\delta_{ijc}^{sdn} + \sum_{i\in\mathcal{V}}\sum_{0\leq c<W}\sum_{j\in\mathcal{V}}\sum_{0\leq a<n_{ij}}\lambda_{ic}\phi_{i,ca}^{sdn}]$$
$$-\sum_{(i,j)}\sum_{0<c\leq n_{ij}}\xi_{ijc} - \sum_{i\in\mathcal{V}}\sum_{0\leq c<W}\lambda_{ic}F_{jc}\}. \tag{3.29}$$

By using the fact that $\delta_{ijc}^{sdn} = \alpha_{sdn}\delta_{ijc}^{sdn}$ and $\phi_{i,ca}^{sdn} = \alpha_{sdn}\phi_{i,ca}^{sdn}$, we can rewrite (3.29) as:

$$\min_{\mathrm{A},\Delta,\Phi}\{\sum_{(s,d)}\sum_{0<n\leq N_{sd}}[P_{sdn}(1-\alpha_{sdn}) + \alpha_{sdn}(\sum_{(i,j)}\sum_{0<c\leq n_{ij}}\xi_{ijc}\delta_{ijc}^{sdn} + \sum_{i\in\mathcal{V}}\sum_{0\leq c<W}\sum_{j\in\mathcal{V}}\sum_{0\leq a<n_{ij}}\lambda_{ic}\phi_{i,ca}^{sdn})]$$
$$-\sum_{(i,j)}\sum_{0<c\leq n_{ij}}\xi_{ijc} - \sum_{i\in\mathcal{V}}\sum_{0\leq c<W}\lambda_{ic}F_{jc}\}. \tag{3.30}$$

Since the last two terms are independent of the decision variables, the problem can be further simplified as:

$$\min_{\mathrm{A},\Delta,\Phi}\{\sum_{(s,d)}\sum_{0<n\leq N_{sd}}[P_{sdn}(1-\alpha_{sdn}) + \alpha_{sdn}(\sum_{(i,j)}\sum_{0<c\leq n_{ij}}\xi_{ijc}\delta_{ijc}^{sdn} + \sum_{i\in\mathcal{V}}\sum_{0\leq c<W}\sum_{j\in\mathcal{V}}\sum_{0\leq a<n_{ij}}\lambda_{ic}\phi_{i,ca}^{sdn})]\}, \tag{3.31}$$

which we shall refer to as RP (the Relaxed Problem).

Note that RP can be further decomposed into Z (the total number of semi-lightpaths) subproblems, each of which corresponds to exactly one semi-lightpath. Each semi-lightpath-level subproblem (denoted as $\mathrm{SP}_{\mathrm{sdn}}$), corresponding to s_{sdn}, is defined as follows:

$$\min_{\alpha_{sdn},\Delta_{sdn},\Phi_{sdn}}\{P_{sdn}(1-\alpha_{sdn}) + \alpha_{sdn}(\sum_{(i,j)}\sum_{0<c\leq n_{ij}}\xi_{ijc}\delta_{ijc}^{sdn} + \sum_{i\in\mathcal{V}}\sum_{0\leq c<W}\sum_{j\in\mathcal{V}}\sum_{0\leq a<n_{ij}}\lambda_{ic}\phi_{i,ca}^{sdn})\}, \tag{3.32}$$

subject to constraints (3.22), (3.23) and (3.25).

3.2.2.2 Using MMCSLP to Solve the Subproblems

In order to solve the subproblem introduced in Section 3.2.2.1, we can rewrite SP_{sdn} in (12) as:

$$\min_{\alpha_{sdn}}\{P_{sdn}(1-\alpha_{sdn})+\alpha_{sdn}\min_{\Delta_{sdn},\Phi_{sdn}}\{D_{sdn}\}\}, \tag{3.33}$$

subject to the constraints (3.22), (3.23) and (3.25), where

$$D_{sdn}=\sum_{(i,j)}\sum_{0<c\le n_{ij}}\xi_{ijc}\delta_{ijc}^{sdn}+\sum_{i\in\mathcal{V}}\sum_{0\le c<W}\sum_{j\in\mathcal{V}}\sum_{0\le a<n_{ij}}\lambda_{ic}\phi_{i,ca}^{sdn}.$$

One can see that the formulation of the subproblem (3.33) is similar to the problem solved by Modified Minimum Cost Semi-lightpath (MCSLP) in [ChFa96]. In order to obtain the optimum for our subproblem, we propose the MMCSLP algorithm, whose detail can be found in the Appendix A.1.

3.2.2.3 Updating Lagrange Multipliers

We now employ the subgradient method to solve the Dual Problem (DP). The multiplier vector $z=(\xi,\lambda)$ are first updated by the following formula:

$$z^{(h+1)}=z^{(h)}+\alpha^{(h)}g(z^{(h)}), \tag{3.34}$$

where $z^{(h)}$ denotes the value of vectors z obtained at the hth iteration, and $\alpha^{(h)}$ denote the step size at the hth iteration. The vector $g(z)$ is the subgradient of q with respect to z, i.e. $g(z)=\{g(\xi), g(\lambda)\}$. The vectors $g(\xi)$, $g(\lambda)$ are in turn composed of $g_{ijc}(\xi)$, $g_{ic}(\lambda)$, respectively, where

$$g_{ijc}(\xi)=\sum_{(s,d)}\sum_{0<n\le N_{sd}}\delta_{jic}^{sdn}-1, \tag{3.35}$$

$$g_{ic}(\lambda)=\sum_{(s,d)}\sum_{0<n\le N_{sd}}\sum_{j\in\mathcal{V}}\sum_{0\le a<n_{ij}}\phi_{i,ca}^{sdn}-F_{ic}. \tag{3.36}$$

The step size is determined by

$$\alpha^{(h)}=\mu\times\frac{q^U-q^{(h)}}{g^T(z^{(h)})g(z^{(h)})}, \tag{3.37}$$

where q^U is an estimate of the optimal solution, and $q^{(h)}$ is the value of q at the hth iteration. In general, we let q^U take on the best value of the objective function J obtained from the feasible RWA scheme. We change the parameters adaptively as the subgradient method evolves, in order to speed up the convergence of the algorithm. In doing so, we also need to change the parameters μ and q^U as the algorithm converges. Specifically, if the value of $q^{(h)}$ remains roughly the same for x iterations, the value of μ is decreased by a factor $p<1$, and if the value of $q^{(h)}$ keeps increasing for y iterations, the value of μ is increased by a factor $1/p$. From our experience, faster convergence is obtained when p=0.95, x=3 and y=5. The value q^U is also updated if a lower J value is obtained.

3.2.2.4 Constructing a Feasible RWA Scheme

Since some of the constraints are relaxed by the Lagrange multipliers, the solution to DP might have an infeasible routing scheme. In other words, some wavelength capacity constraints in (3.24) and converter capacity constraints in (3.26) might have been violated at some links and nodes. On the other hand, other constraints will not be violated because this is guaranteed by the formulation and the solution of SP_{sdn} in (3.32).

To construct a feasible routing scheme, a heuristic algorithm has to be employed to decide which semi-lightpaths should be rejected or rerouted and which paths the rerouted semi-lightpaths should take. The following priority rules are used to determine which semi-lightpaths should occupy the 'critical resources':

1. The semi-lightpaths with the highest rejection penalty are first deployed.
2. For semi-lightpaths with the same rejection penalty, the one with the lowest hop number in the dual solution is deployed first. Further ties are broken randomly.

The heuristic algorithm to search for a feasible route and wavelength assignment for a semi-lightpath demand based on the dual solution can now be described in the following three-step procedure. The algorithm is implemented on the wavelength graph (WG) of the network, which is formed by separating every node into W vertices representing the wavelengths available on the node, and connecting these vertices with arcs representing the wavelength channels and the wavelength converters. The details of the algorithm to form the WG is provided in Appendix A.2.

Step 1 (Check the Solution of the Dual Problem):
Check if the Wavelength Assignment (WA) derived from the dual solution is *viable*, i.e., no wavelength channel in the WA is occupied by some other semi-lightpaths while the wavelength converters used in the WA are not used up.

1.1) If the WA from the dual solution is viable, the WA is assigned to this semi-lightpath. Then go on to the next semi-lightpath demand in the priority list, and repeat *Step 1.*
1.2) If the WA from the dual solution is not viable, go to *Step 2.*

Step 2 (Find the WA in the Same Route):
Find a viable WA for the semi-lightpath in the same route derived from the dual solution, but with a different wavelength assignment. The procedure of this step is the same as MMCSLP

described before. However, the WG is only made of the fibers and nodes traversed by the semi-lightpath in the dual solution.

2.1) If a viable WA can be found, this semi-lightpath shall monopolize this WA. Then go on to the next semi-lightpath demand, and repeat *Step 1*.

2.2) If no viable WA can be found, go to *Step 3*.

Step 3 (Find the WA for One Semi-lightpath with the MEWA):
Apply the Modified Esau-Williams algorithm (MEWA) to find a feasible RWA for the semi-lightpath.

3.1) If a viable WA can be found, this semi-lightpath shall monopolize all the wavelength channels and converters assigned by this WA. Then go on to the next semi-lightpath demand, and repeat *Step 1*.

3.2) If no viable WA can be found, reject the semi-lightpath demand. Then go on to the next semi-lightpath demand, and repeat *Step 1*.

The description of the MEWA algorithm can be found in Appendix A.4. In essence, this heuristic algorithm re-routes the collided semi-lightpaths that might result in a low overall rejection penalty.

3.2.3 Computation Analysis

The complexity of our MMCSLP is the same as MCSLP, i.e., $O((N+W)NW)$. By grouping the subproblems with the same source node, we make use of the Shortest Path Algorithm for Wavelength Graph (SPAWG) [ChFa96] to obtain their $\min_{\Delta_{sdn},\Phi_{sdn}}\{D_{sdn}\}$. (The SPAWG can be found in the Appendix A.3.)

The worst case complexity of *Step 2* in the heuristic algorithm is $O((N+W)NW)$ (see Section 3.2.2.4). Using a straight-forward analysis similar to [KeCh74], we can now obtain the complexity of MEWA as $O((NW)^2)$. The worst case for the whole heuristic algorithm occurs when every semi-lightpath goes through *Step 3* (the MEWA algorithm), and the complexity in this scenario is $O(Z(NW)^2)$, where Z is the number of semi-lightpaths.

3.2.4 Performance Evaluation

In this section, we shall show the computation results on some sample networks. We first show the result of the Max-RWA problem (which is a special case of our formulation) for the NSFNET example. We then compare the results and the time complexities of our algorithm and the algorithm provided in [SwSi02]. Finally, we shall study the fairness of the results and the influence from the number of wavelength converters. We also provide a 22-node example to show the efficiency of our LRSM framework.

3.2.4.1 The NSFNET Network Example

To have a fair comparison, we shall use the same example of NSFNET with 14 nodes and 21 edges (Figure 3.6), and the same traffic demand matrix (Figure 3.7) as [SwSi02]. We also assume the same number of wavelengths in every fiber, and assume every node has unlimited number of wavelength converters with limited conversion degree. This is done by setting F_{ic} to a very large integer. We set all the rejection penalties to the same value to make our objective function the same as [SwSi02].

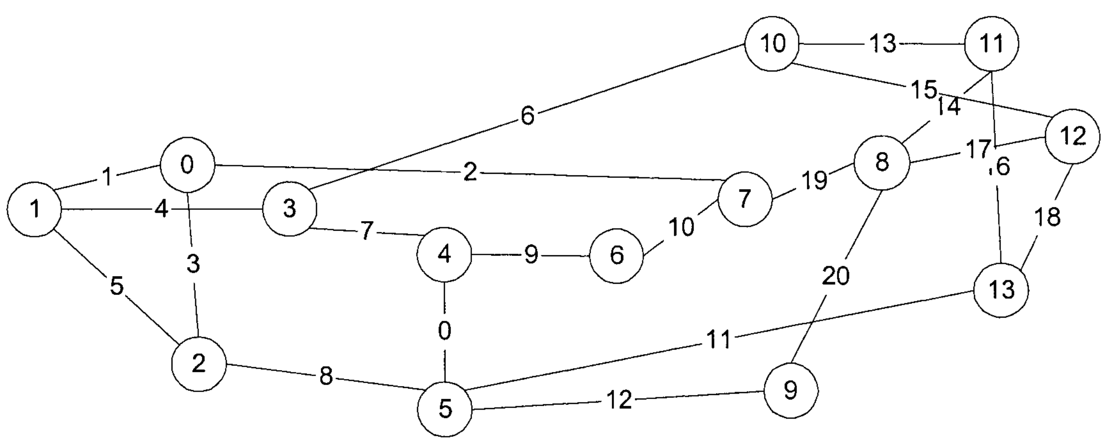


Figure 3.6. NSFNET with 14 Nodes and 21 Edges

0	1	3	1	1	1	3	0	2	0	1	2	0	3
0	0	0	2	2	2	1	1	1	2	1	0	1	3
3	2	0	3	0	1	2	3	1	3	1	2	2	0
3	1	0	0	1	1	2	3	2	2	1	2	1	3
1	3	0	2	0	1	0	2	0	3	0	1	1	3
1	2	1	3	2	0	1	3	3	1	0	1	1	2
2	2	3	1	3	3	0	0	3	1	2	0	3	3
3	1	2	3	1	0	1	0	0	3	2	0	3	0
3	0	1	3	3	3	1	0	0	2	1	1	1	0
0	0	0	1	2	0	2	0	1	0	1	0	0	3
1	0	0	2	0	3	0	1	0	3	0	3	1	3
2	3	1	1	3	2	3	2	2	2	2	0	1	3
2	0	1	2	0	1	2	0	3	0	2	1	0	3
1	1	0	2	1	0	1	3	0	1	2	1	3	0

Figure 3.7. Traffic Demand Matrix for NSFNET

Table 3.1. Number of Connections Rejected in Results of [SwSi02]

	LP bound			*J* obtained by LP		
	Conversion Degree(v)			Conversion Degree(v)		
W | v	1	2	3	1	2	3
10	70	70	70	78	72	72
11	60	60	60	68	64	60
12	50	50	50	61	57	56
13	40	40	40	55	52	47
14	30	30	30	42	35	34
15	20	20	20	34	25	21
16	10	10	10	28	17	12
17	5	5	5	18	11	8
18	1	1	1	13	9	6
19	0	0	0	5	0	0
20	0	0	0	3	0	0
21	0	0	0	1	0	0
22	0	0	0	1	0	0
23	0	0	0	0	0	0

Table 3.2. Number of Connections Rejected by Our Algorithm

	LR bound (q)			*J* obtained by LR		
	Conversion Degree(v)			Conversion Degree(v)		
W | v	1	2	3	1	2	3
10	70	70	70	76	73	72
11	59	59	59	65	63	61
12	50	50	50	56	52	51
13	40	40	40	45	43	42
14	30	30	30	36	34	32
15	20	20	20	26	25	22
16	10	10	10	18	15	12
17	5	5	5	8	7	7
18	1	1	1	4	2	2
19	0	0	0	0	0	0
20	0	0	0	0	0	0
21	0	0	0	0	0	0
22	0	0	0	0	0	0
23	0	0	0	0	0	0

3.2.4.1.1 Comparison with the Max-RWA Problem: the Number of Rejected Demands

The bound and the number of rejected semi-lightpath demands (J) provided in [SwSi02] from LP-relaxation are shown in Table 3.1, and the data obtained from our LR framework are shown in Table 3.2. In both tables, the number of wavelengths is shown in the first column. The next three columns are the bounds generated for different wavelength conversion degrees (v=1, 2, 3). The rightmost three columns show the number of connections rejected (J) for v=1, 2, 3. Note that the LP bound is the optimal solution for the relaxed LP problem, and recall the difference between the final J value and the bound shows how close to the optimum the J value obtained is.

We can see in most cases, the number of demands rejected (J) in Table 3.2 obtained from our algorithm is lower than those (Table 3.1) from [SwSi02], and our algorithm has a greater advantage when applied in the lower range of wavelength conversion degree (v). One can see that the minimum number of wavelengths (W) required to accommodate all the semi-lightpath demands for the cases of $v = 1$ has gone down by nearly 20% (22→18). According to our results in Table 3.2, the higher wavelength conversion degree results in a larger number of semi-lightpath demands that can be routed through the network for a given network resource configuration. However, increasing the wavelength conversion degree v does not have an obvious influence on the minimal number of wavelengths to obtain a feasible RWA scheme for the given traffic demand matrix. In terms of bounds achieved, our algorithm is worse (lower) than the LP bounds by 1 only for the case with W=11.

3.2.4.1.2 Comparison with the Max-RWA Problem: Time Complexity

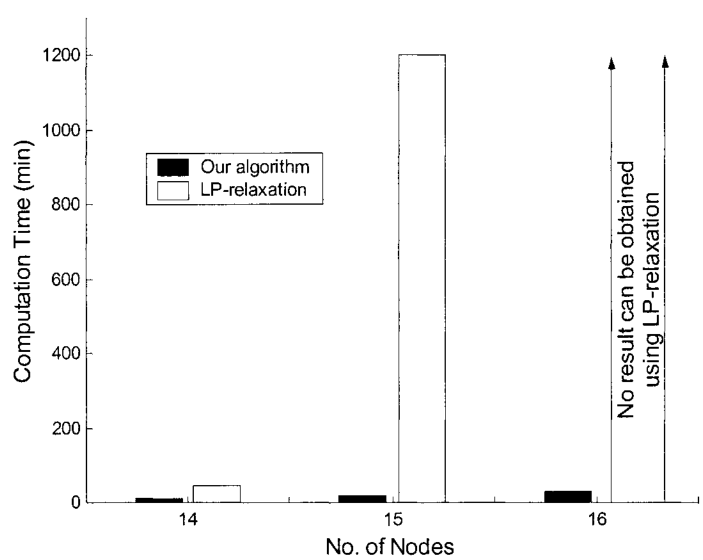


Figure 3.8. Computation Time Comparison of Our Algorithm and LP-relaxation

Most of the results from our algorithm in the NSFNET example can be obtained within 10 minutes. We have also implemented the LP-relaxation based algorithm used in [SwSi02] with the CPLEX® MILP-solver in the same platform, and the time taken to solve the 14-node problem of NSFNET is over 40 minutes. We have also done further experiment on a 15-node network by adding 1 node, 2 links and the corresponding semi-lightpath demands to the NSFNET. The LP-relaxation algorithm takes more than a day to obtain some results, while our algorithm takes less than 20 minutes using our current stopping criterion. The rough comparison of the computation time of our algorithm and the LP-relaxation used in [SwSi02] on some sample networks is shown in Figure 3.8.

Note that we could not use the LP-relaxation method in [SwSi02] to obtain the solution for a sample network of 16 nodes, because the computation time for one additional node (from 15 to 16 node) has increased exponentially to go beyond the capability of our computation facility.

As stated in the Introduction (Chapter 1), one good attribute of our framework is the reusability of the Lagrange multipliers to get a good estimate for the next round of computation. By saving the values of the Lagrange multipliers at the end of the computation, and using these values as the starting values for the computation in the next round for a different network configuration, the time has been much shortened to 4-5 minutes in the NSFNET example here. Since the complexity to solve the dual problem and the complexity of the heuristic algorithms used are both polynomial, the complexity of this algorithm is also polynomial as shown in Section 3.2.2 and thus can be applied to fairly large networks. In general, we expect better results by running our program for a longer time due to the properties of the LRSM [Bert99].

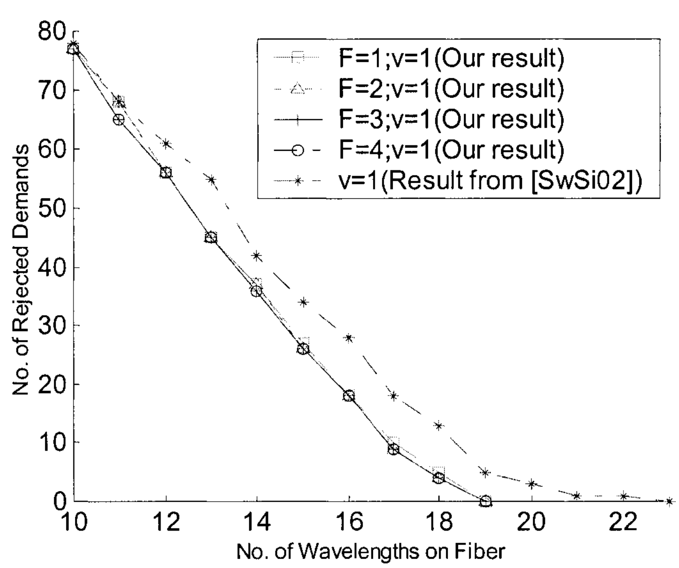


Figure 3.9. Number of rejected connection demands vs. *W* and *F* (*ν*=1)

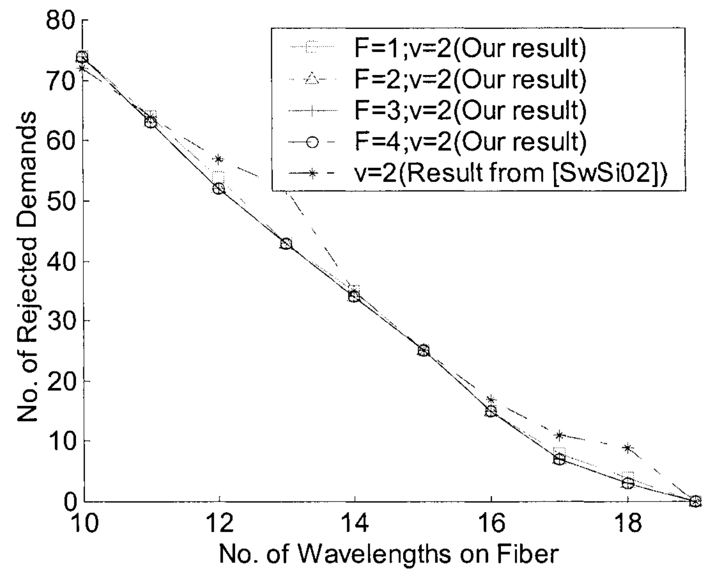


Figure 3.10. Number of rejected connection demands vs. W and F (v=2)

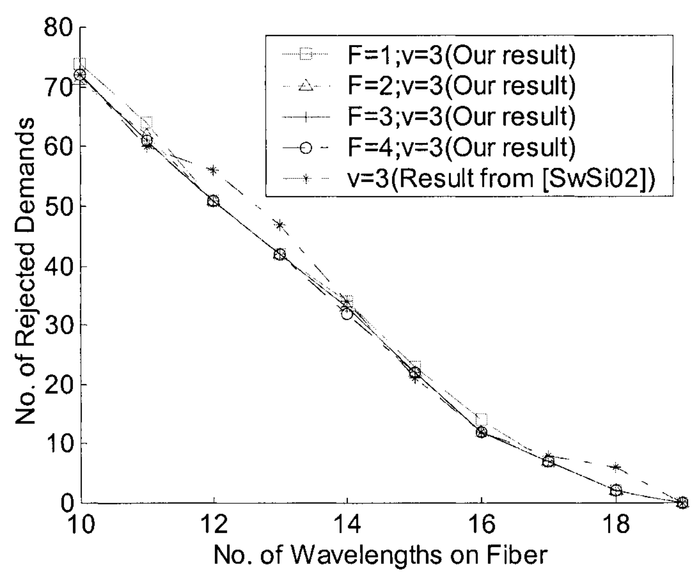


Figure 3.11. Number of rejected connection demands vs. W and F (v=3)

3.2.4.2 Effect of Network Parameters

In the comparison of the Max-RWA problem, the number of wavelength converters F on a node is assumed to be infinite as done in [SwSi02]. Since F is generally very limited, we would like to study its influence. We shall examine the changes in the number of rejected demands with respect to F, W (the number of wavelengths) and v (the wavelength conversion degree). This is done in Table 3.3 and in Figure 3.9 to Figure 3.11. Note that the computation time varies as the number of wavelengths (because the number of variables changes) even for the network with same number of nodes.

Table 3.3. Number of rejected connection demands vs. *W* and *F*

		F=1			F=2			F=3			F=4		
		Conversion Degree(v)			Conversion Degree(v)			Conversion Degree(v)			Conversion Degree(v)		
W	v	1	2	3	1	2	3	1	2	3	1	2	3
10		77	74	74	77	74	71	76	74	72	76	74	72
11		65	64	64	65	63	62	65	63	61	65	63	61
12		56	54	51	56	52	51	56	52	51	56	52	51
13		45	43	42	45	43	42	45	43	42	45	43	42
14		37	35	34	37	34	33	36	34	33	36	34	32
15		26	25	23	26	25	22	26	25	22	26	25	22
16		18	15	14	18	15	12	18	15	12	18	15	12
17		9	8	7	9	7	7	9	7	7	9	7	7
18		4	4	2	4	3	2	4	3	2	4	3	2
19		0	0	0	0	0	0	0	0	0	0	0	0

We can see that the results of F=4 in Table 3.3 are the same as the results in Table 3.2, where $F=+\infty$. This indicates that setting F to values higher than 4 does not further improve the results. In most practical cases, we notice that F=2 is enough to obtain the same results as those with infinite F. To have a graphical view, we plot the data of Table 3.3 in Figure 3.9 to Figure 3.11. The results of [SwSi02] are also included in Figure 3.9 to Figure 3.11 for comparison.

For the v=1 case, one can see from Figure 3.9 that F hardly influences the results. This agrees with the fact that v=1 means no wavelength conversion is allowed. We can also see from Figure 3.10 and Figure 3.11 that the influence from the increasing of the number of wavelength converters is not as strong as that from the increasing of the number of wavelength conversion degree. This is due to the low usage of the wavelength conversion, so that the number of used wavelength converters seldom exceeds F=2. The tiny discrepancy at W=14 may be attributed to the randomness in the performance of our heuristic algorithm. We also notice that the generated bound hardly changes as F and v change. This is again due to the low usage of the wavelength converters, resulting in that most of λ's are zeros or assume very small positive values.

3.2.4.3 Fairness Study

In essence, this measure gives us an idea on the percentage of the source-destination pairs that are totally disconnected. To measure the fairness of the scheduling result, we introduce a performance measure *disconnection ratio*, defined to be D (the number of source-destination pairs that have traffic demands, but are not assigned any semi-lightpath) to T (the overall number of source-destination pairs that have the traffic demands). This is because when T is given, D becomes our major concern. Generally, the rejection penalty P_{sdn} should decrease as n increases, so that the more semi-lightpaths that a source-destination pair (s, d) requires, the easier the demand would be rejected, because that the optimization algorithm always tries to minimize the overall rejection penalty. To achieve the fairness, we would do the following,

1) $P_{sd(n+1)} \leq P_{sdn}$ in order to avoid the problem of total rejection of the semi-lightpath demands from the same (s, d) pair.
2) Set all P_{sdn} with the same n to have the same value in order to achieve fairness among different (s, d) pairs.

One method we adopt is simply to set all P_{sd0} to the same positive value P_0. We also set all the penalty step sizes $S = (P_{sdn}-P_{sd(n+1)})/ P_0$ to a constant value, resulting in $P_{sdn} = P_0 (1 - n \cdot S)$. All of the following cases studied have a conversion degree v=3.

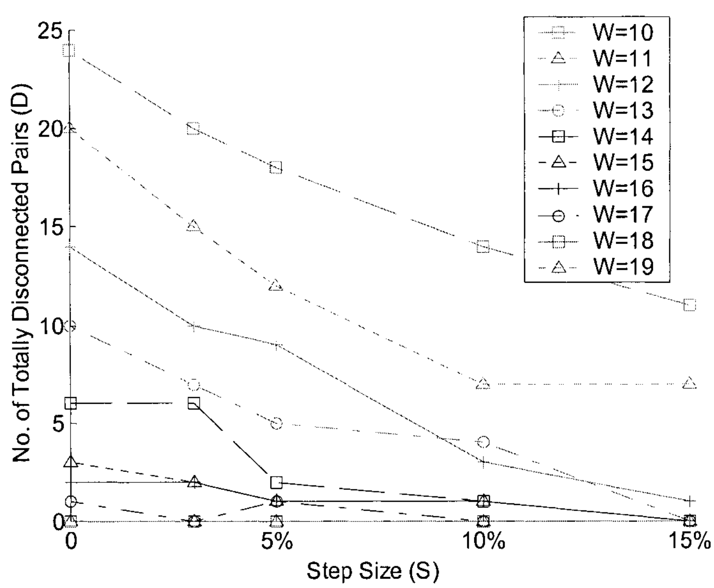


Figure 3.12. Relation between S and D (v=3)

Figure 3.12 depicts the relation between the number of totally disconnected pairs (D) and the penalty step size (S) for different numbers of wavelengths (W). We can see that as S increases, D tends to be lower, which means the rejection tends to be fairer. The influence from S on the fairness at a lower W (where the rejection ratio is higher) is larger than that at a higher W. Note that some abnormal cases in Figure 3.12 (such as W=13 at 10%) can be attributed to our heuristic algorithm, which can stop at different local optima that vary in different cases.

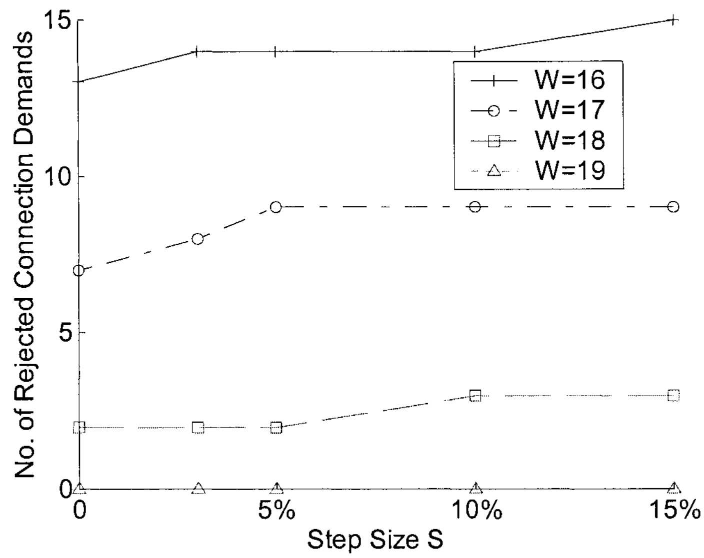


Figure 3.13. ***S*** **vs. Number of rejected connection demands (*ν*=3)**

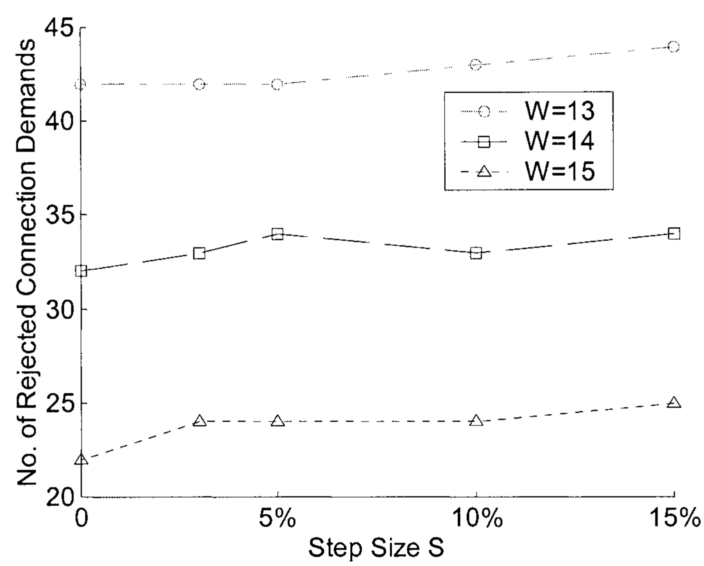


Figure 3.14. ***S*** **vs. Number of rejected connection demands (*ν*=3)**

]
To see more clearly the relationship between S and D, we separate the wavelengths into 3 sets (W = 16 to 19, 13 to 15, and 10 to 12) and re-plot the data in Figure 3.13, Figure 3.14 and Figure 3.15, respectively. We can see from these graphs that as S increases, more connection demands tend to be rejected. The amount of increase in D as S varies from 0% to 15% is about the same for different wavelengths.

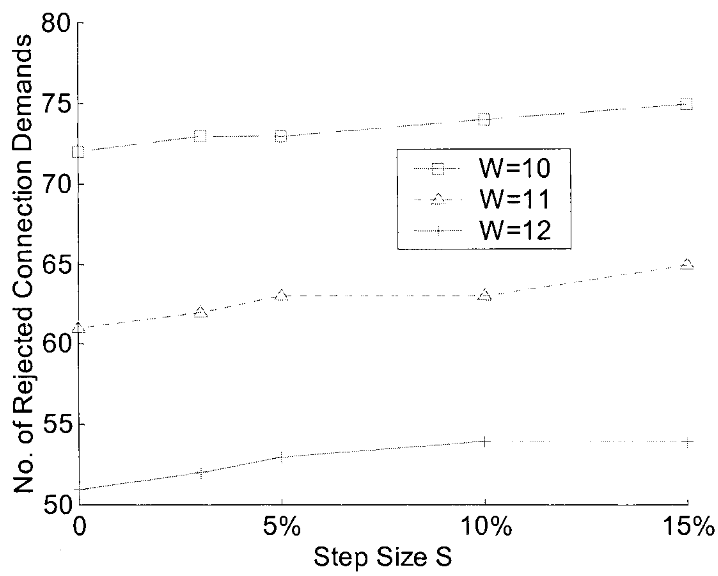


Figure 3.15. ***S*** **vs. Number of rejected connection demands (*v*=3)**

One can see there is a trade-off between the number of rejected connection demands and the disconnection ratio when *S* changes. When *S* is larger, the optimization result tends to be fairer, i.e., more source-destination pairs that have connection demands are assigned at least one semi-lightpath. However, it is at the cost of rejecting more semi-lightpath demands. By selecting a proper *S*, one can determine the disconnection ratio and the overall number of rejected demands to satisfy one's objective by using the trade-off relationship.

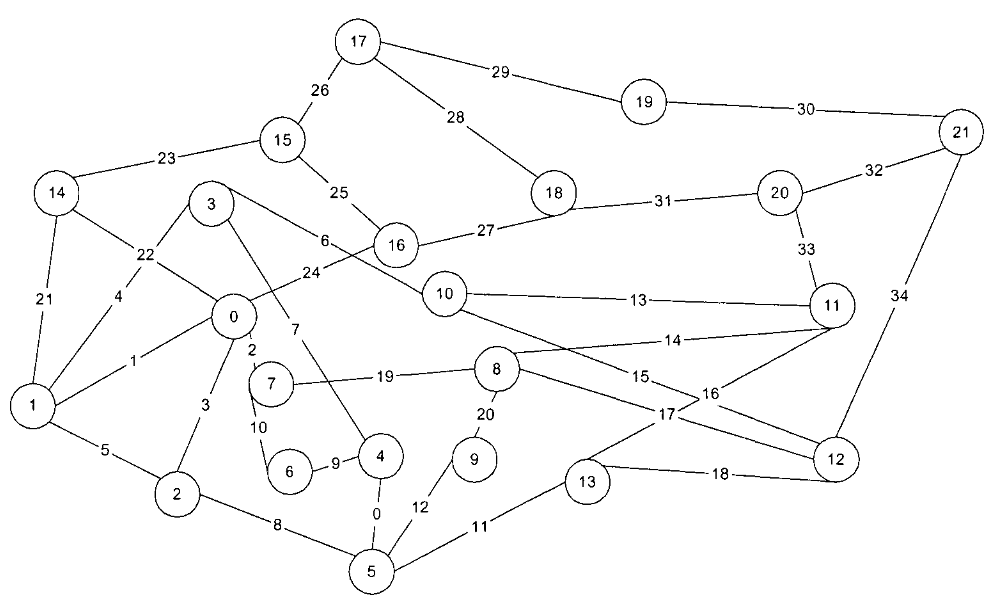


Figure 3.16. Network With 22 Nodes And 35 Edges

```
0 0 2 0 0 0 2 0 2 0 0 2 0 2 2 0 2 0 0 0 2 0
0 0 0 2 2 2 0 0 0 2 0 0 0 2 0 2 0 0 2 2 0 2
2 2 0 2 0 0 2 2 0 2 0 2 2 0 0 2 2 2 0 2 2 2
2 0 0 0 0 0 2 2 2 2 0 2 0 2 2 0 2 2 0 2 2 0
0 2 0 2 0 0 0 2 0 2 0 0 0 2 0 2 0 0 0 2 0 2
0 2 0 2 2 0 0 2 2 0 0 0 0 2 2 2 0 0 2 0 0 2
1 1 2 0 2 2 0 0 2 0 1 0 2 2 2 0 0 1 2 0 0 0
2 0 1 2 0 0 0 0 0 2 1 0 2 0 0 2 0 2 0 2 0 2
2 0 0 2 2 2 0 0 0 1 0 0 0 0 0 2 0 2 2 1 0 2
0 0 0 0 1 0 1 0 0 0 0 0 0 2 0 0 1 0 1 0 0 0
0 0 0 1 0 2 0 0 0 2 0 2 0 2 0 1 0 0 0 2 2 1
1 2 0 0 2 1 2 1 1 1 1 0 0 2 1 0 2 1 2 1 0 0
1 0 0 1 0 0 1 0 2 0 1 0 0 2 2 1 1 1 0 0 0 1
0 0 0 1 0 0 0 2 0 0 1 0 2 0 0 1 0 0 0 0 0 1
2 0 0 0 0 0 1 2 1 1 0 1 0 2 0 0 1 2 0 1 1 0
0 1 0 2 1 0 0 2 2 0 0 0 0 1 2 0 0 0 1 0 0 2
0 2 0 1 0 0 0 1 0 2 0 0 0 2 0 1 0 0 0 2 0 1
1 1 2 0 2 2 0 0 2 0 1 0 2 2 2 0 0 0 2 0 0 0
2 0 1 2 0 0 0 0 0 2 1 0 2 0 0 2 0 2 0 2 0 2
0 0 0 1 0 2 0 0 0 2 0 2 0 2 0 1 0 0 0 0 2 1
0 0 0 0 1 0 1 0 0 0 0 0 0 2 0 0 1 0 1 0 0 0
1 0 0 1 0 0 1 0 2 0 1 0 0 2 2 1 1 1 0 0 0 0
```

Figure 3.17. Traffic Demand Matrix for the 22-node Network

Table 3.4. Number of Connections Rejected Obtained for the 22-node Network(F=4)

	LR bound Degree of conversion(v)				LR RWA result Degree of conversion(v)			
W | v	1	2	3	4	1	2	3	4
11	80	80	80	80	95	94	86	85
12	70	70	70	70	82	75	74	72
13	60	60	60	60	70	67	65	63
14	51	51	51	51	58	57	54	53
15	41	41	41	41	50	46	45	44
16	30	30	30	30	41	38	36	35
17	21	21	21	21	32	31	27	27
18	14	14	14	14	23	20	18	18
19	10	10	10	10	16	15	13	13
20	7	7	7	7	12	10	9	9
21	4	4	4	4	7	5	5	5
22	0	0	0	0	2	1	1	1
23	0	0	0	0	0	0	0	0

3.2.4.4 An Example of a 22-Node Network

In order to show the effectiveness of our algorithm, we have provided a much bigger network example on a network with 22 nodes, 32 links and 352 semi-lightpath demands. The topology and the traffic demand matrix of the network are depicted in Figure 3.16 and Figure 3.17

respectively. The traffic demand between any two nodes is a random integer number ranging from 0 to 2. Table 3.4 shows the results obtained from our algorithm for different parameters. We have also tried other traffic patterns and observed similar results and behaviour. In this example, the average time spent by our algorithm to obtain the final results is only about 30-40 minutes, which is 4-5 times of the 14-node NSFNET network, due to the polynomial time complexity of our approach.

3.3 Concluding Remarks

In this chapter, we have provided the basic LRSM framework to solve the RWA problem in WRNs with full-range and limited wavelength conversion capability. We studied both the minimum-cost problem and the Max-RWA problem. We have also provided a new link-based framework in this chapter. In our framework, the fairness issue has been considered, which has a trade-off with the number of connections being routed through the network. The Max-RWA problem merely becomes a special case of our formulation. The dual function generated by LR can be decomposed into subproblems, which can be solved by the MMCSLP/SPAWG. The testing results of a randomly generated network demonstrated excellent computation effectiveness and fast convergence speed. The influence to the overall routing cost from the number of converters and the number of wavelengths installed has been studied at the same time. Using the Max-RWA problem (as a special case of our formulation) as an example, our results showed that our approach could find more connections being routed through the network than what was obtained by the LP-relaxation approach. Also, due to the polynomial complexity of our algorithm, we have shown that our approach is able to solve the RWA problems for fairly large networks that cannot be solved previously by LP-relaxation (because of its exponential complexity). These comparisons demonstrated excellent computation effectiveness of our algorithm. With the extra constraint on the converter capacity, we were able to carry out more studies: 1) the effect due to the change in the number of converters and we make the observation that excessive number of converters is not necessary; 2) the influence on the fairness of the solution (which has not been studied before) due to the change in the design parameter of our algorithm.

In this chapter, we have succesfully solved different models of all-optical WRN problems with various assumptions, among which an important one is that the minimal bandwidth unit of the traffic demand between any node pairs is one WC (wavelength channel). However, in practical applications, the bandwidth of the traffic demand between a node pair is generally much less than the capacity of one WC. Thus the assumption used in the previous chapters might result in huge bandwidth waste. Therefore in the next chapter, we want to explore the network with time-division multiplexing (TDM), which allows low-speed traffic streams multiplexing/ demultiplexing, and thus improves the bandwidth-using efficicency.

Chapter Four
Integrated Design of Traffic Grooming on WRN Structure

In this chapter, we study the WRN with traffic grooming capability, which allows the low-speed TDM traffic streams to be bundled and routed together to make more efficient use of the network bandwidth. At the same time, we propose a model for network profit, which bridges the gap between minimizing network cost and maximizing network throughput. Two factors need to be considered: the predicted traffic matrix and the expected revenue for each traffic demand. The predicted traffic matrix is defined in the form of traffic demands between node pairs. The traffic grooming model contains two layers of problems closely correlated (i.e., semi-lightpath layer problem and the traffic demand layer problem), and it is thus generally much more difficult to solve than the pure RWA problems.

4.1 Design and Optimization Objective

The ultimate goal for minimizing network cost and maximizing network throughput is to maximize network profit. Optimizing either network cost or network throughput separately cannot achieve the goal. For example, in the network cost minimization, the assumption that all traffic demands should be satisfied is questionable with respect to the profit objective. Some traffic demands may be costly and no efficient provisioning plan exists for them. To provide service to them is an excessive burden to the whole network, and the revenue generated from them cannot recover the provisioning cost. Such non-profitable traffic demands should be rejected instead of being accepted regardless of the provisioning cost. On the other hand, network throughput optimization may not achieve the profit objective, because the implicit evaluation for the network resource utilization of traffic demands may not reflect the cost factor. So far, no study has been reported to link the minimizing network cost and maximizing network throughput so that network profit can be maximized.

The following are some notes on the new design and optimization objective:

1. Representation of the revenue generated by a traffic demand: In order to incorporate the revenue factor into the network cost function, we will use the revenue penalty for rejecting a traffic demand. The amount of revenue penalty is the same as the amount of revenue the

traffic demand will generate if it is accepted. With this modification, the actual optimization objective will be to minimize the summation of the revenue penalty and the total network cost. This objective is in fact minimizing the profit loss, which is equivalent to maximizing profit. For the total network cost, a direct measurement of network cost needs to be used as opposed to any of the previously used indirect cost functions. This is because the network cost has to be measured by a monetary value and will be added to the revenue penalty. This chapter adopts a virtual unit for cost, revenue and profit, which is proportional to the dollar unit.

2. Handling non-profitable traffic demand: If the profit situation for a traffic demand is marginal, i.e., close to break-even, it may be re-considered. This can be done by adjusting the estimated charge after the first round of network planning, when the overall situation of the network becomes clearer than at the beginning. Hence the charge can be estimated more accurately.
3. Limiting network capacity: Because of the extremely high complexity associated with the integrated design of SONET-over-WDM networks, we simplify the problem by fixing the network capacity. The network capacity is set to a limited level such that not all traffic demands can be accommodated. After the LR-based optimization, a subset of traffic demands will be identified as high profitable traffic. A series of trials will then be made by adjusting the network capacity, i.e., by adjusting the number of transmitters, receivers and wavelength converters at some selected nodes, and the number of WCs in each fiber. The results will be presented in Section 4.5.
4. Difference between identifying the profit situation of a traffic demand in the static network planning and in the dynamic network operational stage: The philosophy we use to create the new design and optimization objective is that no such traffic demand should be accepted when it is known to be non-profitable before the network is built. However, after the network is built and all network resources are deployed, the dynamic network operational stage should adopt a different strategy for accepting or rejecting traffic demands. This is because the objective is changed to generate as much revenue as possible with existing network resources. How to utilize spare network resource is beyond the scope of this chapter.

4.2 Assumptions for the Traffic Grooming Network

In addition to the general assumptions stated in Section 2.4, we use the following assumptions in this chapter for the traffic grooming network models.

1. In the SONET layer, the switching fabric has unlimited multiplexing, de-multiplexing and time slot exchanging capability. The granularity of traffic requests can be any of the Synchronous Transport Signal (STS) or the Optical Carrier (OC) levels of 1, 3, 12 or 48, i.e., having a bit rate of 51.84, 155.52, 622.08 or 2488.32 Mbps. A traffic demand may be groomed with other traffic and rounded up to any granularity of OC-3, OC-12 or OC-48.
2. The capacity of SONET switching fabric is unlimited.
3. One fiber connects a pair of neighbour nodes. Every WC on fibers can carry optical signals travelling in both directions, i.e., all WCs are bi-directional.
4. Full-range wavelength conversion is used. Therefore, each input wavelength can be converted into any possible output wavelength if a wavelength converter is available.
5. A traffic demand must be handled as its entirety, and cannot be split.
6. All nodes are equipped with wavelength switching capability. However, the SONET switching capability is optional. When SONET switching is not available at a node, we set the number of transmitters and receivers at that node to zero.

4.3 Problem Formulation

As discussed above, our objective function is to minimize the profit loss, which is equivalent to maximize the profit as follows.

Objective Function:

$$\min_{\mathrm{A},\Delta,\Phi,\Gamma,\mathrm{Y}} \{J\},\ with\ J \equiv \sum_{(p,q)} \sum_{0<z\le Z_{pq}} P_{pqz} G_{pqz}(1-\gamma_{pqz}) + \sum_{(p,q)} \sum_{0<z\le Z_{sd}} E_{pqz} + \sum_{(s,d)} \sum_{0<n\le N_{sd}} D_{sdn} \tag{4.1}$$

We can see that the objective function is composed of three parts: the first summation is the penalty of rejecting traffic demands; the second summation is the cost associated with the electronic domain for routing traffic demands; and the third summation is the cost associated with the optical domain for setting up semi-lightpaths. The *constraints* are organized into two parts.

<u>A) Relation between semi-lightpaths and the optical transmission network</u>

There are five constraints to be considered:

a) Semi-lightpath flow continuity constraint: Semi-lightpath continuity means every semi-lightpath has to be continuous from source to destination. Thus, the balance of entering and exiting semi-lightpaths at all intermediate nodes does not change. Only at the source and destination nodes, the semi-lightpath adds the number of entering and exiting semi-lightpaths by one, respectively. So, we have

$$\sum_{j}\sum_{0<c\le n_{ij}}\delta_{ijc}^{sdn}-\sum_{j}\sum_{0<c\le n_{ij}}\delta_{jic}^{sdn}=\begin{cases}\alpha_{sdn} & \text{if } i=s,\\ -\alpha_{sdn} & \text{if } i=d,\ \forall(s,d),0<n\le N_{sd}.\\ 0 & \text{otherwise.}\end{cases} \tag{4.2}$$

b) Wavelength conversion constraint: A wavelength converter at an intermediate node j is used only when different wavelengths are assigned to the semi-lightpath s_{sdn} for the incoming and outgoing WCs at node j.

$$\phi_j^{sdn}=\begin{cases}1, & \text{if } \exists\, m>0, k\le N, b\ne c,\ \delta_{mjb}^{sdn}=\delta_{jkc}^{sdn}=1,\\ 0, & \text{otherwise.}\end{cases} \tag{4.3}$$

c) WC capacity constraint: These constraints stipulate that no more than one semi-lightpath may be routed through a WC in either direction.

$$\sum_{(s,d)}\sum_{0<n\le N_{sd}}\delta_{ijc}^{sdn}\le 1, \qquad \forall(i,j), 0<c\le n_{ij} \tag{4.4}$$

d) Wavelength converter capacity constraint: The number of converters used in one node cannot exceed the number of available converters, because of the share-per-node wavelength converter configuration (Figure 2.4).

$$\sum_{(s,d)}\sum_{0<n\le N_{sd}}\phi_j^{sdn}\le F_j, \qquad \forall j \tag{4.5}$$

e) Transmitter and receiver capacity constraint: The total number of transmitters and receivers used on a node cannot exceed the corresponding number of transmitters and receivers installed on this node.

$$\sum_{j}\sum_{0<c\le n_{sj}}\sum_{d}\sum_{0<n\le N_{sd}}\delta_{sjc}^{sdn}\le T_s, \quad \forall s \tag{4.6}$$

$$\sum_{i}\sum_{0<c\le n_{id}}\sum_{s}\sum_{0<n\le N_{sd}}\delta_{idc}^{sdn}\le R_d, \quad \forall d \tag{4.7}$$

The parameters T_i, R_i represent respectively the number of transmitters and receivers on switch i. Constraints (4.6) and (4.7) confine that at most $N_{sd} \le \min(T_s, R_d)$ semi-lightpaths can be set up between source-destination pair (s, d). However, in the optimal solution, it is very unlikely that N_{sd} will go very close to $\min(T_s, R_d)$, because if the semi-lightpaths between (s, d) occupy too many of the transceivers and other resources, the semi-lightpaths between s-d, and other switches will likely be blocked and eventually limit the number of the traffic demands being routed through the network. From our computation experience, $N_{sd} = ½ \min(T_s, R_d)$ is adequate enough for most of the network examples to obtain very good near-optimal result. Although not required, we usually set T_i and R_i to the same value.

B) Relation between traffic flows and semi-lightpaths

There are two such constraints:

f) Traffic flow continuity constraint: The explanation of these constraints is similar to the semi-lightpath flow continuity constraints (4.2).

$$\sum_{d} \sum_{0<n\le N_{sd}} \upsilon_{sdn}^{pqz} - \sum_{d} \sum_{0<n\le N_{ds}} \upsilon_{dsn}^{pqz} = \begin{cases} \gamma_{pqz} & \text{if } s = p, \\ -\gamma_{pqz} & \text{if } s = q, \quad \forall (p,q), 0 < z \le Z_{pq} \\ 0 & \text{otherwise.} \end{cases} \tag{4.8}$$

g) Semi-lightpath capacity constraint: The total traffic being routed through a semi-lightpath should be less than the capacity of the semi-lightpath.

$$\sum_{(p,q)} \sum_{0<z\le Z_{sd}} \upsilon_{sdn}^{pqz} G_{pqz} \le C\alpha_{sdn}, \ \forall (s,d) \tag{4.9}$$

, where C is the bandwidth of a semi-lightpath, measured by the number of OC-1 or STS-1 channels. In this thesis, the bandwidth of a semi-lightpath is OC-48, so C= 48. Note that the bandwidth of the semi-lightpath is the same as the bandwidth of a wavelength channel. We use x_{pqz} to denote the zth traffic demand between node pair (p,q). When the traffic demand is accepted, x_{pqz} also denotes the traffic flow. So for simplicity, we use the term of "a traffic", whose meaning may be a traffic demand or flow depending on the context. The variable G_{pqz} denotes the bandwidth of the traffic x_{pqz}, which is measured by the equivalent number of OC-1/STS-1 channels, i.e., $G_{pqz} \in \{1,3,12,48\}$;

4.4 Mathematical Solution Procedure

4.4.1 Overview of the Solution Framework

The traffic grooming problem is comprised of the RWA (Routing and Wavelength Assignment) and the electronic traffic routing sub-problems, which are very complicated by themselves and closely coupled to each other. It is thus very difficult to obtain a good solution considering both sub-problems at the same time. We shall apply the LR method to decouple and solve the two sub-problems independently. Similar to the pricing concept in the marketplace, the LR method can replace "hard" coupling constraints by "soft" prices for the use of resources [Bert99].

In the traffic grooming problem, semi-lightpath capacity is the primary constraint, because it should be less than the capacity of a WC and on the other hand, more than the aggregated traffic onto it. The Lagrange multipliers in the LR method will serve as the "prices" for semi-lightpaths to use optical transmission resources, and the "prices" for traffic to use the semi-lightpaths. The multipliers are updated iteratively to reflect eventually the criticality of the resources. The integration of the RWA and electronic traffic routing sub-problems will be realized by adjusting the multipliers at the outer loop, and the two problems are solved independently as inner problems. This solution framework greatly simplifies the solution method without losing the consideration of the interaction of the two sub-problems. Note that our approach is different from other approaches which, while separating the RWA and electronic traffic routing sub-problems for simplicity, lose the interaction between the two sub-problems. In our solution method, although the two sub-problems are separated in solving the inner problem, they are always associated in the outer loop.

In our LR framework, some constraints of the original optimization problem (1) (i.e., the primal problem) are relaxed and a dual problem is created by using Lagrange multipliers. The derived dual problem is then decomposed into independent RWA and electronic traffic routing sub-problems, and solved independently by the Revised Minimum Cost Semi-lightpath (RMCSLP) and the Revised Shortest Path Algorithm (RSPA). The optimal solution to the dual problem is a lower bound for the primal problem [Bert99]. Since in transforming the original optimization problem into the dual problem, some constraints are relaxed, the solution to the dual problem is generally an infeasible solution to the primal problem, which mean some constraints are violated. We therefore propose a heuristic algorithm to obtain a feasible solution based on the dual solution. The subgradient method [Bert99] will be used in updating Lagrange multipliers. Solving the sub-problems becomes the inner process with regard to updating the multipliers, which can be thought of as an outer loop process. The overall structure of the algorithm will be given in Section 4.4.7.

4.4.2 The LR Solution

According to the LR method, we should relax the WC capacity constraint (4.4), converter capacity constraint (4.5), transmitter capacity constraint (4.6) and semi-lightpath capacity constraint (4.9), by introducing Lagrange multipliers ξ_{ijc}, λ_j, η_s and κ_{sdn}, respectively. This leads to the following Lagrangean dual problem, denoted as DP.

$$\max_{\xi,\lambda,\eta,\theta\geq 0} q = \min_{\mathrm{A},\Delta,\Phi,\Gamma,\mathrm{Y}} \left\{ \sum_{(p,q)} \sum_{0<z\leq Z_{pq}} P_{pqz} G_{pqz} (1-\gamma_{pqz}) + \sum_{(p,q)} \sum_{0<z\leq Z_{sd}} E_{pqz} + \sum_{(s,d)} \sum_{0<n\leq N_{sd}} D_{sdn} + \sum_{(i,j)} \sum_{0<c\leq n_{ij}} \xi_{ijc} \left(\sum_{(s,d)} \sum_{0<n\leq N_{sd}} \delta_{ijc}^{sdn} - 1 \right) \right.$$

$$\left. + \sum_j \lambda_j \left(\sum_{(s,d)} \sum_{0<n\leq N_{sd}} \phi_j^{sdn} - F_j \right) + \sum_s \eta_s \left(\sum_j \sum_{0<c\leq n_{sj}} \sum_d \sum_{0<n\leq N_{sd}} \delta_{sjc}^{sdn} - T_s \right) + \sum_{(s,d)} \sum_{0<n\leq N_{sd}} \kappa_{sdn} \left(\sum_{(p,q)} \sum_{0<z\leq Z_{sd}} \upsilon_{sdn}^{pqz} G_{pqz} - C\alpha_{sdn} \right) \right\}, \tag{4.10}$$

subject to the constraints (4.2), (4.3), (4.7) and (4.8), where ξ, λ, η, κ are the vectors of Lagrange multipliers $\{\xi_{ijc}\}$, $\{\lambda_i\}$, $\{\eta_s\}$ and $\{\kappa_{sdn}\}$, respectively.

After regrouping the relevant terms, DP leads to the following problem:

$$\min_{\mathrm{A},\Delta,\Phi,\Gamma,\mathrm{Y}} \left\{ \sum_{(s,d)} \sum_{0<n\leq N_{sd}} \left[D_{sdn} + \sum_{(i,j)} \sum_{0<c\leq n_{ij}} \xi_{ijc} \delta_{ijc}^{sdn} + \sum_j \lambda_j \phi_j^{sdn} + \sum_j \sum_{0<c\leq n_{sj}} \eta_s \delta_{sjc}^{sdn} - C\kappa_{sdn}\alpha_{sdn} \right] \right.$$

$$\left. + \sum_{(p,q)} \sum_{0<z\leq Z_{sd}} \left[\sum_{(s,d)} \sum_{0<n\leq N_{sd}} \kappa_{sdn} \upsilon_{sdn}^{pqz} G_{pqz} + E_{pqz} + P_{pqz} G_{pqz} (1-\gamma_{pqz}) \right] - \sum_{(i,j)} \sum_{0<c\leq n_{ij}} \xi_{ijc} - \sum_s \eta_s T_s - \sum_j \lambda_j F_j \right\}. \tag{4.11}$$

Since the last three terms are independent of the decision variables, the problem can be simplified as:

$$\min_{\mathrm{A},\Delta,\Phi,\Gamma,\mathrm{Y}} \left\{ \sum_{(s,d)} \sum_{0<n\leq N_{sd}} \left[D_{sdn} + \sum_{(i,j)} \sum_{0<c\leq n_{ij}} \xi_{ijc} \delta_{ijc}^{sdn} + \sum_j \lambda_j \phi_j^{sdn} + \sum_j \sum_{0<c\leq n_{sj}} \eta_s \delta_{sjc}^{sdn} - C\kappa_{sdn}\alpha_{sdn} \right] \right.$$

$$\left. + \sum_{(p,q)} \sum_{0<z\leq Z_{sd}} \left[\sum_{(s,d)} \sum_{0<n\leq N_{sd}} \kappa_{sdn} \upsilon_{sdn}^{pqz} G_{pqz} + E_{pqz} + P_{pqz} G_{pqz} (1-\gamma_{pqz}) \right] \right\}, \tag{4.12}$$

which is referred to as RP.

By using the fact that $\delta_{ijc}^{sdn} = \alpha_{sdn} \delta_{ijc}^{sdn}$, $\phi_j^{sdn} = \alpha_{sdn} \phi_j^{sdn}$, $\upsilon_{sdn}^{pqz} = \gamma_{pqz} \upsilon_{sdn}^{pqz}$, and using constraints (4.2), we can rewrite the problem RP as:

$$\min_{A,\Delta,\Phi,\Gamma,Y}\left\{\sum_{(s,d)}\sum_{0<n\le N_{sd}}\left[\alpha_{sdn}\left(D_{sdn}+\sum_{(i,j)}\sum_{0<c\le n_{ij}}\xi_{ijc}\delta_{ijc}^{sdn}+\sum_{j}\lambda_j\phi_j^{sdn}+\eta_s\right)-C\kappa_{sdn}\alpha_{sdn}\right]\right.$$
$$\left.+\sum_{(p,q)}\sum_{0<z\le Z_{sd}}\left[\gamma_{pqz}\left(\sum_{(s,d)}\sum_{0<n\le N_{sd}}\kappa_{sdn}\upsilon_{sdn}^{pqz}G_{pqz}+E_{pqz}\right)+\left(1-\gamma_{pqz}\right)P_{pqz}G_{pqz}\right]\right\}$$
$$=\min_{A,\Delta,\Phi}\left\{\sum_{(s,d)}\sum_{0<n\le N_{sd}}\left[\alpha_{sdn}\left(\sum_{(i,j)}\sum_{0<c\le n_{ij}}\left(\xi_{ijc}+d_{ijc}\right)\delta_{ijc}^{sdn}+\sum_{j}\left(\lambda_j+c_j\right)\phi_j^{sdn}+\eta_s+t_s+r_d\right)-C\kappa_{sdn}\alpha_{sdn}\right]\right\}$$
$$+\min_{\Gamma,Y}\left\{\sum_{(p,q)}\sum_{0<z\le Z_{sd}}\left[\gamma_{pqz}\left(\sum_{(s,d)}\sum_{0<n\le N_{sd}}\upsilon_{sdn}^{pqz}\left(\kappa_{sdn}G_{pqz}+V_{pqz}\right)\right)+\left(1-\gamma_{pqz}\right)P_{pqz}G_{pqz}\right]\right\}$$
$$=\sum_{(s,d)}\sum_{0<n\le N_{sd}}\min_{\alpha_{sdn}}\left\{\alpha_{sdn}\left[\min_{\Delta_{sdn},\Phi_{sdn}}\left(Q_{sdn}\right)-C\kappa_{sdn}\right]\right\}+\sum_{(p,q)}\sum_{0<z\le Z_{sd}}\min_{\gamma_{pqz}}\left[\gamma_{pqz}\min_{Y_{pqz}}\left(R_{pqz}\right)+\left(1-\gamma_{pqz}\right)P_{pqz}G_{pqz}\right] \tag{4.13}$$

where $Q_{sdn}=\sum_{(i,j)}\sum_{0<c\le n_{ij}}\left(\xi_{ijc}+d_{ijc}\right)\delta_{ijc}^{sdn}+\sum_{j}\left(\lambda_j+c_j\right)\phi_j^{sdn}+\eta_s+t_s+r_d$

$R_{pqz}=\sum_{(s,d)}\sum_{0<n\le N_{sd}}\upsilon_{sdn}^{pqz}\left(\kappa_{sdn}G_{pqz}+V_{pqz}\right).$

In this step, we can see that equation (4.13) is composed of two independent parts. The first minimization is over the variables in the optical domain (A, Δ, Φ), while the second minimization is over the variables in the electronic domain (Y, Γ). This important feature leads to the separation of the sub-problems.

4.4.3 Solving the Sub-problems

4.4.3.1 The decomposition of the dual problem

Because the first and second summations in equation (4.13) are independent of each other, we can separate the problem into two independent sub-problem sets: a set of RWA sub-problems, one for each semi-lightpath; a set of electronic traffic routing sub-problems, one for each traffic demand.

A) RWA Sub-problems (RWAS):

Considering the optical domain subproblems in equation (4.13), we have

$$\sum_{(s,d)}\sum_{0<n\le N_{sd}}\min_{a_{sdn}}\{\alpha_{sdn}(\min_{\Delta_{sdn},\Phi_{sdn}}\{Q_{sdn}\}-C\kappa_{sdn})\},$$

subject to the constraints (4.2), (4.3) and (4.7), with

$$Q_{sdn} = \sum_{(i,j)} \sum_{0<c\le n_{ij}} (\xi_{ijc} + d_{ijc})\delta_{ijc}^{sdn} + \sum_{j} (\lambda_j + c_j)\phi_j^{sdn} + \eta_s + t_s + r_d).$$

Without considering the constraints (4.7), we can decompose the RWAS to $|A|$ sub-problems, each of which corresponds to one semi-lightpath. We shall use $\mathrm{RWA}_{\mathrm{sdn}}$ to denote the sub-problem in RWAS corresponding to the semi-lightpath s_{sdn}. Note that Q_{sdn} represents the minimum routing cost for the semi-lightpath s_{sdn}.

B) Electronic Traffic Routing Sub-problems (ETRS):

Considering the electronic domain subproblems in equation (4.13)

$$\sum_{(p,q)} \sum_{0<z\le Z_{sd}} \min_{\gamma_{pqz}} \{[\gamma_{pqz} \min_{\Upsilon_{pqz}}\{R_{pqz}\} + (1-\gamma_{pqz})P_{pqz}G_{pqz}]\},$$

subject to constraints (4.8), where $R_{pqz} = \sum_{(s,d)} \sum_{0<n\le N_{sd}} \upsilon_{sdn}^{pqz}(G_{pqz}\kappa_{sdn} + V_{pqz})$.

By assuming the cost of the semi-lightpath s_{sdn} is $G_{pqz}\kappa_{sdn} + V_{pqz}$, it is obvious that $\min_{\Upsilon_{pqz}}\{R_{pqz}\}$ represents the cost of the shortest path from nodes p to q for routing the traffic x_{pqz} over the virtual topology $(\mathcal{V}_v,\mathcal{E}_v)$, which is formed by the electronic switches and all the semi-lightpath candidates. We decompose the ETRS into $|Y|$ subproblems, each of which corresponds to one traffic demand. We shall use $\mathrm{ETR}_{\mathrm{pqz}}$ to denote the sub-problem in ETRS corresponding to x_{pqz}.

4.4.3.2 The solution to the sub-problems

The total number of sub-problems is $|A| + |Y|$, where $|\cdot|$ denotes the number of elements in the set. The optimal solutions for RWAS and ETRS may be obtained separately by applying RMCSLP and RSPA, respectively.

A) RMCSLP:

Intuitively, the semi-lightpath s_{sdn} can be thought of as a "broker" in the dual problem, which "negotiates" between cost and "offer". The cost of the semi-lightpath s_{sdn} is $\min_{\Delta_{sdn},\Phi_{sdn}}\{Q_{sdn}\}$, which reflects the criticality of the physical resources that the semi-lightpath uses. The more critical the resources that the semi-lightpath uses are, the higher cost the semi-lightpath has. The "offer" from the electronic traffic demands is $c\kappa_{sdn}$, which reflects the "intention" of the electronic traffic demands to use the semi-lightpath. The acceptance decision of the semi-lightpath is based on whether the semi-lightpath is "profitable". In this way, the RWAS and ETRS sub-problems are linked to each other.

We can see from the expression of Q_{sdn} that the first two summations assume the same form as the well-known problem that has been solved optimally by the MCSLP (Minimum Cost Semi-lightpath) [ChFa96] which is essentially a shortest path algorithm. Therefore, we can first obtain $\min_{\Delta_{sdn},\Phi_{sdn}}\{Q_{sdn}\}$ by applying MCSLP and then add the constants $\eta_s + t_s + r_d$. Afterwards each node i, we do the sorting of the s_{sin}'s to satisfy constraints (4.7). This is captured in our *Revised Minimum Cost Semi-lightpath* (RMCSLP) algorithm below in order to solve RWA_{sdn}:

A.1) Employ the MCSLP to solve $\min_{\Delta_{sdn},\Phi_{sdn}}\left[\sum_{(i,j)}\sum_{0<c\le n_{ij}}(\xi_{ijc}+d_{ijc})\delta_{ijc}^{sdn} + +\sum_{j}(\lambda_j + c_j)\phi_j^{sdn}\right]$, subject to constraints (4.2), (4.3). By assuming the cost of using WC w_{ijc} to be $d_{ijc} + \xi_{ijc}$, and the cost of using one wavelength converter on node j to be $\lambda_j + c_j$, add $\eta_s + t_s + r_d$ to obtain $\min_{\Delta_{sdn},\Phi_{sdn}}(Q_{sdn})$. The solution of the MCSLP algorithm can be found in [ChFa96].

A.2) For any node i, among all the RWA_{sin}'s,

A2.1) Compute $\min_{\Delta_{sin},\Phi_{sin}}(Q_{sin}) - C\kappa_{sin}$ for all RWA_{sin}'s, and select the R_i RWA_{sin}'s with the lowest $\min_{\Delta_{sin},\Phi_{sin}}(Q_{sin}) - C\kappa_{sin}$ values (the ties are broken arbitrarily);

if $\min_{\Delta_{sin},\Phi_{sin}}(Q_{sin}) - C\kappa_{sin} > 0$, set $\alpha_{sin} = 0$;

if $\min_{\Delta_{sin},\Phi_{sin}}(Q_{sin}) - C\kappa_{sin} < 0$, set $\alpha_{sin} = 1$. The tie is broken arbitrarily.

A2.2) If the number of RWA_{sin}'s is more than R_i, then set $\alpha_{sin} = 0$ for the rest of the RWA_{sin}'s.

B) RSPA:

In part A, we have solved RWAS optimally, and now we present the following *Revised Shortest Path Algorithm* (RSPA) to solve ETR_{pqz}.

B.1) Solve $\min_{Y_{pqz}}\{R_{pqz}\}$, subject to constraints (4.8), after applying SPA [Be98], and assuming the cost of using semi-lightpath s_{sdn} is $G_{pqz}\kappa_{sdn} + V_{pqz}$.

B.2) If $\min_{\Upsilon_{pqz}}\{R_{pqz}\} > P_{pqz}G_{pqz}$, set $\gamma_{pqz} = 0$;

If $\min_{\Upsilon_{pqz}}\{R_{pqz}\} < P_{pqz}G_{pqz}$, set $\gamma_{pqz} = 1$. The tie is broken arbitrarily.

4.4.4 Solving the Dual Problem

Because of the integer variables involved in the formulation, the subgradient method [Bert99] can be used to solve the dual problem.

Let ζ, λ, η, κ be the vectors of Lagrange multipliers $\{\zeta_{ijc}\}$, $\{\lambda_i\}$, $\{\eta_s\}$ and $\{\kappa_{sdn}\}$, respectively. The multiplier vector $z = (\xi, \lambda, \eta, \kappa)$ are updated using the following formula:

$$z^{(h+1)} = z^{(h)} + \alpha^{(h)} g(z^{(h)}), \tag{4.14}$$

where $z^{(h)}$ denote the value of vectors z obtained at the hth iteration, and $\alpha^{(h)}$ denote the step size at the hth iteration. The notation $g(z)$ is the subgradient of q with respect to z, i.e. $g(z) =\{g(\xi), g(\lambda), g(\eta), g(\kappa)\}$. The vectors $g(\xi)$, $g(\lambda)$, $g(\eta)$ and $g(\kappa)$ are composed of $g_{ijc}(\xi)$, $g_j(\lambda)$, $g_i(\eta)$ and $g_{sdn}(\kappa)$ respectively, where

$$g_{ijc}(\xi) = \sum_{(s,d)} \sum_{0<n\le N_{sd}} \delta_{jic}^{sdn} - 1, \tag{4.15}$$

$$g_j(\lambda) = \sum_{(s,d)} \sum_{0<n\le N_{sd}} \phi_j^{sdn} - F_j, \tag{4.16}$$

$$g_i(\eta) = \sum_j \sum_{(s,d)} \sum_{0<n\le N_{sd}} \Delta_{w_{jic}}^{sdn} - T_i, \tag{4.17}$$

$$g_{sdn}(\kappa) = \sum_{(p,q)} \sum_{0<z\le Z_{sd}} \upsilon_{sdn}^{pqz} G_{pqz} - C\alpha_{sdn}. \tag{4.18}$$

The step size is given by

$$\alpha^{(h)} = \mu \times \frac{q^U - q^{(h)}}{g^T(z^{(h)})g(z^{(h)})} \tag{4.19}$$

where q^U is an estimate of the optimal solution, and $q^{(h)}$ is the value of q at the hth iteration. Generally, the best value of the objective function J of the feasible routings obtained is used to be q^U. The parameters μ and q^U are changed adaptively as the algorithm converges. The details of the methods to speed up the convergence can be found in previous chapters.

4.4.5 Constructing a Feasible RWA and Routing Scheme

Because some of the constraints are relaxed by the Lagrange multipliers, the solution to the dual problem is generally associated with an infeasible routing scheme, i.e., WC capacity constraints (4.4), converter capacity constraints (4.5), transmitter capacity constraints (4.6) and semi-lightpath capacity constraints (4.9) might be violated. Note that other constraints will be respected because of the way sub-problems of (4.13) are solved. To construct a feasible RWA and a routing scheme from the solution to the dual problem, a heuristic algorithm has to be employed to decide how to re-route the collided semi-lightpaths and the corresponding traffic demands. The main idea of the heuristic algorithm is to select the less profitable semi-lightpaths and traffics to detour through other paths, or to reject them when there are conflicts.

The whole heuristic algorithm can be described as:

1. Construct the wavelength graph (WG). The details of this procedure can be found in [ChFa96].
2. (*RWA Step*): Search for a feasible RWA solution for the semi-lightpaths, based on the solution from the dual problem, using Modified Esau-Williams algorithm (MEWA) in Appendix A.4 to every s_{sdn}.
3. Construct the virtual topology: For all s_{sdn}, if $\alpha_{sdn} = 1$, add a directed arc from node s to node d with capacity of C.
4. (*Routing Step*): The traffic demands are routed through the virtual topology.
    4.1. (*Rough search stage*) For every traffic demand x_{pqz}, apply the *Rough Search Algorithm*. The details of this algorithm can be found in the Appendix A.4.

    4.2. (*Check solvency stage*) For every s_{sdn}, if $\alpha_{sdn} = 1$, calculate the total revenue from traffic $t_{sdn} = \sum_{(p,q)}\sum_{0<z\leq Z_{sd}} \upsilon_{sdn}^{pqz} P_{pqz} G_{pqz} / h_{pqz}$, where $h_{pqz} = \sum_{(s,d)}\sum_{0<n\leq N_{sd}} \upsilon_{sdn}^{pqz}$ is the hop number of x_{pqz}.

    If $t_{sdn} < D_{sdn}$, set $\alpha_{sdn} = 0$.

    For every x_{pqz} with $\upsilon_{sdn}^{pqz} = 1$, set $\gamma_{pqz} = 0$ and apply *Rough Search Algorithm* to x_{pqz}.

    4.3. If the result obtained from step 4.1 (J) > (1 + l) × J*, terminate the heuristic algorithm; otherwise go to the next step;

    4.4. Use another LR based algorithm to search for a better routing scheme, which consumes much more time than the *Rough Search Algorithm*. The details of the algorithm can be found in the Appendix A.5.

In the *RWA Step*, the routing and wavelength assignment of the semi-lightpaths is decided (i.e., variables in A, Δ, Φ are determined). Based on the result from this step, the virtual topology is formed by the semi-lightpaths connecting the nodes. In the *Routing Step*, the routing of the traffic demands over the virtual topology (i.e., variables in Γ, Y are determined). To minimize the computation time, we separate the heuristic algorithm of the *Routing Step* into two stages. In the first stage (*rough search stage*), a rough search of the feasible routing scheme is launched. Only if the result obtained in the first stage is within a certain range ($l > 0$) from the best result obtained (J^*), that the second stage, i.e., the extensive search (*extensive search stage*), is launched to search for a good routing scheme, based on the result from the dual solution and the virtual topology formed in *RWA Step*. In our current implementation, we chose $l = 0.05$.

To decide which semi-lightpath should be deployed first, we define the rank $r_{sdn} = L_{sd} - Q_{sdn}$ for all the semi-lightpaths, where Q_{sdn} represents the dual cost of the sub-problem corresponding to s_{sdn} (see section 4.4.3), and $L_{sd} = \sum_{0<n\leq N_{sd}} t_{sdn}$ represent the "revenue" from the left traffic load between (s, d). At the beginning of the heuristic algorithm, L_{sd} is set to be the total traffic going from nodes s to d in the dual problem solution. If the semi-lightpath s_{sdn} is deployed, L_{sd} is decreased by $m \times C$. If L_{sd} is less than 0, then L_{sd} will be set to 0. In our implementation, $m = 0.8$. The semi-lightpath with the highest rank r_{sdn} is deployed first, and the tie is broken arbitrarily.

The heuristic algorithm to search for a feasible RWA for a semi-lightpath is the *MEWA* in Appendix A.4.

In the *Check solvency stage*, the semi-lightpaths that have a higher cost (D_{sdn}) than the revenue (Pt_{sdn}) are deleted (set $\alpha_{sdn} = 0$) to minimize the overall objective function value. The traffic demands that are routed through those semi-lightpaths are set to the rejected status ($\upsilon_{sdn}^{pqz} = 0$), and re-routed again using *Rough Search Algorithm* to see if there might be some other viable route.

After the semi-lightpaths are deployed, the traffic demand should be routed through the virtual topology formed by the semi-lightpaths. In the *rough search stage*, we simply route the traffic demands with lower hop number in the dual solution. In the *extensive search stage*, a search algorithm employing LRSM is launched. The detailed algorithms for the *rough search stage* and the *extensive search stage* can be found in the Appendix A.6.

4.4.6 Computation Complexity

The computational complexity of step A.1 (See Section 4.4.3.2) in the RMCSLP is the same as the MCSLP provided in [ChFa96], i.e., $O((N+W)N^2W)$. The worst-case computational complexity for the sorting operation in step A.2 is $\Sigma_i O(G_i log(G_i))$, where G_i represents the number of semi-lightpaths that are destined for node i. The value of G_i is generally much smaller than NW, so the computational complexity is generally dominated by step a. There are altogether $|A|$ sub-problems in the sub-problem RWAS, so the computational complexity to solve RWAS is $O(|A|(N+W)N^2W)$.

The computational complexity of the RSPA algorithm is dominated by step B.1 (See Section 4.4.3.2). Since there are altogether $|\Gamma|$ sub-problems in the sub-problem ETRS and the computational complexity for the SPA is $O(|M|^2 N^2)$, so the computational complexity to solve the sub-problem TRS is $O(|\Gamma||M|^2 N^2)$.

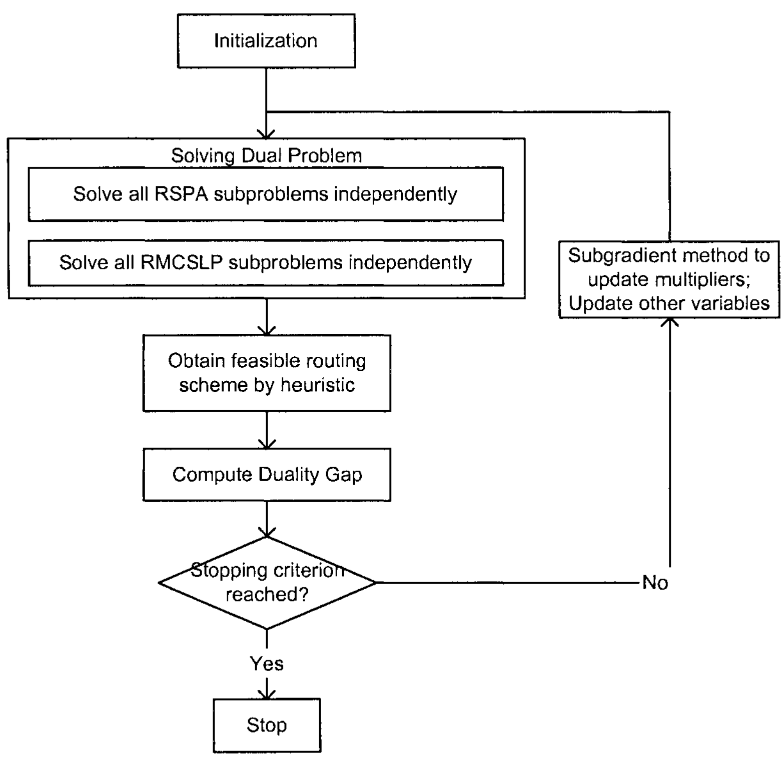


Figure 4.1. The schematic depiction of the overall algorithm

4.4.7 Overall Structure of the Algorithm

Figure 4.1 depicts the overall structure of the algorithm to solve the problem proposed in Section 4.4. We first solve the sub-problems independently using the RMCSLP and RSPA (Section 4.4.3). Note that solving all the sub-problems is to solve the dual problem (Section 4.4.2). After the solution for the dual problem is obtained, a heuristic algorithm (Section 4.4.5) is used to generate a feasible solution based on the solution for the dual problem. If the stopping criterion is not reached, the Lagrange multipliers are updated (Section 4.4.4), and another iteration of computation is launched. There are various criteria that can be used to terminate the algorithm, such as the duality gap, the number of iterations, the time of computation and the objective function value, etc. In the current implementation, we stop the algorithm when the duality gap does not decrease for 1000 iterations.

4.5 Performance Evaluation

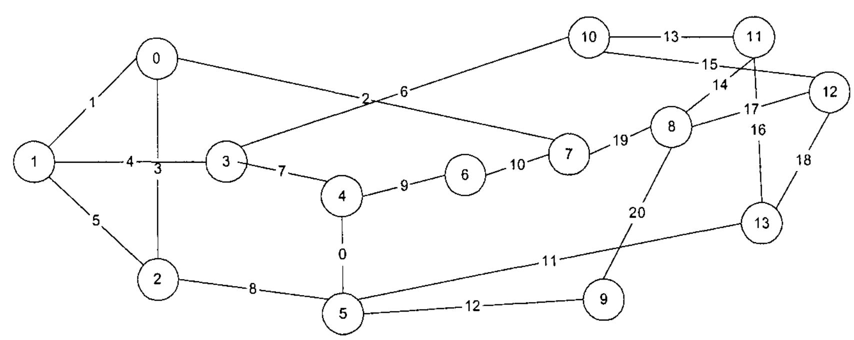


Figure 4.2. 14-node NSFNET topology

To test the performance of our algorithm, we use the 14-node NSFNET topology as shown in Figure 4.2. The randomly generated traffic demands are shown in Table 4.1 where the horizontal and vertical indexes are the source and destination nodes, respectively. The total traffic counts to an equivalence of 1880 times of OC-1 traffic (i.e., 1880 * 51.84 Mbps = 97.46 Gbps). Because under our assumption, an OC-48 traffic occupies a whole semi-lightpath, traffic grooming for OC-48 traffic becomes the well-studied RWA problem. Thus, we use only the lower bandwidth traffic to study the behaviour of the algorithm. The number on the ith row and the jth column represents the number of traffic demands of specific types from node i to j.

We set all penalty coefficients P_{pqz}'s to 1. This means the potential revenue generated by accepting a traffic demand is proportional to its bandwidth requirement. We will demonstrate the influence of different P_{pqz} values reflecting different pricing policy at the end of this section. We also set the transmitter cost t_i and the receiver cost r_i to the same value and refer to it as the transceiver cost. Various computation results for different parameters (W, V_{pqz}, r_i, t_i, d_{ijc}, T_i, R_i, F_i

and c_i) are obtained. To study the influence on the network behaviour from different parameters in the optical domain, we firstly study the cases with V_{pqz}=0, i.e., ignoring the electronic routing cost. Then we study the cases with various V_{pqz} values with electronic routing cost.

Table 4.1. Traffic demand matrices

OC-1 TRAFFIC DEMANDS

0	1	3	1	5	1	3	0	2	0	1	2	0	3
0	0	2	2	2	11	1	1	1	2	1	0	1	3
3	2	0	3	0	1	2	3	1	3	1	2	2	0
3	1	4	0	1	1	2	3	2	2	1	2	1	3
1	3	0	2	0	1	0	2	0	3	0	1	1	3
1	2	1	3	2	0	1	3	3	11	0	6	10	9
2	2	3	1	10	3	0	0	3	1	2	0	3	7
3	10	2	3	1	4	1	0	0	3	2	0	3	0
3	0	12	3	3	3	1	0	0	2	1	1	1	0
0	0	0	1	2	0	2	0	1	0	1	0	0	3
1	0	0	10	0	3	0	1	0	3	0	3	1	3
2	3	1	1	3	2	3	2	10	2	2	0	1	3
13	0	1	2	0	1	2	0	9	0	2	1	0	3
10	14	0	15	9	3	1	3	0	12	2	1	3	0

OC-3 TRAFFIC DEMANDS

0	1	0	1	1	1	0	2	2	0	1	2	0	1
0	0	0	2	2	2	1	1	1	2	1	0	1	1
1	2	0	1	0	1	2	1	1	1	1	0	2	0
1	1	0	0	1	1	2	1	2	2	1	2	1	1
1	1	0	2	0	1	0	2	0	1	0	1	1	1
1	2	1	1	2	0	1	1	1	1	0	1	1	2
2	2	1	1	1	1	0	0	1	1	2	0	0	1
2	1	2	1	1	0	1	0	0	1	2	0	1	0
1	0	1	1	1	1	1	0	0	2	1	1	1	0
0	0	0	1	2	0	2	0	1	0	1	0	0	1
1	0	0	2	0	0	0	1	0	1	0	1	1	2
2	1	1	1	1	2	1	2	2	2	2	0	1	1
0	0	1	2	0	1	2	0	1	0	2	1	0	1
1	1	0	2	1	0	1	1	0	1	2	1	0	0

OC-12 TRAFFIC DEMANDS

0	1	0	1	0	1	0	0	1	0	1	1	0	0
0	0	0	1	0	0	1	0	0	1	1	0	1	0
0	1	0	0	0	1	0	0	1	0	1	1	0	0
0	1	0	0	1	1	1	0	1	1	0	1	1	0
1	0	0	1	0	1	0	1	0	0	0	0	0	0
1	1	0	0	1	0	1	0	0	1	0	1	0	1
1	1	0	1	0	0	0	0	0	1	1	0	0	0
0	1	1	0	1	0	1	0	0	0	1	0	0	0
0	0	1	0	0	0	1	0	0	1	1	0	1	0
0	0	0	1	1	0	0	0	1	0	1	0	0	0
1	0	0	0	0	0	0	1	0	0	0	0	1	0
1	0	0	1	0	1	0	1	0	1	1	0	1	0
0	0	1	0	0	1	1	0	0	0	1	1	0	0
1	1	0	1	1	0	1	0	0	1	0	1	0	0

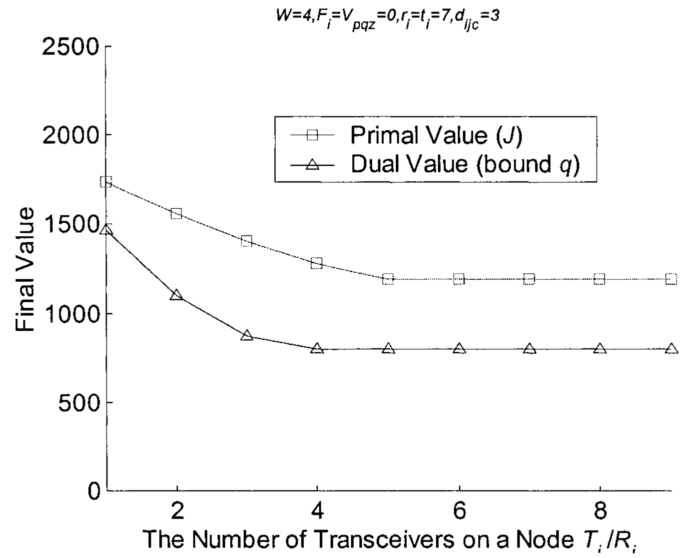


Figure 4.3. *J* and *q* versus T_i/R_i

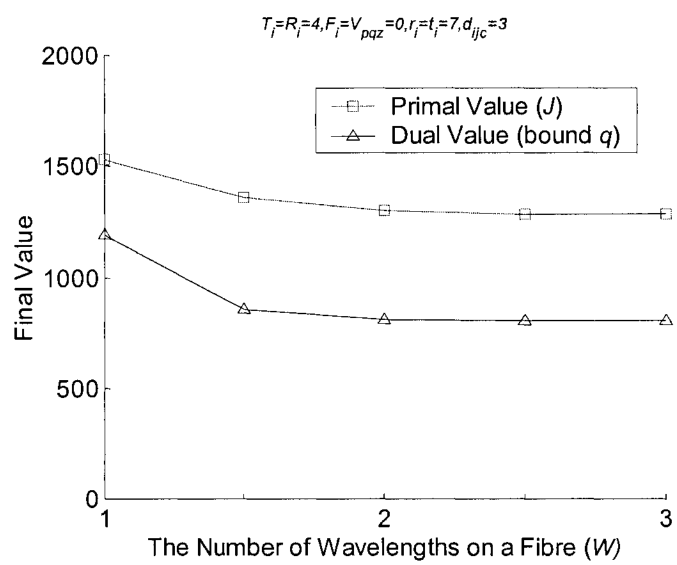


Figure 4.4. *J* and *q* versus *W*

Figure 4.3 shows the influence of the number of transceivers (T_i/R_i) on J and q. The other parameters are set as W=4, V_{pqz}=0, r_i=t_i=7, d_{ijc}=3, and F_i=0, which is indicated on the top of the graph. In the following, the parameters we use are all listed on the top of the graphs, same as Figure 4.3. We can see from Figure 4.3 that as the number of transceivers (T_i/R_i) increases, the final objective value J and the dual function value q decrease. J decreases almost linearly before T_i and R_i reach a certain value (5 in this case). No further improvement can be achieved by increasing the number of transceivers beyond that. It indicates that when a certain number of resources are installed, no further profit can be achieved by installing more resources. This is because that when the resources (fibers, transceivers, and wavelength converters) are abundant, J is limited by the resource cost, instead of the amount of resources. Figure 4.4 reflects the changes of J and q with respect to the number of wavelengths on a fiber (W). We can see from Figure 4.4 that as W increases, J and q decrease. However, when W reaches a certain value (4 in this case), J cannot be further improved by increasing W, for the same reason explained above for Figure 4.3. Most of the results can be obtained in two hours.

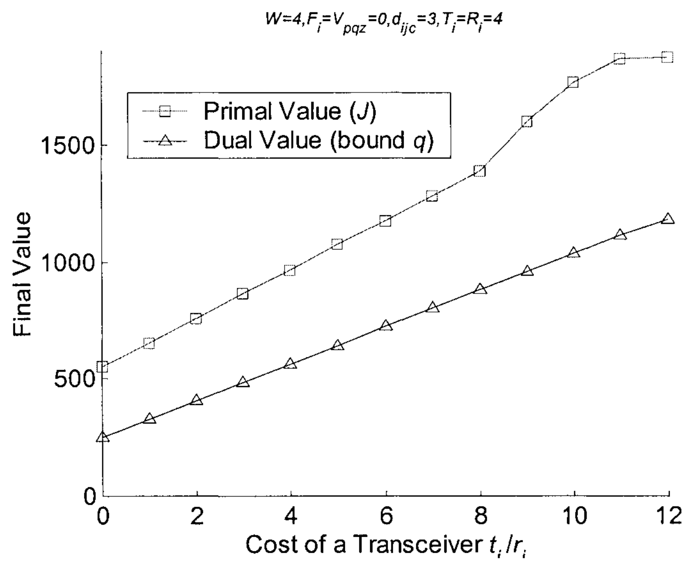


Figure 4.5. J and q versus t_i/r_i

In Figure 4.5 and Figure 4.6, we can see that as the cost of transceivers (t_i/r_i) and cost of WC (d_{ijc}) increases, the objective value J increases almost linearly. In Figure 4.5, we can see that as t_i and r_i reach a certain value (11 in this case), because the cost of setting up the semi-lightpaths is so high that most traffic demands are rejected, J stays close to 1880, which is the penalty value for rejecting all the traffic demands (the total bandwidth of the traffic is OC-1880 and P_{pqz}=1). J cannot become worse when t_i and r_i increase more. The similar phenomenon is observed in Figure 4.6.

Figure 4.7 shows the influence of d_{ijc} on the average hop number of the semi-lightpaths. When the cost of a WC increases, the hop number of a semi-lightpath tends to be lower. We can see from Figure 4.7 that as d_{ijc} increases, the average hop number of the semi-lightpaths decreases, until it reaches the minimum hop number (i.e., 1).

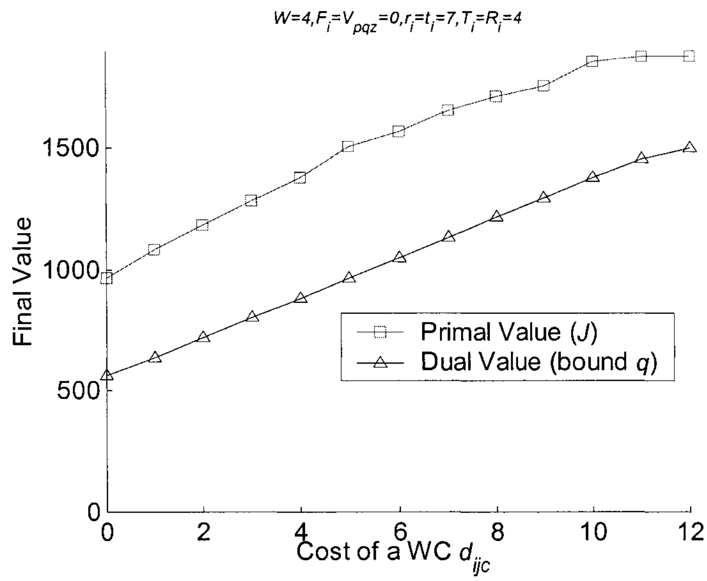


Figure 4.6. *J* and *q* versus d_{ijc}

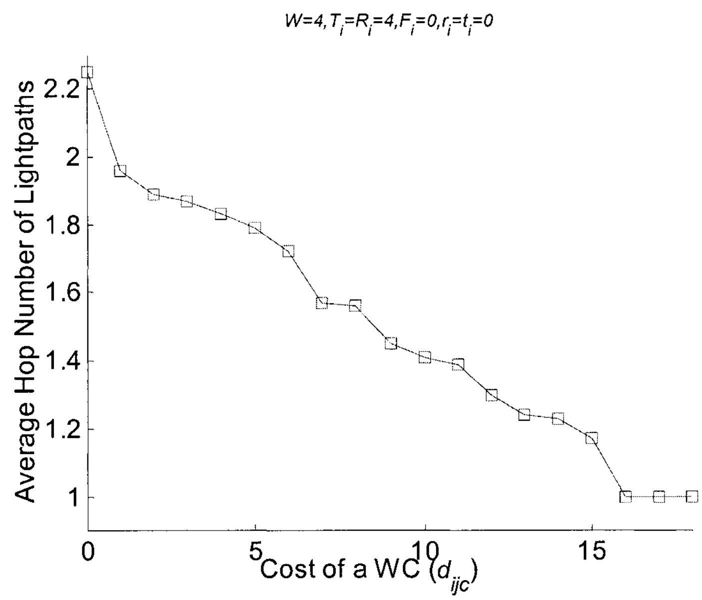


Figure 4.7. Average hop number of semi-lightpaths versus d_{ijc}

To study the influence from V_{pqz} on the hop number of the electronic traffic demands, we set the V_{pqz} value of OC-1 and OC-12 traffic to 0.1 and 1.2, respectively, and change the V_{pqz} value of OC-3 to derive the curve shown in Figure 4.8. We can see that as V_{pqz} increases, the hop number of OC-3 traffic decreases. However, when V_{pqz} reaches 1.5, all semi-lightpaths become single-hop. The reason is that the revenue of an OC-3 stream is 3 when P_{pqz}=1. When V_{pqz} reaches 1.5, the traffic cannot afford to go 2 hops anymore ($1.5 \times 2 = 3$), because the cost for each hop is at least 1.5. The similar behaviour does not appear in Figure 4.7, which is because the revenue of the semi-lightpath is proportional to the total bandwidth of the traffic routed through it. As d_{ijc} increases, more traffic demands are rejected, and thus the revenue of the semi-lightpaths drops slowly. The average hop number that the semi-lightpaths can afford therefore also decreases slowly until reaches 1. Note that the "hop" in Figure 4.7 and Figure 4.8 have different meanings. In Figure 4.7, the hop is the fiber-count over the physical topology, while in Figure 4.8, the hop is the semi-lightpath-count on the virtual topology.

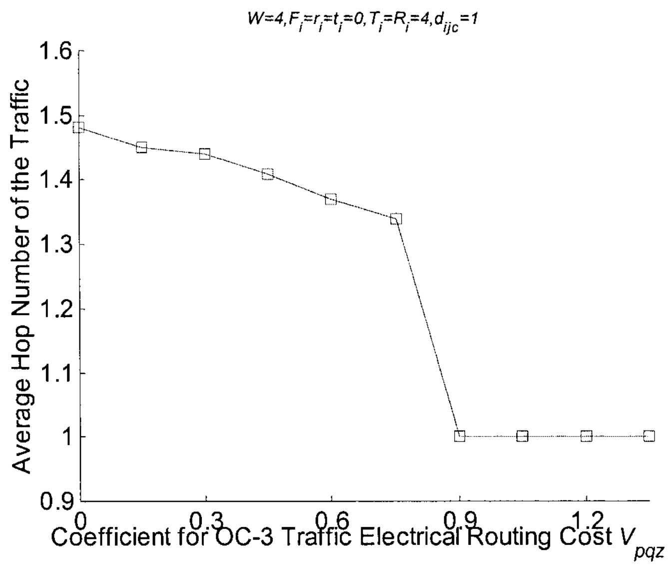


Figure 4.8. Average hop number of traffic versus V_{pqz}

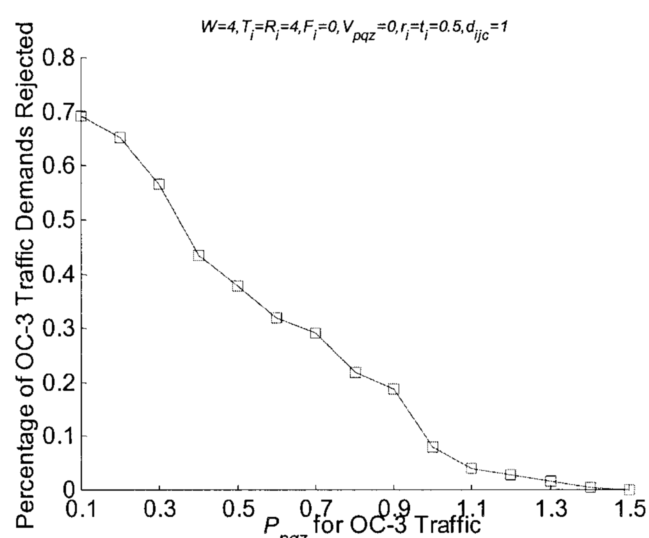


Figure 4.9. Percentage of OC-3 traffic rejected versus P_{pqz}

Figure 4.9 shows the influence of the OC-3 traffic's penalty coefficient (P_{pqz}) on the percentage of the OC-3 traffic rejected, while the P_{pqz} value for the OC-1 and OC-12 traffic remains 1. We show in Figure 4.9 that as P_{pqz} increases, the percentage of OC-3 traffic demands rejected decreases. This is because as the revenue of OC-3 traffic increases (i.e., P_{pqz} goes higher), it is more likely for the network to gain profit to route OC-3 traffic than other traffic (OC-1 and OC-12). Also, as the revenue goes higher, it is more likely to set up profitable semi-lightpaths. After P_{pqz} reaches a certain value, the majority of the OC-3 traffic are accepted, and there are still some residual OC-3 traffic demands that need some expensive semi-lightpaths to be set up, which might not be profitable, when the P_{pqz} is not high enough.

We also extensively study the benefit of using wavelength converters. Some of the results under different parameters are listed in Table 4.2.

Table 4.2. Influence of F_i on J

W	T_i, R_i	d_{ijc}	r_i, t_i	V_{pqz}			F_i	c_i	J
				OC-1	OC-3	OC-12			
4	4	0	0	0	0	0	0	0	227
4	4	0	0	0	0	0	1	0	216
4	4	0.2	0	0	0	0	0	0.01	262
4	4	0.2	0	0	0	0	1	0.01	235
4	4	3	7	0	0	0	0	0	1284
4	4	3	7	0	0	0	1	0	1284
4	4	1	0	0.1	0.3	1.2	0	0.01	580
4	4	1	0	0.1	0.3	1.2	1	0.01	580

We can see that the benefit on J of having wavelength converters is marginal, even if we assume that the wavelength converter is free, by setting the cost of wavelength converters (c_i) to 0. The cases with wavelength conversion capacity (F_i) larger than 1 are not listed in Table II, because we do not observe any improvement by setting F_i to values larger than 1. This observation is in accordance to the results obtained in Chapter 3, which concludes that F_i does not have significant influence on the total RWA cost in optical domain.

<u>Special Case</u>:

When we set all the costs for the semi-lightpath routing, i.e., r_i, t_i, d_{ijc}, and c_i, to 0, and set the F_i to 0. The problem we solve is exactly the same as the multi-hop case studied in [ZhMu02].

To have a fair comparison with the existing algorithm, we used the same 6-node network and same traffic matrix used in [ZhMu02]. The topology and the traffic matrices are shown in Figure 4.10, and Table 4.3, respectively.

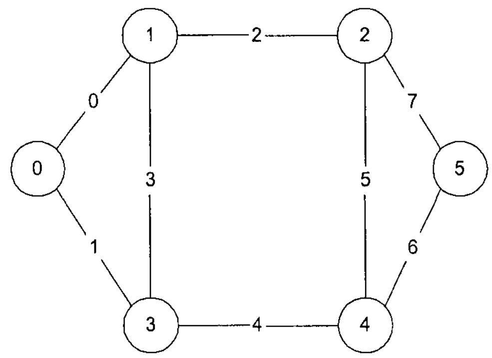


Figure 4.10. A 6-node network

Table 4.3. Traffic matrices for the 6-node network (Equivalent to 988 OC-1 bandwidth in total)

OC-1 TRAFFIC DEMANDS

0	5	4	11	12	9
0	0	8	5	16	6
14	12	0	9	6	16
4	11	15	0	1	5
10	2	3	3	0	9
2	1	8	15	13	0

OC-3 TRAFFIC DEMANDS

0	6	2	1	5	4
8	0	8	6	7	8
1	3	0	0	2	7
5	7	3	0	2	6
6	4	5	0	0	2
5	4	4	2	0	0

OC-12 TRAFFIC DEMANDS

0	1	1	1	0	0
1	0	1	1	0	2
0	1	0	2	1	0
2	0	2	0	2	0
1	2	0	2	0	1
1	1	2	2	2	0

Table 4.4. The comparison between the results in [ZhMu02] and our results

	Result from CPLEX [ZhMu02]	Result from Heuristic 1 [ZhMu02]	Result from Heuristic 2 [ZhMu02]	Result from our algorithm	Upper bound from our algorithm
T=2,W=3	--	--	--	516	576
T=3,W=3	738	701	666	751	851
T=4,W=4	927	883	925	930	976
T=5,W=3	967	933	933	969	987
T=7,W=3	967	933	933	969	987
T=3,W=4	738	701	666	751	851
T=4,W=4	933	920	925	930	976
T=5,W=4	988	988	988	988	988

The comparison between the results provided in [ZhMu02] and our algorithm is shown in Table 4.4. Results are numbers of equivalent OC-1 bandwidth routed by different algorithms. We can see that the amount of traffic that can be routed by our algorithm is much more than either of the heuristic algorithms provided in [ZhMu02], and in most cases, is even better than the result generated from the CPLEX. This proves that the results from CPLEX in [ZhMu02] are not optima, because after a period of time the optimal value were not yet able to be found and the computation was terminated.

4.6 Concluding Remarks

In this chapter, we have provided an integrated design of both layers in the traffic grooming network problems, which offers better results than the two-step design for the two layers, although the optimization problem is more complex and thus solving the optimization problem is more challenging. By properly relaxing some of the constraints, we have solved the problem by using LRSM. The overall dual function was decomposed into two sets of sub-problems. The solutions to the two subproblems are independent and the Lagrange multipliers associate the two subproblem sets. This procedure simplifies the solution and the overall time complexity to solve the dual problem is polynomial. An effective heuristic algorithm has also been proposed to generate the feasible result based on the solution to the dual problem. As a special case of our formulation, the comparison between our results and some other research has shown the high effectiveness of our algorithm.

We have also proposed a novel objective function taking into account the rejection penalty, the routing cost associated with the electronic and optical domains. By optimizing this cost function, the potential profit of network operation is maximized. Compared to accepting traffic demands without considering the network cost, the new objective is adjusted to accepting only the most profitable traffic demands with a given network capacity. As a result, the traffic demands can be identified as profitable or not. By doing a series of computations for different parameters, the impact of network resource allocation on the network profit has been obtained.

In all the previous chapters, we do not consider the existing network configuration at the time of reconfiguration. In the next chapter, we will study the network reconfiguration issue that is critical to the network real-time operations.

Chapter Five
Semi-dynamic Reconfiguration Problem in WRN Network

In all the previous chapters, the network scenario we studied is pure static optical network, which assumes that every time when the network is reconfigured, all the existing connections are disconnected or rerouted. However, in a real operating network, any reconfiguration of the existing connection will result in the disruption of the upper level traffic, thus aggravating the quality of service. It is important to minimize the rerouting of the existing connections. Therefore, we shall address the semi-dynamic network reconfiguration problem by optimizing the demand rejection, the reconfiguration, and the network congestion in the network, as well as the fairness concern. In this chapter we only consider the optical layer of the network, which is an extension to the RWA problems studied in Chapter Three.

5.1 Problem Formulation

In formulating the network design problem in this chapter, we assume that once scheduled, all the semi-lightpath connection matrices and node connection matrices are assumed to be fixed over time until the next reconfiguration session. In the next session, the demand matrix may have changed and therefore network may have to be reconfigured accordingly, but any reconfiguration must consider the existing network configuration (i.e., the routings and wavelength assignments of the existing connections) [GhDi00].

As introduced before and detailed later, the relationship between the number of demands (N_{sd}) and the number of existing connections (X'_{sd}) is of great importance to the problem formulation. We need a preformulation procedure as described below relating each s_{sdn} with one t'_{sdn} to formulate the network reconfiguration and to facilitate a decompositional solution approach. After the preformulation, we can then re-formulate the RWA problem to form a relationship (according to the rules that will be presented later) between the *existing connections* (from the previous reconfiguration session) and the traffic demands for the coming session in order to decide the new network configuration, with respect to certain optimization objectives.

5.1.1 Pre-formulation Processing

We refer to the new connection demands to be established (might as well be rejected) in the new session as *new connection demands*; the connections established in the previous session as *existing connections*. To facilitate the illustration, we introduce the variable s_{sdn}, representing the nth new connection demand of source-destination pair (s, d). Let t'_{sdn} represent the nth existing connection of (s, d). Let α_{sdn} be a 0-1 integer variable, representing the rejection status of s_{sdn}. It has a value of 0, if s_{sdn} is rejected, and 1 if admitted. Note that all existing connections are the connection demands that are accepted in the previous session (α'_{sdn}=1).

Before formulating the problem, we should first understand the relationship between the capacities of the two sessions. To ensure that the 'promised' capacity from the previous session does not suddenly 'disappear' in the new session, we should use the following rules:

Rule 1: If there are more or the same number of new connection demands between (s, d) as the number of existing connections (i.e., $N_{sd} \geq X'_{sd}$), the connection demands accepted in the new session have to be more or equal to the previous session.

Rule 2: If there is less capacity demand in the new session between (s, d), i.e., $N_{sd} < X'_{sd}$, the connection demands accepted in the new session has to be equal to N_{sd}.

These two rules (formulated in Constraint (5.9)) avoid the drastic reconfiguration in the optical layer, which also results in less disturbance in the upper layer traffics, thus stabler network performance can be achieved [SaLu05]. Note that when N_{sd}=0, all existing connections between (s, d) will be disconnected in the new session, which means we can just ignore any existing connection between those (s, d) pairs.

One major difficulty of the formulation lies in the fact that for the same (s, d), X'_{sd} (the number of existing connections) is generally different from N_{sd} (the number of new connection demands), and we want to associate every new connection demand with an existing connection to relate the two consecutive sessions. At the same time, the rejection penalty assigning scheme (for the fairness consideration) in Section 3.2 (which assigns different rejection penalties to every s_{sdn} for the same (s, d) without differentiation) cannot be used any more, because all s_{sdn}'s for the same (s, d) in Section 3.2 are identical, while s_{sdn}'s (for the same (s, d)) in this thesis might be associated with existing connections, which take non-identical RWAs.

To formulate the problem we want to study, we hereby introduce a variable H_{sd}=max(N_{sd}, X'_{sd}) to relate the two adjacent sessions and to simplify the description. Because the new connection demands (s_{sdn}'s) between (s, d) are all identical, without losing generality when $N_{sd} \geq X'_{sd}$, we relate the new connection demand s_{sdn} to an existing connection t'_{sdn}, by using the same index n. Specifically, every connection demand (s_{sdn}), with the index $0<n \leq X'_{sd}$, corresponds to the existing connection t'_{sdn}; for s_{sdn}'s with $X'_{sd}<n\leq N_{sd}$, and they are not associated with any existing connection (H_{sd}=max(N_{sd}, X'_{sd})=N_{sd} new connection demands in total). However, if $N_{sd}<X'_{sd}$, the choice of which existing connections to disconnect has to be optimized. In this case, we will create totally H_{sd}=max(N_{sd}, X'_{sd})=X'_{sd} new connection demands between (s, d) in the formulation ($f=X'_{sd}-N_{sd}$ new connection demands are 'dummy' demands), and because that Constraint (5.9) confines that the eventual number of connections accepted have to be equal to N_{sd}, the optimization result will not be influenced by introducing more 'dummy' demands, and we can, on the other hand, make the formulation much easier by just associating every existing connection with a new connection demand. Based on the procedure above, we can simply use H_{sd}=max(N_{sd}, X'_{sd}) new connection demands for every (s, d) in the formulation.

5.1.2 Objective and Constraints

Our formulation is penalty-based, penalising the rejection of demands, the reconfiguration and the congestion. We adopt the same rejection penalty system introduced in [ZhYa04] to consider the fairness, i.e., assigning decreasing rejection penalties to the s_{sdn}'s of the same (s, d). Let $P_{sd}(k)$ be the penalty for rejecting k connection demands of (s, d), Q be the penalty for reconfiguring an existing connection (in other words, the connection is rerouted), and G be the penalty for the congestion. We use the overall penalty to formulate our objective function as shown in the following section.

5.1.2.1 Objective Function

Our objective function is

$$\min_{\mathrm{A,B,\Delta,\Phi}} \quad \{J\}, \text{ with } J \equiv \sum_{(s,d)} \left[P_{sd}\left(\sum_{0<n\leq H_{sd}} [1-\alpha_{sdn}] \right) + \sum_{0<n\leq H_{sd}} Q \cdot \beta_{sdn} \right] + G \cdot \gamma \,. \tag{5.1}$$

This function consists of the summation of three penalties: the rejection penalty, the reconfiguration penalty and the congestion penalty. The summation of $P_{sd}\left(\sum_{0<n\leq H_{sd}}[1-\alpha_{sdn}]\right)$ is the total rejection penalty for the overall traffic demand in the network (See the following paragraph for more detailed explanations). The second term (i.e., the summation of $Q\cdot\beta_{sdn}$) in the objective function penalizes the overall reconfiguration, where the variable β_{sdn} represents the reconfiguration status of an existing connection t'_{sdn} corresponding to s_{sdn}, which equals one if t'_{sdn} is rerouted, and 0 otherwise (see Constraint (5.8)). The third term ($G\cdot\gamma$) in the objective function represents the penalty for the congestion, where the variable γ ($0\leq\gamma\leq1$) represents the congestion, i.e., the largest percentage of the WCs used on all links (see Constraint (5.7)).

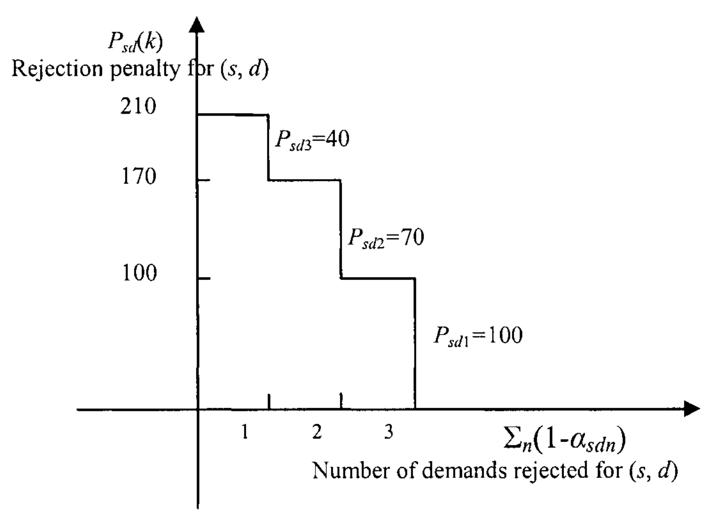


Figure 5.1. The illustration graph of the rejection penalty for (s, d)

Figure 5.1 shows a typical rejection penalty assigning scheme we used, it shows an example of an (s, d) with 3 connection demands. The rejection is penalized increasingly. Specifically, rejecting the first connection demand is the easiest (P_{sd3}=40 in the example); rejecting the second connection demand is more penalized than the first one (P_{sd2}=70); rejecting the last connection demand reaches the highest penalty (P_{sd1}=100). Because the optimization algorithm always tries to obtain the lowest overall penalty, the rejection of the demands can thus be balanced among all the (s, d)'s. We use the same idea as stated above in this chapter. However, due to the procedure of associating s_{sdn} with the non-identical t'_{sdn}, we cannot penalize every s_{sdn}, because how to assign penalties to every s_{sdn} is an optimization problem itself. We have to therefore penalize the number of rejections for every (s, d), which is denoted as $P_{sd}(k)$ in Figure 5.1 as the summation of the penalties for all s_{sdn}'s. More precisely, $P_{sd}(k)$ is the penalty coefficient for rejecting k connection demands of (s, d), and $P_{sd}(0) = 0$. If no fairness is considered, the penalty for rejecting any connection demand should be all the same, say P, which means $P_{sd}(k) = k \times P$. When the fairness is considered, the penalty to reject one connection demand should decrease as the number of demands of (s, d) increases to make the rejection easier, so that the rejection of the connection demands from all source-destination pairs can be balanced in the optimization procedure. Specifically, the penalty to reject one more connection demand, when there are k connection demands already rejected, is $P_{sdk}=(P-(N_{sd}-k)S)$, and $P_{sd}(k) = \sum_{h=1}^{k} P_{sdh}$, ($S$ is a constant rejection penalty step size). The term $\sum_{0<n\le H_{sd}} [1-\alpha_{sdn}]$ is the number of connection demands rejected for (s, d). To accommodate the source-destination pairs with $N_{sd}<X'_{sd}$ (for which there are totally $f = X'_{sd}-N_{sd}$ 'dummy' demands), we can simply set $P_{sd1} = P_{sd2} = \ldots = P_{sdf} = 0$, so that f demands will be naturally rejected, because the rejection will result in the least overall cost. Considering both situations, we can now represent P_{sdk} as follows: $P_{sdk}=(P-(H_{sd}-k)\cdot S)$. If $f>0$, $P_{sd1} = P_{sd2} = \ldots = P_{sdf} = 0$.

5.1.2.2 Constraints

The constraints can be classified into *General Constraints* and *Session Relation Constraints*.

5.1.2.2.1. General Constraints

General Constraints confines the network operation in an independent session, which are the same as classic static RWA problems. Note that the existing connections are the accepted connection demands from the previous session and conform to the General Constraints too.

a) Semi-lightpath flow continuity constraints:

Semi-lightpath continuity means that if a demand is admitted, the semi-lightpath assigned to it has to be continuous along the path between the source-destination pair. Since the assigned semi-lightpaths terminate at the two end nodes, we have

$$\sum_{j\in\mathcal{V}} \sum_{0<c\le n_{ij}} \delta_{ijc}^{sdn} - \sum_{j\in\mathcal{V}} \sum_{0<c\le n_{ij}} \delta_{jic}^{sdn} = \begin{cases} \alpha_{sdn} & \text{if } i = s, \\ -\alpha_{sdn} & \text{if } i = d, \ \forall (s,d), 0<n\le H_{sd}, \\ 0 & \text{otherwise,} \end{cases} \tag{5.2}$$

where δ_{ijc}^{sdn} is a 0-1 integer variable, which equals to one if wavelength channel w_{ijc} is used by s_{sdn}, and zero otherwise. Note that δ_{ijc}^{sdn} equals 0, if $\alpha_{sdn} = 0$. If s_{sdn} is accepted, (i.e. $\alpha_{sdn} = 1$) at the source node ($i = s$), there is one unit of flow going out of this node, thus equation (5.2) equals 1. At the destination node ($i = d$), there is a flow of 1 coming into this node, thus equation (5.2) equals –1. Finally, at the intermediate nodes, equation (5.2) equals 0 due to the conservation of flows. If the s_{sdn} is rejected (i.e., $\alpha_{sdn} = 0$), equation (5.2) equals 0 at any node.

b) Wavelength conversion constraints:

$$\phi_{j,ab}^{sdn} = \begin{cases} 1, & \text{if } \exists m,k \in \mathcal{V} \text{ and } b \neq a, \delta_{mja}^{sdn} = \delta_{jkb}^{sdn} = 1, \\ 0, & \text{otherwise.} \end{cases} \tag{5.3}$$

where $\phi_{j,ab}^{sdn}$ is a 0-1 integer variable representing the use of wavelength converters by s_{sdn} at node j to convert a signal from wavelength a to wavelength b. The value '1' means in-use and '0' not-in-use. These constraints are due to the semi-lightpath flow continuity constraints in (5.2). They stipulate that a converter with index a on an intermediate node j is used only when different wavelengths are assigned to s_{sdn} for the incoming and outgoing signals at this node.

c) Wavelength channel capacity constraints:

$$\sum_{(s,d)} \sum_{0<n\leq H_{sd}} \delta_{ijc}^{sdn} \leq 1 \qquad \forall (i,j), 0<c\leq n_{ij} \tag{5.4}$$

These constraints mean every wavelength channel on a fiber allows only one connection routed in the same direction.

d) Limited wavelength conversion degree constraints:

$$\delta_{ijc}^{sdn} = \sum_{y\in\mathcal{V}} \sum_{w_{jyx}\in I_j(c)\cap W_{jy}} \delta_{jyx}^{sdn} \ \forall (s,d), 0<n\leq H_{sd}, 0<c\leq n_{ij}; \ \forall (i,j), j\neq d, i\neq s \tag{5.5}$$

These constraints stipulate that a wavelength can only be converted to a certain set of wavelengths that are allowed by the wavelength converters and physical links.

e) Converter capacity constraints:

$$\sum_{(s,d)} \sum_{0<n\leq H_{sd}} \sum_{j\in\mathcal{V}} \sum_{0\leq a<n_{ij}} \phi_{i,ca}^{sdn} \leq F_{ic} \qquad \forall i, 0\leq c<W \tag{5.6}$$

These constraints mean that the number of occupied converters with an index c in node i cannot be more than the number of the converters available on this node.

f) Link congestion constraints:

$$\sum_{(s,d)} \sum_{0<n\leq H_{sd}} \sum_{0\leq c<n_{ij}} \delta_{ijc}^{sdn} \leq \gamma |W_{ij}|, \tag{5.7}$$

where W_{ij} denotes the wavelength set available in the physical link e_{ij}, γ represents the congestion (note that $0 \le \gamma \le 1$) and $|\cdot|$ denotes the number of elements in the set. In this chapter, we assume that all links have the same number of wavelengths, i.e., $|W_{ij}|=W$. These constraints have the same meaning as $W\gamma = \max_{(i,j)}\left\{\sum_{(s,d)}\sum_{0<n\le H_{sd}}\sum_{0\le c<n_{ij}}\delta_{ijc}^{sdn}\right\}$. We use this formulation to facilitate the mathematical solution.

5.1.2.2.2. Session Relation Constraints

Session Relation Constraints stipulate the relationship between the two consecutive sessions. In other words, the relationship between the existing connections and the new traffic demands.

g) Reconfiguration constraints:

$$\beta_{sdn} = \begin{cases} 1, \text{if } \Delta_{sdn} \neq \Delta'_{sdn} \text{ and } 0 < n \le X'_{sd} \\ 0, \qquad\qquad\qquad \text{otherwise} \end{cases} \tag{5.8}$$

These constraints mean that only if an existing connection t'_{sdn} is rerouted, β_{sdn} equals 1, and β_{sdn} equals 0 otherwise. Note that β_{sdn} equals 0, if s_{sdn} is not associated with any existing connection.

h) Non-disconnection constraints:

$$\begin{cases} \sum_{0<n\le N_{sd}} \alpha_{sdn} = N_{sd}, \text{ if } N_{sd} < X'_{sd} \\ \sum_{0<n\le N_{sd}} \alpha_{sdn} \ge X'_{sd}, \text{ if } N_{sd} \ge X'_{sd} \end{cases}. \tag{5.9}$$

These constraints ensure that the bandwidth promised in the previous session is observed in the new session, by confining that the existing connections (say t'_{sdn}) can be disconnected in the new session only when there are less connection demands between (s, d) in the new session. If there are more or the same number of connection demands between (s, d) in the new session, i.e., $N_{sd} \ge X'_{sd}$, the number of accepted new connections cannot be less than the number of existing connections ($\sum_{0<n\le N_{sd}} \alpha_{sdn} \ge X'_{sd}$); if there are less connection demands between (s, d) in the new session, i.e, $N_{sd} < X'_{sd}$, only $(X'_{sd} - N_{sd})$ connections should be disconnected ($\sum_{0<n\le N_{sd}} \alpha_{sdn} = N_{sd}$).

5.2 Solution Methodology

By properly relaxing some constraints, we will derive in this section the DP (Dual Problem), which can be decomposed into independent subproblems.

5.2.1 The LR Solution Procedure

We first use the Lagrange multipliers ξ_{ijc}, λ_{ic}, π_{ij} to relax respectively the wavelength channel capacity constraints (5.4), converter capacity constraints (5.6) and link congestion constraints (5.7). This leads to the following Lagrangean dual problem, which we shall denote as **DP** (Dual Problem):

$$\max_{\xi,\lambda,\pi\geq 0} q = \min_{A,B,\Delta,\Phi,\gamma}\left\{\sum_{(s,d)}[P_{sd}\left(\sum_{0<n\leq H_{sd}}[1-\alpha_{sdn}]\right)+\sum_{0<n\leq H_{sd}}Q\beta_{sdn}+G\gamma]+\sum_{(i,j)}\sum_{0<c\leq n_{ij}}\xi_{ijc}(\sum_{(s,d)}\sum_{0<n\leq H_{sd}}\delta_{ijc}^{sdn}-1)\right.$$

$$\left.+\sum_{i\in V}\sum_{0\leq c<W}\lambda_{ic}(\sum_{(s,d)}\sum_{0<n\leq H_{sd}}\sum_{j\in V}\sum_{0\leq a<n_{ij}}\phi_{i,ca}^{sdn}-F_{ic})+\sum_{(i,j)}\pi_{ij}(\sum_{(s,d)}\sum_{0<n\leq H_{sd}}\sum_{0\leq c<n_{ij}}\delta_{ijc}^{sdn}-\gamma W)\right\},$$

subject to the constraints (5.2), (5.3), (5.5), (5.8) and (5.9), where ξ, λ, π are respectively the vectors of Lagrange multipliers $\{\xi_{ijc}\}$, $\{\lambda_{ic}\}$, $\{\pi_{ij}\}$.

After regrouping the relevant terms, the dual function leads to the following problem:

$$\min_{A,B,\Delta,\Phi,\gamma}\left\{\sum_{(s,d)}\left[P_{sd}\left(\sum_{0<n\leq H_{sd}}[1-\alpha_{sdn}]\right)+\sum_{0<n\leq H_{sd}}Q\beta_{sdn}+\sum_{(i,j)}\sum_{0<c\leq n_{ij}}\delta_{ijc}^{sdn}(\xi_{ijc}+\pi_{ij})\right.\right.$$

$$\left.\left.+\sum_{i\in V}\sum_{0\leq c<W}\sum_{j\in V}\sum_{0\leq a<n_{ij}}\lambda_{ic}\phi_{i,ca}^{sdn}\right]\right\}-\sum_{(i,j)}\sum_{0<c\leq n_{ij}}\xi_{ijc}-\sum_{i\in V}\sum_{0\leq c<W}\lambda_{ic}F_{jc}+\gamma(G-W\sum_{(i,j)}\pi_{ij})\}. \quad (5.10)$$

By using the fact that $\delta_{ijc}^{sdn}=\alpha_{sdn}\delta_{ijc}^{sdn}$, $\phi_{i,ca}^{sdn}=\alpha_{sdn}\phi_{i,ca}^{sdn}$ and $\beta_{sdn}=\alpha_{sdn}\beta_{sdn}$, we can rewrite equation (5.10) as:

$$\min_{A,B,\Delta,\Phi,\gamma}\left\{\sum_{(s,d)}\left[P_{sd}\left(\sum_{0<n\leq H_{sd}}[1-\alpha_{sdn}]\right)+\sum_{0<n\leq H_{sd}}\alpha_{sdn}(Q\beta_{sdn}+\sum_{(i,j)}\sum_{0<c\leq n_{ij}}(\xi_{ijc}+\pi_{ij})\delta_{ijc}^{sdn}+\sum_{i\in V}\sum_{0\leq c<W}\sum_{j\in V}\sum_{0\leq a<n_{ij}}\lambda_{ic}\phi_{i,ca}^{sdn})\right]\right.$$

$$\left.+\gamma(G-W\sum_{(i,j)}\pi_{ij})-\sum_{(i,j)}\sum_{0<c\leq n_{ij}}\xi_{ijc}-\sum_{i\in V}\sum_{0\leq c<W}\lambda_{ic}F_{jc}\right\}.$$

Since the last two terms are independent of the decision variables, the problem can be further simplified as:

$$\min_{A,B,\Delta,\Phi}\left\{\sum_{(s,d)}\left[P_{sd}\left(\sum_{0<n\leq H_{sd}}[1-\alpha_{sdn}]\right)+\sum_{0<n\leq H_{sd}}\alpha_{sdn}(Q\beta_{sdn}+\sum_{(i,j)}\sum_{0<c\leq n_{ij}}(\xi_{ijc}+\pi_{ij})\delta_{ijc}^{sdn}+\sum_{i\in V}\sum_{0\leq c<W}\sum_{j\in V}\sum_{0\leq a<n_{ij}}\lambda_{ic}\phi_{i,ca}^{sdn})\right]\right\}$$

$$+\min_{\gamma}\left\{\gamma(G-W\sum_{(i,j)}\pi_{ij})\right\}$$

$$= \sum_{(s,d)} \min_{\mathrm{A}_{sd}} \left\{ P_{sd} \left(\sum_{0<n \le H_{sd}} [1-\alpha_{sdn}] \right) + \sum_{0<n \le H_{sd}} \{ \alpha_{sdn} \cdot \min_{\beta_{sdn}, \Delta_{sdn}, \Phi_{sdn}} (Q\beta_{sdn} + \sum_{(i,j)} \sum_{0<c \le n_{ij}} (\xi_{ijc} + \pi_{ij}) \delta_{ijc}^{sdn} + \sum_{i \in \mathcal{V}'} \sum_{0 \le c < W} \sum_{j \in \mathcal{V}'} \sum_{0 \le a < n_{ij}} \lambda_{ic} \phi_{i,ca}^{sdn}) \} \right\}$$

$$+ \min_{\gamma} \left\{ \gamma (G - W \sum_{(i,j)} \pi_{ij}) \right\}, \tag{5.11}$$

which we shall refer to as **RP** (Relaxed Problem).

5.2.2 RWSS and CGSS

We can see that RP is composed of two minimization subproblem sets. The first subproblem set **RWSS** (RWA Subproblem Set) is

$$\sum_{(s,d)} \min_{\mathrm{A}_{sd}} \left\{ P_{sd} \left(\sum_{0<n \le H_{sd}} [1-\alpha_{sdn}] \right) + \sum_{0<n \le H_{sd}} \{ \alpha_{sdn} \cdot \min_{\beta_{sdn}, \Delta_{sdn}, \Phi_{sdn}} (Q\beta_{sdn} + \sum_{(i,j)} \sum_{0<c \le n_{ij}} (\xi_{ijc} + \pi_{ij}) \delta_{ijc}^{sdn} + \sum_{i \in \mathcal{V}'} \sum_{0 \le c < W} \sum_{j \in \mathcal{V}'} \sum_{0 \le a < n_{ij}} \lambda_{ic} \phi_{i,ca}^{sdn}) \} \right\}, \tag{5.12}$$

subject to the constraints (5.2), (5.3), (5.5), (5.8) and (5.9). RWSS can be decomposed into source-destination-level sub-problems (denoted as $\mathbf{SDS_{sd}}$), corresponding to every (s, d):

$$\mathrm{SDS_{sd}} = \min_{\mathrm{A}_{sd}} \left\{ P_{sd} \left(\sum_{0<n \le H_{sd}} [1-\alpha_{sdn}] \right) + \sum_{0<n \le H_{sd}} \alpha_{sdn} \cdot I_{\mathrm{sdn}} \right\}, \tag{5.13}$$

where I_{sdn} (,corresponding to every s_{sdn}) is defined as:

$$I_{\mathrm{sdn}} = \min_{\beta_{sdn}, \Delta_{sdn}, \Phi_{sdn}} \left\{ Q\beta_{sdn} + \sum_{(i,j)} \sum_{0<c \le n_{ij}} (\xi_{ijc} + \pi_{ij}) \delta_{ijc}^{sdn} + \sum_{i \in \mathcal{V}'} \sum_{0 \le c < W} \sum_{j \in \mathcal{V}'} \sum_{0 \le a < n_{ij}} \lambda_{ic} \phi_{i,ca}^{sdn} \right\}, \tag{5.14}$$

subject to the constraints (5.2), (5.3), (5.5), and (5.8). In RWSS, there are altogether Z semi-lightpath-level subproblems (I_{sdn}'s).

In the subproblem set **CGSS** (Congestion Subproblem Set), there is only one subproblem:

$$\min_{0 \le \gamma \le 1} \{ \gamma (G - W \sum_{(i,j)} \pi_{ij}) \} \tag{5.15}$$

5.3 The LRSM Framework

The overall LRSM framework is shown in Figure 5.2. We have modified and extended LRSM framework by solving two independent subproblem sets in the DP, and developed a heuristic algorithm to obtain a feasible solution based on the dual solution. We will provide these details while summarizing the essentials.

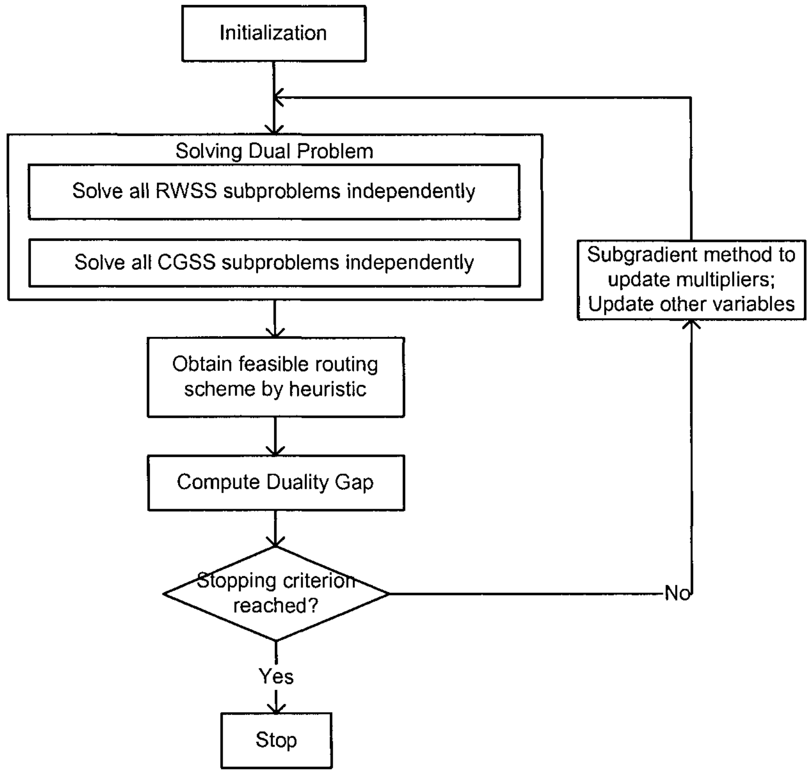


Figure 5.2. Schematic depiction of the overall algorithm

5.3.1 Solving RWSS and CGSS

The overall dual problem is optimized, when the two subproblem sets (i.e., RWSS and CGSS) are optimized separately. Each subproblem in RWSS corresponds to one connection demand. CGSS has only one subproblem, solving the congestion problem.

5.3.1.1 Solving RWSS

We first solve I_{sdn}'s in (5.14) without considering Constraint (5.9) or α_{sdn}. After we obtain the results for I_{sdn}'s for every (s, d), we launch a sorting process for every SDS_{sd} to comply with Constraint (5.9) and optimize on α_{sdn}.

In order to solve the subproblem introduced in Section 5.2.2, we can rewrite I_{sdn} in (5.14) as:

$$I_{sdn} \equiv \min_{\beta_{sdn}}\{ Q\beta_{sdn} + \min_{\Delta_{sdn},\Phi_{sdn}}\{D_{sdn}\} \}, \tag{5.16}$$

subject to the constraints (5.2), (5.3), (5.5), and (5.8), where $D_{sdn} = \sum_{(i,j)} \sum_{0<c\le n_{ij}} (\xi_{ijc} + \pi_{ij})\delta_{ijc}^{sdn}$

$+ \sum_{i\in\mathcal{V}} \sum_{0\le c<W} \sum_{j\in\mathcal{V}} \sum_{0\le a<n_{ij}} \lambda_{ic}\phi_{i,ca}^{sdn})\}$.

Some similar formulation to I_{sdn} has already been solved in Appendix A.1 by Modified Minimum Cost Semi-lightpath (MMCSLP) on Wavelength Graph (WG) [ChFa96]. We can thus adapt MMCSLP to solve I_{sdn}. The overall solution to RWSS can be then described as:

First solve all I_{sdn}'s:

A.1) Construct the wavelength graph of the network (see Appendix A.2).

A.2) Find a shortest path between the source and the destination columns using the MCSLP algorithm (see Appendix A.3) to get $\min_{\Delta_{sdn},\Phi_{sdn}}\{D_{sdn}\}$.

A.3) Check the dual cost $D'_{sdn} = \sum_{(i,j)}\sum_{0<c\le X'_{ij}}(\xi_{ijc}+\pi_{ij})\delta'^{sdn}_{ijc} + \sum_{i\in\mathcal{V}}\sum_{0\le c<W}\sum_{j\in\mathcal{V}}\sum_{0\le a<X'_{ij}}\lambda_{ic}\phi'^{sdn}_{i,ca})\}$ for the t'_{sdn}.

A.4) If $D'_{sdn}<Q+\min_{\Delta_{sdn},\Phi_{sdn}}\{D_{sdn}\}$, set $\Delta_{sdn}=\Delta'_{sdn}$ and $\beta_{sdn}=0$ (i.e., use the semi-lightpath of the existing connection). Set $I_{sdn}=\min_{\Delta_{sdn},\Phi_{sdn}}\{D_{sdn}\}$

A.5) Otherwise, set $\beta_{sdn}=1$ and $I_{sdn}=Q+\min_{\Delta_{sdn},\Phi_{sdn}}\{D_{sdn}\}$. The tie is broken arbitrarily (in our implementation, we choose $\beta_{sdn}=0$).

After the I_{sdn} value for every s_{sdn} is obtained, we do the following sorting process for every (s, d) to solve SDS_{sd}, to comply with Constraint (5.9) and optimize on α_{sdn} (we use an intermediate variable k in the following process, representing the number of demands rejected for (s, d)):

B.1) For (s, d), sort all s_{sdn}'s of (s, d) regarding the I_{sdn} value; initialize all α_{sdn}=0, k=0

B.2) If $N_{sd}\ge X'_{sd}$, continue with Step B.3).
Go to Step B.5) otherwise

B.3) Set the X'_{sd} s_{sdn}'s with the lowest I_{sdn} values accepted (set α_{sdn}=1)

B.4) If all s_{sdn}'s have been calculated, go to Step B.6).

- B.4.1.) otherwise, for the rest s_{sdn}'s, find the one with the highest I_{sdn} value (A tie is broken arbitrarily.)
- B.4.2.) if $I_{sdn} > P_{sdk}$, set k=k+1 and α_{sdn}=0.
Go to Step B.4)
- B.4.3.) otherwise set α_{sdn}=1.
Go to Step B.4)

B.5) If $N_{sd}<X'_{sd}$,

- B.5.1) set the N_{sd} s_{sdn}'s with the lowest I_{sdn} values accepted (set α_{sdn}=1)
- B.5.2) set other s_{sdn}'s rejected (set α_{sdn}=0)

B.6) If not all source-destination pairs are calculated, go to B.1) for the next (s, d); otherwise finish the DP minimization.

Through the above procedure, we can solve all SDS_{sd}'s, and RWSS can thus be optimized.

5.3.1.2 Solving CGSS

To solve CGSS, γ is set to 0, when $G-W\Sigma_{(i,j)}\pi_{ij} < 0$, and γ is set to 1, when $G-W\Sigma_{(i,j)}\pi_{ij} > 0$. If $G-W\Sigma_{(i,j)}\pi_{ij} = 0$, γ is set to 0 or 1 arbitrarily.

5.3.2 Updating Lagrange Multipliers

Since there are integer variables involved in the formulation, we now employ the subgradient method [Be99] to solve the DP maximization.

The multiplier vector $z = (\xi, \lambda, \pi)$ are first updated by the following formula:

$$z^{(h+1)} = z^{(h)} + \alpha^{(h)} g(z^{(h)}),$$

where $z^{(h)}$ denotes the value of vectors z obtained at the hth iteration, and $\alpha^{(h)}$ denote the step size at the hth iteration. The vector $g(z)$ is the subgradient of q with respect to z, i.e. $g(z) = \{g(\xi), g(\lambda), g(\pi)\}$. The vectors $g(\xi)$, $g(\lambda)$, and $g(\pi)$ are composed of $g_{ijc}(\xi)$, $g_{ic}(\lambda)$, and $g_{ij}(\pi)$ respectively, where

$$g_{ijc}(\xi) = \sum_{(s,d)} \sum_{0<n\le N_{sd}} \delta_{jic}^{sdn} - 1,$$

$$g_{ic}(\lambda) = \sum_{(s,d)} \sum_{0<n\le N_{sd}} \sum_{j\in V} \sum_{0\le a<n_{ij}} \phi_{i,ca}^{sdn} - F_{ic},$$

$$g_{ij}(\pi) = \sum_{(s,d)} \sum_{0<n\le N_{sd}} \sum_{0\le c<n_{ij}} \delta_{ijc}^{sdn} - W\gamma .$$

The step size is determined by

$$\alpha^{(h)} = \mu \times \frac{q^U - q^{(h)}}{g^T(z^{(h)}) g(z^{(h)})}, \tag{5.17}$$

where q^U is an estimate of the optimal solution, and $q^{(h)}$ is the value of q at the h^{th} iteration. Generally, the best value of the objective function J of the feasible routings obtained is used to be q^U. The parameters μ and q^U are changed adaptively as the algorithm converges.

5.3.3 Constructing a Feasible RWA Scheme

Since some of the constraints are relaxed by the Lagrange multipliers, the solution to DP might have an infeasible routing scheme. In other words, some wavelength channel capacity constraints in (5.4), converter capacity constraints in (5.6), and link congestion constraints in (5.7) might have been violated at some links and nodes. On the other hand, other constraints will not be violated because this is guaranteed by the formulation and the solution to RWSS in (5.12) and the solution to CGSS in (5.15).

To construct a feasible routing scheme, a heuristic algorithm has to be employed to decide which connection demands should be rejected or rerouted and which paths the rerouted connection demands should take. Although more complicated, the heuristic algorithm proposed in this chapter has the similar structure as the heuristic algorithm in Section 3.2.2.4, which can be described by a flowchart shown in Figure 5.3.

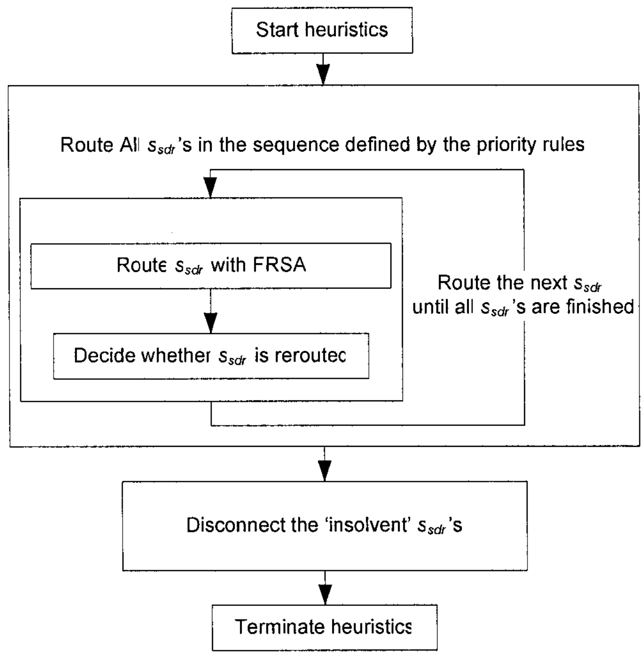


Figure 5.3. The flowchart of the heuristic algorithm to obtain a feasible solution

For the collided connection demands, the following *priority rules* are used to determine which connection demands should occupy the 'critical resources'. The most prioritized semi-lightpath will be deployed first to use the critical resources. Given two collided connection demands $c1$ and $c2$, assuming the corresponding source destination pairs are ($s1$, $d1$), ($s2$, $d2$) (For the convenience of the notation, if $c1 < c2$, it means $c1$ has a higher priority than $c2$, and vice versa.),

1. If ($s1$, $d1$) is different from ($s2$, $d2$), then use *rule 2*. Otherwise, if $c1$ has lower dual cost than $c2$, then $c1 < c2$; and vice versa. The tie is broken arbitrarily.
2. If ($s1$, $d1$) has more 'must-accept' demands (i.e., $\min(N_{sd}, X'_{sd}) - \sum_{0<n\le N_{sd}} \alpha_{sdn}$) to satisfy constraints (5.9) than ($s2$, $d2$), then $c1 < c2$; and vice versa. If $c1 = c2$, then use *rule 3*.
3. If ($s1$, $d1$) has the same or more connections accepted than the demands (i.e., $\sum_{0<n\le N_{sd}} \alpha_{sdn} \ge N_{sd}$) while ($s2$, $d2$) does not, then $c1 > c2$; and vice versa. If $c1 = c2$, then use *rule 4*.
4. If $c1$ is associated with an existing connection while $c2$ is not, then $c1 < c2$; and vice versa. If $c1 = c2$, then use *rule 5*.
5. If $c1$ is accepted in the dual solution while $c2$ is not, then $c1 < c2$; and vice versa. The tie is broken arbitrarily.

The similar procedure as the heuristic algorithm in Section 3.2.2.4 is launched to find a feasible route for every s_{sdn} based on the dual solution, which is denoted as *Feasible Route Searching Algorithm* (FRSA). The algorithm is implemented on the wavelength graph (WG) of the network, which is formed by separating every node into W vertices representing the wavelengths available on the node, and connecting these vertices with arcs representing the wavelength channels and the wavelength converters. The details of the algorithm to form the WG are provided in Appendix A.2. The values of the arcs in the WG are assigned in the following way: the weight (WE_{ijc}) of every arc corresponding to WC w_{ijc} is estimated by $WE_{ijc}=\sigma\cdot(G\cdot M_{ij})/W+\rho\cdot(P\cdot m_{ijc})$, where M_{ij} and m_{ijc} are the total number of connections routed through e_{ij} and w_{ijc}, respectively; σ and ρ are the estimate coefficient for the congestion and the rejection respectively. We use σ=0.01 and ρ=0.005 in the current implementation.

The Feasible Route Searching Algorithm (FRSA):

Step 1 (*Check the Solution of the Dual Problem*):
Check if the Wavelength Assignment (WA) derived from the dual solution is *viable*, i.e., no wavelength channel in the WA is occupied by some other connection demands while the wavelength converters used in the WA are not used up.

1.1) If the WA from the dual solution is viable, the WA is assigned to this semi-lightpath. Then go on to the next connection demand in the priority list, and repeat *Step 1*.
1.2) If the WA from the dual solution is not viable, go to *Step 2*.

Step 2 (*Find the WA in the Same Route*):
Find a viable WA for the connection demand in the same route derived from the dual solution, but with a different wavelength assignment. The procedure of this step is the same as MMCSLP (Appendix A.2). However, the WG is only made of the fibers and nodes traversed by the connection demand in the dual solution.

2.1) If a viable WA can be found, this connection demand shall monopolize this WA. Then go on to the next connection demand demand, and repeat *Step 1*.
2.2) If no viable WA can be found, go to *Step 3*.

Step 3 (*Find the WA for One Connection Demand with the MEWA*):
Apply the Modified Esau-Williams algorithm (MEWA) to find a feasible RWA for the connection demand.

3.1) If a viable WA can be found, this connection demand shall monopolize all the wavelength channels and converters assigned by this WA. Then go on to the next connection demand, and repeat *Step 1*.

3.2) If no viable WA can be found, reject the connection demand. Then go on to the next connection demand, and repeat *Step 1*.

After we obtain the RWA for s_{sdn}, if s_{sdn} is associated with an existing connection (t'_{sdn}), we have to decide whether s_{sdn} should be rerouted (or use the RWA of the existing semi-lightpath), by using the following estimation procedure:

1) If there is a feasible route found in FRSA, we define the estimated cost of s_{sdn} as $C_{sdn} = \sum_{(i,j)} \sum_{0<c\le X'_{ij}} WE_{ijc}\delta_{ijc}^{sdn}$; if there is no feasible route found, $C_{sdn}=\infty$.
2) We calculate the estimate cost of the existing route $C'_{sdn} = \sum_{(i,j)} \sum_{0<c\le X'_{ij}} \delta'^{sdn}_{ijc} WE_{ijc}$. The 'gain' for rerouting s_{sdn} is thus defined as $G_{sdn} = C'_{sdn} - C_{sdn}$.
3) If $\varsigma \cdot G_{sdn} > Q$, then use the newly found route ($\beta_{sdn} = 1$); otherwise use the existing route ($\beta_{sdn} = 0$), where ς is the estimated coefficient and we use $Q/6$ in the current implementation.

In the last stage of the heuristic algorithm, we disconnect the 'insolvent' s_{sdn}'s. To calculate the insolvency of every s_{sdn}, we use the following Insolvency Disconnection Procedure:

Step.1) If $\gamma = 0$, stop the heuristic algorithm. Otherwise, calculate the number of the most congested links MC, which is the number of links having the same congestion γ. The 'congestion cost' of going through every most-congested link is calculated as $CC=G/(W\cdot MC)$. The 'congestion cost' of every s_{sdn} is defined as $CC_{sdn}=NM_{sdn}\cdot CC$, where NM_{sdn} is the number of most congested links s_{sdn} goes through. The 'revenue' of every s_{sdn} (denoted as R_{sdn}) is equal to its rejection penalty P_{sdk}. (See the definition of P_{sdk} in Section 5.1.2)

Step.2) For each of the MC most congested links e_{ij}, find one 'insolvant' ($R_{sdn} < CC_{sdn}$) s_{sdn} that goes through e_{ij}, and store it into a list L.

Step.3) If each of the MC most congested links has one 'insolvant' s_{sdn}, then disconnect all the s_{sdn}'s stored in list L. Go to *Step.1*; otherwise terminate the heuristic algorithm.

5.3.4 Computation Complexity

The dual solution is minimized when both RWSS and CGSS are minimized (See Section 5.3.1). The complexity of solving each I_{sdn} is the same as the complexity of MMCSLP (Appendix A.1), i.e., $O((N+W)NW)$. By grouping the I_{sdn}'s with the same source node, we can solve all I_{sdn}'s with $O((N+W)N^2W)$. The worst-case time complexity for the sorting operation is $\Sigma_{(s,d)}O(H_{sd}\log(H_{sd}))$ [Bert99], and thus the overall complexity to solve DP is $O((N+W)N^2W) + \Sigma_{(s,d)}O(H_{sd}\log(H_{sd}))$.

The worst case complexity for the Insolvency Disconnection Procedure is $O(EW\log(E) + ZW^2)$, where E is the number of links, and Z is the number of s_{sdn}'s. The complexity of the rest part of the heuristic algorithm is $O(Z(Z\log Z+(NW)^2))$, which has been analyzed in Section 3.2. Therefore the overall complexity of the heuristic algorithm is $O(Z^2\log Z+Z\,(NW)^2+EW\log(E))$.

5.4 Performance Evaluation

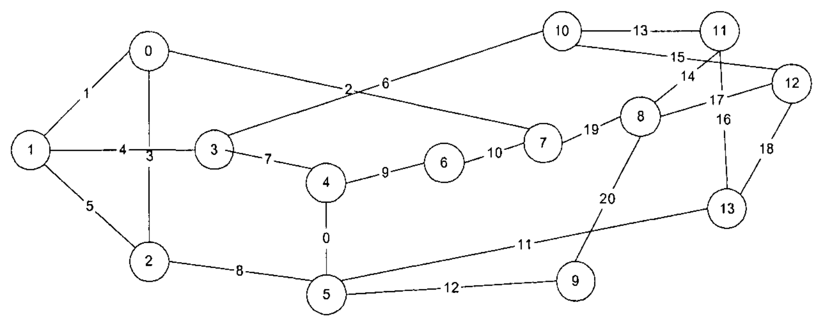


Figure 5.4. NSFNET with 14 Nodes and 21 Edges

To test the performance of our algorithm, we use the 14-node NSFNET topology as shown in Figure 5.5 to test our proposed algorithm. The traffic demands are shown in Figure 5.6, where the horizontal and vertical indexes are the source and destination nodes, respectively. Specifically, the number on the i^{th} row and the j^{th} column represents the number of traffic demands from node i to j (N_{ij}), which takes random values between 0 and 3. We also use the traffic matrix shown in Figure 5.6 and use the same optimization procedure as in Section 3.2 for 100 iterations to generate the existing connections from the previous session. Most of the results from our algorithm in the NSFNET example can be obtained within 5 minutes running on our testing computer (see Section 1.5). The readers are referred to Section 3.2.4 for the computation time testing, the large network application and the computation time reduction by reusing Lagrange multiplier values.

0	1	3	1	1	1	3	0	2	0	1	2	0	3
0	0	0	2	2	2	1	1	1	2	1	0	1	3
3	2	0	3	0	1	2	3	1	3	1	2	2	0
3	1	0	0	1	1	2	3	2	2	1	2	1	3
1	3	0	2	0	1	0	2	0	3	0	1	1	3
1	2	1	3	2	0	1	3	3	1	0	1	1	2
2	2	3	1	3	3	0	0	3	1	2	0	3	3
3	1	2	3	1	0	1	0	0	3	2	0	3	0
3	0	1	3	3	3	1	0	0	2	1	1	1	0
0	0	0	1	2	0	2	0	1	0	1	0	0	3
1	0	0	2	0	3	0	1	0	3	0	3	1	3
2	3	1	1	3	2	3	2	2	2	2	0	1	3
2	0	1	2	0	1	2	0	3	0	2	1	0	3
1	1	0	2	1	0	1	3	0	1	2	1	3	0

Figure 5.5. Traffic Demand Matrix for NSFNET

0	1	0	3	1	2	1	0	2	0	0	2	1	0
1	0	0	2	0	3	1	2	1	3	0	0	0	3
1	2	0	3	0	1	2	3	1	3	1	2	2	0
3	2	0	0	1	2	0	1	1	2	0	2	2	3
1	3	0	0	0	0	0	3	0	2	0	1	2	3
1	2	1	2	2	0	1	1	3	1	0	1	1	2
2	2	3	2	0	3	0	0	3	1	2	0	0	0
3	1	2	1	1	0	1	0	0	3	2	0	3	0
3	0	1	2	3	0	1	0	0	2	1	1	1	0
1	0	1	1	3	0	2	0	1	0	2	0	1	1
1	0	0	2	0	2	0	2	0	2	0	3	2	2
1	0	3	3	0	1	0	3	2	0	1	0	1	3
3	0	2	1	0	0	3	0	1	0	0	3	0	1
2	0	3	1	2	0	1	1	0	0	1	0	3	0

Figure 5.6. Simulated Traffic Demand Matrix for NSFNET at Previous Session

5.4.1 Effect of Different Factors in the Semi-dynamic Reconfiguration

To simplify our study (not required), we set the rejection penalty of the last connection demand for every (s, d) to the same value, i.e., $P_{sdk}=P$, if $k=N_{sd}$ (see Section 5.1.2). We first use number of wavelengths W=11, rejection penalty P=100, rejection penalty stepsize S=2, and congestion penalty G=100 to study the network behaviors under heavy traffic.

The new connection demands and the existing connections are correlated through Constraint (5.8) and (5.9), as well as the reconfiguration penalty Q. Figure 5.7 shows the trade-off between the number of rejected connection demands and the number of reconfigured connections, as Q changes with number of wavelength converters F_{ic}=0 (the parameters are shown on the top of the graph). We can see that as Q increases (before Q reaches 400), more connection demands are rejected, and fewer connections are reconfigured. After Q reaches 400, both values stopped changing, due to the constraints that have to be satisfied. Note that when Q=0, the existing connections can be freely rerouted, which is the same scenario as the static RWA problems studied in Chapter Three.

Figure 5.8 shows the influence of Q for different F_{ic} (the number of converters) values. Note that when Q=0, the situation is the same as the problems studied in Chapter 3 and [SwSi02], which means the existing connection can be freely rerouted with no penalty. We can see in Figure 5.8 that the contribution of the wavelength conversion around Q=0 is marginal (i.e., the curves of F_{ic}=0 and F_{ic}=1 are not far apart), which conforms to the results of Chapter 3 and [SwSi02]. However, as Q increases, the contribution of the wavelength conversion is gradually highlighted. All curves increase with Q, until the limitation of other hard constraints is reached (around Q=400). Increasing the conversion capacity (from F_{ic}=1 to F_{ic}=2) or increasing the conversion degree v (see Figure 5.9) does not have further impact. (Note that v=1 means no wavelength conversion capability.) We can thus come to a conclusion that the real value of the wavelength conversion lies in less network reconfigurations. For the wavelength conversion architecture we use in this chapter, more wavelength converters (F_{ic}>1) or higher conversion degree (v>2) does not generate better results than the lowest configuration (F_{ic}=1, v=2).

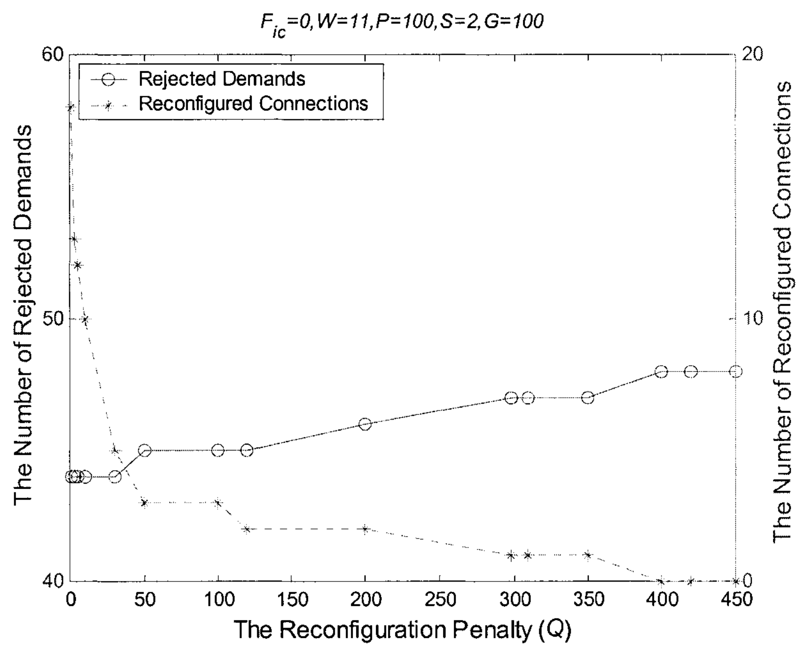


Figure 5.7. The number of rejected connection demands and the number of reconfigured connections

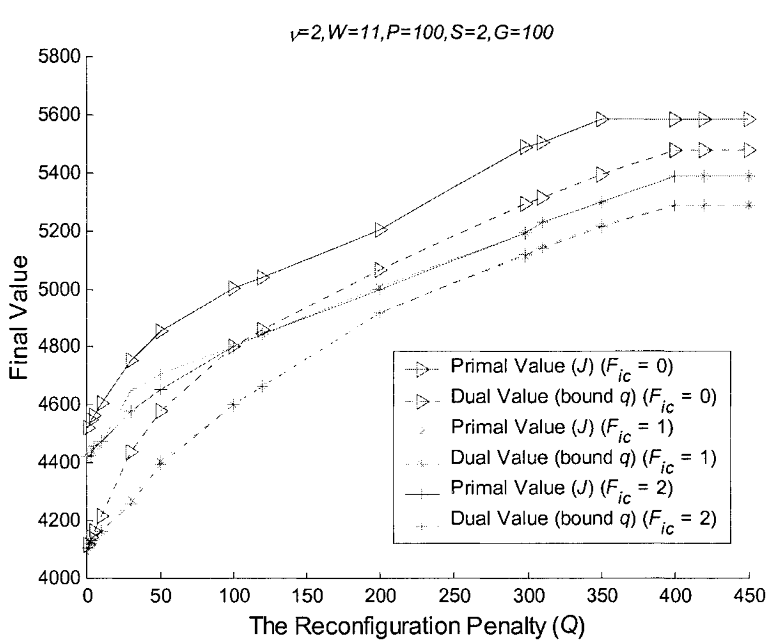


Figure 5.8. The comparison of different number of wavelength converters

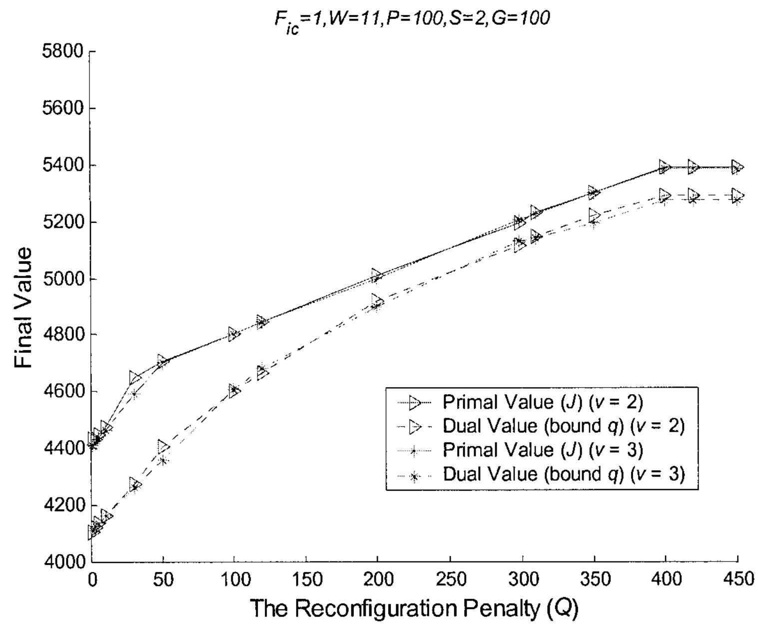


Figure 5.9. Influence of different conversion degree (v)

For the lightly-loaded network, we want to minimize the congestion. We can simply use the same traffic matrix and increase the number of wavelengths (W) on the links to simulate the lightly-loaded network. Figure 5.10 and Figure 5.11 show the results obtained for W=20. We can see from Figure 5.11 that our algorithm generates very good and consistent near-optimum results. The results are mostly within 3% of the optima.

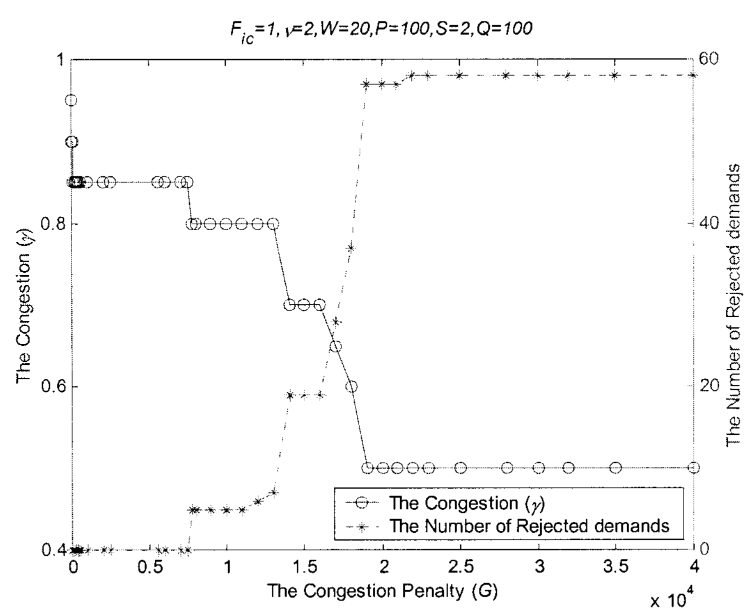


Figure 5.10. The congestion (γ) and the number of rejected connection demands

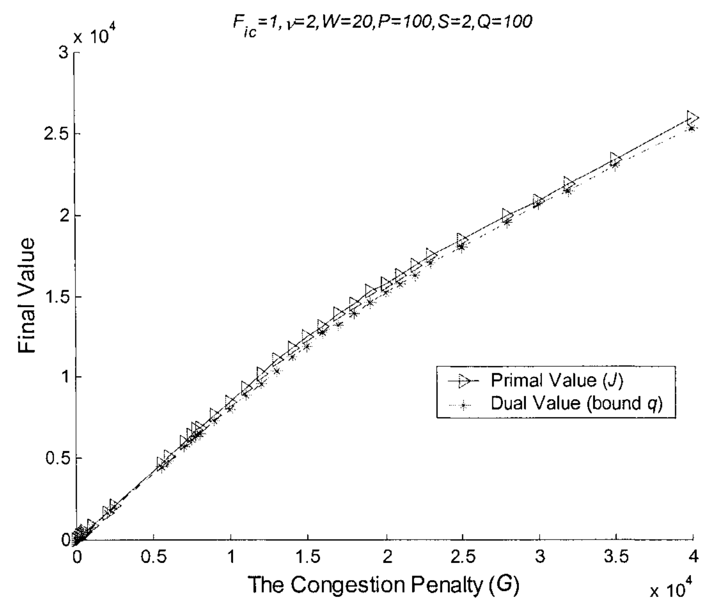


Figure 5.11. The congestion penalty (*G*) v.s. the final results of *q* and *J*

Figure 5.10 shows the trade-off between the congestion (γ) and the number of rejected connection demands. We can see that the congestion goes down and the number of rejected connection demands goes up as congestion penalty G increases. We can see the values vary in stages, because optimization algorithm tends to distribute the traffic, and there are generally several congested links (of congestion ratio γ). So every time when the congestion goes down, there are connections on several links being disconnected at the same time. We can see in Figure 5.11 that J and q increase linearly with respect to G. Around G=18,000, the slope of the curves changes, this is because before G=18,000, J and q increase because of both the increase of the rejections and the increase of G. After this point, no more connection can be rejected (see Figure 5.10) because the non-disconnection constraints (5.9). Therefore only the increase of G contributes to the increase of final values, and thus the slope of the curves become less steep.

5.4.2 Fairness Study

Figure 5.12 to Figure 5.14 demonstrate the results of the fairness study for W=20, P=100, G=0, and Q=100. We assign the penalty coefficient G to zero here in order to minimize the influence from the congestion penalty. Similar to Chapter 3, when S (rejection penalty stepsize) increases, the algorithm tends to make the rejection fairer (e.g., less totally disconnected pairs D), at the cost of more rejections. Figure 5.13 shows J and q decrease linearly as S increases. This is mostly because when S increases, the overall rejection penalty decreases linearly at the same time. Figure 5.14 again demonstrated that our algorithm is highly stable in obtaining the near-optimal results.

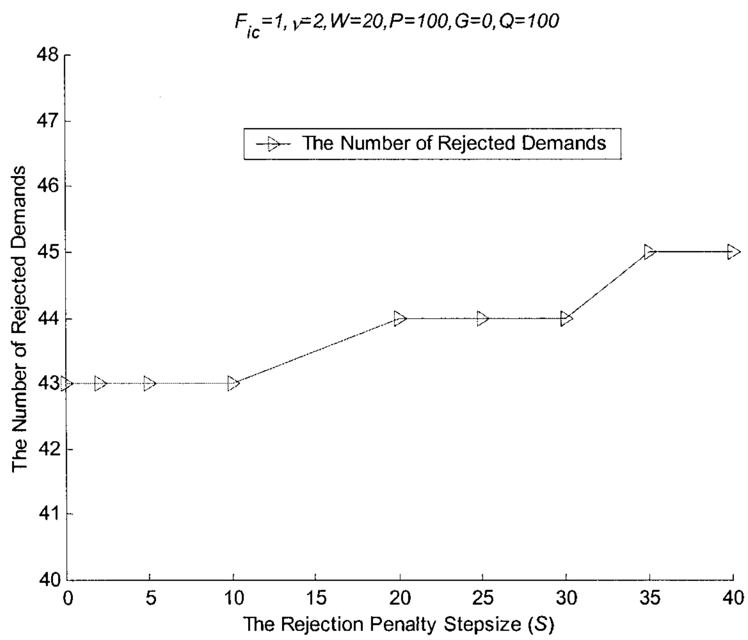


Figure 5.12. Rejection penalty stepsize (S) v.s. number of rejected connection demands

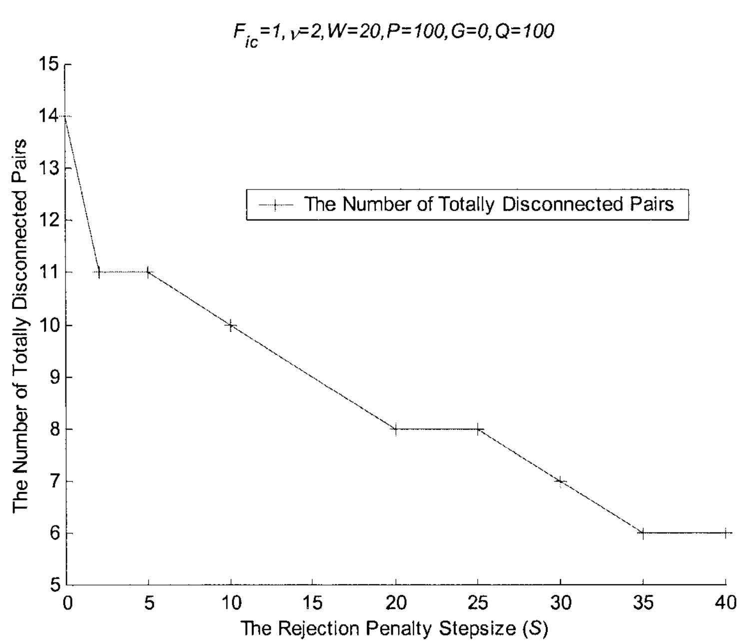


Figure 5.13. Rejection penalty stepsize(S) v.s. number of totally disconnected pairs (D)

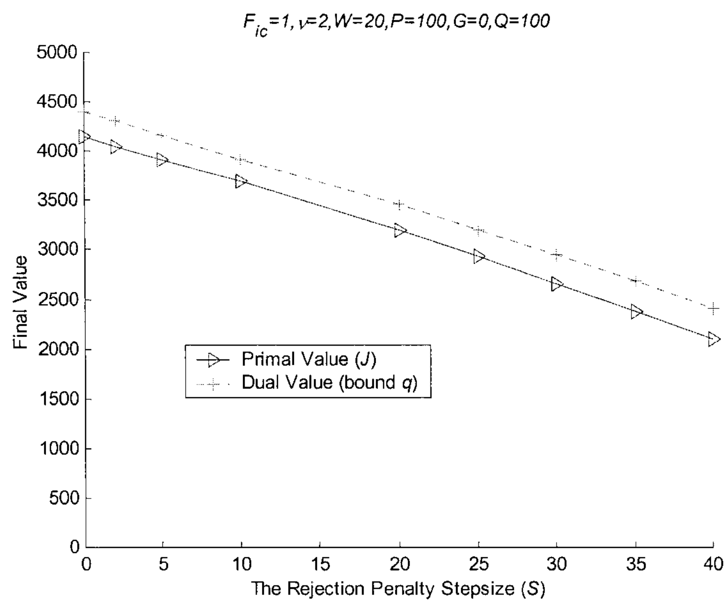


Figure 5.14. Rejection penalty stepsize (S) v.s. final results of q and J

5.5 Concluding Remarks

In this chapter, we are successfully to formulate and solve the semi-dynamic network reconfiguration problem, which has not been systematically and mathematically solved previously due to formulation difficulties and the lack of mathematical tools. To relate the current session and the previous session in the optimization, we exerted extensive effort to formulate the problem in a way that it can be solved mathematically. We proposed a penalty-based objective function, i.e., penalizing the reconfiguration of a connection, the rejection of a connection demand and the most congested link. We employed the LRSM to solve the complicated formulation with polynomial computation complexity. We have discovered many novel network behaviors in the NSFNET computation example, including the real contribution of the wavelength conversion, the trade-offs between the rejection and the reconfiguration/congestion, and the influence of various parameters on the final optimization results. These behaviors have important bearings for the successful operation of a semi-dynamic network. At the same time, the correctness of our formulation and the high efficiency of our solution methodology have been again demonstrated.

Chapter Six

Design Guideline and Recommendations

In the previous chapters, we have studied our optimization framework and presented its applications on several optical network optimization cases. Based on our own computation experiences, we will provide in this chapter guidelines to use our optimization software package and to further apply it to other network optimization problems.

6.1 Range of Applicability

All the optimization algorithms in this thesis assume the network resource configuration is already given. Therefore they cannot be used directly to optimize the network resource planning. The test sample points for different network configurations (note that the optimization algorithms in this thesis deal with only one single point for a given configuration) should be obtained to form the performance curve with respect to different configurations, and then the optimal network resource configuration can be obtained.

All connection demands are static/semi-dynamic in the networks we studied. Since the computation is centralized, the algorithm cannot be used for distributed traffic routing without coordination. The traffic from the users will be assembled at the access nodes, and the bandwidth demand between the source-destination node pairs will be generated from these traffic or traffic prediction. This topic is not covered in this thesis.

The polynomial complexities of our algorithms allow us to have a fairly big applicable network size. With our hardware configuration (see Section 1.5 for the details), the networks we have tried in Chapter Three can go up to 35 nodes, and in Chapter Four, our examples can go up to 32 wavelengths in the 14-node network. The algorithm only slows down as the number of nodes and wavelengths go up (e.g., wavelengths in Figure 3.8). The limitation depends on the available computer speed and the computation time allowed.

6.2 Selection of Rejection Penalties

We have provided in this thesis (introduced in Chapter Three) a simple penalty assignment scheme to evenly distribute the rejection/acceptance of the traffic demands among all the source-destination pairs. However, the network operators can use any other penalty assignment scheme to address the specific network operation purpose. For example, if the traffic demand between two nodes (say (i, j)) is very important and must be accepted in the RWA scheduling, then the rejection penalty for this specific traffic demand can be set to a very high value (e.g., 100 time higher than the normal traffic demands), so that the optimization algorithm will choose to accept this demand to minimize the overall rejection penalty.

The Stepsize (S) defined in Section 3.2 is a very important parameter in our penalty assignment scheme. The trade-off between the number of rejected commands and the fairness of the scheduling can be achieved by adjusting the value of S. However, to set up the proper value for S to serve the network operation purpose, some computational experience is needed. We just provide two scenarios here as examples in order to provide a better explanation:

Scenario 1: In a network application where no fairness is considered, i.e., all connection demands are considered equally and the only objective is the number of connections that can be established, all the rejection penalty coefficients should be set to the same value (e.g., 100).

Scenario 2: In a network application where no source destination pair ((s, d)) with connection demands can be disconnected, i.e. at least one connection demand between (s, d) has to be accepted if there are connection demands between (s, d), the rejection penalties for the first connection demand can be set to a high value (e.g., P_{sd1}=100) while the rejection penalties for the rest of the connection demand can be set to a much lower value (e.g., $P_{sdN} = 0.1\ P_{sd1}$).

6.3 Stopping Criteria

One important implementation issue for our algorithm is the decision on the stopping criteria used for the termination of our algorithm. It can come in several forms as listed below:

1. Manual: rely on human decision to terminate the algorithm.
2. Duality Gap Threshold: when the duality gap is less than a certain value (Note that the value setting is problem-dependant and hence experience-dependant, because the intrinsic duality gaps of the problems differentiate largely and the optimization goal is very subjective).
3. Running Time Threshold: when the running time is more than a certain threshold.

4. Running Iterations Threshold: when the limit on the number of iterations is reached. Generally the optimizer can have a test run of the algorithm and estimate the rough number of iterations till the optimization results do not improve.
5. No Change In Result: this is the time duration when the optimization result does not improve.
6. No Change In Duality Gap: for a period of time or for a number of iterations when the duality gap does not decrease.
7. Hybrid: the combination of two or several stopping criteria mentioned above.

The selection of the stopping criteria is very application-dependent. We use two scenarios as examples to have a better explanation:

Scenario 1: In a network application where the reconfiguration is done very frequently, it requires the algorithm to generate a usable result within a very short time (e.g., 2 seconds), but because the obtained network configuration only lasts for a very short time, the optimality of the results does not have as much impact. We can thus use the Running Time Threshold criterion to terminate the algorithm when a usable result is obtained (might not be much optimized) to facilitate the fast reconfiguration.

Scenario 2: In a network application with rare reconfiguration, there is generally abundant time for the algorithm to run, but very good optimized results are important for the network operation, because the network configuration lasts for a long time and the optimality has a higher impact. Thus Duality Gap Threshold, No Change In Result, or No Change In Duality Gap criteria should be used to ensure the optimality of the results. Generally the final duality gap within 10% is considered to be of fairly good results.

Generally speaking, more optimized result will be obtained if the optimization algorithm is run for a longer time, but the algorithm can be stopped at any time to obtain a sub-optimal solution which can be better or worse. A working configuration scheme can be sent out before the session starts, as long as the algorithm can be run for at least one iteration.

6.4 Multiplier Reuse

One major advantage of our framework in the implementation is the reusability of the multiplier values. This attribute implies much faster convergence in the computation (see Section 3.2.4). In most of the implementations, the backbone network configuration/traffic demand does not change drastically in two continuous configuration sessions, which facilitates the reusing of the old results from the previous session.

Specifically, the computation result for every session can be recorded and the values of the Lagrange multipliers (each one of them corresponds to exact one network resource) are saved. In the next session when the optimization algorithm starts, the Lagrange multipliers are initialized to the same recorded values from the previous session. In our computation experiments, this can save 30% - 50% of the computation time for the algorithm to converge. Generally, in the first computation session, the multipliers are initialized to zero, and for the following sessions, the multipliers can simply be initialized to the values saved from the previous session.

6.5 Recommendations for Employing LRSM Framework

The application of LRSM framework in network optimization models is generally a very complicated mathematical process, which includes problem formulation, problem relaxation, dual problem solution, heuristic algorithm construction, and software implementation. The problem formulation, problem relaxation and dual problem solution are the most important three parts that have to be considered together in a big picture. The formulation has to facilitate the relaxation and the dual problem generated from the relaxation has to be optimally solvable. The solution generally has to be of polynomial time complexity (otherwise it might defeat the important benefit of LRSM framework). The core skill here is to select proper constraints to relax in order to make the dual problem obtained decomposable. After that, the solution to the decomposed subproblems is generally much easier than solving the original primal problem, thus the dual problem is solved as long as all the subproblems are solved. The performance our framework is sometimes problem-dependant, which means that the algorithm is very efficient for some formulations but it may not work very well for some others. It takes a lot of experience to envisage the applicability of our framework.

In our computation experience, we found that the capacity constraints in the form of ($\sum_i x_i \leq Y$) are the best to relax, and generally speaking, the less the constraints are relaxed, the closer the lower bound is to the primal value, thus if we have the choice not to relax some of the constraints, it is better to keep it in the original form in the dual problem to minimize the duality gap. The relaxation of the capacity constraints is usually to disconnect the correlation between the network entities (such as semi-lightpaths and traffic demands) competing for the same network resources. After relaxation, the correlation confined by these capacity constraints can be replaced by the Lagrange multipliers, and the subproblems corresponding to the network entities can then be solved separately. The SPA (Shortest Path Algorithm) and its variations have significant roles in solving the subproblems of the dual problem. The SPA should be carefully adapted to solve the special formulation of the subproblems.

Although it is not encountered in this research work, it is possible that the subgradient method can take many iterations to converge (therefore a very long time) in some extreme cases. The readers may consult the alternative solutions listed/suggested in [Bert99].

Chapter Seven
Conclusion

In this thesis, we have provided an efficient optimization framework called the LRSM, using the new link-based integer linear programming (ILP) formulations. We adopted a new formulation system dealing with the fairness consideration in accepting new traffic demands. This has been reflected in our formulations of pure static RWA problems, traffic grooming optimization problems and semi-dynamic reconfiguration optimization problems. Our formulation also addressed many other various aspects of the practical network operation and design concerns, including the network congestion, the limited wavelength conversion, the network operation cost and the network revenue. Our solution framework was proved to be highly efficient and well adaptable when studying optimization problems for various WDM mesh networks with different assumptions and different design objectives. We were also able to provide tight bounds at the same time of obtaining good near-optimal solutions, with reasonable time complexity. Compared with many other methodologies proposed in some previous research, our framework demonstrates higher efficiency in terms of both computation results and time complexity.

We have also studied the influences of various network resources and network design parameters that have not been studied before. In the grooming network optimization problem, we proposed a new objective function maximizing the network 'profit', which answered the question of whether adopting a new traffic is profitable.

In our study on the semi-dynamic reconfiguration problem, we have been able to associate the existing connections with the new traffic demands. We have also analysed the influence of the wavelength conversion and discovered the real value of using wavelength converters, which has only been suspected by some researchers before. Moreover, we have successfully formulated and studied the trade-off between the semi-dynamic reconfiguration and rejection/congestion/ fairness.

Nonetheless, we have also encountered some problems in our computation experiments, i.e., the convergence of the subgradient method is sometimes unsatisfactory, and some further adaptation of the subgradient method might be helpful to this problem.

7.1 Future work

One further research that can be done following this thesis is the improvement on the subgradient method, i.e., a faster convergence speed in some stiff cases, especially when close to the top of the solution space. One research direction might be the "surrogate-subgradient method" [ZhLu93, HiLe93, ZhLu99], which reduces the complexity of computing the subgradient in every iteration.

We have applied the LRSM in the solution of the network optimization problems, which has demonstrated excellent efficiency and adaptability to different network design objectives and constraints, such as considering different types of costs. However, the research applying this framework has only been started in our research here, and there will be a lot of future work to improve the algorithms and to apply this framework to more network applications.

References

[AlAy99] M. Alanyali, E. Ayanoglu, "Provisioning Algorithms for WDM Optical Networks," *IEEE/ACM Trans. On Networking*, Vol. 7, No. 5, Oct. 1999

[AlEl03] M. Ali and D. Elie-Dit-Cosaque, "Routing of 40Gb/s Traffic in Heterogeneous Optical Networks", *Proc. of 2003 IEEE International Conference on Communications* (ICC '03), May 12-14, 2003, Anchorage, Alaska, pp. Page(s): 1391-1396 vol.2

[BaBa96] S. Baroni and P. Bayvel. "Wavelength requirements in arbitary connected wavelength-routed optical networks", *Proc. IEEE Globecom.*, London, November 1996

[BaCh96] S. Banerjee and C. Chen, "Design of Wavelength-Routed Optical Networks for Circuit Switched Traffic", *Proc. Globecom 1996*, London, England, November 20 and 21, 1996

[BaMu96] D. Banerjee and B. Mukherjee, "A Practical Approach for Routing and Wavelength Assignment in Large Wavelength-Routed Optical Networks," *IEEE JSAC*, vol. 14, no. 5, pp. 903-908, June 1996.

[BaMu00] D. Banerjee and B. Mukherjee, "Wavelength-routed optical network: linear formulation, resource budgeting tradeoffs and a reconfiguration study", *IEEE/ACM Trans. on Networking*, pp. 598-607, Oct. 2000

[BaRo01] I. Baldine and G.N. Rouskas, "Traffic Adaptive WDM Networks: A Study of Reconfiguration Issues", *IEEE/OSA Journal of Lightwave Technology*, Vol.19, No.4, April 2001, pp.433-455

[BeCo03] J. Bermond and D. Coudert, "Traffic Grooming in Unidirectional WDM Ring Networks Using Design Theory," *Proc. of 2003 IEEE International Conference on Communications* (ICC '03), May 12-14, 2003, Anchorage, Alaska, pp. Page(s): 1402-1406 vol.2

[BoLa05] E. Bouillet, J.F. Labourdette, Ramamurthy, R. and Chaudhuri S., "Lightpath re-optimization in mesh optical networks", *IEEE/ACM Transactions on Networking*, Vol.13, No.2, April 2005, pp.437-447

[Bert99] D. P. Bertsekas, *Nonlinear Programming,* Athena Scientific, Athena Scientific, 1999

[BiWa03m] A.R.B. Billah, B. Wang and A.A.S. Awwal, "Multicast Traffic Grooming in WDM Optical Mesh Networks," *Proc. of IEEE 2003 Global Communications Conference* (Globecom'03), Dec. 1-5, 2003, San Francisco, CA, pp.2755-2760

[BiWa03e] A.R.B. Billah, B. Wang and A.A.S. Awwal, "Effective Traffic Grooming Algorithms in SONET/WDM Ring Networks", *Photonic Network Communications*, Vol.6, No.2, pp.119-138, Sept. 2003

[Boui98] E. Bouillet, "Monte Carlo techniques for design of wavelgnth-routed all-optical networks with blocking", *Internal Report, New York: Center for Telecommunications Research, Columbia University*, 1998

[BrPo98] "Broadband optical access systems based on passive optical networks (PON)", *ITU-T Recommendation* G.983.1, October 1998.

[Cana03] http://www.canarie.ca/advnet/canet4_rfi.html

[Chin04] ELG6104 Course Notes, Prof. John Chinneck, Systems and Computer Engineering, Carleton University, http://www.sce.carleton.ca/faculty/chinneck/po.html

[ChBa95] C. Chen and S. Banerjee, "Optical switch configuration and lightpath assignment in wavelength routing multihop lightwave networks", *Proc. of IEEE INFOCOM*, pp 1300-1307, 1995

[ChFa96] I. Chlamtac, A. Farago and T. Zhang, "Lightpath (Wavelength) Routing in Large WDM Networks", *IEEE JSAC*, Vol. 14, No. 5, June 1996.

[ChGa92] Imrich Chlamtac, Aura Ganz, and Gadi Karmi. "Lightpath communications: An approach to high bandwidth optical WAN's", *IEEE/ACM Trans. on Networking*, 40(7):1171-1182, 1992.

[ChGa93] I. Chlamtac, A. Ganz, "Lightnets: Topologies for High-Speed Optical Networks," *J. of Lightwave Tech.*, Vol. 11, No. 5/6, May/June, pp 951-961 1993

[ChLi05] X. Chu and B. Li, "Dynamic Routing and Wavelength Assignment in the Presence of Wavelength Conversion for All-Optical Networks", *IEEE/ACM Transaction on Networking*, Vol.13, No.3, June 2005, pp.704-715

[ChMo00] A. Chiu and E.H. Modiano, "Traffic Grooming Algorithms for Reducing Electronic Multiplexing Costs in WDM Ring Networks", *J. of Lightwave Technology*, Vol.18, No.1, Jan. 2000, pp.2-12

[ChRo04] B. Chen, G.N. Rouskas and R. Dutta, "Traffic Grooming in WDM Ring Networks with the Min-Max Objective," *Proc. of 3rd Int'l IFIP-TC6 Networking Conference* (Networking'04), Lecture Notes in Computer Science, Vol.3042, May 9-14, 2004, Athens, Greece, pp.174-185

[ChWa02] B. Chen and J. Wang, "Efficient Routing and Wavelength Assignment for Multicast in WDM Networks", *IEEE JSAC*, pp.97-109, Vol. 20, No. 1, Jan. 2002.

[Cink03] T. Cinkler, "Traffic and Grooming", *IEEE Network*, Vol.17, No.2, pp 16-21, March/ April 2003

[CiMa00] T. Cinkler, D. Marx, et al., "Heuristic Algorithms for Joint Configuration of the Optical and Electricla Layer in Multi-Hop Wavelength Routing Networks", *Proc. Inforcom 2000*, New York, March 2000

[CoSa01] L.A. Cox, Jr. and J. Sanchez, "Cost Savings from Optimized Packing and Grooming of Optical Circuits: Mesh versus Ring Comparisons", *Optical Networks Magazine*, Vol.2, No.2, May/June 2001, pp.72-90

[DuRo02t] R. Dutta and G.N. Rouskas, "Traffic Grooming in WDM Networks: Past and Future", *IEEE Network*, Vol.16, No.6, pp 46-56, Nov/Dec 2002

[DuRo02o] R. Dutta and G.N. Rouskas, "On Optimal Traffic Grooming in WDM Rings", *IEEE JSAC*, Vol.20, No.1, pp 110-121, Jan. 2002

[Dutt01] Rudra Dutta, "Virtual topology design for traffic grooming in WDM networks", Ph.D. Dissertation, 2001 North Carolina State University.

[DuHu03] R. Dutta, S. Huang and G.N. Rouskas, "On Optimal Traffic Grooming in Elemental Network Topologies," *Proc. of OptiComm'03*, Oct. 15-18, Dallas, pp.13-24

[DuHu03a] R. Dutta, S. Huang and G.N. Rouskas, "Traffic Grooming in Path, Star and Tree Networks: Complexity, Bounds and Algorithms," *Proc. of SIGMETRICS'03*, June 10-14, 2003, San Diego, California, USA, pp.298-299

[DuRo00] R. Dutta and G.N. Rouskas, "A Survey of Virtual Topology Design Algorithms for Wavelength Routed Optical Networks," *Optical Networks Magazine*, Vol.1, No.1, Jan. 2000, pp.73-89

[DuRo02t] R. Dutta and G.N. Rouskas, "Traffic Grooming in WDM Networks: Past and Future", *IEEE Network*, Vol.16, No.6, pp 46-56, Nov/Dec 2002

[DuRo02o] Rudra Dutta, and George N. Rouskas, "On Optimal Traffic Grooming in WDM Rings", *IEEE JSAC*, VOL. 20, NO. 1, JANUARY 2002, P110-121

[DuRo02] R. Dutta and G.N. Rouskas, "On Optimal Traffic Grooming in WDM Networks", *IEEE JSAC*, Vol.20, No.1, pp 110-121, Jan. 2002

[ElMo00] J.M.H. Elmirghani and H.T. Mouftah, "All-Optical Wavelength Conversion: Technologies and Applications in DWDM Networks", *IEEE Communications Magazine*, Vol.38, No.3, March 2000, pp.86-92

[Ermo66] Y. M. Ermol'ev, "Methods for solving nonlinear extremal problems", *Cybernetics*, 16/1 (1966) 1-14.

[Falk67] J.E. Falk, "Lagrange Multipliers and Nonlinear Programming", *Journal of Mathematical Analysis*, 19, 141-159, 1967.

[Fish81a] M. L. Fisher, "The Lagrangean relaxation method of solving integer programming problems". *Management Science* 27(1981) 1-18.

[Fish81b] M. L. Fisher, "The Lagarangian Relaxation Method for Solving Integer Programming Problems," *Management Science*, Vol. 27, No. 1, pp, 1-18, Jan.

[GaWa94] A. Ganz and X. Wang, "Efficient algorithm for virtual topology design in multihop lightwave networks," *IEEE/ACM Trans. Networking*, vol. 2, pp. 217-225, June 1994.

[GaHa83] B. Gavish and S. L. Hantler, "An Algorithm for Optimal Route Selection in SNA Networks," *IEEE Trans. on Comm.*, COM-31, pp. 1154-1160, Oct. 1983.

[GaJo79] Michael R. Garey and David S. Johnson. "*Computers and Intractability. A guide to the theory*"

[GaNe89] B. Gavish, I. Neuman, "A System for Routing and Capacity Assignment in Computer Communication Networks", *IEEE Trans. On Comm.*, pp.360-366, Vol. 37, No. 4, 1989

[GaNe92] B. Gavish and I. Neuman, "Routing in a Network with Unreliable Component", *IEEE Trans. On Comm.*, Vol. 40, No. 7, pp. 1248-1258, Vol. 40, No. 7, Jul. 1992.

[GeMu03] A. Gencata and B. Mukherjee, "Virtual-Topology Adaptation for WDM Mesh Networks Under Dynamic Traffic," *IEEE/ACM Trans. on Networking*, Vol.11, No.2, April 2003, pp.236-247

[Geof70] Geoffrion, A.M. "Elements of Large-Scale Mathematical Programming", *Management Science*, 10, 652-691, 1970.

[Geof70a] A.M. Geoffrion, "Lagrangian Relaxation for Integer Programming", *Mathematical Programming Study*, 2, M.L.Balinski (Ed.), North-Holland, Amsterdam, 82-114, 1974. Trees", *Operations Research*, 18, 1138-1162, 1970.

[GeRa98] O.Gerstel, R.Ramaswami,and G.Sasaki, "Cost effective traffic grooming in WDM rings", *Proc. IEEE Inforcom* ,pages 69 -77,1998.

[GeRa00] O. Gerstel, R. Ramaswami and G.H. Sasaki, "Cost-Effective Traffic Grooming in WDM Rings," *IEEE/ACM Trans. on Networking*, Vol.8, No.5, Oct. 2000, pp.618-630

[GhDi00] N. Ghani; S. Dixit; T. Wang, "On IP-over-WDM integration", *IEEE Comm. Magazine*, pp.72 - 84, ,Vol. 38 , Issue: 3 , March 2000

[Gree91] P. E. Green, "The future of fiber-optic computer networks," *IEEE Compter*, vol. 24, no. 9, pp. 78-87, 1991.

[Gree96] P.E. Green, "Optical Networking Update", *IEEE JSAC* ,14(5):764 -779,Jun 1996.

[GoBo04] W. Golab and R. Boutaba, "Policy-Driven Automated Reconfiguration for Performance Management in WDM Optical Networks", *IEEE Communications Magazine*, Vol.42, No.1, Jan. 2004, pp.44-51

[Haye84] Jeremiah F. Hayes, *Modeling and Analysis of Computer Communications Network*, Plenum Pub Corp, 1984

[HeKa70] M. Held, and R.M. Karp. "The Traveling Salesman Problem and Minimum Spanning trees: Part II", *Mathematical Programming*, 1:6--25, 1971

[HiLe93] Hiriart-urruty, J.B., and Lemarechal, C., *Convex Analysis and Minimization Algorithms I and II*, Springer-Verlag, Berlin, Germany, 1993.

[HoLu93] D.J. Hoitomt, P.B. Luh and K.R. Pattipati, "A practical approach to job-shop scheduling problems", *IEEE Transactions on Robotics and Automation*, Vol. 9, No. 1, Feb. 1993

[Hu02] J. Hu, "Traffic Grooming in Wavelength-Division-Multiplexing Ring Networks: A Linear Programming Solution," *Journal of Optical Networking*, Vol.1, No.11, Nov. 2002, pp.397-408

[Hu02a] J. Hu, "Optical Traffic Grooming in Wavelength-Division-Multiplexing Rings with All-to-All Uniform Traffic," *Journal of Optical Networking*, Vol.1, No.1, Jan. 2002, pp.32-42

[KaAy98] E. Karasan and E. Ayanoglu, Performance of WDM Transport Networks, *IEEE JSAC*, Vol.16, No.7, Sept. 1998, pp.1081-1096

[KoCh01] V.R. Konda and T.Y. Chow, "Algorithm for Traffic Grooming in Optical Networks to Minimize the Number of Transceivers," *Proc. of 2001 IEEE Workshop on High Performance Switching and Routing* (HPSR'01), May 29-31, 2001, Dallas, Texas, pp.218-221

[KeCh74] A. Kershenbaum and W. Chou, "A Unified Algorithm for Designing Multidrop Teleprocessing Networks", *IEEE Trans. on Comm.*, pp. 1762-1772, Vol. 22, No. 11, Nov., 1974

[KrSi98] R. Krishnaswamy and K. Sivarajan, "Design of logical topologies: A linear formulation for wavelength routed optical networks with no wavelength changers," in *Proc. IEEE INFOCOM*, vol. 2, Mar. 1998, pp. 919-927.

[KrSi01] R. M. Krishnaswamy and N. Sivarajan. "Algorithms for Routing and Wavelength Assignment Based on Solutions of LP-Relaxations", *IEEE Comm. Letters.* Pp.435-437, Vol. 5, No. 10, Oct. 2001

[KuPu03] J. Kuri, N. Puech, M. Gagnaire, E. Dotaro and R. Douville, " Routing and Wavelength Assignment of Scheduled Lightpath Demands", *IEEE JSAC*, vol. 21, no. 8, Oct. 2003

[Labo97] J.F.P. Labourdette, “Performance Impact of Partial Reconfiguration on Multihop Lightwave Networks”, *IEEE/ACM Trans. on Networking*, Vol.5, No.3, June 1997, pp.351-358

[LeeS99] S. W. Lee, "Path Planning and Routing in Survivable Communication Networks", Doctoral Dissertation of Electronic Engineering in Chung Cheng University, 1999.

[LeLi93] K. C. Lee and V. O. K. Li, "A Wavelength-Convertible Optical Network," *J. of Lightwave Tech.*, Vol. 11, No. 5/6, pp962-970, 1993

[LeMe00] E. Leonardi, M. Mellia and M.A. Marsan, "Algorithms for the Logical Topology Design in WDM All-Optical Networks," *Optical Networks Mag.*, Vol.1, No.1, Jan. 2000, pp.35-46

[LeYu04] S.S.W. Lee; M.C. Yuang,; M.C. Tien; S.-H. Lin, “A Lagrangean relaxation-based approach for routing and wavelength assignment in multigranularity optical WDM networks:, *IEEE JSAC*, Vol. 22, No. 9, pp. 1741- 1751, Nov. 2004

[LeZh03] K. Lee, L. Zhang and C.H. Youn, "An Adaptive Virtual Topology Reconfiguration Policy in Multi-Wavelength Optical Internet", *European Transactions on Telecommunications*, Vol.14, No.5, Sept./Oct. 2003, pp.417-422

[Luen84] D.G. Luenberger, "Linear and Nonlinear Programming, second edition", Addison-Wesley, MA, 1984

[Mukh97] B. Mukherjee, "Optical Communication Networks" McGraw-Hill, July 1997.

[MaPi03] S.D. Maesschalck, M. Pickavet, D. Colle and P. Demeester, "Multi-Layer Traffic Grooming in Networks with an IP/PLS Layer on Top of a Meshed Optical Layer," *Proc. of IEEE 2003 Global Communications Conference* (*Globecom'03*), Dec. 1-5, 2003, San Francisco, CA, pp.2750-2754

[MoEr03] G. Mohan, P.H.H. Ernest and V. Bharadwaj, "Virtual Topology Reconfiguration in IP/WDM Optical Ring Networks", *Computer Communications*, Vol.26, No.2, Feb. 2003, pp.91-102

[MoLi01] E. Modiano and P.J. Lin, "Traffic Grooming in WDM Networks", *IEEE Comm. Mag.*, Vol.39, No.7, pp.124-129, July 2001

[Mukh92a] B. Mukherjee, "WDM-based local lightwave networks part I: single-hop systems," *IEEE Network Mag.*, Vol. 3, pp. 21-27, 1992.

[MuRa96] B. Mukherjee, S. Ramamurthy, D. Banerjee, and A. Mukherjee, "Some principles for designing a wide-area optical network," *IEEE/ACMTrans. Networking*, vol. 4, pp. 684-696, Oct. 1996.

[OkSh03] E. Oki, K. Shiomoto, M. Katayama, W. Imajuku and N. Yamanaka, "Performance of Dynamic Multi-Layer Routing Schemes in IP+Optical Networks", *Proc. of 2003 Workshop on High Performance Switching and Routing* (*HPSR'03*), June 24-27, 2003, Torino, Italy, pp.233-238

[OuZh03] C. Ou, K. Zhu, H. Zang, L.H. Sahasrabuddhe and B. Mukherjee, "Traffic Grooming for Survivable WDM Networks: Shared Protection", *IEEE JSAC*, Vol.21, No.9, Nov. 2003, pp.1367-1383

[OzBe03] A. E. Ozdaglar and D. P. Bertsekas, "Routing and wavelength assignment in optical networks," *IEEE/ACMTrans. Network*, vol. 11, no. 2, pp. 259–272, Apr. 2003

[Poly69] Polyak, B. T, "Minimization of unsmooth functionals", *USSR Computational Mathematics and Mathematical Physics*, Vol. 9, 1969, pp. 14-29.

[PaMa03] Z. Patrocinio Jr. and G.R. Mateus, "A Lagrangean-based Heuristic for Traffic Grooming in WDM Optical Networks", *Proc. of IEEE 2003 Global Communications Conference* (*Globecom'03*), Dec. 1-5, 2003, San Francisco, CA, pp.2767-2771

[PaRa88] R. G. Parker and L. R.Rardin, *Discrete Optimization*. Academic Press, Inc., London, 1988.

[PrSt03] P. Prathombutr, J. Stach and E.K. Park, "An Algorithm for Traffic Grooming in WDM Optical Mesh Networks with Multiple Objectives", *Proc. of 12th Int'l Conference on Computer Communications and Networks* (*ICCCN'03*), Oct. 20-22, 2003, Dallas, Texas, pp.405-411

[PrSt04] P. Prathombutr, J. Stach and E.K. Park, and S.W. Tak, "Model for Series of reconfigurations in wavelength-routed optical networks", *Proceedings of 13th International Conference on Computer Communications and Networks*, (*ICCCN* 2004), October 11-13, 2004 Chicago, IL USA, pp.437 - 443

[RaMu02] R. Ramamurthy and B. Mukherjee, "Fixed-Alternate Routing and Wavelength Conversion in Wavelength-Routed Optical Networks", *IEEE/ACM Transactions on Networking*, Vol.10, No.3, June 2002, pp.351-367

[RaSi95] R. Ramaswami and K. N. Sivarajan, "Routing and wavelength assignment in all-optical networks", *IEEE/ACM Transactions on Networking*, 3(5):489-500, 1995.

[RaSi01] R. Ramaswami and K. Sivarajan, *Optical Networks: A Practical Perspective, Morgan Kaufmann Publishers*, 2nd edition (October 15, 2001)

[RiSa02] F. Ricciato, S. Salsano, A. Belmonte, and M. Listanti, "Off-Line Configuration of a MPLS over WDM Network under Time-Varying Offered Traffic", *Proceedings of IEEE INFOCOM 2002*, The 21st Annual Joint Conference of the IEEE Computer and Communications Societies, New York, USA, June 23-27, 2002, pp.57-65

[Rous01] G. Rouskas, "Routing and wavelength assignment in Optical WDM networks", *Wiley Encyclopedia of Telecommunications*, (John Proakis, Editor), John Wiley & Sons, 2001.

[SaLu02] M. E.M Saad, Zhi-Quan Luo, "A Lagrangean decomposition approach for the routing and wavelength assignment in multifiber WDM networks", *Proc. GLOBECOM '02*. IEEE, Vol. 3, pp: 2818 -2822, 17-21 Nov. 2002.

[SaLu05] M. Saad and Z. Luo, "Reconfiguration With No Service Disruption in Multifiber WDM Networks", *J. Of Lightwave Technology*, VOL. 23, NO. 10. pp. 3092 - 3104, Oct. 2005

[Shor64] N.Z. Shor, "On the structure of algorithms for the numerical solution of optimal planning and design problems," Ph.D. dissertation, Cybernetics Inst., Academy of Sciences, USSR, 1964

[Shor85] N. Z. Shor, *Minimization Methods for Non-differentiable Functions*, Springer Series in Computational Mathematics. Springer, 1985.

[SiSu00] K. M. Sivalingam and Subramaniam, *Optical WDM networks: principles and practice,* 2000 - Boston: Kluwer Academic Publishers, Chapter 4, pp. 79 - 102

[SrSi01] N. Sreenath, C.A.R. Murthy, B.H., Gurucharan, and G. Mohan, "A Two-Stage Approach for Virtual Topology Reconfiguration of WDM Optical Networks", *Optical Networks Magazine*, Vol.2, No.3, May/June 2001, pp.58-71

[SoZh04] T. Song, H. Zhang, Y. Guo and X. Zheng, "Optimal Design of WDM Ring Networks to Minimize SDH ADMs", *J. of Optical Communications*, Vol.24, No.3, 2004, pp.144-148

[SrSo04] R. Srinivasan and A.K. Somani, "Analysis of Optical Networks with Heterogeneous Grooming Architectures", *IEEE/ACM Trans. on Networking*, vol. 12, no. 5, October 2004, pp. 931-943

[SrSo01] M. Sridharan, A.K. Somani and M.V. Salapaka, "Approaches for Capacity and Revenue Optimization in Survivable WDM Networks," *Journal of High Speed Networks*, Vol.10, No.2, 2001, pp.109-125

[StBa99] T. E. Stern and K. Bala, *Multiwavelength Optical Networks*, Addison-Wesley, 1999

[SwSi02] M.D. Swaminathan, K.N. Sivarajan, "Practical Routing and Wavelength Assignment algorithms for All Optical Networks with Limited Wavelength Conversion", *Proc. 2002. ICC* (*IEEE International Conference on Communications*)

[SuSo03] C.S. Sung and S.H. Song, "Branch-and-Price Algorithm for a Combined Problem of Virtual Path Establishment and Traffic Packet Routing in a Layered Communication Network", *J. of the Operational Research Society*, Vol.54, No.1, Jan. 2003, pp.72-82

[ThSo01] S. Thiagarajan and A.K. Somani, "Capacity Fairness of WDM Networks with Grooming Capabilities," *Optical Networks Magazine*, Vol.2, No.2, May/June 2001, pp.24-32

[ToMa02] M. Tornatore, G. Maier and A. Pattavina, "WDM Network Optimization by ILP Based on Source Formulation", *Proc. Inforcom 2002*

[WaLu97] J. Wang, P.B. Luh, X. Zhao, and J. Wang, "An Optimization-based Algorithm for Job Shop Scheduling", *Sadhana*, Vol. 22, Part 2, pp. 241-256, 1997.

[WeHe03] H. Wen, R. He, L. Li and S. Wang, "Dynamic Traffic-Grooming Algorithms in Wavelength-Division-Multiplexing Mesh Networks," *Journal of Optical Networking*, Vol.2, No.4, April 2003, pp.100-111

[WiVe03] W. L. Winston and M. Venkataramanan, *Introduction to Mathematical Programming*, 4th edition, Thomson, Brooks/Cole, 2003

[XiLe99] G. Xiao and Y.W. Leung, "Algorithms for Allocating Wavelength Converters in All-optical Networks", pp 545-557, *IEEE/ACM Trans. On Networking*, Vol. 7, No. 4, Aug. 1999

[XiQi03p] C. Xin and C. Qiao, "Performance Analysis of Multi-Hop Traffic Grooming in Mesh WDM Optical Networks," *ICCCN'03*, Oct. 20-22, 2003, Dallas, pp.237-242

[XiQi03t] C. Xin, C. Qiao and S. Dixit, "Traffic Grooming in Mesh WDM Optical Networks: Performance Analysis", *Proc. of IEEE 2003 Global Communications Conference* (*Globecom'03*), Dec. 1-5, 2003, San Francisco, CA, pp.3732-3736

[YeFr01] H. Yen and F. Y. Lin, 2001 Apr., "Near-optimal Delay Constrained Routing in Virtual Circuit Networks", *Proc. IEEE INFOCOM* 2001, Vol. 2, pp. 750-756.

[YeLi01a] H. Yen and F. Y. Lin, 2001 Feb., "Near Optimal Design of Lightpath Routing and Wavelength Assignment in Purely Optical WDM Networks", *Proceeding of IFIP Optical Network Design and Modeling* 2001 (*ONDM* 2001).

[ZaJu00] H. Zang, J. P. Jue, and B. Mukherjee, "A review of routing and wavelength assignment approaches for wavelength-routed optical WDM networks", *Optical NetworksMagazine*, 1(1):47-60, January 2000.

[ZaVu03] A. Zalesky, H. L. Vu, M. Zukerman, C. Leckie and I. Ouveysi, "Routing and Wavelength Assignment over Multi-fiber Optical Networks with No Wavelength Conversion", *Proc. ONDM* 2003

[ZhAc95] Z. Zhang and A S. Acampora, "A Heuristic Wavelength Assignment Algorithm for Multihop WDM Networks with Wavelength Routing and Wavelength Re-Use," *IEEE/ACM Trans. Networking*, Vol. 3, No. 3, pp281-288, June 1995

[ZhHa03] D. Zhemin and M. Hamdi, "Traffic Grooming in Optical WDM Mesh Networks Using the Blocking Island Paradigm", *Optical Networks Magazine*, Vol.4, No.6, Nov./Dec. 2003, pp.7-15

[ZhMu02] K. Zhu and B. Mukherjee, "Traffic Grooming in an Optical WDM Mesh Network", *IEEE JSAC*, pp.122-133, Vol. 20, No. 1, Jan. 2002.

[ZhMu03] K. Zhu and B. Mukherjee, "A Review of Traffic Grooming in WDM Optical Networks: Architectures and Challenges," *Optical Networks Magazine*, Vol.4, No.2, March/April 2003, pp.55-64

[ZhLu99] X. Zhao , P. B. Luh , J. Wang, "Surrogate gradient algorithm for Lagrangean relaxation", *Journal of Optimization Theory and Applications*, v.100 n.3, p.699-712, March 1999 of NP-completeness. Freeman, 1979.

[ZhRa03] S. Zhang and B. Ramamurthy, "Dynamic Traffic Grooming Algorithms for Reconfigurable SONET over WDM Networks," *IEEE JSAC*, Vol.21, No.7, Sept. 2003, pp.1165-1172

[ZhQi00] X. Zhang and C. Qiao, "An Effective and Comprehensive Approach for Traffic Grooming and Wavelength Assignment in SONET/WDM Rings," *IEEE/ACM Trans. on Networking*, Vol.8, No.5, Oct. 2000, pp.608-617

[ZhZa03] K. Zhu, H. Zang and B. Mukherjee, "A Comprehensive Study on Next-Generation Optical Grooming Switches," *IEEE JSAC*, Vol.21, No.7, Sept. 2003, pp.1173-1186

[ZhZa03a] H. Zhu, H. Zang, K. Zhu and B. Mukherjee, "A Novel Generic Graph Model for Traffic Grooming in Heterogeneous WDM Mesh Networks," *IEEE/ACM Trans. on Networking*, Vol.11, No.2, April 2003, pp.285-299

[ZhZh02] B. Zhou, J. Zheng and H.T. Mouftah, "Dynamic reconfiguration based on balanced alternate routing algorithm (BARA) for all-optical wavelength-routed WDM networks," *Proceedings of IEEE Global Telecommunications Conference*, 2002. (*GLOBECOM* '02), Vol.3, Taipei, Taiwan, 17-21 Nov. 2002, pp.2706 – 2710, Digital Object Identifier 10.1109/GLOCOM.2002

[ZhZh03] K. Zhu, H. Zhu and B. Mukherjee, "Traffic Engineering in Multigranularity Heterogeneous Optical WDM Mesh Networks through Dynamic Traffic Grooming," *IEEE Network*, Vol.17, No.2, March/April 2003, pp.8-15

[ZhZh04] J. Zheng, B. Zhou, and H. T. Mouftah, "Virtual Topology Design and Reconfiguration of Virtual Private Networks (VPNs) over All-Optical WDM Networks", *Photonic Network Communications*, May 2004, vol. 7, iss. 3, pp. 255-266(12), Kluwer Academic Publishers

[ZhYa03] Y. Zhang and O. Yang, "An Effective Approach To The Connection Routing Problem Of All-Optical Wavelength Routing DWDM Networks With Wavelength Conversion Capability", pp. 1370 - 1374 vol.2, *Proceed. of IEEE ICC* (International Conference on Communications), Alaska, USA, May 11–15, 2003

[ZhYa04] Y. Zhang, O. Yang and H. Liu, "A Lagrangean Relaxation And Subgradient Approach For The Routing And Wavelength Assignment Problem In WDM Networks", Nov. 2004, *IEEE JSAC*, pp. 1752- 1765, Vol. 22, Issue 9

[ZhYa05] Y. Zhang and O. Yang, "A Proposal for a Semi-dynamically Reconfigurable Optical Network Optimization", *Proceed. of SPIE Photonic North*, 12 -14 September, Toronto, Canada

[ZiDo04] S. Ziyu, Y. Dongbin, L. Zhengbin, W. Ziyu and X. Anshi, "Performance Analysis of Multi-Fiber WDM Network with Limited-Range Wavelength Conversion", *Photonic Network Communications*, Vol.7, No.3, May 2004, pp.301-312

Appendix A
Algorithms

This chapter provides the complete description of some algorithms used throughout the thesis. Section A.1 to A.3 supply the details of the heuristic algorithm used in Section 3.2.2.4, and Section A.3 and Section A.5 provide the details for the algorithm in Section 4.4.5.

A.1. MMCSLP Algorithm

The MMCSLP (Modified Minimum Cost Semi-lightpath) algorithm proposed here is to solve the subproblem (3.33) optimally to obtain the solution for the dual problem. MMCSLP has 4 steps. The first 3 steps are similar to the MCSLP algorithm [ChFa96]. To get the optimum for $\min_{\Delta_{sdn},\Phi_{sdn}}\{D_{sdn}\}$, subject to constraints (3.22), (3.23) and (3.25), we can modify the MCSLP [ChFa96] by assuming the cost of using wavelength channel w_{ijc} is ξ_{ijc}, and the cost of using one wavelength converter corresponding to wavelength c on node i is λ_{ic}. The procedure of constructing a wavelength graph (WG) is similar to MCSLP. The difference is that in the construction of the column edges, for the wavelength conversion from the wavelength with index c to the wavelength with index a in node i, if the conversion is not available, i.e., $a \notin I_i(c)$, a very large positive value is assigned to the weight of this edge. Then by using the Shortest Path Algorithm for Wavelength Graph (SPAWG) [ChFa96], we can obtain the optimal routing and wavelength assignment for $\min_{\Delta_{sdn},\Phi_{sdn}}\{D_{sdn}\}$, subject to constraints (3.22), (3.23) and (3.25). The algorithm MMCSLP can be described as follows:

A.6) Construct the wavelength graph (see Part B) of the network.

A.7) Find a shortest path between the source and the destination columns using the SPAWG algorithm (see Part C) to get $\min_{\Delta_{sdn},\Phi_{sdn}}\{D_{sdn}\}$.

A.8) If $P_{sdn}>\min_{\Delta_{sdn},\Phi_{sdn}}\{D_{sdn}\}$, set $\alpha_{sdn}=1$ and map the found wavelength assignment back to the network to obtain the minimum cost semi-lightpath.

A.9) If $P_{sdn}<\min_{\Delta_{sdn},\Phi_{sdn}}\{D_{sdn}\}$, set $\alpha_{sdn}=0$. (A tie is broken arbitrarily.)

A.2. Constructing the WG (Wavelength Graph) [ChFa96]

We can construct the wavelength graph in the following steps (step 1 of MMCSLP):

B.1) Take $Z = WN$ vertices, where W is the number of wavelengths and N is the number of nodes in the original network.

B.2) Arrange the vertices in a matrix-like structure with W rows and N columns. Each column corresponds to a node of the network and each row corresponds to a wavelength.

B.3) In the ith row, $i=1,...,W$, draw a directed edge from the column j *to* column h. If w_{jhi} exists, assign the weight $WE_{jhi}= \xi_{jhi}$ to this edge. Otherwise, set $WE_{jhi}=\infty$.

B.4) In column j, $j=1,\ldots,n$, draw a directed edge from row i to l, if at node j wavelength conversion is available from the wavelength with index i to the wavelength with index l. If the conversion is not available, i.e., $l \notin I_j(i)$, a very large positive value is assigned to the weight of this edge. Assign the weight $CE_{ilj}= \lambda_{ji}$ to this edge otherwise.

The difference between this algorithm and what proposed in [ChFa96] is the 4^{th} step, where we assign a very large positive value to the weight of the edge, if the conversion is not available.

A.3. The SPAWG (Shortest Path Algorithm for Wavelength Graph) [ChFa96]

We apply the SPAWG to every subproblem with the same source node in order to obtain the optimal solution for $\min_{\Delta_{sdn},\Phi_{sdn}} \{D_{sdn}\}$, subject to constraints (3.22), (3.23) and (3.25). For the subproblems with the same source node, the same procedure of SPAWG is used. To minimize the computation time, we can obtain the optimal solution for $\min_{\Delta_{sdn},\Phi_{sdn}} \{D_{sdn}\}$ of the subproblems with the same source node together.

In this algorithm, the vertex corresponding to wavelength j of node i in the WG is labeled by "ij". The variable $a_{ij,mn}$ denotes the weight of the edge (ij, mn). The label u_{ij} is used to mark the vertex ij corresponding to wavelength j of node i. The variable R_i is associated to the row corresponding to node i and C_j is a variable associated to the column corresponding to wavelength j. The variables R_i and C_j represent the row and column minimum values, respectively. For simplicity, the column corresponding to the source node is labeled by one. Sets P and T represent permanently and tentatively labeled vertices, respectively. The following are the essential steps, and readers are referred to [ChFa96] for details.

Step 1 (Initialization):

1.1) $u_{1j}=0\,(\forall j)$, $u_{ij}=a_{1j}\,(\forall j, i \neq 1)$;

1.2) $R_i=\min_j\{u_{xj}: x=i, j\neq 1\}$, $(\forall i)$

1.3) $C_j=\min_i\{u_{ix}: x=j, i\neq 1\}$, $(\forall j)$

1.4) $P=\{11,12,\ldots,1W\}$; $T=\{21,22,...,2W,\ldots,N1,N2,\ldots,NW\}$

Step 2 (Designation of a New Permanent Label)

2.1) Find the minimum of R_i, $C_j\,(\forall i, j)$.

2.2) Find a vertex $ij \in T$ with the minimum u_{ij} in the row or column which gave the minimum in Step 2.1 (ties are broken arbitrarily).
2.3) T=T-$\{ij\}$; P=$P \cup \{ij\}$
2.4) If T=Φ then stop.

Step 3 (Updating Row and Column Minimum):
3.1) If vertex ij found in *Step 2* is in row i and column j, then $R_i = \min_j\{u_{xj}: x=i,\ xj \in T\}$, and $C_j = \min_i\{u_{ix}: x=j,\ ix \in T\}$

Step 4 (Revision of Tentative Labels):
4.1.) If vertex ij, found in *Step 2*, is in row i, and column j, then for all $mn \in T$ in row i and in column j (either i=m or j=n), set $u_{mn} = \min\{u_{mn}, u_{mn}+a_{ij,mn}\}$.
Go to *Step 2*.

A.4. The MEWA (Modified Esau-Williams Algorithm)

The basic idea of the MEWA algorithm is to find a viable RWA with a fairly low cost in the whole network for one semi-lightpath. The MEWA algorithm can be described as follows. The notations used here are the same as those in *PART A* to *C*.

Step 0 (Constructing WG):
0.1) Constructing the wavelength graph of the network in the same procedure as Appendix A.2. The source node is assumed to be at the first column in the wavelength graph.

Step 1 (Initialization):
1.1) Set u_{1j}=0, $(\forall j)$ and n=false.
1.2) If $i \neq 1$, then $u_{ih} = WE_{1ih}, (\forall i, h)$.

Step 2(Search All Vertices):
2.1) Node m= next node (the first node for the first run; until all nodes have been searched).
2.2) For all wavelengths with index l of node m, execute Step 3.

Step 3 (Find the Improvement):
3.1) For all wavelengths with index q, if $u_{mq} > (u_{ml} + CE_{lqm})$ and the converter l of node m is not all used, then $u_{mq} = (u_{ml} + CE_{lqm})$, n=false.
3.2) For all nodes p, if $u_{pl} > (u_{ml} + WE_{mpl})$ and the wavelength l on the link between nodes m and p is not used by other semi-lightpaths, then $u_{pl} = (u_{ml} + WE_{mpl})$, n=false; If node p is the destination node of the semi-lightpath considered, terminate the algorithm and assign the RWA in the WG to this semi-lightpath.
3.3) If $!n$, set n=true, and go to *Step 2*
3.4) Else end the algorithm and set the semi-lightpath rejected.

Note that "n = false" means that there is no more possible label change.

A.5. Rough Search Algorithm:

The *Rough Search Algorithm* is a simple heuristic algorithm to obtain an estimate of the overall cost in the electronic domain (the switching cost and the rejection penalty) for the routing of the traffic demands, i.e., $\sum_{(p,q)}\sum_{0<z\le Z_{pq}} P_{pqz}G_{pqz}(1-\gamma_{pqz}) + \sum_{(p,q)}\sum_{0<z\le Z_{sd}} E_{pqz}$, subject to constraints (4.8) and (4.9).

u_i is the label associated with node i. To simplify the description, the source node is set to node 1, and u_1 is set to 0.

Step 1 (Initialization):

1.3) u_1:=0; n:=false;

1.4) If $i\neq1$ and there exist an arc (an admitted semi-lightpath) from node 1 to i, then u_i:= $1, (\forall i)$.

Step 2 (Search all vertices):

2.3) Node m:= next node (the node 1 for the first run; until all nodes have been searched).

Step 3 (Find the improvement):

3.5) For all every node $q\neq1$, if there exists an arc from node m to node q, $u_q>(u_m+V_{pqz})$ and the capacity of this arc is not all used, then $u_q:=(u_m+V_{pqz})$, n:=false; If node q is the destination node of the traffic demand considered, terminate the algorithm and assign the route to this traffic demand;

3.6) If !n, n:=true; Go to *Step 2*; else end the algorithm and set the traffic demand rejected ($\gamma_{pqz} = 0$).
n=true means that there is no more possible label change.

A.6. The Extensive Search Algorithm:

Only if the estimate from *Rough Search Stage* is within a certain range from the best result obtained, the *Extensive Search Stage* is launched. *Extensive Search Stage* generally takes much more time than the *Rough Search Stage*. In *RWA Step*, the variables in A, Δ, Φ are decided. The *Extensive Search Stage* is to minimize $\sum_{(p,q)}\sum_{0<z\le Z_{pq}} P_{pqz}G_{pqz}(1-\gamma_{pqz}) + \sum_{(p,q)}\sum_{0<z\le Z_{sd}} E_{pqz}$, subject to constraints (4.8) and (4.9), with Γ, Y as variables.

The Lagrangean relaxation and subgradient method is again used to solve this problem. We can simply relax the semi-lightpath capacity constraints (4.9) by using Lagrangean multiplier ω_{sdn}, and obtain the dual problem (DP1):

$$\max_{\omega\ge0} p \equiv \min_{\Gamma,\mathrm{Y}}\{\sum_{(p,q)}\sum_{0<z\le Z_{pq}} [P_{pqz}G_{pqz}(1-\gamma_{pqz}) + E_{pqz}] + \sum_{(s,d)}\sum_{0<n\le N_{sd}} \omega_{sdn}(\sum_{(p,q)}\sum_{0<z\le Z_{sd}} \upsilon_{sdn}^{pqz}G_{pqz} - C\alpha_{sdn})\}$$

, subject to (4.8).

By using the fact that $\upsilon_{sdn}^{pqz} = \upsilon_{sdn}^{pqz}\gamma_{pqz}$, DP1 can be rewritten as:

$$\max_{\omega \geq 0} p \equiv \min_{\Gamma, \mathrm{Y}}\{\sum_{(p,q)} \sum_{0<z\leq Z_{pq}} [(1-\gamma_{pqz})P_{pqz}G_{pqz} + \gamma_{pqz} \sum_{(s,d)} \sum_{0<n\leq N_{sd}} (V_{pqz} + G_{pqz}\omega_{sdn})\upsilon_{sdn}^{pqz}]\} - \sum_{(s,d)} \sum_{0<n\leq N_{sd}} C\omega_{sdn}\alpha_{sdn}$$

$$= \sum_{(p,q)} \sum_{0<z\leq Z_{pq}} \min_{\gamma_{pqz}, \mathrm{Y}_{pqz}} \{(1-\gamma_{pqz})P_{pqz}G_{pqz} + \gamma_{pqz} \sum_{(s,d)} \sum_{0<n\leq N_{sd}} (V_{pqz} + G_{pqz}\omega_{sdn})\upsilon_{sdn}^{pqz}\} - \sum_{(s,d)} \sum_{0<n\leq N_{sd}} C\omega_{sdn}\alpha_{sdn}$$

$$= \sum_{(p,q)} \sum_{0<z\leq Z_{pq}} \min_{\gamma_{pqz}}\{(1-\gamma_{pqz})P_{pqz}G_{pqz} + \gamma_{pqz} \min_{\mathrm{Y}_{pqz}}\{\sum_{(s,d)} \sum_{0<n\leq N_{sd}} (V_{pqz} + G_{pqz}\omega_{sdn})\upsilon_{sdn}^{pqz}\}\} - \sum_{(s,d)} \sum_{0<n\leq N_{sd}} C\omega_{sdn}\alpha_{sdn},$$

subject to (4.8).

Note that α_{sdn}'s are already decided in *RWA Step*, so the *Routing Step* in the heuristic algorithm sees only the virtual topology formed by all the nodes and the semi-lightpaths with $\alpha_{sdn} = 1$.

To obtain the minimum for every sub-problem $S_{pqz} = \min_{\gamma_{pqz}}\{(1-\gamma_{pqz})P_{pqz}G_{pqz} + \gamma_{pqz} \min_{\mathrm{Y}_{pqz}}\{\sum_{(s,d)} \sum_{0<n\leq N_{sd}} (V_{pqz}$, we $+ G_{pqz}\omega_{sdn})\upsilon_{sdn}^{pqz}\}\}$ can simply use the same Revised Shortest Path Algorithm (RSPA) provided in Section 5.4, assuming the cost of using semi-lightpath s_{sdn} by traffic demand x_{pqz} is $V_{pqz}+\omega_{sdn}G_{pqz}$.

The dual solution is thus obtained by solving all the sub-problems. The semi-lightpath constraint (4.9) are relaxed in the dual problem, so that the solution to the dual problem is generally not feasible. To obtain the feasible result, we can simply apply the same *Rough Search Algorithm*.

Subgradient method is used to maximize the dual solution.

The multiplier vector ω is updated using the following formula:

$$\omega^{(i+1)} = \omega^{(i)} + \beta^{(i)} g(\omega^{(i)}),$$

where $\omega^{(i)}$ denote the value of vectors ω obtained at the ith iteration, and $\alpha^{(i)}$ denote the step size at the ith iteration. The notations $g(\omega)$ is the subgradients of p with respect to ω. The vector $g(\omega)$ is composed of $g_{sdn}(\omega)$, where

$$g_{sdn}(\omega) = \sum_{(p,q)} \sum_{0<z\leq Z_{sd}} \upsilon_{sdn}^{pqz} G_{pqz} - C\alpha_{sdn}. \tag{A.1}$$

The step size is given by

$$\beta^{(i)} = \theta \times \frac{p^U - p^{(i)}}{g^T(\omega^{(i)}) g(\omega^{(i)})},$$

where p^U is an estimate of the optimal solution, and $p^{(i)}$ is the value of p at the ith iteration. Generally, the best value of the objective function of the feasible routings obtained is used to be p^U. The adaptation of the parameters in the algorithm is the same as Section 3.2.2.3. In the current implementation, we terminate the *Extensive Search Stage* after 2000 iterations.