TEXT and rectangles in blue will NOT show on printed copy

Type final title of thesis or dissertation (M.S. and Ph.D.) below . If your title has changed since your submitted an Application for Graduate Degree, notify Graduate Office.

Resource Assignment for Fiber Optic Networks

A Dissertation

Presented to

the faculty of the School of Engineering and Applied Science

University of Virginia

---

in partial fulfillment

of the requirements for the degree

Doctor of Philosophy

by

Name

Xu Wang

Month degree is awarded December

Year 2014

APPROVAL SHEET

The dissertation

is submitted in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Xu Wang

AUTHOR *signature*

The dissertation has been read and approved by the examining committee:

Please insert committee member names below:

Maite Brandt-Pearce

Advisor

Stephen G. Wilson

Suresh Subramaniam

Malathi Veeraraghavan

Andreas Beling

Donald Brown

Accepted for the School of Engineering and Applied Science:

Dean, School of Engineering and Applied Science

Month degree is awarded December

Year 2014

Print Form

Resource Assignment for Fiber Optic Networks

Xu Wang

A Dissertation
Presented to the Faculty
of University of Virginia
in Candidacy for the Degree
of Doctor of Philosophy

Recommended for Acceptance
by the Department of
Electrical and Computer Engineering
Adviser: Professor Maïté Brandt-Pearce

Nov, 2014

Abstract

Transport optical networks form the backbone of the information infrastructure worldwide. Current systems use wavelength division multiplexing (WDM) technology to accommodate large traffic volumes. In the near future elastic optical networks (EON) are expected to replace WDM networks to further increase the network capacity. This dissertation examines the problem of assigning physical and spectral resources to WDM networks and EON for efficient design and use of these systems.

The physical resource assignment problem, often referred to as the routing and wavelength assignment (RWA) problem, is very important part of the design of fiber optic WDM networks. Objectives of this problem include minimizing the total capital investment in the static design phase and maximizing the throughput in the dynamic operation phase. We develop strategies both in heuristic algorithms and in mixed-integer linear programming (MILP) not just for RWA but also to include considerations of physical impairments, traffic grooming for both static and dynamic networks. For heuristic algorithms, we develop both centralized and distributed algorithms based on the information sharing and assignment decision making. The distributed heuristic algorithms are based on ant colony optimization (ACO) which is a meta-heuristic method that is inspired by the foraging behavior of ants and has been widely implemented in solving discrete optimization problems. Simulation results show although the centralized algorithm shows better efficiency in terms of blocking probability, our ACO shows great robustness and adaptivity to varying network and traffic conditions. We also show implementing technologies such as traffic grooming and signal regeneration will greatly reduce the blocking probability of calls.

Elastic optical networks (EON) have added flexibility to network deployment and management. We propose a link-based MILP formulation for EON to implement signal regeneration as well as wavelength conversion and modulation conversion. We then propose a recursive model in order to either augment existing network deploy-

ments or speed up the resource allocation computation time for larger networks with higher traffic demand requirements than can be solved using an MILP. We show through simulation that systems equipped with signal regenerators or wavelength converters require a notably smaller total bandwidth, depending on the topology of the network. We also show that the suboptimal recursive solution speeds up the calculation and makes the running-time more predictable, compared to the optimal MILP. We compare the two approaches, namely path-based (PB) and link-based (LB) MILP formulations, in their implementation, optimality, and complexity for EONs. We show using simulation that it is beneficial to use LB formulation when including the signal regeneration and that the network topology and traffic demand affect the difference in performances between the two formulations. We combine the MILP formulation for static network and time-slot concept to solve a real-time traffic scenario so the overall network throughput is maximized. The throughput is maximized given the current network state.

Impact of technologies such as signal regeneration, wavelength conversion, and modulation conversion on network performance metrics such as the amount of spectrum needed is also investigated by analytical modeling. Analytical modeling provides the desirable method for network designers to begin with a fast coarse estimate of network performance implementing such technologies without requiring computationally-burdensome and detailed algorithms.

Acknowledgements

First and foremost, I would like to thank my advisor Professor Maite Brandt-Pearce. Her determination and commitment have always been an inspiration to me. Her strict requirement towards the quality of publication and presentation also helps me to hold myself with a higher standard. Throughout the years of my Ph.d program, Professor Brandt-Pearce has been working closely with me, instructing me and giving great advice. She always keeps an open mind in terms where our work should direct. I also feel grateful that she gives me a relative open space for me to try out ideas and develop my own interest. I would like to thank Professor Suresh Subramaniam. We have worked closely on many publications. He always gave great advice on how to make our work better. His expertise also helps me progress in my research.

Next, I would like to thank my family, my parents have always been supporting me and encouraging me both emotionally and financially. Then I would thank my lovely wife, who always tells me how good I am when I feel stuck any time along the way and takes care of my daily life.

Finally, I would thank my colleagues and friends who have given me great support during the years and during the preparation of the thesis. I would thank Professor Stephen Wilson, Professor Malathi Veeraraghavan and Professor Andreas Beling for serving in my defense committee and Professor Brown for help at my proposal.

I gratefully acknowledge the financial support of the National Science Foundation.

To my family.

Contents

List of Tables

List of Figures

List of Acronyms

ACO Ant colony optimization.

ASE Amplified spontaneous emssion.

ASP Adaptive shortest path routing.

BER Bit error rate.

CapEx Capital expenditure.

DH-SP Distance-hop shortest path.

DPSK Differential phase shift keying.

DQPSK Differential quaternary phase shift keying.

DWDM Dense wavelength division multiplexing.

EDFA Erbium doped fiber amplifier.

EON Elastic optical network.

FA Fixed-alternate routing.

FEC Forward error correction.

FF First fit.

FTTx Fiber-to-the-x.

FWM Four-wave mixing.

GASP Grooming-adaptive shortest path algorithm.

GRWA Grooming, routing and wavelength assignment.

ILP Integer linear programming.

LB Link-based.

LP Linear programming.

LSA Link state advertisement.

MC Modulation conversion.

MILP Mixed-integer linear programming.

MLR Mixed-line-rate.

OEO Optical-electrical-optical.

OFDM Orthogonal frequency-division multiplexing.

OOK On-off keying.

PB Path-based.

PI Physical impairments.

QAM Quadrature amplitude modulation.

QoS Quality-of-service.

QoT Quality of transmission.

QPSK Quaternary phase shift keying.

RN Regeneration node.

RSA Routing and spectrum assignment.

RWA Routing and wavelength assignment.

SNR Signal to noise ratio.

SP Shortest path routing.

SPM Self-phase modulation.

SQP Signal quality prediction based.

TDM Time division multiplexing.

TR Transmission reach.

TTL Time-to-live.

WC Wavelength conversion.

WDM Wavelength division multiplexing.

XPM Cross-phase modulation.

Chapter 1

Introduction

Our society's ability to access the massive amounts of information which needs to grow and prosper depends on the availability of technically and economically efficient optical networks. This dissertation examines the problem of assigning physical and spectral resources to large-scale optical networks to improve the design and use of future generations of these systems.

The routing and wavelength assignment (RWA) problem for WDM networks is a problem in the field of network design. It is often approached as a multi-commodity flow problem with unique constraints such as wavelength continuity, etc., for wavelength routing networks. The unique problem caused by physical impairments for long-haul transport network requires a RWA algorithm to consider signal regeneration and also regeneration resource assignment. Different design stages require different problem definition and design objectives. For example, for static design problems the network traffic matrix is given, and the objective is to minimize the resources requirement; for the dynamic operation problem, resources have already been allocated, and the objective becomes to minimize the call blocking probability. The network control mechanism handling the RWA algorithms can either be *centralized* or *distributed*, differing in their information collection and decision making functions.

As other technologies for fiber optic networks emerge, the RWA also evolves. Mixed-line-rate networks allow different wavelength channels in WDM network to support different bit rate transmission, which implies different physical impairments situations. Elastic optical networks (EON) break down the fixed-grid wavelength deployment into a finer and more flexible spectrum deployment, and require new resource allocation strategies (referred to as *routing and spectrum assignment*, RSA). The difference lies in the constraints imposed on the spectrum assignment. Additional functionality such as wavelength conversion, multiple modulation schemes, modulation conversion, etc., also impose new constraints on the design.

In this thesis, we develop a set of solutions to cover all aforementioned issues. These types of problems can be approached using different methods. Based on knowledge of the relationship between network resources and problem objectives, many heuristic algorithms have been developed, as discussed below. The problem can also be framed as a mixed-integer linear programming (MILP) problem with constraints, where the objective function states the design goal. Finally, analytical models are sometimes used to investigate the relationship between network parameters and provide a coarse estimation of network performance. From Chapters 3 to 8, we examine the full RWA/RSA problem from these different perspectives and develop novel powerful solutions.

1.1 Fiber Optic Networks

Optical fibers, introduced decades ago, only became an acceptable channel for long-haul communications after the fiber manufacturing technology matured and overcame the large attenuation of early systems [2]. Through this advancement and the advent of low-loss optical amplifiers, optical signals are now able to travel through fibers for distances of thousands of kilometers without being severely degraded [3].

Fiber optic networks have the advantage of high speed, low attenuation, and stronger resilience to interference, compared to other systems such as copper and wireless communication systems. They have quickly become the only reasonable choice for long distance, backbone communication. Copper networks are still being used for end-user access, such as home access networks deployed by cable television companies. Still, they are all connected to the optical fiber backbone. Almost every cable in the telephone system is fiber-optic, and the internet is all fiber. For mobility purposes, wireless networks have become the technology of choice as access networks for commercial buildings, campuses, etc. As the cost of optical devices continues to decrease, fiber optic networks become more available directly to end-users, creating fibre-to-the-x (FTTx) networks, where x can be a neighborhood curb, a business, or even an individual home. Fibers are also expected to penetrate into more aspects of our daily networking usage, such as for household appliances management systems, surveillance systems, etc. [4].

1.2 Spectrum Partitioning

The available spectrum of a single fiber strand can be as much as ten terahertz (THz). For such a huge bandwidth to be assigned a single user is very wasteful. WDM separates the whole bandwidth into channels and assign them to different users, thereby improving the spectrum usage efficiency.

Compared with the older coarse WDM standard, dense WDM (DWDM) has a much tighter channel spacing, usually 50 or 25 GHz per channel. DWDM allows more channels to be packed into a single fiber and therefore further improves the usage efficiency. Each user is usually assigned one wavelength channel, occupying the channel exclusively from the source node to the destination node. This is called the *wavelength continuity constraint*. Switching the signal from one wavelength channel

to another mid-path is only possible if wavelength conversion is performed along the path, which is an expensive operation typically requiring optical-electrical-optical conversion.

Current communication traffic consists of a diverse set of bit rate requests. To handle such diversity, fiber optic networks can be designed to handle various wavelength channel capacities, called mixed-line-rate (MLR) networks. Yet as the traffic becomes more heterogenous, chunky fixed-grid spectrum assignment becomes inefficient. Exclusive channel occupancy causes wavelength channels to quickly become congested with traffic streams carrying bit rates much smaller than the channel capacity. EONs, as compared to legacy fixed-grid WDM, have the advantage of finer and more flexible spectrum granularity, together with more sophisticated signal processing such as bandwidth-variable modulation. Thanks to new bandwidth-variable transponders (which are devices that send and receive the optical signal from a fiber) that allow flexible spectrum tuning and bandwidth allocation, flexible spectrum assignment has become available. For example, implementation using orthogonal frequency-division multiplexing (OFDM) allows expanding and shrinking spectrum allocation simply by assigning or terminating subcarriers to new traffic.

1.3 Resource Allocation Problem

Our research interests focus on the design of fiber optic network infrastructure, such as component allocation and managing physical resource assignment.

Algorithms that plan and manage the assignment of physical resources to traffic demands in wavelength routing networks are called RWA algorithms. With an expansion that allows low bit-rate sub-streams to be multiplexed into higher bit-rate streams (i.e., traffic grooming), the algorithm is referred to as grooming, routing and wavelength assignment (GRWA). Traffic grooming introduces a cost-efficient balance

between network capacity increase and initial capital expenditure (CapEx). It also exhibits a dramatic improvement in adaptivity for heterogeneous traffic patterns.

Much research has been devoted to addressing the GRWA problem [5, 6, 7, 8]. Solutions can be categorized from several perspectives. For the design phase, a deterministic traffic matrix is assumed to reflect current and foreseen future evolving trends and GRWA aims to minimize the total CapEx of the networks, the so-called offline or static design problem. For the network operation phase, in which a certain existing network structure and deployment of equipment are assumed, GRWA aims to minimize the traffic demand that cannot be served due to shortage of network resources (or maximize the total accepted traffic demand).[1] The network operational algorithms are important because, unlike assumptions made for static design where traffic arrives synchronously, in real-time networks traffic is highly dynamic, which then leads to a highly dynamic network state (e.g., channel availability, optical equipment usage, etc.). Based on the network state, dynamic GRWA optimization needs to accommodate new and unpredictable connection requests.

Increasing traffic volume and growing heterogeneity of bandwidth requirements have pushed the development of optical transport networks. Using WDM technology, spectrum usage has greatly increased by allowing multiple-line-rates and traffic grooming, which is discussed in Chapter 3. Yet WDM is unable to handle increasing traffic heterogeneity because of the coarse wavelength grid employed. EONs [9] provide flexibility in both bandwidth assignment (using sub channel granularity and super channel assignments) and lightpath reconfigurability not available in WDM. The problem of assigning route and spectrum to calls in EON is called the RSA problem [10, 11]. As the technology matures, additional functionality such as modulation selection and conversion can be added, with the hope of further increasing the spec-

[1]Connection blocking probability is an important measure for quality of service for network operators. Although the practical requirement of blocking probability is quite strict, for the purpose of examining GRWA algorithms, a comparison over a relatively wider range of blocking probabilities shows performance differences.

tral efficiency. When major additions in physical layer features are being considered, the networks design should be re-examined to determine the realized benefit gained by their implementation.

1.4 Thesis Contribution

Our research acknowledges that for RWA to be realistic for our targeted long-haul fiber optic networks, the consideration of physical impairments is important. We then develop our algorithms (from heuristics to MILP formulations) around the incorporation of physical impairments.

In our RWA algorithms for WDM networks, we also incorporate traffic grooming and mixed-line-rate transmission to bridge the gap between legacy fixed-grid WDM networks and ever increasingly heterogeneous traffic. Numerical results show the improvement in network performance obtained by implementing these technologies. We develop two categories of algorithms based on the management and control mechanism, namely centralized algorithms and distributed algorithms. We first acknowledge that both mechanisms have advantages and disadvantages that make one of them more preferable than the other in certain scenarios. Then we test our designs through numerical simulation and support our conclusions by comparing many aspects, such as call blocking probability, call setup delay, network control overhead, RWA decision domain, information sharing domain and computational complexity.

We then build an MILP formulation for the implementation of technologies mentioned in our heuristics such as wavelength conversion, multiple modulation, modulation conversion for EON networks. The simulation results show the differences in network performance under different technology setups. We also acknowledge that although MILP provides an optimized RSA solution, it is highly un-scalable with the size of the network and traffic matrix. Therefore, we develop a recursive model

of the same formulation and greatly reduce and stabilize the calculation time, albeit sacrificing some performance. By doing this, we provide a way to balance the need of complexity and performance for algorithm users and also a way to investigate problems such as dynamic resource assignment and network expansion over existing structures.

We investigate the popular acceptance of path-based MILP formulations in RWA and RSA problems. We use our link-based formulation to compare the existing path-based formulation and show that when it comes to networks that involve signal regeneration and multiple modulation schemes, path-based formulations are ineffective in optimizing regeneration resources, and therefore often lead to poor spectrum usage efficiency.

In order to explore the intrinsic relationship between network parameters such as network order, size, scale, network regeneration resources availability, and network performance metrics such as total spectrum usage, highest link spectrum requirement, and network capacity, we develop an analytical model based on probabilistic assumptions and mathematics enumeration. Our analytical model readily shows the dependence of the network performance on resource availability and network structure. We also use our model to test technologies such as signal regeneration and modulation conversion.

1.5 Thesis Organization

The thesis is organized as follows. Chapter 2 introduces the network topology and traffic profile based on the networks we consider in our research. In Chapter 3 we propose a centralized GRWA algorithm. In Chapter 4 we propose a distributed GRWA algorithm based on ant colony optimization. Chapter 5 solves the RSA problem using mixed-integer linear programming for EON. Chapter 6 compares the link-based MILP

formulation with the path-based MILP formulation. In Chapter 7 we implement a similar MILP model on a dynamic WDM network. We approximate the impact of several technologies with an analytical model in Chapter 8. We summarize the dissertation in Chapter 9.

Chapter 2

Network Description

In this chapter we introduce the network topologies and characteristics that we consider in our research work. The scale of the network and practical physical-layer constraints (physical impairments, wavelength capacities) and technologies (mixed-line-rates, regeneration, traffic grooming) affect many aspects of the network design.

2.1 Network Topology

Most of our work is illustrated by implementation on the NSF-24 network, shown in Fig. 2.1. It is a cross country scale network with 24 nodes and 43 bi-directional links (each consists of two uni-directional links in opposite directions). The physical length of the links are shown next to the links, yielding a graph diameter of 6650 km. An earlier variant of this topology is the NSF-14 network, shown in Fig. 2.2, which has 14 nodes and 21 bi-directional links (i.e., 42 uni-directional links). We consider one of the two topologies as our more realistic scenario, which among the two is tested depending on calculation complexity concerns.

In Chapter 5 we compare the NSF-24 network with a same scale but more connected and symmetric network shown in Fig 2.3, which we call the Symmetric-24 network. The diameter is the same as the NSF-24 network. All links are of the same

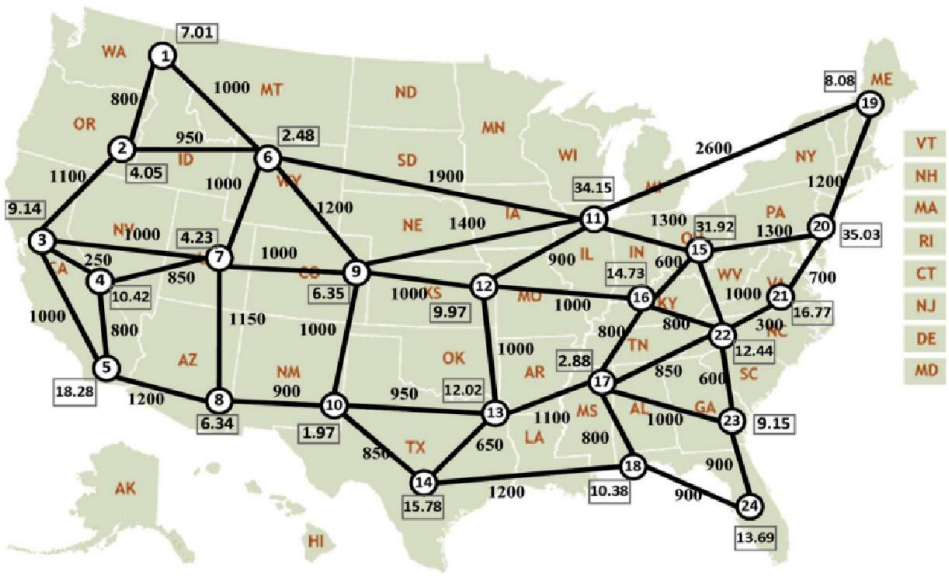

Figure 2.1: NSF-24 network map, Edge labels are physical link lengths in km [1]

length, 1330 km. In Chapter 8, we demonstrate our analytical model first on a larger mesh network shown in Fig. 2.4, with 36 nodes and 60 bi-directional links (i.e., 120 uni-directional links). The link length is set to 800 km, chosen for transmission reach considerations.

The reason that we test our algorithms on many topologies is that the impact on the network performance of networking technologies differs between topologies. This effect becomes obvious, for example, in Chapters 5 and 6. When path-based routing algorithms are implemented, the improvement gained by increasing the number of candidate paths is not the same for a symmetric topology and for an asymmetric topology. In Chapter 5, the improvement obtained by allowing modulation conversion is also different between topologies.

2.2 Channel Spacing and Modulation Techniques

For classic systems where the transmission rate is 10 Gbps, a simple modulation called on-off keying (OOK) is used. Each symbol encodes 1 bit of data. In order to increase

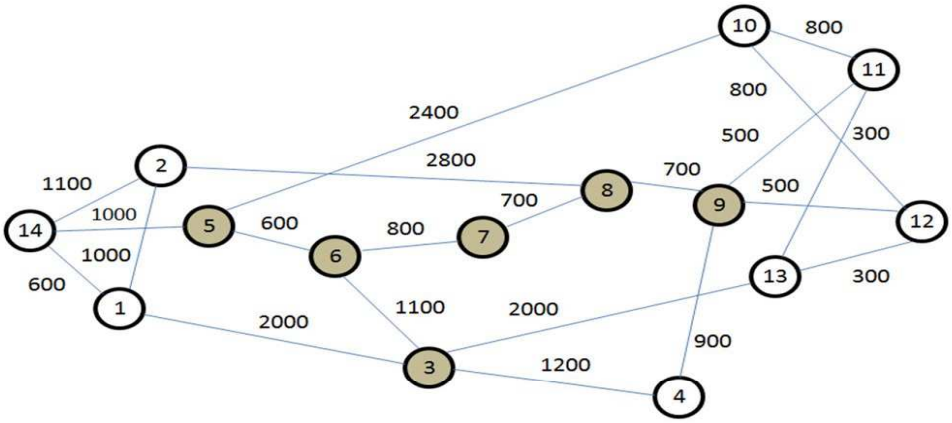

Figure 2.2: NSF-14 network, shaded circles represent signal regeneration nodes

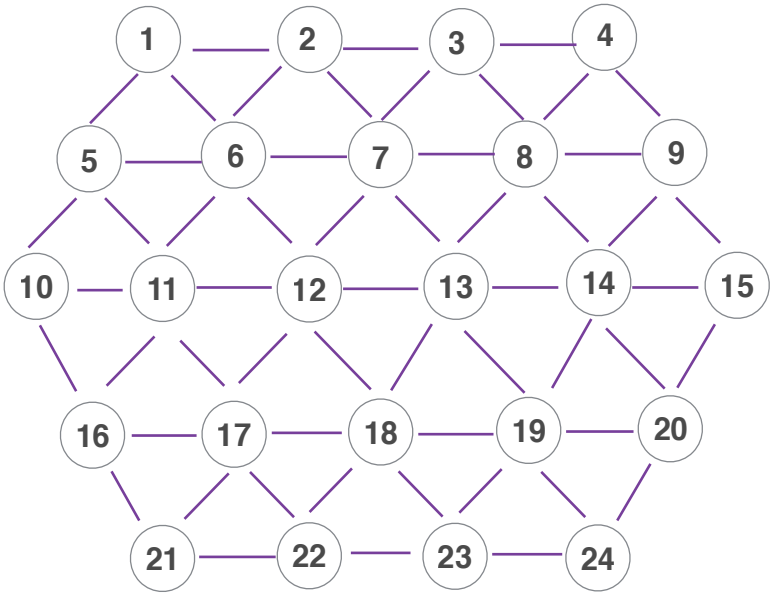

Figure 2.3: Symmetric-24 network

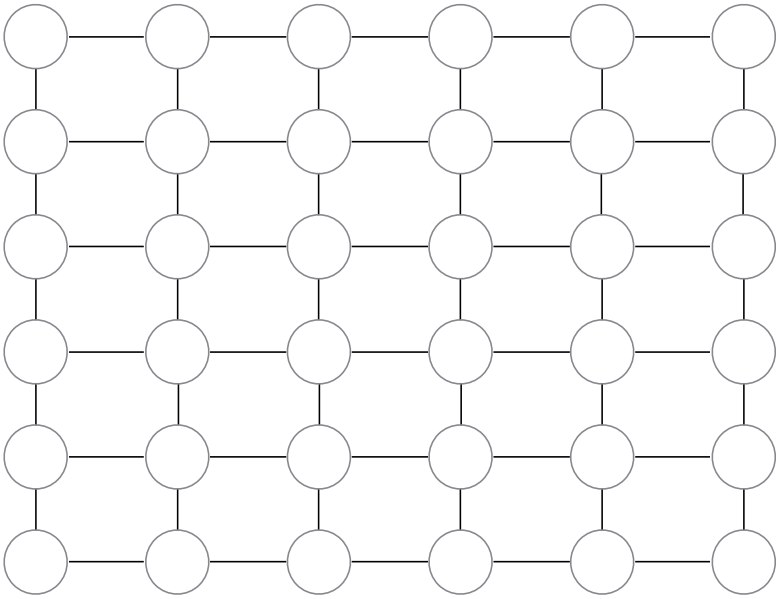

Figure 2.4: 36-node mesh network

the transmission rate, if the transmitting symbol rate cannot be increased, we can increase the amount of data carried on each symbol by implementing more advanced modulation techniques. For example, differential phase shift keying (DPSK) and differential quaternary phase shift keying (DQPSK) are used for transmission rates of 40 and 100 Gbps, typically one stream per polarization. (Note that for such high speeds, OOK would require much more spectrum than is allowed by the 50 GHz channel spacing). For data rates of 100 Gbps and higher, coherent modulation and detection is used, based on some form of quadrature amplitude modulation (QAM). Future systems could either use higher symbol rate QAM, lower-rate channels multiplexed into super-channels, or OFDM modulation [12, 13, 14, 15].

Early generation fiber optic networks with channel capacity of 10 Gbps had channel spacing of 50 or 100 GHz. As networks evolved, one important requirement was to make networks backward compatible. The legacy 10 Gbps fiber optic networks are well equipped. So the process of increasing channel data-rates therefore leads to the coexistence of several line-rates on a single fiber. Such networks are called mixed-line-

rate (MLR) networks [16, 17, 1]. In MLR networks, different channels have different susceptibility to physical impairments, as discussed below.

In EON, the central frequency and spectrum allocation is more flexible than WDM network. We model our EON with channel capacity of any real value within the limit. This is a relaxation of the current standards for EON, which have a frequency slotted model where each channel consists of one or more frequency slots with fine granularity, such as 12.5 GHz.

2.3 Physical Impairments

Long-haul optical fibers suffer degradations originating from many different physical effects, as described here.

Signal attenuation refers to signal loss from fiber absorption and signal scattering. Attenuation makes signals more susceptible to noise by decreasing the signal to noise ratio (SNR) at the receiver. In order to increase the signal strength, optical amplification is employed. The most common optical amplifier is the Erbium doped fiber amplifier (EDFA). Signal amplification, however, has its own issues. The spontaneous emission from the amplifiers behaves as noise, named amplified spontaneous emission (ASE) noise, to the original signal.

Linear distortion can degrade the quality of transmission (QoT), including dispersion and crosstalk. Dispersion simply means "to spread out" and in fibers is mainly from two effects. The first is called chromatic dispersion, which is caused by different frequencies of light propagating at different speeds. The second is called polarization dispersion and this is due to the imperfect shape of the fiber that causes one polarization of light propagating faster than the other. Dispersion causes optical pulses to become indistinguishable from overlapping with neighboring pulses. Dispersion compensation can be used by including a piece of fiber with opposite dispersion char-

acteristics along the path. Linear cross-talk often occurs at the multiplexer found in intermediate nodes and at the receiver. It is due to leakage from neighboring channels and is caused by imperfect channel isolation.

Nonlinear effects mainly include self-phase modulation (SPM), cross phase modulation (XPM), and four-wave mixing (FWM). A nonlinear phase shift occurs from the dependency of the refractive index on optical intensity. As the pulse propagates along the fiber, its spectrum changes because of SPM. Similarly, as signals of two neighboring channels propagate in parallel, the refractive index also depends on the optical intensity of the other wave. FWM is a third-order nonlinear process that transfers energy of one signal to another. SPM causes pulse broadening, while XPM and FWM cause inter-channel cross-talk. Nonlinear effects depend on the intensity of the signal, which implies that as signal launch power increases, so does the nonlinear degradation. Although increasing the launch power augments the SNR, beyond certain threshold the nonlinear effects are so strong that they become the dominant effects that limits the QoT [2].

As degradation of the signal increases, it starts to cause bit errors. The signal quality is then measured by the bit error rate (BER). The targeted BER before forward error correction (FEC) is 10^{-3}. FEC adds redundant information to the transmission so that the receiver can correct a small number of errors.

Once the optical signals are distorted, measures need to be taken to restore the signal quality, called signal regeneration. The aforementioned signal amplification is part of this process. In addition to increasing the signal strength, the original pulse shape and timing between pulses needs to be restored. Re-amplification, re-shaping and re-timing together are called 3R regeneration. Since re-timing is done in the electrical domain, it usually involves an optical-electrical-optical (OEO) conversion. This implies that for optical signals to be fully regenerated, they have to undergo

OEO conversion along the path, usually at intermediate nodes. In this thesis we consider regeneration to be implemented only at the network nodes.

In all topologies shown above, all nodes are considered as edge nodes, meaning traffic arrive at those nodes, and also intermediate nodes, meaning transient traffic traverses them. They are then divided into two types of nodes: nodes equipped with and without OEO conversion circuits. As 3R regeneration involve OEO conversion, and since OEO conversion requires costly high speed electronic devices, it is not economical to have all nodes perform 3R regeneration. Therefore, a sparsely regenerated network (also called a translucent network) only selects a few nodes to equip with OEO conversion. We call such nodes regeneration nodes (RN). In Fig. 2.2 we illustrate the selection of RN by marking them with a darker shade.

The physical impairments are complicated to predict because they depend on many factors such as signal power, modulation techniques, neighboring channel state, etc. In order to design the fiber optic network taking these impairments into consideration, a conservative measure is often made assuming the worst case degradation. Taking all the effects as a single entity, we can quantify the physical layer effects in term of one measure: the need for regeneration after a certain propagation distance. The transmission reach (TR) measures the distance that optical signals can travel without needing 3R regeneration [1]. Obviously, for any traffic, the TR is not fixed, and can be optimized for achieving best overall performance.

In Chapters 3 and 4, we incorporate the physical impairments (PI) into our GRWA algorithm by imposing the same limits on the TR as were proposed in [1]. Using a heuristic method, they estimate the optimal TRs for different data rate signals: $TR = \infty$ for 10 Gbps calls, $TR = 2500$ km for 40 Gbps calls, and $TR = 2000$ km for 100 Gbps calls. We use their regeneration algorithm, which is optimized for when the TR's are fixed for channel bandwidth and regeneration is performed as needed.

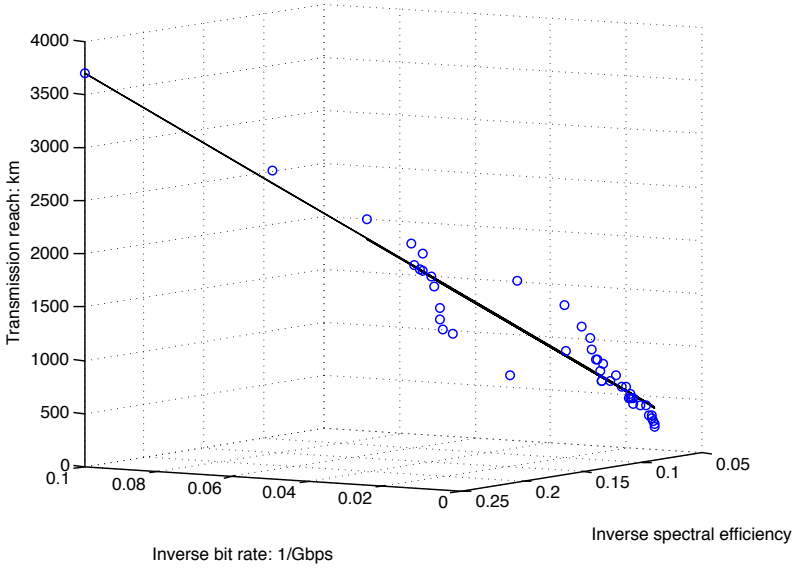

Figure 2.5: Transmission reach based on bit rate and spectral efficiency using polynomial fitting over experimental result data from [18].

In later chapters, we acknowledge that the transmission reach for a traffic demand d depends on many factors such as bit rate of the demand R_d and its spectral efficiency η_d ($\eta_d = R_d/B$, where B is the bandwidth needed). We formulate a linear relationship between transmission reach and these two factors as

$$T_r = \alpha R_d^{-1} + \beta \eta_d^{-1} + \gamma \tag{2.1}$$

where, $\alpha = 18600, \beta = 8360, \gamma = -250$ are coefficients derived from polynomial curve fitting based on experimental data in [18] when T_r is in unit of km, R_d is in unit Gbps and η_d in unit b/s/Hz.

2.4 Wavelength and Modulation Conversion

Wavelength continuity requires that each traffic demand only use one wavelength for the whole lightpath. As the network load increases, when wavelength continuity is enforced the spectrum becomes fragmented causing inefficient wavelength usage.

Sometimes, in order to defragment the spectrum, we want to convert wavelength along the lightpath, so that it can be re-used by other demands. Such conversion is usually done in the electrical domain, implying OEO conversion. In this thesis we assume that wavelength conversion can only be implemented at the RNs.

In EON, each node is equipped with bandwidth variable transponder which support multiple modulation schemes. It is then possible to convert the modulation scheme to fit the TR requirement. (We discuss this further in Chapter 5.) Modulation conversion is also assumed to be performed only at the RNs.

2.5 Traffic Profile

When designing fiber optic networks, traffic can be categorized in several ways. For static networks, the traffic is described as a demand matrix representing the statistical traffic demand between node pairs. For dynamic networks, the traffic can be an arrival-termination model where call demands arrive at the network edge nodes, and last for a period of time before termination. It can also be in the form of a lasting data stream with variable volume.

In our work we assume a uniformly distributed traffic demand assignment throughout the nodes. Unless stated otherwise, each demand has bit rate uniformly distributed from 1 to 100 Gbps. For static traffic, all demands are assumed to arrive at the same time and hold forever. For dynamic traffic, we use a Poisson arrival-termination model and each demand arrives at a particular time and lasts for an exponentially-distributed period of time. The network load for dynamic networks is measured in Erlangs, which is the product of the mean arrival rate and the mean call duration time, assuming the call arrivals are Poisson distributed and call durations are exponential distributed. For simplicity, we set the mean call duration to be one second and adjust the load by adjusting the arrival rate.

2.6 Traffic Grooming for Optical Networks

We point out that optical networks have a huge bandwidth: a single optical fiber strand has over ten THz bandwidth and a single wavelength channel has 50 GHz. Assigning a single wavelength channel exclusively to traffic with a low bit-rate such as several Gbps or lower causes a waste of optical spectrum.

Traffic grooming is an operation that merges low bit-rate new traffic with other new traffic or existing traffic into higher bit-rate stream that can be accommodated as a single lightpath using a RWA algorithm. By grooming new traffic onto existing lightpaths, the available spectrum in the wavelength channel can be better used, thereby increasing the spectrum usage efficiency and reducing the blocking probability. Another benefit of traffic grooming is to reduce the number of costly optical transponders.

Traffic grooming has also attracted research interest in EON architectures (discussed in the later chapters) [19]. In EON the spectrum allocation is flexible and can be tuned small enough to fit 'small' traffic requests. Nevertheless, if too many lightpaths are established, spectrum is wasted on guard bands that separate the neighboring channels. Traffic grooming is realized by assigning separate subcarriers assuming an orthogonal frequency-division multiplexing (OFDM) modulation at the optical transponders used by existing lightpaths. In [19], the authors assign different weights to the optical spectrum, transponders, and existing lightpaths in their weighted Dijkstra's shortest-path algorithm for grooming, routing, and spectrum assignment.

Common traffic grooming is performed in the electrical domain at nodes in the network. This implies that traffic grooming can be done at either the source and destination of the lightpath, or at any intermediate node where traffic undergoes OEO conversion. For other transparent intermediate nodes, traffic grooming is not possible. Traffic grooming is also presumably not allowed to affect existing traffic:

those signals do not undergo new OEO conversion for the sole purpose of grooming new traffic. New traffic, however, can be configured to purposely undergo OEO conversion at certain points to groom with other new traffic.

When implementing traffic grooming in RWA (we call it grooming, routing and wavelength assignment, GRWA), depending on the location of the traffic grooming, it can be categorized as source grooming, destination grooming and intermediate node grooming. Note that a connection may go through multiple new and existing lightpaths, and each lightpath can go through several fiber links with spectrum continuity, which is called multi-hop traffic grooming in [20]. If the grooming is not between traffic demands that share the same source and destination pair, rerouting of the new demand is necessary. In order to balance the excessive spectrum usage from rerouting and the saved spectrum from grooming, certain strategies have to be adopted, as discussed in Chapter 3 where we design a heuristic GRWA algorithm that takes this into account.

Chapter 3

Centralized Heuristic GRWA Algorithm for Translucent WDM Networks

3.1 Introduction

Routing and wavelength assignment (RWA) algorithms for dynamic networks aim to reduce the blocking probability of call demands often caused by the lack of available network resources and unacceptable signal quality. As the traffic volume on fiber optic networks continues to increase, the RWA becomes more important, since adding more fibers and devices is not economically attractive.

Traffic also becomes increasingly heterogeneous and low bit-rate traffic reduces the spectrum utilization efficiency for traditional fixed-grid optical networks. Even with wavelength division multiplexing (WDM) and mixed line rates (MLR), channel resources may still be poorly utilized, i.e., some call requests may be blocked due to unavailable high-quality channels when the network is congested. For the 10-100 GHz

bandwidth on each channel to be efficiently used, lower rate traffic must be combined into higher rate transmissions, a process called traffic grooming.

Grooming allows several calls to share a channel when their data rate requests combined is less than the channel capacity. Therefore, network resources are more efficiently used and the overall throughput is increased. Using currently deployed technology, grooming can only be done in the electrical domain, so regeneration-capable nodes (RNs), where signals undergo regeneration, have the added functionality of providing opportunities for grooming. We present an algorithm that jointly performs the grooming, routing, and wavelength assignment (GRWA) functions for dynamic traffic, where call requests arrive and must be configured in real-time, to improve the overall throughput of the physically impaired translucent optical networks.

In this chapter we consider a heterogeneous network using multiple data rates, depending on the transceiver equipment that has been installed at the nodes. These types of networks have been the subject of several studies [16, 17, 1]. Our work addresses this joint GRWA for dynamic networks with physical impairments.

3.2 Regenerator Placement and GRWA Algorithm

We incorporate the physical impairments (PI) into our GRWA algorithm by imposing a limit on the transmission reach (TR), as was proposed in [1]. Using a heuristic method, they estimate the optimal TR's for different data rate signals: $TR = \infty$ for 10 Gbps calls, $TR = 2500$ km for 40 Gbps calls, and $TR = 2000$ km for 100 Gbps calls. We use their regeneration algorithm, which is optimized for when the TR's are fixed for channel bandwidth and regeneration is performed as needed.

For the sake of consistency in TRs, we use as an example for our technique the same network topology as used in [21], shown in Fig. 3.1. This topology consists of 24 nodes and 43 bidirectional links, each of which is marked in the figure with its physical

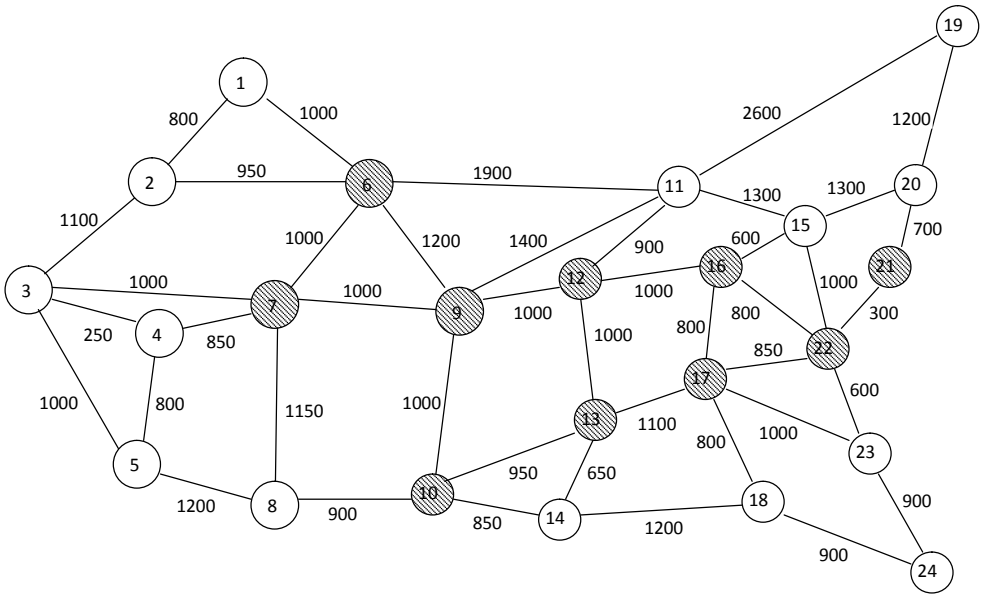

Figure 3.1: NSF-24 network with marked regeneration nodes

length in km. RNs are marked using a shadowed circle. The network contains no other regeneration except for standard periodic amplification and dispersion compensation. OEO resources are available to each wavelength for each port at the RNs.

Due to their high cost, regenerators typically cannot be used at every node. However, a network without any RNs (such a network is called *opaque*) cannot satisfy the QoT requirement for long-distance high-data-rate traffic. We consider a translucent network, where only a selected few nodes have regeneration capability. For these networks two problems exist: the regenerator placement problem (choosing which nodes to assign regenerators) and the regenerator allocation problem (if and which regenerator to use for a call's regeneration needs). In [21], four different regenerator placement algorithms are compared. The most powerful is a heuristic algorithm called *signal quality prediction based* (SQP), which takes the TR as a constraint to search for regenerator locations. We use this algorithm in our work, and, for consistency, also use their QoT constraint (TRs). For example, if we assign 40% of the nodes as RNs, using this algorithm nodes $\{6, 7, 9, 10, 12, 13, 16, 17, 21, 22\}$ in Fig. 3.1 are selected as RNs. To solve the regenerator allocation problem, [22] uses a *distance-hop shortest*

path (DH-SP) for routing, and regenerates calls as needed. In addition to the DH-SP, our work also considers rerouting the call to an alternate path if there is no available regeneration resources along its shortest path. In this case, the algorithm searches for all RNs and reroutes to one for which there is an available wavelength and the TR is satisfied. Rerouting allows calls to be serviced that would otherwise be blocked either by regenerator shortage or PIs.

Grooming must typically be done electrically so that different calls can be multiplexed into (groomed onto) one signal channel and travel (through one or more links) as one call by, for example, time division multiplexing (TDM). They can eventually be de-multiplexed at either another RN or at the receiving node.

The grooming algorithm we propose is as follows:

Step 1. Look for existing calls that share the same source and destination with the new call. Merge the two calls if the free capacity is larger than the data rate request of the new call.

Step 2. If Step 1 fails, look for existing calls that regenerate (or groom) at RNs with enough free capacity (same criteria as in Step 1). If there is one, then adjust the link distance of those links that connect to the RN to $D_{adj} = D * (\frac{\mu^{r_a} - \mu^{r_b}}{\mu^{r_a}})$ to encourage rerouting to the grooming link in the shortest-path algorithm. D_{adj} is the adjusted link distance, D is the original link distance in km, μ^{r_a} is a cost assigned to the lightpath that the new call is being groomed on, μ^{r_b} is the same cost assigned to the lowest-rate lightpath for the new call. The costs capture the cost of interfaces and, based on industry trends, the volume discount (large data rate interfaces cost less per Gbps than low data rate interfaces). We use $\mu = 1$ for 10 Gbps, $\mu = 3$ for 40 Gbps, and $\mu = 5$ for 100 Gbps links. Note that D_{adj} is only a temporary setting (used to calculate the shortest path), and is reset to its original value D for new calls.

Step 3. Calculate the DH-SP for the call. If the DH-SP travels through one of the RNs that was adjusted in Step 2, then grooming occurs on that node.

Step 4. Traffic is groomed on all or part of the lightpath (from an RN to the destination). For the non-groomed part, we look for a RN with available resources, so we can groom on those segments (from RN to RN).

Step 5. After exhausting all possible grooming opportunities, we look for a new wavelength for the not-yet-groomed portion of the route, using a successive first fit (FF) scheme. TR is checked, rerouting the call if the TR constraint is not satisfied.

A call is accepted if it is groomed (partially or entirely) and ungroomed segments are assigned a new wavelength. It is blocked if neither grooming nor a new wavelength is available.

3.3 Numerical Results

In our simulations, we generate 10,000 call requests between each node pair randomly using a uniform distribution. The call requests are Poisson distributed in time. In order to simulate different network applications' requirements for data rates, we model 20% of the calls requesting 10 Gbps, 60% requesting 11 Gbps, and the remaining 20% requesting 60 Gbps. We use the network shown in Fig. 1, with 8 wavelength channels in each fiber. In order to meet different traffic volume needs, two of the eight channels can support 10 Gbps, four channels support 40 Gbps, and the remaining two can carry up to 100 Gbps.

The goodput is defined as the data rate successfully admitted onto the network divided by the data rate requested. In Fig. 3.2, the goodput of the networks is given for both grooming and no-grooming cases, for different loads. As the load increases, the goodput decreases due to the system becoming congested; yet the goodput for

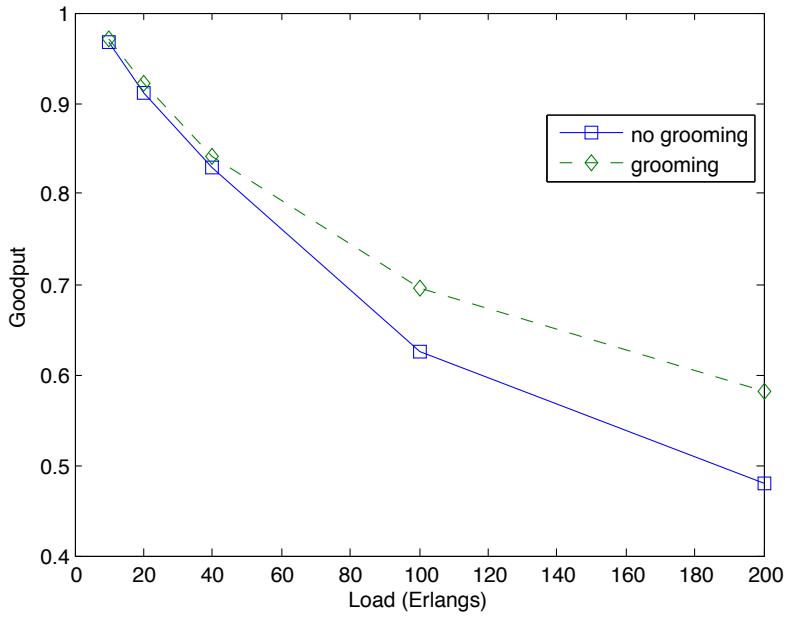

Figure 3.2: Goodput using grooming and no grooming versus load

the grooming case decreases slower than for the no-grooming case. Grooming avoids wasting free channel capacity by allowing new calls access to it. In Fig. 3.3 we show that for any given time slot, the grooming case can fit more calls into the network than the non-grooming case. Instead of simply blocking a new call because all channels are being used, the grooming algorithm looks for available free capacity in used channels, so the blocking probability decreases, as shown in Fig. 3.4.

In Figs. 3.5 and 3.6, we investigate the effect of the number of RNs on goodput and blocking probability, respectively. As the number of RNs increases, the goodput increases because more grooming opportunities are created and regeneration requirements are being satisfied. But marginal improvement decreases when the number of RNs reaches around 10 nodes (10% of the maximum goodput). The remaining nodes not selected as RNs are used by few lightpaths, and would therefore rarely be selected for grooming and regeneration.

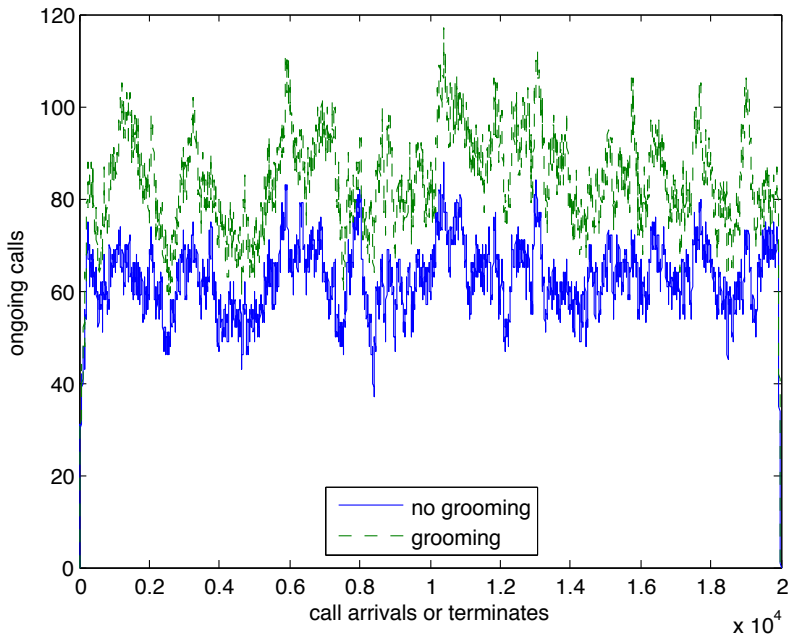

Figure 3.3: Ongoing calls during simulation using grooming

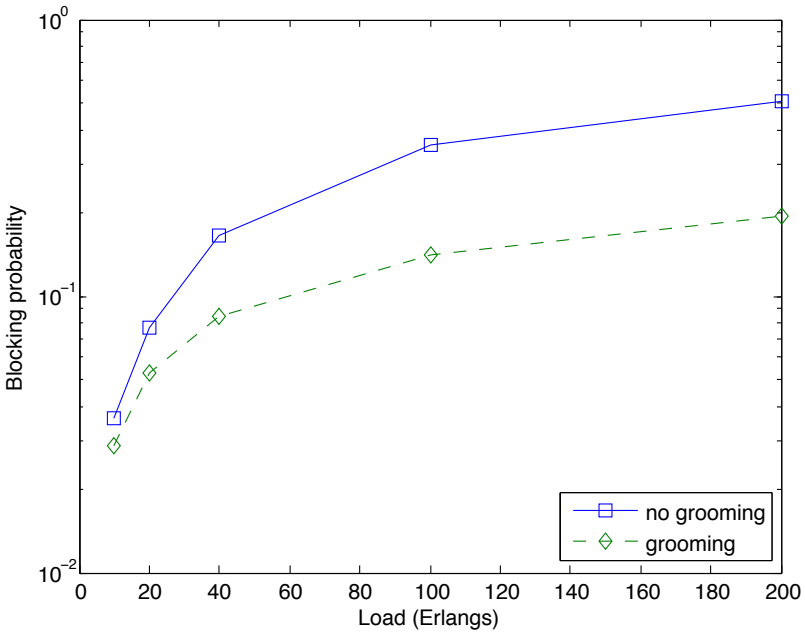

Figure 3.4: Blocking probability using grooming and no-grooming versus load

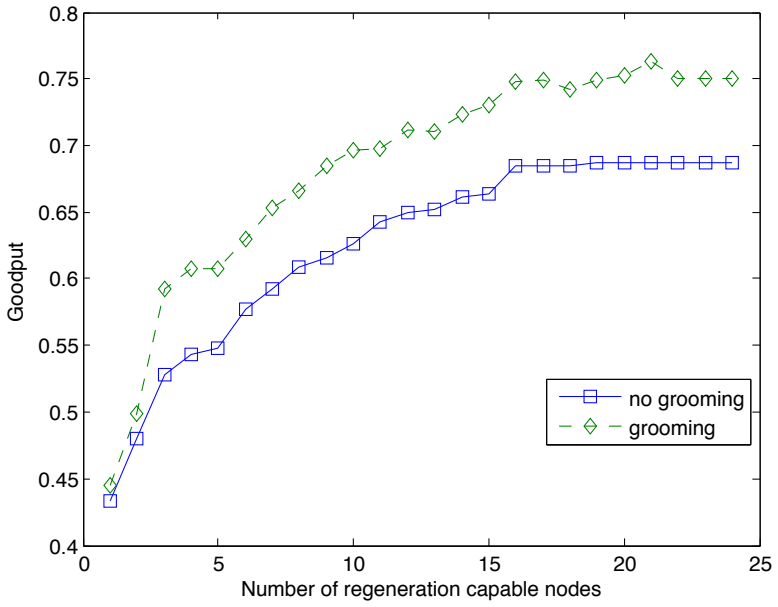

Figure 3.5: The effect of number of regeneration-capable-nodes on the goodput, load is 100 Erlangs

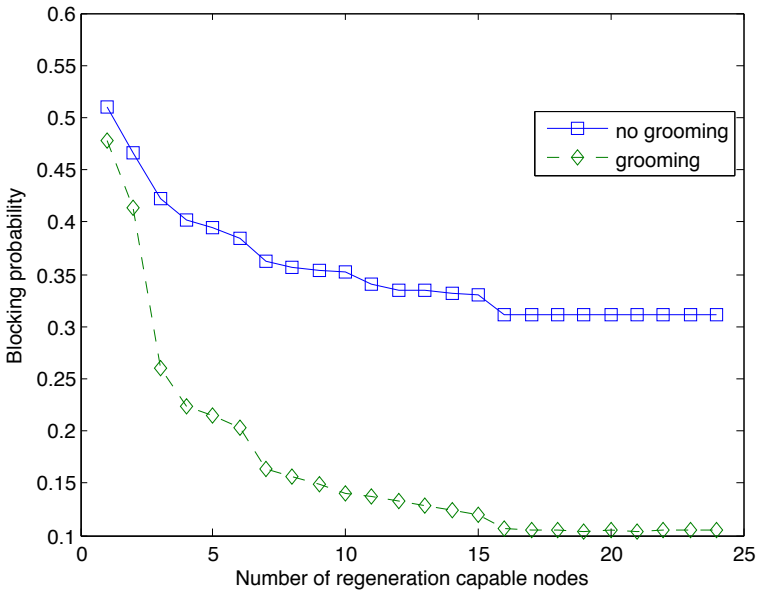

Figure 3.6: The effect of number of regeneration-capable-nodes on the blocking probability, load is 100 Erlangs

3.4 Summary

In this chapter, we propose an algorithm to dynamically make grooming and RWA decisions for real-time traffic so that the goodput (overall throughput) of the network is maximized and the blocking probability is minimized. The benefits of grooming are quantified. The sensitivity of these metrics to the number of regeneration nodes is also discussed. This approach can be used by networks designers to trade-off regenerator cost with network performance.

Chapter 4

Distributed GRWA for Dynamic WDM Networks Using Ant Colony Optimization

4.1 Introduction

In the previous chapter, we propose a centralized GRWA algorithm. Although centralized algorithms often benefit from global information and yield better GRWA solutions, a distributed algorithm may be preferable when information sharing is limited (e.g., multi-domain information segregation). Because distributed algorithms do not require a central node or control message flooding (current popular routing algorithms require network state information disseminated throughout the entire network whenever its state changes), they benefit from quick connection setup and control information isolation. Also, distributed algorithms show improved scalability in case of network expansion or multiple domains. In this chapter, we compare a highly optimized distributed algorithm to a centralized algorithm and show the difference in their blocking performance.

Ant colony optimization (ACO) is a meta-heuristic method inspired by ants' foraging behavior in nature. 'Real' ants are able to find the shortest path between their nest and a food source. This behavior is a fundamental combinatorial optimization problem – optimization based on varying certain parameter(s) to maximize the goodness of the solution. Ant colonies achieve this goal with simple individual agent complexity and limited indirect interaction. Colorni and Dorigo first proposed the ACO algorithm for solving the traveling salesman problem after noticing ant colonies' capability [23]. ACO has been widely applied to solving RWA problems in optical networks. For optical packet-switched networks, Di Caro et al. proposed an algorithm named AntNet [24]. Pedro et al. proposed an ACO algorithm that solves the RWA problem for optical burst-switched networks [25]. Garlick et al. [26] and Ngo et al. [27] implemented an ACO on the dynamic RWA problem, in which path length and wavelength availability are used to measure the goodness of lightpaths. Bhaskaran also solved the dynamic RWA problem using ACO with separate ants during the RWA selection phase to test if the previously collected information is still up-to-date [28]. Kim designed an ACO model to include multiple virtual sub-colonies of ants where each sub-colony holds a different quality-of-service (QoS) requirement [29]. Mapisse et al. created a variation of ACO that launches additional child ants to perform a local search based on the solutions found at that point [30]. Wang et al. implemented an ACO to solve the routing and spectrum assignment (RSA) problem for elastic optical networks [31]. Pavani et al. proposed to extend AntNet in order to solve network restoration problems for WDM networks [32].

To the best of our knowledge, no ACO algorithm has been proposed to handle traffic grooming and physical impairment issues. In Li et al. [33], although they claim their ACO includes traffic grooming, it is not included in the ACO mechanism but only considered when lightpaths are constructed. Physical impairments, which are not considered in the previous literature, are critical in long-haul fiber optic networks since

they affect the quality of transmission (QoT); therefore, they should be included in the ACO as well. In this chapter we show physical impairments and traffic grooming can be incorporated directly in ACO; dynamically varying parameters can adapt to the traffic characteristics, and the optimization of parameters shows the desired balance between performance and control overhead.

In most previous ACO work such as [24, 25, 26, 27, 28], shortest path routing (SP) and adaptive shortest path routing (ASP) are used to compare with ACOs. Since a centralized algorithm that includes physical impairments and traffic grooming does not seem to exist in the literature, we also propose an algorithm that we call the grooming-adaptive shortest path algorithm (GASP). It incorporates the network state (with traffic grooming information) into a dynamic logical topology and optimizes over global information.

In order to deploy ACO, a comprehensive analysis of the effect of various parameters is necessary. Our simulations show insight into the performance of ACO in solving GRWA for fiber optic networks. Although a great amount of research has been done on how to implement ACO in solving the RWA problem for optical networks, not much work has been done on how to configure the ACO system to optimize the process itself. Despite the fact that there are some general guidelines on how different configurations affect ACO behavior, when implementing ACO for a particular task, such as GRWA, the problem needs to be put into that unique perspective to determine how the eventual solution is affected by various configurations. This chapter aims to investigate this relationship in depth.

The chapter is organized as follows: Section 4.2 explains the details of our ACO algorithm and its innovation over previous ACO algorithms. In Section 4.3 we propose the ASP algorithm with traffic grooming. Section 6.4 includes comprehensive simulation results that compare the algorithm with other algorithms and show how different

system configurations affect network performance. Finally, Section 4.5 summarizes the chapter.

4.2 Ant Colony Optimization

Ant colony optimization is a meta-heuristic algorithm that is inspired by ants in nature. An ant colony is seen as being able to find the shortest path between the nest and a food source in food foraging behavior, without either sophisticated individual or central control units. Such behavior also shows high adaptivity to changes in the environment, such as a sudden obstacle along its path. Because of the simple processing of each individual ant, no direct contact and exchange of information between them is supported. Instead, ants communicate with each other indirectly by changing the environment using pheromones (such behavior is called "Stigmergy"). Ants sense the pheromones left by other ants and make their movement decisions based on this information. In algorithm design, this movement is modeled as a probabilistic behavior [23].

In communication systems, agents mimicking the ants can be used to survey network state information, which is then used for GRWA decisions. The simplicity of ACO agents along with their simple method of communication reduces the complexity of the control and management mechanism compared with centralized algorithms and is therefore preferable for designing a transport-sized network GRWA algorithm. Throughout network operation, ants travel through the network between neighboring nodes. At each node, the ants leave a numeric marker mimicking the pheromone that indicates the quality of its last move, such as the physical length, free spectrum, available regeneration resources, etc. The pheromone then affects the movement of ants that arrive in the future.

The pheromone levels are represented by $\tau_{i,j}^{(k)}$ where, k is the node that the ant is currently located at, i is a neighboring node of k, and j is the ant's destination node. The value of $\tau_{i,j}^{(k)}$ represents the goodness of choosing link (k, i) on its path to destination j. In Table 4.1, a pheromone table is shown for node 8 in Fig. 5.2. The column indices refer to the destination of the path from node 8, and the row indices are the neighboring nodes of node 8. Since the ant's movement follows a probabilistic behavior, for each destination the probability of choosing neighboring nodes should sum to one, which means the sum of each column in the pheromone table is one.

Table 4.1: Pheromone Table at Node 8 of Fig. 3.1.

	1	2	...	24
5	$\tau_{5,1}^{(8)}$	$\tau_{5,2}^{(8)}$	...	$\tau_{5,24}^{(8)}$
7	$\tau_{7,1}^{(8)}$	$\tau_{7,2}^{(8)}$	...	$\tau_{7,24}^{(8)}$
10	$\tau_{10,1}^{(8)}$	$\tau_{10,2}^{(8)}$	...	$\tau_{10,24}^{(8)}$

4.2.1 General ACO Approach to RWA

We begin by describing a general ACO used for RWA [28, 34], which does not consider traffic grooming and physical impairments, and then describe our enhanced design that takes these factors into consideration.

In ACO, ants are sent from each node to each destination simultaneously. The ants follow different paths by making a statistical decision from one node to the next. They collect network state information, including wavelength usage, grooming opportunities, etc., along the way. Using this network state information to calculate the 'goodness' of the solution lightpath they find when they arrive at the destination, they then travel back to their sources and update the pheromone levels according to this 'goodness'. Pheromone levels at each node also evaporate after all ants finish

their round-trip. This multi-ant round-trip process is considered as one ACO launch cycle and is performed many times before an optimal solution emerges.

At the start of the algorithm, an initial pheromone level τ_0 is given to each node as $\tau_0 = 1/(ND)$, where N is the total number of the nodes and D is the diameter of the network (the length of the longest shortest-path among all node pairs). This initial pheromone is chosen because the new pheromone laid by ants is based on the physical distance of the lightpaths. At time t an ant heading for node j that arrives at node k makes its next move decision using

$$p_{ij} = \tau_{ij}^{(k)}(t) \tag{4.1}$$

as the probability that the ant picks node i for its next move. Prior research has included instantaneous wavelength usage information at node k to help the ant decide on its next move, mimicking ants' limited visibility [23, 33]:

$$p_{ij} = (1-\beta)\tau_{ij}^{(k)}(t) + \beta\omega_{k,i} \tag{4.2}$$

where, $\omega_{k,i}$ is the fraction of wavelengths that are free on link (k, i), and β is a parameter used to weigh the importance of wavelength usage information.

When an ant reaches its destination, it then traverses backwards following the reverse route back to its source. As the ants return home, they update the pheromone table at each node as (subscripts representing the nodes are omitted here):

$$\tau(t+1) := \rho\tau(t) + (1-\rho)\Delta\tau \tag{4.3}$$

where

$$\Delta\tau = \frac{1}{L}(1+\alpha\omega) \tag{4.4}$$

is the pheromone modification made by ants, L is the distance of the route in 100 km, ω is the percentage of free wavelengths on that route, and (ρ, α) are design parameters that can be tuned to optimize the system performance. Equation (4.4) captures the main metrics of "goodness" of a lightpath, which are the physical distance L of the lightpath and the wavelength availability ω, and shows that a lightpath becomes more preferable if it has a shorter length and more free wavelengths compared to others. If a uniform link length is assumed, using L is equivalent to using the number of hops in the lightpath. In our study, the physical distance has greater importance because we consider physical impairments, which depend on the distance that the optical signal traverses. In WDM networks, wavelength availability measures the congestion level of a path.

After all ants finish their trip back to their source nodes and perform (4.3), the pheromones evaporate at all nodes according to:

$$\tau(t+1) := (1-\rho)\tau(t) + \rho\tau_0 \tag{4.5}$$

to avoid stagnation. Equation (4.3) is called the global pheromone update and is carried out at only those nodes along the ants' paths. It has the effect of reinforcing the pheromone for each successful path for future ants to follow. Equation (4.5) is called the local pheromone update and is carried out on all nodes. It has the effect of lowering the pheromone towards the initial level so that ants have a chance to explore new paths. These two pheromone update processes combined with the exploration of ants compose one cycle of the ACO algorithm. It takes a number of cycles for the ants to cover the entire network and find the best solution.

In addition to updating the pheromone levels at the nodes, each surviving ant also stores its successful path at its source node and all intermediate nodes to form a candidate route list. The order of this list is based on $\Delta\tau$ in (4.4), which depends

on the distance and free wavelengths on the route. As the network state changes this list is updated continuously by the ants.

With the general structure of the ACO algorithm established, we modify the pheromone update mechanism to consider physical impairments and traffic grooming next. The performance of ACO is also improved by modifying the ants' behavior.

4.2.2 Proposed Enhancements

Based on the general ACO structure described above, we extend the algorithm to fit our problem requirements and increase its efficiency. We make the following changes to the general structure:

Mixed line rates

In our MLR network model, each wavelength is assigned a different fixed line-rate (10, 40, or 100 Gbps). This fact is addressed by having three sub-colonies of ants, with each sub-colony representing one line-rate. This also means that we have three different pheromone tables, one for each line-rate, and eventually three different candidate route lists. Ants belonging to different sub-colonies are launched separately, and they do not affect and are not affected by other sub-colonies.

Physical impairments

Ants representing different line-rates enforce different TR requirements. Each ant keeps a record of the distance it has traveled since its last regeneration, and chooses only among those neighbors that do not exceed its TR for its next move. At intermediate nodes that are equipped with 3R regenerators, the ants by default consider themselves regenerated, and therefore their distance traveled is reset. This assumption is reasonable as OEO converters are usually available; we showed in Chapter 7 that for a network with a similar scale as in Fig. 3.1, the number of regenerators needed is reasonably low.

Random walk

To further increase the randomness of the ants' movement to avoid pheromone stagnation, we allow ants to make a uniform selection for its next hop from neighboring nodes with a given probability. The probability that an ant chooses any neighbor node i for next move at node k becomes:

$$p_{ij} = (1 - r)[((1 - \beta)\tau_{ij}^{(k)}(t) + \beta\omega] + \frac{r}{N_k} \tag{4.6}$$

instead of (4.2), where N_k is the number of nodes neighboring k without counting the node the ant just visited, and r is a parameter denoting the probability that an ant makes a random walk.

Backtracking

During its exploration, an ant may reach a dead-end such that all choices for the next move have already been visited, there are no available wavelengths towards any neighbor, or no next move can satisfy the TR requirement. In such situations, we can either simply remove the ant from the network or allow it to backtrack to its previously visited node and choose a different path. In [25] no backtracking is allowed. In [32] a random next hop is selected when a dead end is formed by a loop. In our study, we consider the launching of ants as a costly operation since it introduces control overhead, and try to make each launched ant successfully reach its destination. Therefore we encourage ants to search for a viable solution by introducing backtracking, a method similar to [32] but with some critical differences. Whenever an ant encounters a dead end, it backtracks to the previously visited node and makes a new next-hop decision without considering the node that it moved back from. This process repeats until a viable path is found or all options are exhausted. In the latter case, it then retreats one step further to the next upstream node, until it either finds a viable solution, moves back to its source node, or surpasses its time-to-live (TTL); in

the latter two cases, the ant is removed. Note that pheromone values are considered even after backtracking, which shows an effort to use long-term information as much as possible. By doing this, we reduce the number of ants being wasted and improve our algorithm's efficiency. It may be argued that backtracking may lead to 'bad' solutions with unnecessarily long paths. We show that this does not happen because the pheromone update depends on the goodness of the path, so long paths do not greatly affect pheromone tables.

Traffic grooming

When ongoing traffic is being transmitted in the network, resource usage information, such as wavelength and routing assignment and connection duration, is recorded at each node that is on the lightpath. Each node keeps a list of ongoing traffic that includes connections initiated at that node and transient traffic. Since this information is essential to traffic grooming, ants can consider this information when exploring the network.

If the ant is at a regeneration node, grooming opportunities can be noted and picked up by the ant through the following steps:

Step 1 : Before performing (4.6), the ant first checks for destination grooming opportunities (i.e., an existing lightpath that has the same destination and enough free capacity). The segment that leads the ant directly to its destination is always chosen for its next move, which is effectively destination grooming.

Step 2 : In order to introduce randomness for exploration purposes, ants make a random choice to either perform this step or not at each regeneration node. If step 1 fails, the ant randomly chooses a lightpath segment that leads to a regeneration node which is not its destination for grooming opportunities. If the available capacity of the chosen segment fits the connection request's bit-rate requirement and no nodes along that segment have been visited by the ant (to

avoid loops, which cause a waste of physical resources by re-visiting the same node), it picks that segment as its next move.

Step 3 : If no grooming opportunity is found, the ant looks for a non-grooming move according to the probability assignment in (4.6).

Since traffic grooming does not assign new resources to the newly groomed connection, it is highly desirable when performing GRWA. Therefore, the ability of grooming new traffic with our existing connections should count towards the goodness of a lightpath. Recall in (4.4) the goodness is measured based on the lightpath length and the number of free wavelengths on it. In order to include grooming into our calculation of goodness, we add two modifications to the calculation of how much pheromone to deposit: 1) a grooming segment is considered as having all wavelengths free (i.e., $\omega = 1$), due to the capability of wavelength conversion, and 2) a grooming segment is considered to have zero length, i.e., $L = 0$ (if the whole lightpath is a grooming opportunity, its length is set to be the length of its shortest link). Based on (4.4) these two modifications greatly encourage the selection of lightpaths with grooming opportunities.

4.2.3 Algorithms

To summarize our ACO algorithm we show the pseudo-code, which consists of two parts. Algorithm 1 shows an ant foraging cycle, made up of a lightpath exploration stage followed by a pheromone update stage. In the first stage, ants follow the rules described above to explore the surrounding network state; in the second stage, ants travel back to their source nodes, update pheromone levels along the way based on the lightpaths' goodness measure, and store route information in the GRWA candidate lists. Foraging is performed repeatedly with an ant launch rate that depends on the traffic load. Since the ant foraging process takes a short time compared to the interval

between connection request arrivals, during this process the network state is assumed to be unchanged.

Algorithm 2 shows what happens when a connection request arrives. A GRWA decision for the connection request is made based on the network state information provided by the ACO in the previous section. The grooming and routing approach uses the candidate list provided by the ACO. The wavelength assignment employed is the first fit algorithm (FF), where the lowest indexed wavelength available is tried first. Lower line-rate channels are given lower wavelength indices, which is important because assigning an unused channel of a high line-rate to to low data-rate connection is both a waste of resources and introduces unnecessary TR constraints. When a connection request arrives, the source node picks the route from the top (highest $\Delta\tau$) of the candidate route list created by the ants and inquires from the nodes along that path whether physical resources (wavelength and OEO circuits) are still available. If so the lightpath is reserved and the connection request is accepted. If the first solution is not available at that time for any wavelength, the next route is checked. If a route is found to be unavailable, it is removed from the list so that the list can be repopulated by future ant foraging cycles. If a connection chooses a route that is marked as containing grooming opportunities, it tries grooming first. If the grooming opportunities no longer exist, it then tries a non-grooming solution. If a non-grooming route is chosen, then wavelength assignment proceeds, favoring the lower-indexed (lower line-rate) channels first.

4.3 Grooming Adaptive Shortest Path Algorithm

Since we have not seen a dynamic centralized algorithm that incorporates regeneration, physical impairments and traffic grooming published in the literature, in this section we propose one that we call the grooming adaptive shortest path algorithm

Algorithm 1: Ant Foraging Cycle;

for *each line-rate wavelength (total of 3, i.e., 10/40/100 Gbps) of each node pair in network* (s, d) **do**
  Launch one ant from s to d with probability P_L;
end
for *each ant (in parallel)* **do**
  while *current node* $i \neq d$ *and* $TTL > 0$ **do**
    if *found a next hop* **then**
      Move to next hop and decrease TTL;
    else
      Backtrack to its previous visited node;
    end
    if $TTL=0$ *or* $i = s$ **then**
      Kill ant, exit loop;
    end
  end
  After $i = d$, travel back to s
  while *current node* $i \neq s$ **do**
    Follow the same route back;
    Update the pheromone at node i using (4.3);
    Store sub-route information at i;
  end
  Update the pheromone at node s using (4.3) and store route information at node s.
end
for *each line-rate wavelength at each node* **do**
  Evaporate pheromone using (4.5);
end

(GASP). This centralized algorithm requires complete current network state information, so must be computed on-line as call requests arrive.

In the Dijkstra shortest path (SP) routing algorithm [35], a shortest path tree is created for each node towards all other nodes in the network. The branches of the tree are the routes connecting the node pairs. The SP branches are created by following the rule that only the shortest route is kept, and thus the shortest path between all node pairs can be found. When used in a dynamically changing topology, the shortest path routing algorithm becomes the adaptive shortest path (ASP) algorithm.

Unlike the traditional ASP algorithms, where one shortest path is calculated for each wavelength based on the current network state and then a shortest path among

```
Algorithm 2: GRWA;
for each demand request do
    while no lightpath found do
        Select route stored at the top of the candidate list;
        if resources available on chosen route then
            if grooming lightpath is chosen then
                groom traffic if grooming opportunity still available
            else
                Select the first available wavelength;
                Set up a lightpath;
            end
        else
            Remove route from candidate list;
        end
    end
    if all routes have been tried then
        Consider the request blocked;
    end
end
```

all wavelengths is used, our GASP algorithm considers all wavelength at once in order to account for wavelength conversion functionality (a side effect of regeneration). For dynamic systems, not all wavelengths are available, and grooming opportunities need to be represented. We modify the Dijkstra algorithm by introducing additional checks for wavelength continuity, as well as TR and grooming opportunities. In each step of the Dijkstra algorithm, the distance between a leaf node and the root is only updated if all three criteria are met for the new route:

1: It has a shorter distance compared to the previous one.

2: It has at least one free wavelength (including possibly a grooming wavelength).

3: It does not violate the TR constraint.

The first criterion is the same as the traditional SP except that the portions of the lightpath groomed onto existing traffic are assigned zero length; the second is implemented by having each node keep a record of the current wavelength mask with respect to the root. When a new route is about to be created, the mask of the previous

branch node and mask for the link that connects to it are used to make sure there exists an available wavelength. Once a regeneration node is selected, the mask is reset. Finally the third criterion is implemented by having each node keep a record of the current TR distance since the last regeneration. The TR distance is the length of the link that connects to this node plus the TR distance of the previous branch node. The TR distance is the path distance over a transparent segment (no regeneration), so it is reset to zero once a regeneration node is traversed. Only transparent segments with TR distance within the TR limit are selected.

After a route is selected using GASP, wavelength selection for each transparent segment follows the FF rule. Note that the wavelength could vary along the route depending on the wavelength conversion operation at regeneration nodes.

The static shortest path (SP) and fixed-alternate (FA) algorithms [36] are often used to compare with newly proposed ACO algorithms [24, 25, 26, 27, 28]. Since no existing dynamic algorithm includes TR constraints and traffic grooming, we also provide a comparison with these two algorithms. In SP, the shortest path (in number of hops) between each node pair is always used for traffic between them; in FA, while using the shortest path as the first choice, a second path (usually link-disjoint) is used as the alternative choice if the shortest path is unavailable because it is fully occupied by existing traffic. Although we consider regeneration and traffic grooming in both algorithms, we show below that they underperform against the two proposed algorithms due their lack of flexibility compared to adaptive routing.

4.4 Numerical Results

We test our ACO algorithm using the NSF-24 network shown in Fig. 3.1. There are 24 nodes; links are considered as a pair of uni-directional fibers. Each fiber carries 32 wavelengths (eight 10 Gbps, sixteen 40 Gbps, and eight 100 Gbps channels).

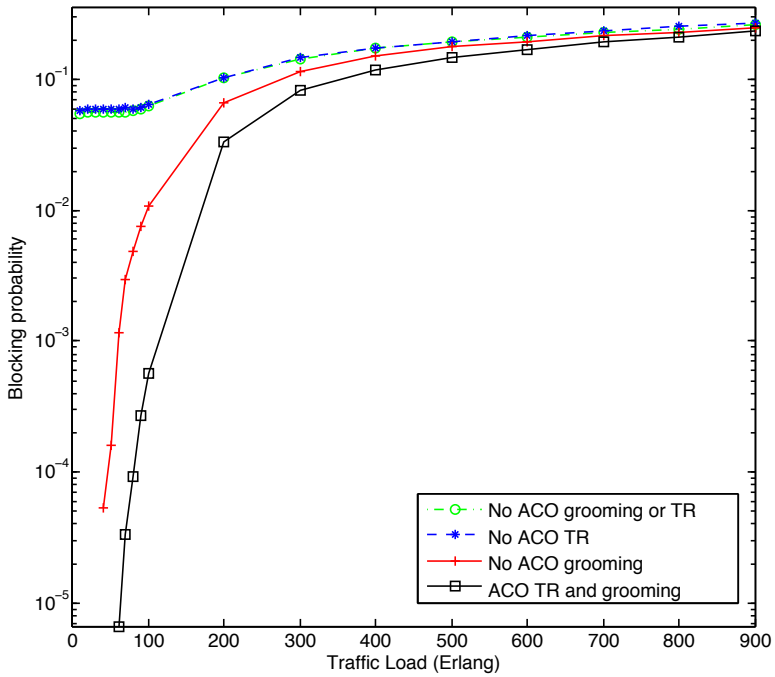

Figure 4.1: Blocking probability of ACO with implementation of traffic grooming and TR ($N_R = 3$, $N_C = 1$, $P_L = 1$, $T_L = 30$, $\rho = 0.7$).

The 10 shaded nodes in the topology are chosen as regeneration nodes following the algorithm in [21]. During our simulation, connection arrivals are modeled as a Poisson process. The bit-rate of the requests ranges from 1 to 60 Gbps uniformly. The connection duration is exponentially distributed with an average of one arbitrary time unit ($h = 1$). The traffic load (E, in Erlang) then is adjusted by changing the connection request arrival rate, λ, with the formula $E = \lambda h$. In this section we first show the dependence of the ACO performance on various parameters to optimize it for our topology, then compare it with the centralized GASP algorithm.

4.4.1 Optimizing ACO configuration

In order to implement ACO with a certain confidence, we should understand its many configurations and how they affect the overall system performance collectively and separately. We perform tests to show the impacts of several aspects of the design on system performance.

In this chapter our main goal is to include new mechanisms such as a TR limit and traffic grooming into the design of ACO. TR limits guarantee the QoT for all node pairs and can be enforced by allowing the signal to be regenerated along the chosen lightpath whenever necessary. However, this also requires the candidate lightpaths provided by ACO to include the necessary regeneration nodes. Traffic grooming improves the physical resource usage efficiency but also requires the candidate lightpaths to include such opportunities.

In Fig. 4.1 we see the necessity of implementing new mechanisms into ACO by comparing our algorithm with previously proposed ACO approaches. When ACO does not check for TR and traffic grooming in the foraging cycle, the figure shows clearly that although the system still tries to find regeneration sites and traffic grooming opportunities along candidate lightpaths in the GRWA step, without such information considered in the ants' exploration, ACO cannot provide good candidate lightpaths to the GRWA to deploy these mechanisms. Note that the TR limit has a much greater impact on network performance compared to traffic grooming (the curves representing the cases where ACO ignores the TR and where it ignores both the TR and grooming are indistinguishable). Without knowledge of the TR, candidate lightpaths are often those with shortest end-to-end physical distance and many will fail the TR requirement, thereby blocking requests. Due to this effect, certain node pairs appear to be unreachable and cause a non-negligible blocking probability even when the load is low and wavelength blocking is rare (shown as the flat region of the graph when the load is less than 100 Erlangs). Even though traffic grooming opportunities are searched in all cases, candidate lightpaths with traffic grooming information tend to provide more chances of enabling traffic grooming, as shown from the improvement achieved between not checking traffic grooming and checking traffic grooming. If grooming is not considered by the ACO, those lightpaths often do not make the candidate list because they tend to be longer.

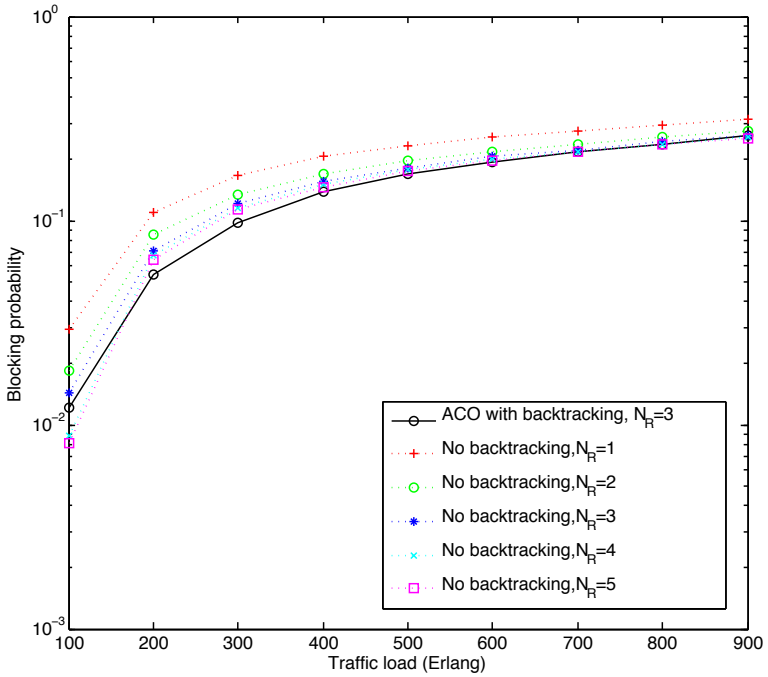

Figure 4.2: Blocking probability of ACO with backtracking ($N_R = 3$) and without backtracking (N_R = 1-5).

As mentioned in Section 4.2, backtracking may be considered undesirable in that it can lead to poor solutions. We argue that such poor solutions, if poor in quality at all, do not affect the system performance because the pheromone update rule is based on the goodness of solutions. We also argue that by allowing backtracking, we can keep the number of ants low, thereby reducing the control overhead (which is often measured as the ratio of control and management message exchanges to real traffic) thanks to the effort to help each ant find a solution. In Fig. 4.2 we show the comparison between algorithms with backtracking and without. In this figure we also show the effect of having different numbers of candidate routes (N_R) in the solution pool on each source node. If no backtracking is allowed, by increasing the numbers of candidate routes, we increase our options for connection requests, and therefore lower the chances of having the requests blocked. We can see that due to the improved efficiency of having ants backtrack, the blocking performance is improved, even with a smaller solution pool.

Algorithm 1 in Section 4.2 is carried out multiple times every fixed time interval, denoted as T. The number of cycles (N_C) determines the level of convergence to an optimal solution. The duration of T measures how frequently the network state information is collected and determines how up-to-date the information is. Since for a given traffic load the average number of connections that arrive in T is fixed to be ET/h, we establish this average number of connections as the relative launch repetition interval (T_L, e.g., when $E = 100, h = 1$, $T_L = 2$ leads to $T = 0.02$ time unit, which means on average, two connections arrive before the next launch; Given fixed traffic load, by adjusting the average of number of connection arrivals, we can adjust the interval between two launches, therefore reflecting the launching frequency). The launch probability (P_L) determines the number of ants launched per cycle. These three parameters together determine the number of ants in the network, which is a measure of control overhead of ACO. The average number of ants per source node is calculated using:

$$\text{Average number of ants} = N_C \times P_L \times \frac{E}{T_L h} \tag{4.7}$$

In Fig. 4.3 we show the blocking probability with different N_C for each ACO launch, T_L, as well as P_L, while keeping the total number of ants launched fixed. Intuitively, if the traffic load is high, the network state changes rapidly and requires ants to be launched more frequently to keep the information up-to-date. When we use many cycles, we allow pheromones to converge to an optimal solution. T_L and P_L show the tradeoff between performance and control overhead.

As can be seen in Fig. 4.3, in general, the blocking probability decreases with more ants in the system. Also, if we set two of the three parameters (N_C, P_L, T_L) fixed (shown as connected lines in the figure), increasing the number of ants by changing only the third parameter, we can lower the blocking probability. This shows that no matter how one chooses to do it, increasing the number of ants decreases the blocking

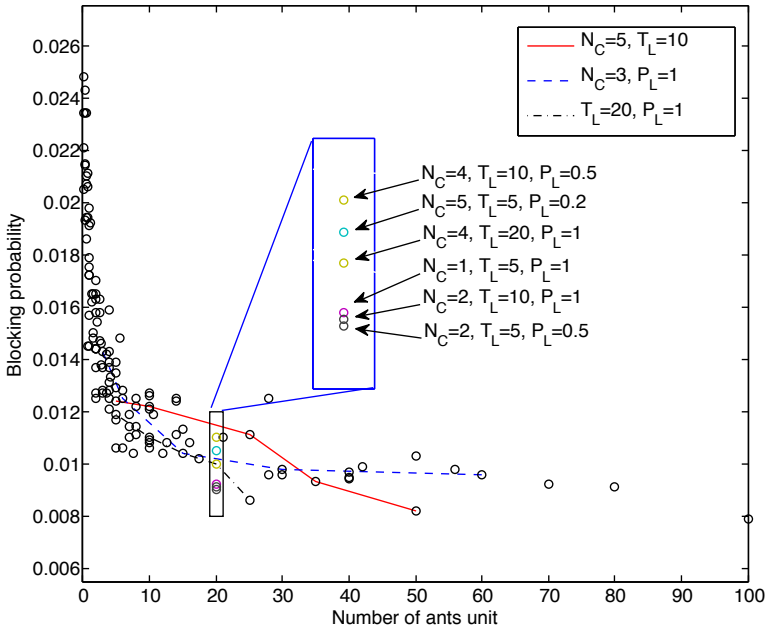

Figure 4.3: Blocking probability of ACO with different ants behavior parameters: number of cycles N_C, launch interval T_L and launch probability P_L ($E = 100$ Erlangs, $h = 1$ time unit).

probability. We also show in the inset that for a fixed number of ants, foraging can be organized differently to achieve different performance. It suggests ACO is more sensitive to launch interval and launch probability than to the number of cycles.

Since the ants foraging step is a process of exploring possible solutions, in order to avoid the emergence of sub-optimal (i.e., locally optimal) solutions, certain randomness is preferred. We test the ants survey with different levels of randomness, as measured by the probability of ant choosing uniformly among neighboring nodes from an intermediate node (parameter r) rather than following the pheromones. Note that this parameter only affects the ants' movement but does not affect the pheromone update or the relationship between the pheromones and the GRWA decision. In Fig. 4.4 we show the blocking probability of ACO with different random walk probabilities. We test different launch probabilities to have different numbers of ants in the system. We notice that a fairly large level of randomness is preferable since it reduces the

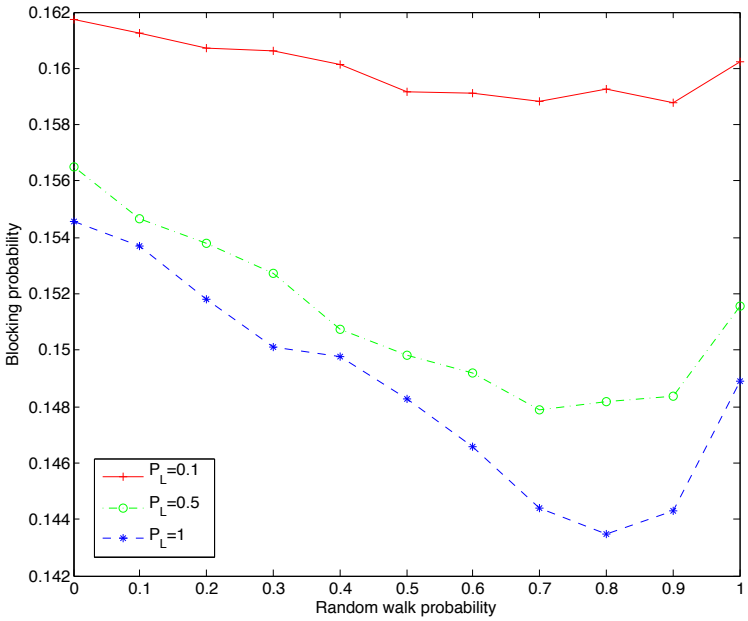

Figure 4.4: Blocking probability of ACO with different random walk probabilities r for various ant launch probabilities P_L ($N_R = 3$, $N_C = 1$, $E = 400$ Erlangs)

blocking probability. For example, for $P_L = 1$, the blocking probability is minimized for $r = 0.8$. A total random movement ($r = 1$) leads to worse performance.

In the previous literature, there are two other different implementations of ACO, one assuming a single ant per cycle [25, 26, 27], and the other using multiple ants per cycle per source node [23]. Both implementations can have the same control overhead in terms of total number of ants in the network; for example, one cycle of ACO with three ants per cycle uses the same number of ants as three cycles of ACO with one ant per cycle. We compare the two implementations with the same number of ants in the network in Fig. 4.5 where we show the blocking probability of the network. There is no discernible difference between the implementations. This is because we model our ACO to have no network state changes between cycles for each launch. It also shows that $N_C = 1$ with a single ant per cycle is enough to minimize the blocking probability, especially at high loads.

The performance of ACO is also affected by the rules used for updating the pheromone levels. In (4.3) and (4.5) we give global and local pheromone update

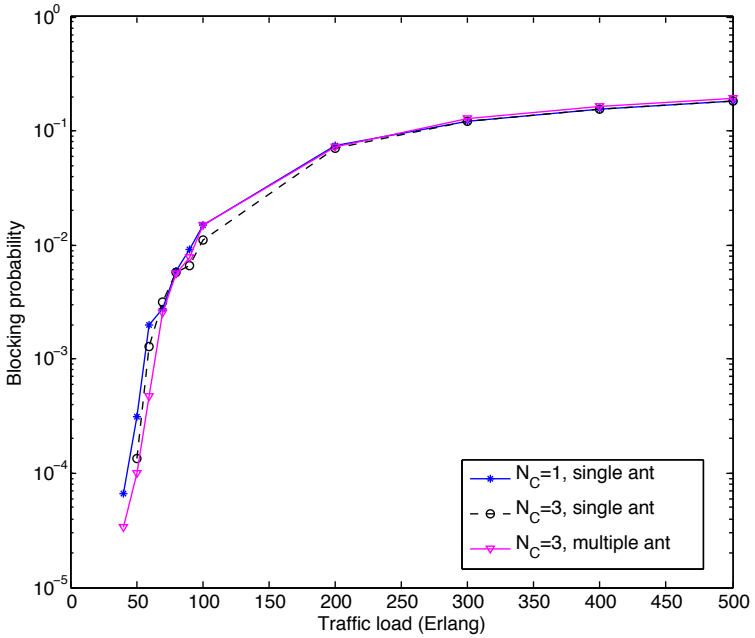

Figure 4.5: Blocking probability of ACO with different configurations: 1 ant per node pair with $N_C = 1$ and $N_C = 3$; 3 ants per cycle and $N_C = 1$ ($N_R = 3$, $P_L = 1$, $T_L = 20$, $\rho = 0.7$)

rules, which are both affected by the parameter ρ. Intuitively, if ρ is large, the pheromone changes slowly; when $\rho = 1$ the pheromone does not update at all. If ρ is small, the pheromone changes drastically, and when $\rho = 0$ the pheromone is always replaced by the newly deposited amount. In Fig. 4.6 we show the dependence of the ACO performance on the parameter ρ. Note that for different numbers of ants in the system, as determined by the launch probability P_L, the best choice for ρ is different: for $P_L = 1$, the best ρ is 0.8, for $P_L = 0.1$, the best ρ is 0.1. We also notice that for some ρ, launching more ants does not guarantee better performance. This is due to a rapid convergence to a suboptimal solution.

4.4.2 Comparison to Centralized Algorithms

After optimizing our ACO over the algorithm's parameters, we compare it with the proposed GASP algorithm, which is a centralized optimization algorithm. In Fig. 4.7 we show that GASP benefits from more network state information and provides

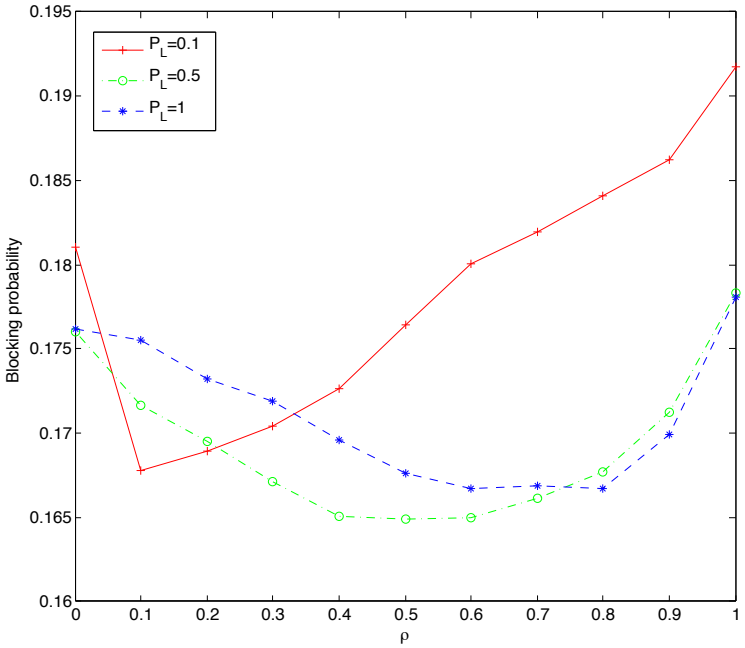

Figure 4.6: Blocking probability of ACO vs. parameter ρ (P_L =[0.1,0.5,1], $N_R = 3$, $N_C = 1$, $T_L = 30$ $E = 400$ Erlangs)

better GRWA decisions, which causes its blocking probability to be lower than that of ACO. The ACO's blocking probability is higher, which shows that with limited information, the ACO provides worse blocking performance. However, the tradeoff of blocking performance and complexity of control mechanism should be considered by network providers. We also show the blocking probabilities of static shortest path (SP) routing and static fixed alternate (FA) routing (with one alternate route) for comparison. The paths in the SP and FA algorithms do not necessarily naturally satisfy the TR constraints. Therefore, even when regeneration is performed whenever possible, some node pairs are still unreachable due to the limited routing choices. When global state information is not available (such as in multi-domain networks) when GASP cannot be used, ACO should be used instead of SP or FA routing.

In Fig. 4.8, we test both algorithms with varying traffic load, which simulates the scenario of a changing traffic pattern throughout the day. The figure shows ACO can adapt to the traffic load change quickly even without a strategy specifically

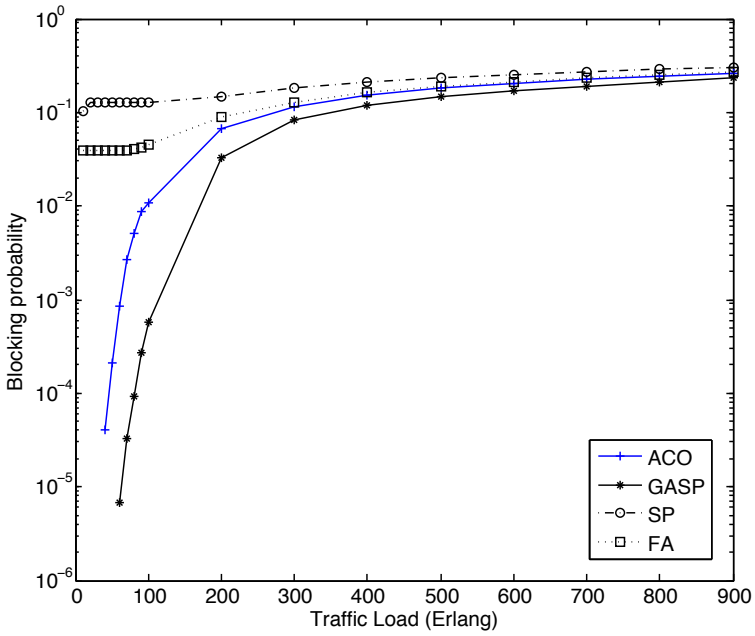

Figure 4.7: Blocking probability of ACO vs. GASP and static SP, static FA ($N_R = 3$, $N_C = 1$, $P_L = 0.5$, $T_L = 10$, $\rho = 0.5$)

defined for such a change. Again, GASP, relying on complete and accurate instantly-updated global information, shows larger capacity and equally fast adaptivity to traffic changes.

The robustness of ACO in case of failure is also an important criterion. In Fig. 4.9 we show a scenario in which link $(7, 9)$ in the network topology of Fig. 3.1 is cut during operation. This fiber cut affects connections in both directions. The affected connections are fed back to the system in the incoming queue and assigned possible alternative routes. As we can see from Fig. 4.9, at the moment of the fiber cut (time 50) the number of ongoing connections drops abruptly due to the removal of the affected connections. Then the inset graph details the recovery of the network, including the reassignment of affected connections and new connections. The ACO is able to adapt to the failure immediately without a separate restoration mechanism. Fig. 4.9 also shows several different configurations of ACO, such as different launch probabilities and whether to restrict candidate routes to be link-disjoint. It shows

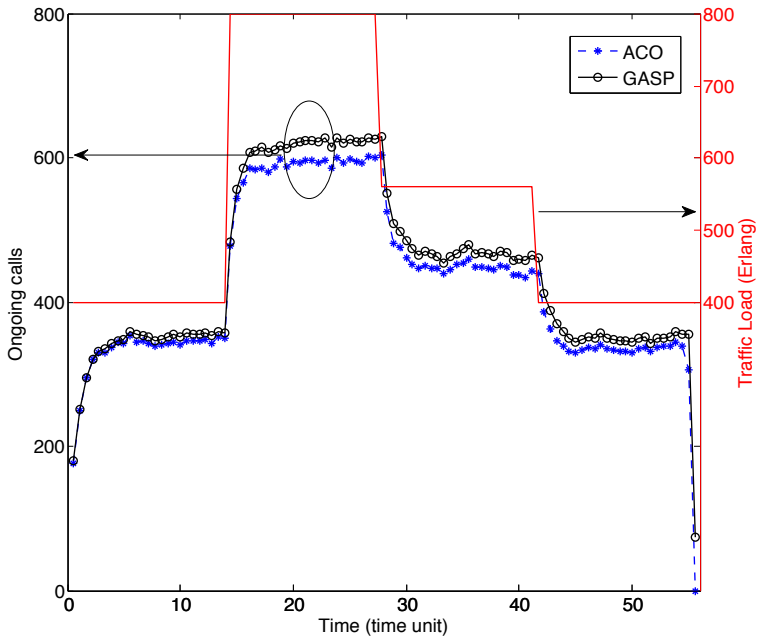

Figure 4.8: One realization of real-time ongoing connections using ACO vs. GASP ($N_R = 3$, $N_C = 1$, $P_L = 0.5$, $T_L = 10$, $\rho = 0.5$, averaged over 60 trials)

again that with higher launch probability, the number of ongoing connections is larger and recovery is faster due to more ants in the system. It also shows that with a link-disjoint candidate route pool, there is a better chance for ACO to find an alternative route that is not affected by the fiber cut, and it therefore recovers faster.

Finally we evaluate the computational complexity and the control overhead of the ACO algorithm by comparing with the proposed GASP algorithm. Suppose network state information is collected at each node using ants survey for ACO and link state advertisement (LSA) for GASP. Upon connection request arrival, each source node needs perform GRWA decision based on certain algorithm. For GASP the complexity is similar to traditional Dijkstra shortest path algorithm which is derived as $O(L + N \log N)$ [37], where L is the number of links and N is the number of nodes. For ACO, however, the solution set is already populated and source node simply needs to look up this table. Control overhead of an algorithm is defined as the bandwidth occupied by control messages for GRWA purposes. Control messages can

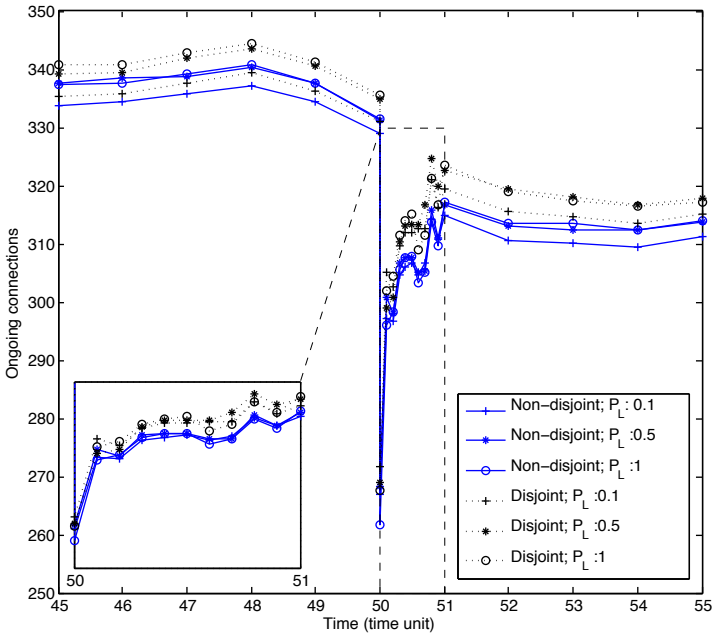

Figure 4.9: Ongoing traffic when fiber cut occurs ($N_R = 3$, $N_C = 1$, $P_L = 0.5$, $T_L = 10$, $\rho = 0.5$, average over 60 trials)

be separated into two categories: routing messages and connection setup messages. Since both algorithms use a similar connection setup messaging mechanism, we do not consider its impact here. Assume link state advertisement (LSA) flooding is used to update routing information for GASP. For the whole network status to be updated once, it takes on the order of N^2 hops. The time interval between exchanging messages is T_F. The ACO routing message depends on the time the ants spend on surveying the network, the ant life-time h. The ratio of control overhead is:

$$\Theta = \frac{\langle h \rangle P_L N(N-1)}{N^2} \frac{\text{Packet Size}_{ACO} N_C}{\text{Packet Size}_{LSA}} \frac{T_F}{T_L} \tag{4.8}$$

where $\langle\rangle$ means average value. The size of LSA packet is around $\langle\delta\rangle W$ excluding packet header, where $\langle\delta\rangle$ is the average node degree, which is 3.5 for the NSF-24 network; the packet size of ACO packet is larger, around $\langle h \rangle W$. So the ratio can be expected to be approximately $\frac{\langle h \rangle^2 P_L N_C T_F}{\langle\delta\rangle T_L}$ when the number of nodes is large. We show that the ratio of control overhead between ACO and GASP depends on the relative

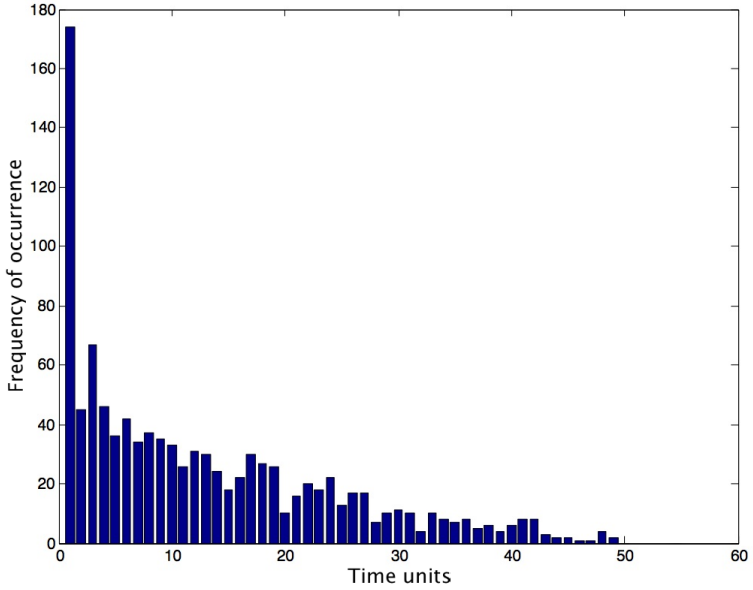

Figure 4.10: Life time of ants, average trip length ≈ 10

launch interval (T_L/T_F), number of cycles per launch (N_C), and launch probability (P_L). Fig. 4.10 shows a histogram of h, which ranges from 1 to 50 hops, with an average $\langle h \rangle \approx 10$ hops. If we assume LSA is updated whenever a connection is created and we set ACO as launching ants every 20 connection requests which means $T_L = 20T_F$, one cycle per launch which means $N_C = 1$, and launch probability of 50% which means $P_L = 0.5$, the ratio is approximately 0.71.

Although not shown in the simulation, ACO reduces the connection setup delay by performing the route optimization calculation in the foraging process before a connection request even arrives. When a new request arrives, the edge node simply looks up the candidate solution table to assign the optimized lightpath. On the other hand, GASP needs to calculate the best lightpath based on the instantaneous network state information and this computation introduces a setup delay.

4.5 Summary

In this chapter we introduce an ACO algorithm for GRWA on a real-time dynamic-traffic fiber-optic network considering features such as mixed-line-rate, physical impairments and traffic grooming. We show the necessity of implementing TR and traffic grooming into the ACO foraging cycle. We discuss in detail how different configurations of the model affect the overall performance of the system. We also compare our ACO to a centralized heuristic method. We validate the robustness of our ACO by testing it on several scenarios, such as non-stationary traffic statistics and device failures. In the future, we plan to include considerations of limited hardware resources at nodes, such as transponders and wavelength switches into the ACO algorithm, implement it for elastic optical networks, and optimize it for energy efficiency as well.

Chapter 5

RSA for Elastic Optical Networks using MILP

Compared to legacy wavelength division multiplexing networks, elastic optical networks (EON) have added flexibility to network deployment and management. EONs can include previously available technology, such as signal regeneration and wavelength conversion, as well as new features such as finer-granularity spectrum assignment and modulation conversion. Yet each added feature adds to the cost of the network. In order to quantify the potential benefit of each technology, we present a link-based mixed-integer linear programming (MILP) formulation to solve the optimal resource allocation problem. We then propose a recursive model in order to either augment existing network deployments or speed up the resource allocation computation time for larger networks with higher traffic demand requirements than can be solved using an MILP. We show through simulation that systems equipped with signal regenerators or wavelength converters require a notably smaller total bandwidth for a given data rate, depending on the topology of the network. We also show that the suboptimal recursive solution speeds up the calculation and makes the running-time more predictable, compared to the optimal MILP.

5.1 Introduction

Increasing traffic volume and growing heterogeneity of bandwidth requirements have pushed the development of optical transport networks. Using wavelength division multiplexing (WDM) technology, spectrum usage has greatly increased by allowing multiple-line-rates and traffic grooming, which we discussed above. We have introduced several algorithms to implement traffic grooming for WDM in Chapters 3 and 4. Yet WDM is unable to handle increasing traffic heterogeneity because of the coarse wavelength grid employed. Elastic optical networks (EONs), on the other hand, provide flexibility in both bandwidth assignment (using sub channel granularity and super channel assignments) and lightpath reconfigurability not available in WDM. As the technology matures, additional functionality such as modulation selection and conversion can be added, with the hope of further increasing the spectral efficiency. When major additions in physical layer features are being considered, the networks design should be re-examined to determine the realized benefit gained by their implementation. This chapter presents an optimal routing, regeneration, and spectrum allocation formulation that is then used to evaluate the merit of wavelength and modulation conversion on EONs affected by physical layer impairments.

The design of transport networks includes the placement and assignment of all physical resources, such as optical fiber and electronic devices (transponders, high speed optical-electrical-optical conversion circuits, etc.). The goal is usually to minimize the capital expenditure while fulfilling certain traffic accommodation expectations. One common way to solve this problem is to address it as a multi-commodity assignment by pairing the physical resources with traffic demands in order to minimize the resources used by each demand. For example, the traffic demand can be assigned the shortest route in order to reduce the cost. Such design principles have been used to develop many heuristic algorithms for network design [38, 39, 40]. Although these algorithms are computationally simple, they often yield poor performance when the

problem becomes complex and consideration cannot be given to all influencing factors. Another approach is to formulate the resource allocation as an optimization problem with physical and network layer constraints and use linear programming (LP) to solve it. The available network resources become the LP design variables. Unlike arbitrary multi-commodity assignment problems, network design often requires its variables to be integer or Boolean, which leads to a mixed-integer linear programming (MILP) formulation. This significantly increases the computational complexity, not providing an approach that can scale to address larger networks. However, for small networks and few traffic demand, the MILP can be solved in reasonable time, and results in an optimal solution, unlike heuristic algorithms. It does this without requiring a complete understanding of the relationship between the multiple design factors, as heuristic algorithms often do [11].

In this chapter we develop an MILP design method for EONs. Our formulation can implement modulation scheme selection, mid-lightpath modulation conversion (MC) and/or wavelength conversion (WC), and regeneration circuit allocation (to satisfy either a quality of service constraint or conversion function). MILP has previously been used to solve the resource assignment optimization problem in optical networks [10, 41, 42] and many other optical networking's key issues, such as: minimizing capital investment [1], [43], the regenerator placement problem [44], [45], network restoration [46], and network service capability maximization [47]. However, to the best of our knowledge, no published MILP solution has included these flexibilities in an optimal way for designing EONs.

Acknowledging the limitations of the MILP approach for solving realistically scaled problems due to its computational complexity, we envision the following two direct uses for our model. The MILP can be solved for a small network to quantify the potential benefit that can be obtained by implementing a particular feature, such as modulation conversion, without introducing artificial limitations imposed by a sub-

optimal resource allocation algorithm. Our approach can also be used on realistic-size networks to solve for the optimal resource allocation of only a few traffic demands at a time. For the off-line resource allocation problem (static network), we can partition the whole traffic matrix into small sub-matrices, and solve the assignment problem for the sub-matrices in a sequential manner. By doing this, we are able to greatly reduce the overall execution time, in exchange for obtaining a suboptimal solution. In the chapter we discuss the tradeoff between complexity and optimality for this approach that we call the recursive solution. For dynamic networks, we can use the recursive MILP to allocate resources for one or a few new connection requests given the current state of the deployed network, as we introduce for WDM systems in Chapter 7.

The rest of the chapter is organized as follows: Section 5.2 explains how we implement the new functionalities with our MILP formulation; Section 5.3 develops our recursive MILP implementation that balances optimality and complexity; Section 6.4 presents numerical simulation results collected by solving the design problem using our formulation. Finally, conclusions are given in Section 7.5.

5.2 MILP

In this section, we first introduce a basic link-based MILP formulation that solves a simple routing and spectrum assignment (RSA) problem, and then extend it to implement signal regeneration and multiple modulation schemes. Lastly we implement wavelength and modulation scheme conversion. Our general objective is to minimize the spectrum required by the system, as measured by the maximum frequency allocated over all links. We also examine the impact of simultaneously optimizing the spectral use and regeneration resources using a multi-objective function.

The network is modeled as a graph $G(\mathcal{N}, \mathcal{L})$ with N nodes and L uni-directional links. We summarize the set notation used by our model in Table 5.1. The model also

Table 5.1: Sets used by Basic ILP

$\mathcal{N}$	Set of nodes in the network.
$\mathcal{L}$	Set of unidirectional links in the network. Each link L_{ij} is represented by its source and destination node, $L_{ij} \in \mathcal{L}$.
$\mathcal{D}$	Set of unidirectional traffic demands. Each demand D_{sd} is represented by its source node s and destination node d, $D_{sd} \in \mathcal{D}$.

Table 5.2: Parameters used by Basic ILP

b_{sd}	Bit rate requested by demand D_{sd}.
η_{sd}	Spectrum efficiency according to particular modulation scheme (e.g., 2 for QPSK).
$S_{n,sd}$	Relationship between nodes and demands: $S_{n,sd} = -1$ if node n is the source node of demand D_{sd} (i.e., $n = s$); $S_{n,sd} = 1$ if node n is the destination node of demand D_{sd} (i.e., $n = d$); $S_{n,sd} = 0$ otherwise (i.e., $n \neq s, n \neq d$).
G	Guard band in GHz.

depends on parameters specific to the network configuration and the traffic demands. The notation for the independent parameters needed is given in Table 5.2.

The objective function of the MILP is to minimize the highest frequency required to support the network traffic:

$$\min_{F_{sd}, V_{ij,sd}, \delta_{sd,s'd'}} c, \tag{5.1}$$

where the optimization variables are defined in Table 7.3. The optimization requires several constraints, listed below:

- Highest required spectrum:

$$c \geq F_{sd} + B_{sd} \quad \forall D_{sd} \in \mathcal{D}, \tag{5.2}$$

where B_{sd} is the bandwidth required by D_{sd} for a given η_{sd}, $B_{sd} = b_{sd} \times \eta_{sd}^{-1}$.

- Flow conservation constraints:

$$\sum_{L_{ij}\in\mathcal{L},j=n} V_{ij,sd} - \sum_{L_{ij}\in\mathcal{L},i=n} V_{ij,sd} \;=\; S_{n,sd} \qquad \forall n \in \mathcal{N},\; D_{sd} \in \mathcal{D} \tag{5.3}$$

- No spectrum overlap constraints, $\forall D_{sd}, D_{s'd'} \in \mathcal{D}$:

$$\delta_{sd,s'd'} + \delta_{s'd',sd} \;= 1 \tag{5.4}$$

$$F_{sd} - F_{s'd'} \;\leq T(1 - \delta_{sd,s'd'} + 2 - V_{ij,sd} - V_{ij,s'd'}) \tag{5.5}$$

$$F_{sd} - F_{s'd'} + B_{sd} + G \;\leq (T+G) \times (1 - \delta_{sd,s'd'} + 2 - V_{ij,sd} - V_{ij,s'd'}) \tag{5.6}$$

  where T is the total spectrum required by the network traffic, $T = \sum_{D_{sd}\in\mathcal{D}} b_{sd} \times \eta_{sd}^{-1}$.

Eqs. (5.1) to (5.6) define a general link-based RSA formulation for EON. Together (5.4)-(5.6) enforce a contiguous spectrum assignment to each demand. Eq. (5.4) says that for any two demands sd and $s'd'$ that share a link, one demand has to have a starting frequency lower than the other, and therefore one of the ordering variables is zero and the other is one. Eq. (5.5) enforces the necessary relationship between starting frequencies of the two demands based on the variable $\delta_{sd,s'd'}$. Then (5.6) forces the starting frequency of the demand with the higher starting frequency to be far enough away from the starting frequency of the lower adjacent channel, i.e., provides room for the signal bandwidth and guard band. These expressions can be modified to implement the more sophisticated signal processing we consider in this paper. Each functionality is discussed below, together with the additional variables and constraints needed.

Table 5.3: Variables used by Basic ILP

F_{sd}	Starting frequency index of demand D_{sd}.
$V_{ij,sd}$	Link assignment: $V_{ij,sd} = 1$ if link L_{ij} is assigned to demand D_{sd}; $V_{ij,sd} = 0$ otherwise.
$\delta_{sd,s'd'}$	Order of the starting frequency index of demand D_{sd} and $D_{s'd'}$.[1] $\delta_{sd,s'd'} = 1$ if $F_{sd} \leq F_{s'd'}$, $\delta_{sd,s'd'} = 0$ if $F_{sd} > F_{s'd'}$.
c	Highest frequency index required by the network traffic.

Table 5.4: Parameters Used by transmission reach constraint

ℓ_{ij}	Length of link L_{ij} in km.
R_{sd}	Transmission reach of demand D_{sd} according to particular spectral efficiency, e.g., in (2.1)
$\mathcal{N}^r$	Set of regeneration nodes.

5.2.1 Multiple Modulation Schemes

When each demand has different spectral efficiency, their transmission reach also varies. This is implemented by making η_{sd}^{-1}, the inverse spectral efficiency of demand D_{sd} according particular a modulation scheme, a variable instead of a constant parameter. In our model we relax this value from its normal discrete nature to be a real number bounded by the largest and smallest inverse spectral efficiencies allowed: $\eta_{sd,\text{MIN}}^{-1} \leq \eta_{sd}^{-1} \leq \eta_{sd,\text{MAX}}^{-1}$.

5.2.2 Signal Regeneration

Signal regeneration can be used to increase the length of a lightpath beyond the transmission reach. The following constraints, using additional parameters and variables defined in Tables 5.4 and 5.5, respectively, must be satisfied so that the QoT requirements are fulfilled for all demands.

[1] This relationship between two demands is only of interest if they share a link. We use this relationship in the following constraints to guarantee no overlapping between spectra assigned to multiple demands.

Table 5.5: Variables used by transmission reach constraint

$Y_{n,sd}$	$Y_{n,sd} = 0$ if node n is not on the lightpath assigned to demand D_{sd}. Otherwise, $Y_{n,sd}$ is the physical distance from node n on the lightpath to the beginning of that transparent segment for demand D_{sd}.
$U_{ij,sd}$	$U_{ij,sd} = 0$ if the entire link L_{ij} is not assigned to demand D_{sd} ($V_{ij,sd} = 0$). Otherwise, $U_{ij,sd}$ is the physical distance from node i to the beginning of the transparent segment for demand D_{sd}. Equivalently, if we were not restricted to linear functions, we could have defined $U_{ij,sd} = V_{ij,sd}Y_{i,sd}$.

We consider two cases. In the first case, the nodes in the network that are equipped with regeneration circuits has been pre-selected. There has been quite some research recently on how to select regeneration nodes, including [45]. The constraints that the MILP must satisfy for all $D_{sd} \in \mathcal{D}$ and $L_{ij} \in \mathcal{L}$ are as follows:

$$U_{ij,sd} \leq V_{ij,sd}R \tag{5.7}$$

$$U_{ij,sd} \leq Y_{i,sd} \tag{5.8}$$

$$Y_{i,sd} - U_{ij,sd} \leq R(1 - V_{ij,ld}) \tag{5.9}$$

$$Y_{n,sd} = \begin{cases} \sum_{L_{ij} \in \mathcal{L}: j=n} \ell_{ij} V_{ij,sd} & \text{if } \exists L_{ij}, i \in \mathcal{N}^r \\ & \text{and } V_{ij,sd} = 1 \\ \sum_{L_{ij} \in \mathcal{L}: j=n} U_{ij,sd} + \ell_{ij} V_{ij,sd} & \text{otherwise} \end{cases} \tag{5.10}$$

where $R = R(b_{sd}, \eta_{sd})$ from (2.1). For cases when node i is an intermediate node but link L_{ij} is not an intermediate link for demand D_{sd} (i.e., $V_{ij,sd} = 0$), the distance $U_{ij,sd} = 0$ but $Y_{i,sd}$ is not necessarily zero but has to be less than the reach. In this case Eq. (8) reduces to $Y_{i,sd} \leq R$.

The second case we consider is one where the regeneration nodes are not pre-selected. We then use the MILP to optimize the placement of regeneration equipment on the network. We treat the regeneration node assignments as binary variables I_n

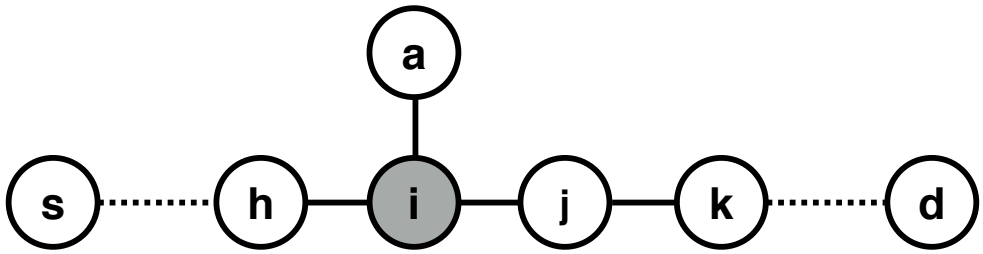

Figure 5.1: Illustration used to explain the variables defined for constraining the transmission reach

and, with the help of additional variables defined in Table 5.6, we use Eqs. (5.7)-(5.9) from above and replace (5.10) by (5.11):

We consider Fig. 5.1 to illustrate the TR constraints. There is a demand D_{sd} and a path from node s to node d, while link L_{ia} does not belong to the path of demand D_{sd}. Assume node i is the only regeneration node on the path. Inequality (5.7) says that for any link on the path (e.g., L_{ij}), $U_{ij,sd} \leq R$. Since node i is the only regeneration node, U_{ij} for link L_{ij} and $Y_{i,sd}$ both represent the distance from node s to node i. Since link L_{ia} is not on the path for demand D_{sd} (i.e., $V_{ia,sd} = 0$), $U_{ia,sd} = 0$. Eq. (5.10) says that, for node i, since no link that leads to it starts with a regeneration node, $Y_{i,sd}$ is the sum of U's for all links that lead to node i plus the link length of the link that is on the path. Since all links except for link L_{hi} have U equal to zero, $Y_{i,sd}$ is the sum of $U_{hi,sd}$ and link length of link L_{hi}. But, for node j, since node i is a regeneration node, $Y_{j,sd}$ is just the link length of link L_{ij}. Inequalities (5.8) and (5.9) say that $Y_{i,sd} < R$, $U_{ia,sd} < Y_{i,sd}$, and $U_{ij,sd} = Y_{i,sd}$. When the allocation of regeneration nodes is unknown, we use the variable $X_{ij,sd}$ to differentiate the cases when node i is a regeneration or not, as in (5.11):

$$Y_{n,sd} = \sum_{L_{ij} \in \mathcal{L}: j=n} X_{ij,sd} + \ell_{ij} V_{ij,sd}, \tag{5.11}$$

Table 5.6: Variables used for Regenerator Circuit Assignment

I_n	Regeneration nodes: $I_n = 1$ if node n is used as a regeneration node; $I_n = 0$ otherwise.
$N_{n,c}$	Number of regeneration circuits used on node n.
$I_{n,sd}$	Regeneration at node n: $I_{n,sd} = 1$ if demand D_{sd} is regenerated at node n, $I_{n,sd} = 0$ otherwise. For $I_{n,sd} = 1$, node n has to be a regeneration node, i.e., $I_n = 1$, but also its regeneration circuit has to be used by demand D_{sd}.
$X_{ij,sd}$	Distance used to calculate $Y_{n,sd}$ based on whether regeneration occurs at node i: $X_{ij,sd} = U_{ij,sd}$ if $I_{i,sd} = 0$, $X_{ij,sd} = 0$, otherwise.

Constraints that limit the number of OEO circuits per regeneration node can also be included using:

$$N_{n,c} = \sum_{D_{sd} \in \mathcal{D}} I_{n,sd} \tag{5.12}$$

$$I_n N_{n,c\mathrm{MAX}} \geq N_{n,c} \tag{5.13}$$

where $N_{n,c\mathrm{MAX}}$ is the largest number of regeneration circuits that can be equipped on a regeneration node.

When the cost of regeneration resources is a concern, we can build a multi-objective function to balance the cost of regeneration and spectrum resources:

$$\min \left\{ ac + (1-a) \sum_{n \in \mathcal{N}} I_n \right\} \tag{5.14}$$

where the coefficient $a \in [0, 1]$ represents the cost relationship between using the two resources. This objective function minimizes the total cost of all resources together, according to their relative costs. While we do not presume to know the exact cost relationship among the two, a network designer can base their objective function on realistic requirement, and use (5.14) to determine what resources are needed and where.

Table 5.7: Variables used by Wavelength and Modulation Conversion

$F_{ij,sd}$	Starting frequency index of demand D_{sd} on link L_{ij}.
$\eta^{-1}_{sd,ij}$	Inverse spectral efficiency of demand D_{sd} on link L_{ij}.

5.2.3 Wavelength and Modulation Conversion

When WC is available, the frequencies used for a demand can be different on the links entering a regeneration node and exiting it. In order to represent this flexibility we define the starting frequency on a link-by-link basis, as shown in Table 5.7, and re-write (5.5) and (5.6), the constraint that guarantees no spectrum overlap, for all $n \in \mathcal{N}$ as:

$$\begin{aligned} \sum_{L_{ij}\in\mathcal{L}:j=n} F_{ij,sd} - \sum_{L_{ij}\in\mathcal{L}:i=n} F_{ij,sd} &\geq -T \times (I_{n,sd} + |S_{n,sd}|) \\ \sum_{L_{ij}\in\mathcal{L}:j=n} F_{ij,sd} - \sum_{L_{ij}\in\mathcal{L}:i=n} F_{ij,sd} &\leq T \times (I_{n,sd} + |S_{n,sd}|) \end{aligned} \tag{5.15}$$

This constraint requires that if node n is an intermediate node for demand D_{sd}, i.e., $n \neq s, n \neq d$ and n is not used as a regeneration node for demand D_{sd}, then the starting frequency assignments entering node n equals the starting frequency assignments exiting node n. For other cases, this constraint does not apply.

Similar to WC, when MC is available the spectral efficiency of each demand on each link can be different than its immediate uplink or downlink if the joining node is used as a regeneration node. We must define the spectral efficiency on a link-by-link basis, as listed in Table 5.7. The MC constraint can be written as:

$$\begin{aligned} \sum_{L_{ij}\in\mathcal{L}:j=n} \eta^{-1}_{sd,ij} - \sum_{L_{ij}\in\mathcal{L}:i=n} \eta^{-1}_{sd,ij} &\geq -\eta^{-1}_{sd,\mathrm{MAX}} \times (I_{n,sd} + |S_{n,sd}|) \\ \sum_{L_{ij}\in\mathcal{L}:j=n} \eta^{-1}_{sd,ij} - \sum_{L_{ij}\in\mathcal{L}:i=n} \eta^{-1}_{sd,ij} &\leq \eta^{-1}_{sd,\mathrm{MAX}} \times (I_{n,sd} + |S_{n,sd}|) \end{aligned} \tag{5.16}$$

Similar to the WC constraint, this constraint requires that spectral efficiency only be converted at nodes n where the demand is regenerated, i.e., where $I_{n,sd} = 1$.

5.3 Recursive MILP

The computational complexity of the MILP formulation for EON restricts its implementation to offline calculation only. And even then, it does not scale well as the size of the network or the number of traffic demands increase. An effective alternative would be to reduce the problem size to such an extent that the result can be found in acceptable time and with reasonable computational resources. Many works have attempted to break the whole RSA problem into sub-problems with or without losing some degree of optimality [48], [49]. In this section, we propose a different way to reduce the problem size by splitting the traffic matrix into sub-matrices and solving them sequentially, a technique we call *recursive MILP*.

The recursive MILP approach is motivated by the understanding that the complexity of the problem is greatly affected by the number of traffic demands that need to be accommodated at once. By separating them into subsets and allocating those in sequential iterations, the overall runtime as well as other computational resources, such as memory, can be reduced. The solution from the previous iteration forms new MILP constraints for the new iteration. In particular, the first iteration can be viewed as a subproblem with the same constraints but with fewer demands. In the new iterations, the constraints (e.g., non-overlapped spectrum assignment) apply to both the assigned and unassigned resources. In this manner, the original problem can be solved, albeit not optimally, after all iterations are done.

Another advantage of using recursive MILP is that the complexity is easy to estimate. For example, in the aforementioned MILP formulation, the complexity-dominating variable is $\delta_{sd,s'd'}$, which grows with the number of demands $|\mathcal{D}|$ squared,

Table 5.8: Complexity of One Iteration of Recursive and Non-recursive Basic MILP

Number of Variables	
non-recursive	recursive
$(1 + L + \|\mathcal{D}\|) \times \|\mathcal{D}\| + 1$	$(1 + L + \|\mathcal{D}\|/S) \times \|\mathcal{D}\|/S$
Number of Constraints	
non-recursive	recursive
$\|\mathcal{D}\| + N\|\mathcal{D}\| + (1 + 2L) \times \|\mathcal{D}\|^2$	$\|\mathcal{D}\|/S + N\|\mathcal{D}\|/S + (1 + 2L) \times (\|\mathcal{D}\|/S)^2$

i.e., $O(|\mathcal{D}|^2)$. By solving the same problem recursively, the number of variables of each calculation is reduced. If the number of subsets is S, the last subproblem (which has the highest number of variables) has about $\frac{|\mathcal{D}|^2}{S^2}$ many $\delta_{sd,s'd'}$. Running the MILP in recursive mode does not require reformulating the problem: the constraints remain the same but the variables that represent demands from previous iterations become constants.

It can be expected that the recursive MILP suffers loss of optimality compared to the non-recursive counterpart. The gap between the sub-optimal solution from the recursive MILP and the optimal solution depends on the size of the subset and the grouping and ordering of traffic demands. Since the complexity is easy to estimate, network designers can base the implementation of the formulation on the complexity they can accept. In Table 5.8 we show a comparison between the order of computational complexity for non-recursive (single-run) and recursive MILP solutions.

The complexity and optimality are affected not just by the size of the demand subsets, but also the selection of demands in each iteration. The grouping of demands can be done in many ways such as sorting them randomly or based on their characteristics such as volume, locality etc. In Section 6.4 we show a comparison of the required spectrum using different ordering schemes.

Another use for the recursive MILP implementation is to help accommodate network expansion with existing infrastructure. The existing assignments of physical

resources (links, regeneration devices, spectrum) can be input into the MILP as constraints, in the same way as was done for results from earlier iterations of the recursive MILP. In our simulation results below we show a progressively increasing number of traffic demands, each demand set including the same demands from previous sets. The spectrum predicted shows the resources required for network expansion. This problem is similar to the so-called *dynamic* resource allocation problem [50], [51], the difference being whether connections are also torn-down. In that scenario, the above MILP can also be used as a similar recursive way as was shown for WDM systems in Chapter 7.

5.4 Numerical Results

We test our MILP formulations on the NSF-24 mesh network, shown in Fig. 5.2, which is often used as a benchmarking topology in literature. In addition, to gauge the sensitivity of our results on network topology we simulate a symmetric network illustrated in Fig. 2.3, with 24 edge nodes that also serve as intermediate nodes for routing traffic. The NSF-24 network has 43 by-directional links (or uni-directional link pairs of 86 links), while the symmetric-24 network has 55 by-directional links (or 110 uni-directional links). In our simulation, we assume traffic demands are generated between random selected node pairs and have random bit-rate requests ranging uniformly from 1 to 100 Gbps in order to represent the heterogeneity of Internet traffic. The traffic is assumed static, and no traffic grooming or reverse grooming (i.e. traffic splitting) is considered. In all cases we collect simulation results over 20 independent random demand sets and report on the average among them.

In the numerical results, we investigate one or two features at a time in each section. Unless otherwise stated, the algorithm tested is an optimum MILP (non-recursive) assuming multiple modulations are available with $\eta_{sd} \in [1, 10]$ and signal

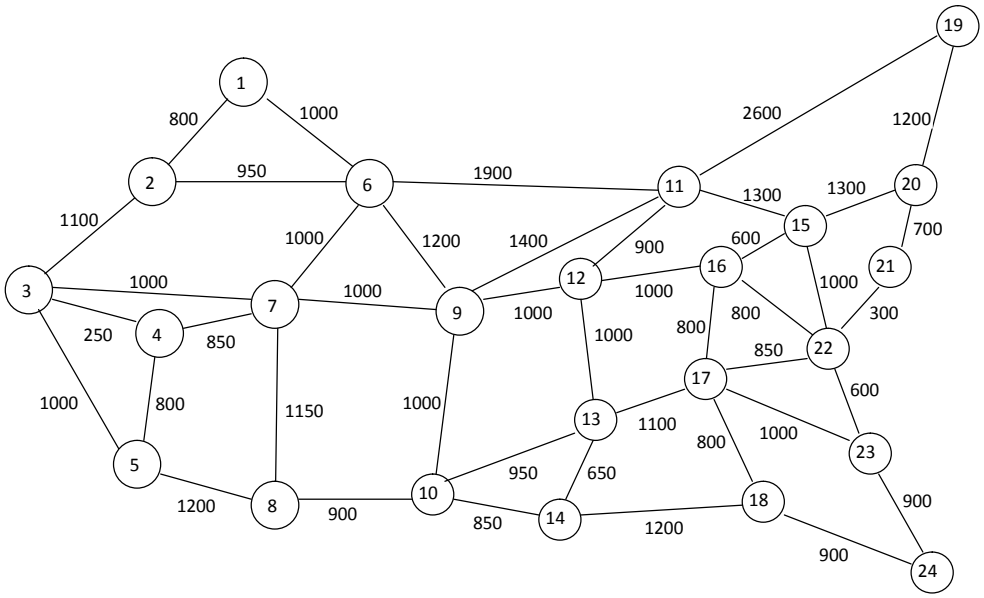

Figure 5.2: NSF-24 network

regeneration capability at all nodes, but no modulation or wavelength conversion is used so that each demand uses the same modulation and spectrum from source to destination.

5.4.1 Recursive and Non-recursive MILP

We first verify the applicability of the MILP solution to the network sizes we have chosen, and compare the optimality and computational complexity between the recursive and non-recursive approaches. We do not consider physical impairments here (no transmission reach constraint); therefore, there is no need for signal regeneration. In Fig. 5.3 we show the required spectrum for both single and multiple modulation schemes. We also plot the standard deviation for our results. As the number of demands allocated increases, the spectral usage increases approximately linearly. For the single modulation case, $\eta_{sd} = 2\ \forall D_{sd}$ (QPSK) so that all demands can reach their destinations with regeneration, while for the multiple modulation case η_{sd} is optimized. When the MILP assigns resources to all demands together (the "single solve" approach), the performance is notably better than the recursive approach (assuming a

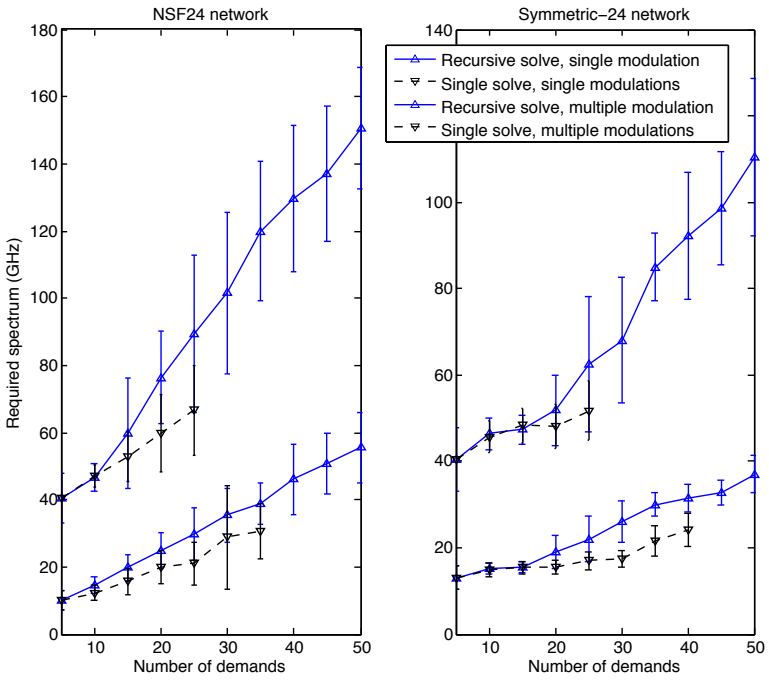

Figure 5.3: Required spectrum using the recursive MILP and single solve MILP for a single modulation scheme ($\eta = 2$) and multiple modulation schemes ($1 \leq \eta \leq 10$)

random partition of the demands into sets of 5), but too computationally burdensome for more than 30-40 simultaneous traffic demands. We also conclude that the added flexibility of optimizing the spectral efficiency for each demand more than halves the required spectrum. Both networks show similar results.

We compare of the complexity of the four approaches from Fig. 5.3 in Fig. 5.4. We show the histogram of running times when the total number of demands is 25. The running time for the optimal MILP varies considerably between trials (we set a time limit of 3000 seconds), while the running time for the recursive approach are uniformly small.

In Fig 5.5 we show that the partitioning and ordering of traffic demands in the recursive solution has a small but non-negligible impact on the required spectrum. When only a few demands have been assigned resources, accommodating high data-rate demands first leads to a lower required spectrum. Also, assigning demands that

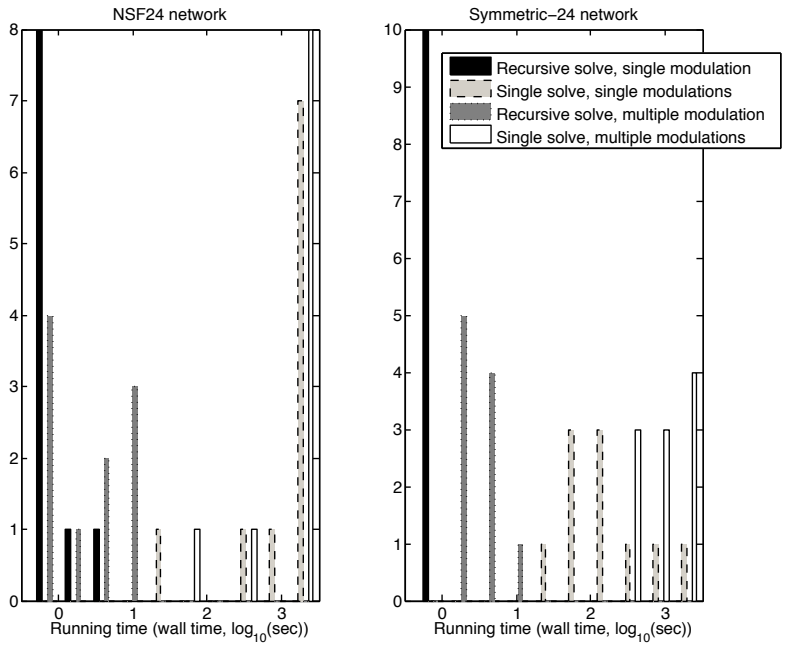

Figure 5.4: Running time for the recursive MILP and single solve MILP for a single modulation scheme ($\eta = 2$) and multiple modulation schemes ($1 \leq \eta \leq 10$)

have the shortest shortest-path routing (labeled "SP") first typically requires less spectrum than assigning the longer-distance connections first. The differences are slightly more pronounced on the symmetric network.

We then investigate the effect of the demand subset size on the required spectrum obtained by the recursive MILP. The single solve approach, which finds the globally optimal result for all the traffic demands together, must always yield the smallest required spectrum. In Fig. 5.6, we see that as the subset size decreases, the required spectrum increases. When few demands are assigned per iteration, the required spectrum appears to be stair-stepped, since new connections can often use gaps in the allocated spectrum left by fragmentation induced by the sub-optimal resource allocation.

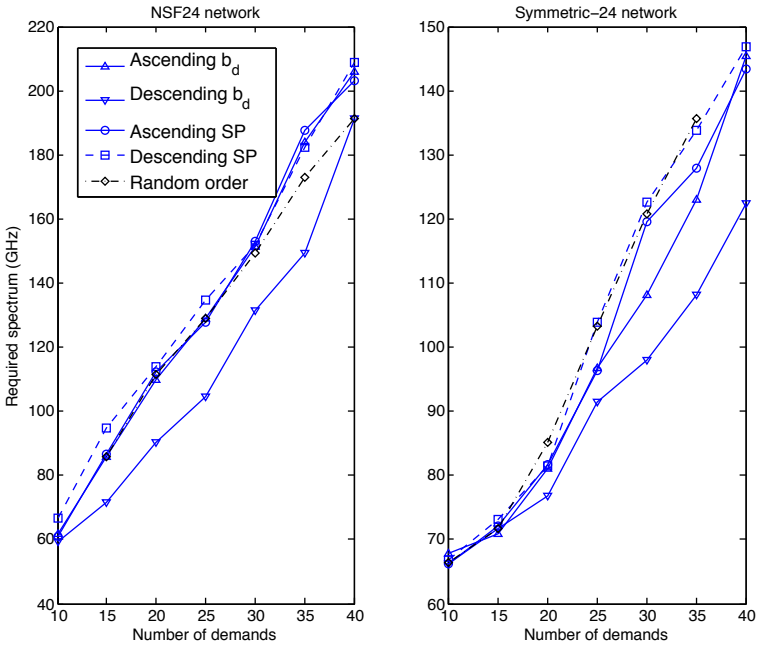

Figure 5.5: Required spectrum for the recursive MILP with different ordering schemes for the same 25 demands, as the resource assignment for the demand subsets progresses.

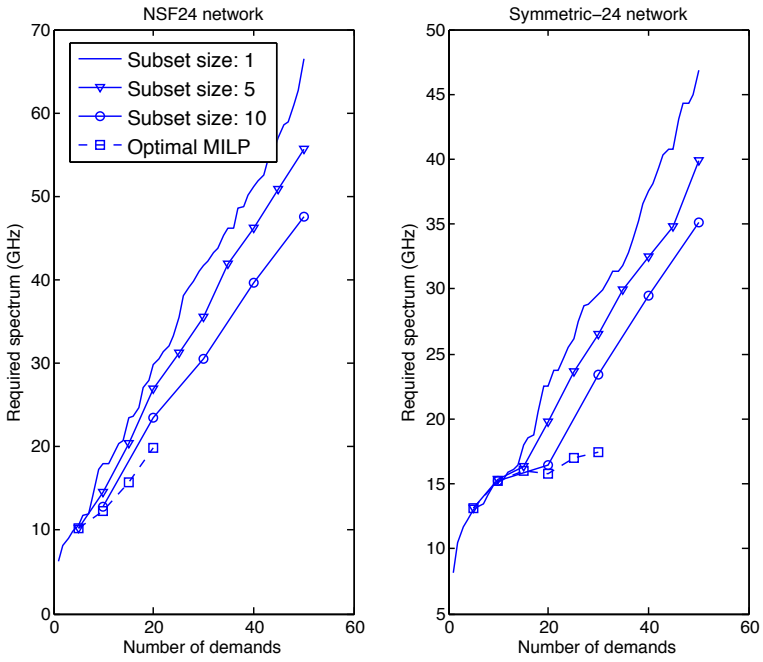

Figure 5.6: Required spectrum by solving the same recursive formulation with different demand subset sizes

5.4.2 Multi-objective Formulation

We investigate the ability of the MILP optimization to effectively trade-off spectral usage with regenerator usage by including the cost of regeneration resources (namely the number of regeneration nodes) in our objective function according to (5.14). The resulting spectrum and regeneration node requirement based on different values of the cost coefficient a are shown in Figs. 5.7 and 5.8, respectively. Assuming *a priori* knowledge of the cost relationship between spectrum and regeneration resources, network designers can choose the cost coefficients accordingly. The results show that when the spectrum cost is not considered (i.e., $a = 0$), the number of regeneration nodes is minimized and the required spectrum is large.[2] The required spectrum in this case is also highly irregular, as it is entirely unconstrained. On the contrary, when the regeneration cost is not considered (i.e.: $a = 1$) the required spectrum is minimized but the number of regeneration nodes is high. It is interesting to note that by assigning even a relative small coefficient to regeneration cost (i.e.: $a = 0.99$), we are able to maintain a similar required spectrum but greatly reduce the number of regeneration nodes needed.

5.4.3 Wavelength and Modulation Scheme Conversion

Since regeneration involves OEO conversion, it can improve the signal quality, so as to extend the TR, and can also provide an opportunity to change the spectrum and modulation assigned starting from that node. We run simulations to show the impact on the spectrum requirements of using the capability to convert the wavelength and/or modulation at the regeneration nodes. In Fig. 5.9, we solve the resource allocation problem with our recursive MILP formulation, since the added flexibility of WC and MC increases the complexity of the problem considerably. Both WC and MC reduce

[2]By configuring the demand volume and spectral efficiency, we make sure there is a spectral efficiency available so that it is possible that no demand requires regeneration. Therefore, in the case of $a = 0$, the number of regeneration nodes is always minimized to zero.

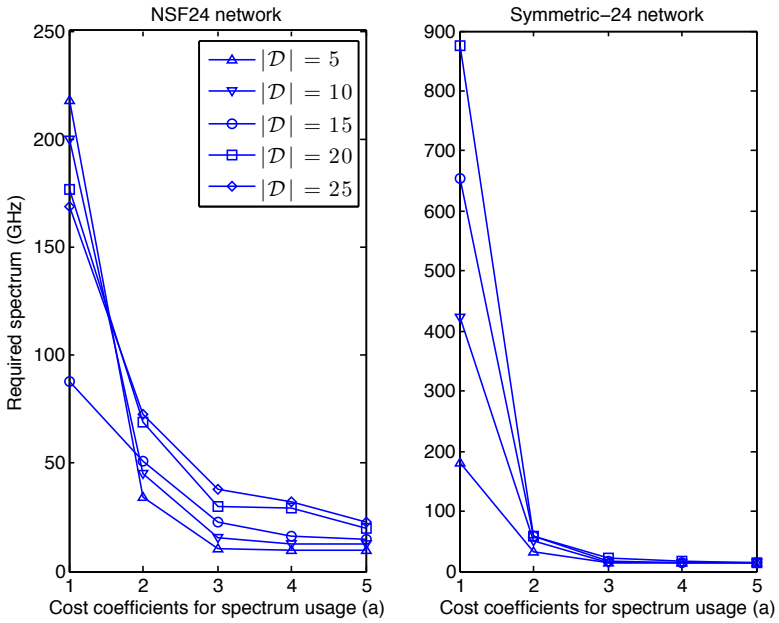

Figure 5.7: Required spectrum as the coefficient a in the objective function (5.14) varies. Note that the horizontal axis is not drawn to scale.

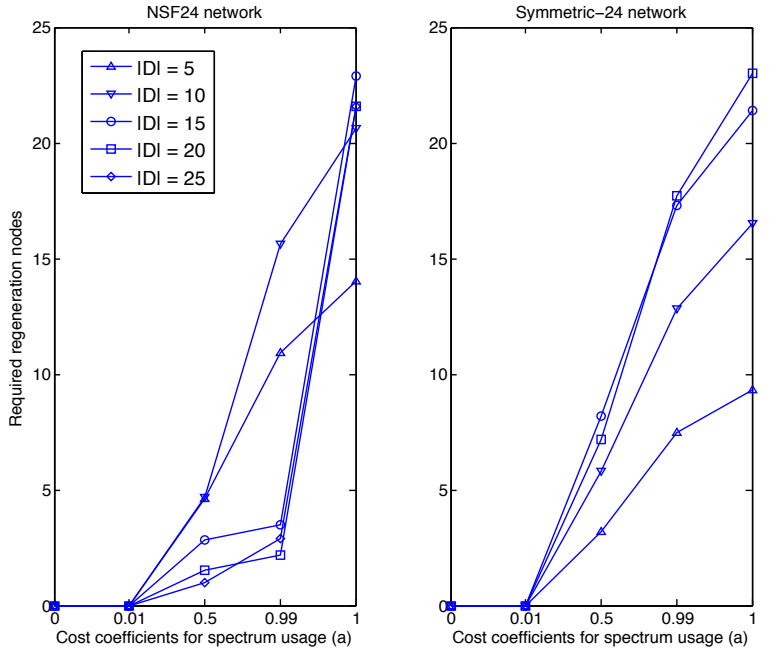

Figure 5.8: Required number of regeneration nodes as the coefficient a in objective function (5.14) varies. Note that the horizontal axis is not drawn to scale.

the amount of spectrum required to support the traffic demand. WC allows signals to be re-allocated on some links so as to fill gaps left by other traffic demands. The improvement made by wavelength conversion depends on the fragmentation condition of the network. When one link is heavily congested, it then becomes a bottleneck of the network creating unnecessary fragmentation on other links. MC takes advantage of the difference in the length of transparent segments for each lightpath. If they are significantly different, the spectrum saved by optimizing the spectral usage based on each segment becomes significant. The NSF-24 network has well-documented bottleneck paths, and thus benefits more from WC than a more symmetric topology. The symmetric network also has equal link lengths, and can therefore not exploit MC as much as the more heterogeneous NSF-24.

In Fig. 5.9, when we compare the WC case with the case when both wavelength and modulation scheme conversion are available, we expect the latter to always outperform the former, since it has more flexibility. However, the results (which are averaged over 20 trials) for the symmetric network show that this relationship is not guaranteed. Solving the MILP recursively optimizes the solution in each iteration according to its constraints, yet does not necessarily leads to the optimal solution for the whole traffic matrix when looking at multiple iterations together. Each iteration can only find the local optimum for its sub-problem and the local optimal solution of a sub-problem may not be one part of the global optimal for the whole problem. To put it in terms of resource assignment, in the previous iterations resources may be assigned to demands according to the sub-problem, but such assignment may be not optimal when one considers the whole traffic matrix, which leads to an increased spectrum requirement. The same limitation exists in dynamic real-time RSA where no future traffic information is available. The RSA algorithm cannot assign physical resources to current traffic demands taking unknown future demands into consideration. One

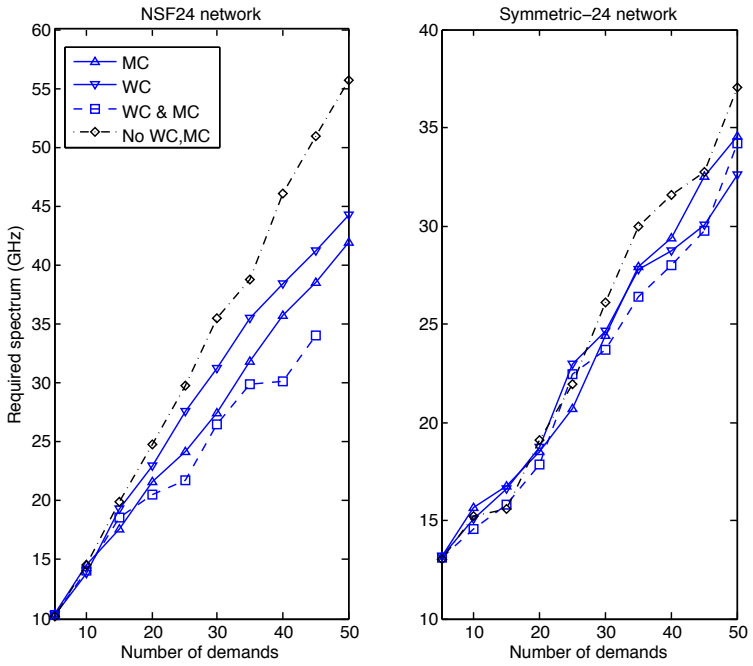

Figure 5.9: Spectrum usage comparison using the recursive MILP with and without wavelength and/or modulation conversion

could introduce a conservative rule based on long-term statistics that prevents over-assigning physical resources to current demands.

5.4.4 Regeneration Node Placement

When regeneration resources are scarce, careful network planing is important to minimize capital expenditure. In order to show the tradeoff between the number of regeneration nodes and the required spectrum, we simulate a case where the network has a limited number of regeneration nodes. In order to show the change of performance by adding additional regeneration nodes we keep the existing regeneration nodes unchanged once assigned. We use our results in Fig. 5.7 for cost coefficient $a = 0.5$ to find and rank the most often used regeneration node locations on average (over 20 trials). After allocating a limited number of nodes as regenerating nodes according to this ranking, we minimize the required spectrum. Our results in Fig. 5.10

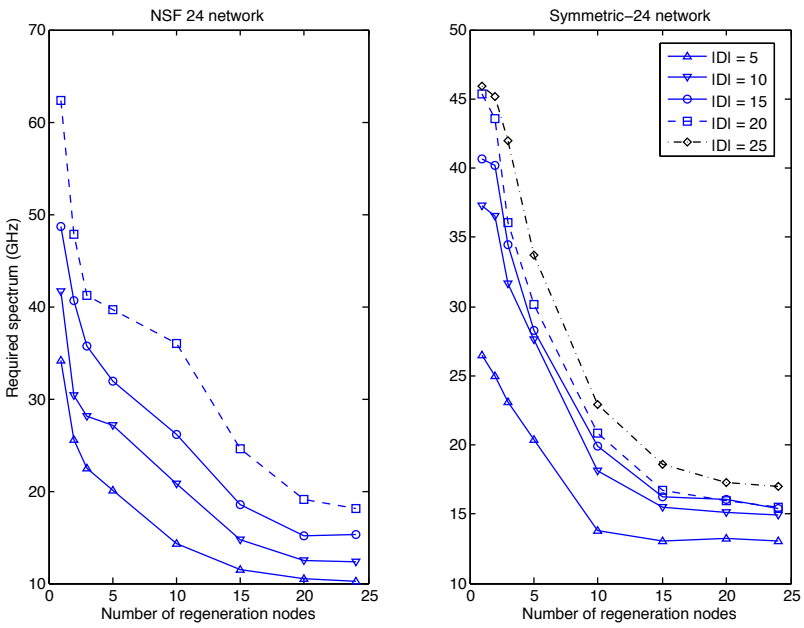

Figure 5.10: Spectrum usage with a limited number of regeneration nodes

show that for the NSF-24 network, the required spectrum reaches its minimum when 20 nodes have been allocated as regeneration nodes; for the symmetric 24 network, the required spectrum is close to minimum already when the number of regeneration nodes reaches 15. We believe this is due to the symmetric structure of the symmetric 24 node network, where most node pairs share joint intermediate nodes.

5.5 Summary

In this chapter we propose an MILP formulation to investigate the impact of technologies such as allowing multiple modulation schemes, signal regeneration, wavelength conversion, and modulation conversion on the required spectrum of the EON. We show through simulation that equipping systems with signal regeneration that controls physical impairments and allow for modulation and/or wavelength conversion reduces the amount of spectrum required. Such improvements depend on the topology of the network. We also show the impact of having a limited number of regenera-

tion nodes and different topology structures. In order to balance the optimality and complexity we propose a recursive MILP formulation that yields a suboptimal yet comparable solution to the MILP by using a lower and more consistent running time. The performance of the recursive model depends on factors such as heuristic demand ordering and iteration size.

Chapter 6

Link-based vs. Path-based MILP formulation for RSA in EON

6.1 Introduction

Elastic optical networks (EON) have brought flexibility to network deployment and operation compared to traditional wavelength division multiplexing (WDM) networks. Although linear programming optimization techniques for resource allocation common to WDM systems have been applied to EON, the common belief that a path-based formulation is sufficiently detailed for EONs requires revisiting when signal regeneration and modulation adaptation are used to combat physical impairments in the network. In this chapter we compare the two approaches, namely path-based and link-based mixed-integer linear programming (MILP) formulations, in their implementation, optimality, and complexity for EONs. We show using simulation that the network topology and traffic demand affect the difference in performance between the two formulations.

Designing EONs includes determining physical resource placement and assignment to the system. Such physical resources include optical spectrum as well as

physical equipment, such as transponders and regenerators. This is often referred to as the *routing and spectrum assignment (RSA) problem*. The goal is to minimize resource usage, implying low capital expenditure and low operational cost, while satisfying expected traffic needs. The problem can be approached as a mixed integer linear program (MILP) or via a heuristic algorithm. MILP formulations, compared to heuristic algorithms, often require longer computation time, limited by their complexity; however, they do not require insight into system behavior assumed by heuristic algorithms. In this chapter two different MILP formulations, path-based and link-based, are evaluated and compared in terms of performance and complexity when applied to EONs.

MILP approaches have been extensively studied in the literature. Most previous works use a simple formulation that only considers a subset of viable solutions (e.g., a small number of possible paths instead of all source-destination routes). These algorithms are referred to as *path-based* (PB), rather than the more general *link-based* (LB) approach to resource allocation that finds a globally optimal solution. The first comprehensive PB MILP formulation for RSA for EON appeared in [52]. In [15], a PB MILP formulation is used to design traffic grooming for EONs. In [10], an MILP formulation is proposed in which the contiguous spectrum assignment is represented by narrow frequency slots. A third approach [53] is to implement a partial LP relaxation method to handle integer (or Boolean) variables. (Such methods can also be implemented on LB formulations.) The only LB technique reported in literature, other than our own, is a low complexity method that considers a localized subset of all links [54].

All of the above papers ignore physical impairments. In very recent work [55], a transmission reach constraint was proposed by the authors. Since their formulation is PB, the location of regeneration nodes (RN) along each candidate path is predetermined. They also include transponder costs in their objective function to balance

the need to minimize spectrum usage with transponder usage. The PB formulation we use in this article is based on this work. To the best of our knowledge, the only LB MILP RSA approach for flexible EON that includes physical impairment constraints is our own work, presented in Chapter 5 and in [56].

It is well understood that with a sufficiently large candidate path list, the PB MILP solution can approach the optimal solution given by the LB MILP. However, results given in the literature often use a small candidate pool, which is sufficient to show the formulation implementation but not enough for comparison with other methods. Furthermore, when signal regeneration is considered, in PB methods there is no effective way to include regeneration resource assignment within the optimization. In this chapter we discuss the effect of the size of the candidate path pool on the optimality of PB MILP solutions. We then compare PB to LB MILP formulations and show the advantages of the LB approach when traffic must be regenerated. Due to space limitations, only a cursory description of the two MILPs is given here; the whole formulation can be found in [55] and [56] for the PB and LB algorithms, respectively.

The chapter is organized as follows. Section 6.2 explains the principle of PB and LB MILP formulations and their complexity and optimality. Section 6.3 shows how each algorithm addresses the regeneration resource assignment problem. In Section 6.4 we show the advantages and disadvantages of using an LB MILP formulation through numerical simulations. We conclude the chapter in Section 7.5.

6.2 Path-based and Link-based MILP Formulations

The objective of any RSA algorithm for EONs is to minimize the total physical resource cost. Since the optical spectrum is an important investment (determines the number of optical fibers needed), it is often used to represent the cost. Other costs,

such as optical/electronic devices required, can also be added to the objective function; in this chapter, we consider these instead as design constraints. As variants of a multi-commodity flow problem, all MILP formulations satisfy common constraints: the traffic assigned to physical resources cannot exceed their capacity (capacity constraints), traffic needs to flow from source through intermediate network nodes to destination (flow conservation constraints), and the traffic demand matrix has to be fully accommodated (demand satisfaction constraints).

Although the principles of the formulations are the same for both PB and LB approaches, their implementations differ in their variable setup and constraint expressions, which directly affect their optimality and complexity. Both approaches yield a resulting path that connects the source and destination nodes for each traffic demand. In PB formulations, the set of candidate paths is often pre-selected based on metrics such as shortest path or minimal resource cost. A Boolean variable is assigned as an indicator function to each path for each demand; path assignment and uniqueness (flow conservation) is then achieved by constraining the sum of these variables to equal unity. In LB formulations, all legitimate link combinations that form a loop-free path between source and destination nodes are considered as candidate paths, which implies a globally optimal solution. Here a Boolean variable is assigned to each link for each demand, instead of each path. The flow conservation constraints are satisfied through node input/output flow constraints. The wavelength continuity constraint is implemented in the same way in both formulations by assigning a variable that represents the starting frequency for each demand throughout the path and not allowing the bandwidth used by each demand to vary link-by-link. The capacity and demand satisfaction constraints are also similar in the two approaches.

We compare the number of variables and constraints between the PB and LB methods in Table 6.1 (ignoring the same constraints that appear in both formulations). L is the number of links, K is the number of candidate paths for each demand

Table 6.1: Complexity of Formulations

Number of Variables		Number of Constraints	
PB	LB	PB	LB
$KD+D+D^2$	$LD+D+D^2$	$D+2K^2D^2$	$ND+2LD^2$

in the PB formulation, D is the number of demands, and N is the number of nodes. The number of constraints for the PB formulation is a worst case number, since it depends on how many lightpaths share a link instead of the total number of demand pairs. The complexity of the LB formulations depends on the number of links and is fixed.

6.3 Regeneration Resources Assignment

For long-haul networks, physical impairments make signal regeneration necessary along the lightpaths. The distance that an optical signal can travel without being regenerated is called the transmission reach (TR). In order to maintain acceptable signal quality beyond the TR, regenerators have to be assigned to the lightpath. Signal regeneration operates in the electrical domain at intermediate nodes, thus costly optical-electrical-optical converters are needed at these locations. Both the total number of regenerators needed and the number of nodes that must be equipped with regenerators, denoted N_r, introduce additional cost. We show that with careful design an LB MILP formulation can reduce N_r compared with the PB method, thereby lowering the total cost. The number of converters needed for a demand depends on the path length, and is therefore difficult to reduce via RSA optimization.

Transmission reach T_r for a traffic demand d depends on many factors such as bit rate of the demand R_d and its spectral efficiency η_d. To fit within our MILP model, we formulate a linear relationship between transmission reach and these two factors

as

$$T_r = \alpha R_d^{-1} + \beta \eta_d^{-1} + \gamma \tag{6.1}$$

where, $\{\alpha, \beta, \gamma\} = \{18600, 8360, -250\}$ are coefficients derived from polynomial curve fitting based on experimental data in [18]. For computational convenience, we relax the spectral efficiency variable η by allowing it to be a real number.

When determining the assignment of regeneration resources, PB MILP formulations are ineffective because candidate lightpaths are pre-determined before solving the MILP. In order to guarantee that each path meets the TR constraint, a minimum number of regenerators are assigned to each path to satisfy (2.1), such as in [55]. Then each candidate path together with its RN becomes one Boolean variable. Although such assignment is optimal to each lightpath, it might not be optimal for the entire network in terms of N_r. However, using the LB approach the RN assignments can be optimized along with routing and spectrum assignment, by treating them as variables separate from links. By enforcing a constraint on the distance that the signal has traveled without being regenerated, we can ensure that all resulting lightpaths satisfy the TR constraint.

6.4 Numerical Results

We implement the PB and LP MILP formulations on the NSF24 topology and a symmetric 24-node mesh network shown in Fig. 5.2 and 2.3. For the NSF24 network, the numbers noted on the links are the link lengths (in km), while for the symmetric 24-node network all link lengths are set to 1330 km, resulting in similar network diameters. The simulation results show the average over 20 instances of each traffic demand scenario, ranging from 5 to 25 random source-destination pairs each, as labeled, with bit-rates uniformly distributed from 1 to 100 Gbps. The PB MILP uses

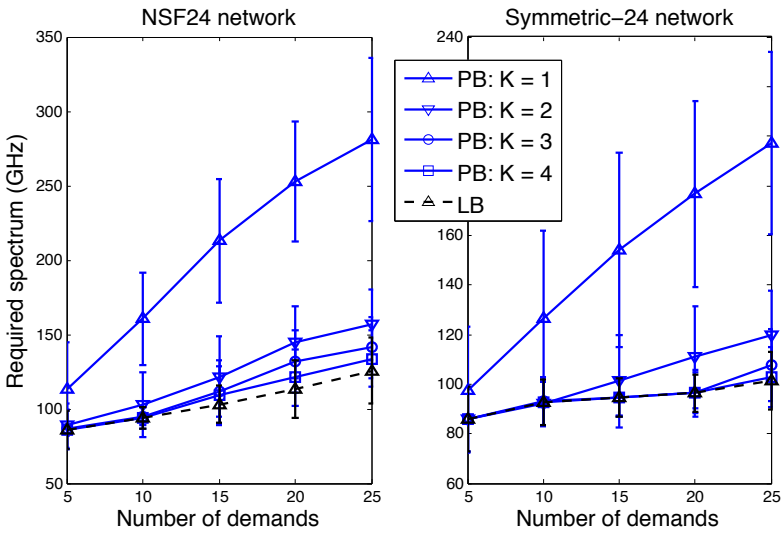

Figure 6.1: Required network spectrum for PB and LB MILP formulations, no TR constraints, $\eta = 1$ bit/symbol.

a link-disjoint (whenever possible) candidate-path-set based on shortest physical path length.

In Fig. 6.1, we show the spectrum required (maximum over all links in the network) to satisfy all traffic demands without considering physical impairments when there is only one modulation scheme available at the transponders with spectral efficiency $\eta = 1$. PB MILP requires significantly more spectrum when $K = 1$ compared to larger candidate-set sizes. This is due to many source-destination node pairs sharing the same links for their shortest path, which creates congestion. However, this phenomenon can be effectively mitigated by adding one additional path, i.e., $K = 2$. For the NSF24 network, the spectrum saved by further increasing K is small because additional candidate paths often share some of the same links, and thus the traffic load cannot be effectively balanced. The savings are more obvious with the symmetric 24-node network since the paths are naturally more link-disjoint. The LB MILP shows the best performance in terms of spectrum requirement,most notable when D is large. Network designers need to select which approach to use with network topology in mind.

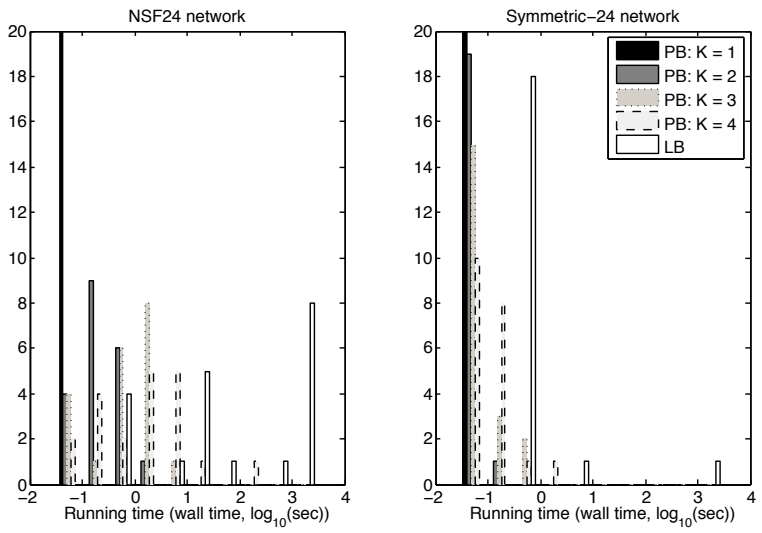

Figure 6.2: Histogram of running time (clock time) for PB and LB MILP formulations, no TR constraints, $\eta = 1$ bit/symbol.

Figure 6.2 shows the running time (wall clock time) for solving the MILP in Fig. 6.1.[1] For the PB method with $K = 1$, the problem becomes strictly one of spectrum assignment (SA), which is a sub-problem of RSA and requires less time than when routing assignment is also involved. The PB MILP shows great scalability: as the number of demands increases, the running time remains relatively small. The LB MILP, however, requires much longer running time and does not scale well with an increase in the number of demands. This is because the algorithm must consider many alternatives in order to minimize the objective function, even if that decrease is small. The runtime of LP MILP varies widely as a function of the particular traffic demand matrix. The calculation time can be shortened by increasing the tolerance in accepting a solution as optimal (i.e., when the improvement in the objective function falls below this tolerance, the algorithm stops and takes the current best solution).

We compare the PB and LB MILP formulations when TR constraints are imposed and observe the same pattern in terms of required spectrum. Figure 6.3 compares the

[1]The simulation is run on Matlab with CPLEX v12.51 [57] as MILP solver engine and YALMIP [58] as interface on a single-core of Intel(R) Xeon(R) CPU X5550 at 2.67GHz with 16 GB of allocated memory.

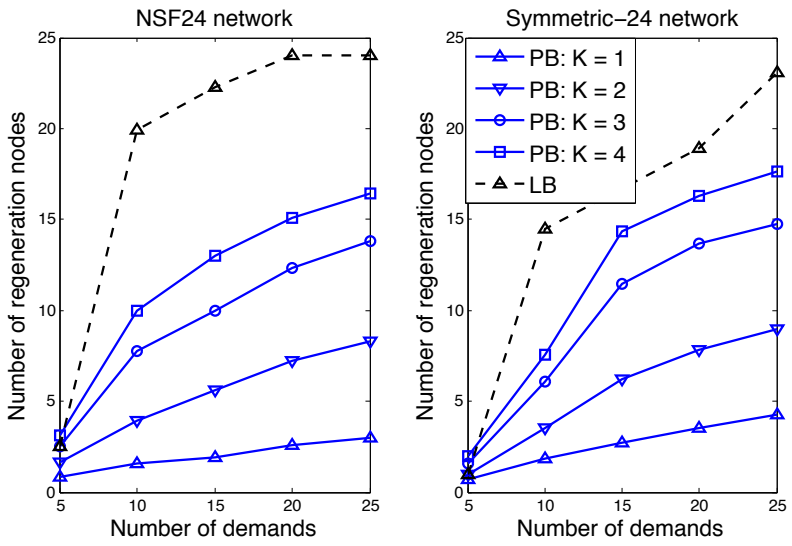

Figure 6.3: Required number of RNs for PB and LB MILP formulations with TR constraint, $\eta = 1$ bit/symbol.

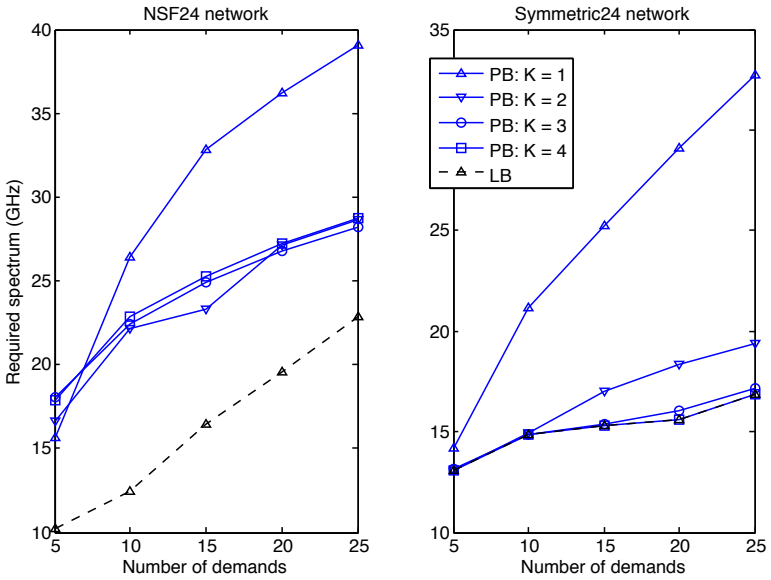

Figure 6.4: Required spectrum for PB and LB MILP formulations with optimized spectral efficiency $1 < \eta < 10$ bits/symbol, with TR constraints, $N_r = 24$

required N_r for the two formulations. We can see that when the traffic load increases, the LB approach employs more RNs to reduce the required spectrum, while the PB approach can only use as many RNs as were preselected for each path (the minimum number needed to satisfy (2.1), typically).

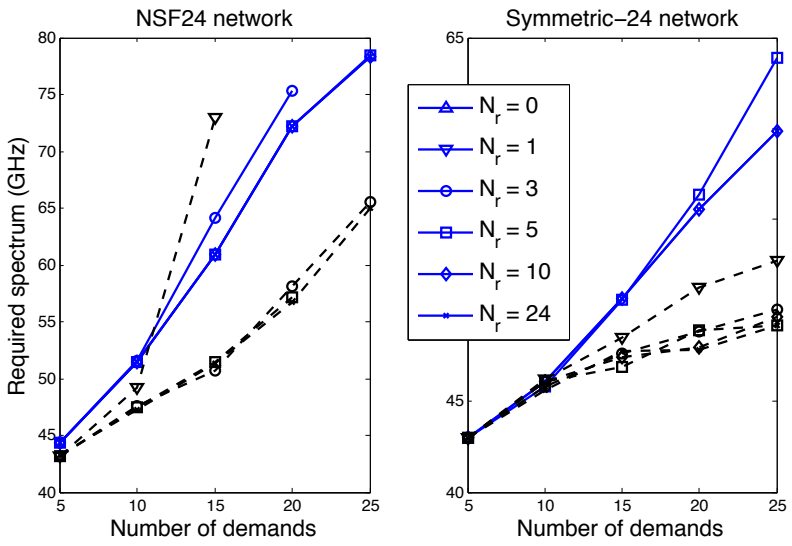

Figure 6.5: Required spectrum for PB and LB MILP formulations with limited regeneration resources, with optimized spectral efficiency $\eta = 1$ bit/symbol, with TR constraints (dash lines represent LB MILP)

When multiple modulation schemes are available at the transponders, the LB formulation can take advantage of this flexibility by adjusting the routing, as seen in Fig. 6.4. In the PB MILP, for each demand each modulation scheme results in a separate candidate path, with its pre-set number of RNs. Since the PB method is limited in the number of candidate paths it can consider, it cannot balance the traffic load as effectively as the LB approach.

The LB approach uses more RNs but achieves lower spectrum usage, and the PB approach results in the opposite (Figs. 6.3 and 6.4). Since it is not clear if spectrum usage or RNs are more expensive, in the next figure we compare the spectrum usages of the two approaches with a given number of allowed RNs. Figure 6.5 shows that when N_r is small, the RSA algorithm chooses a small spectral efficiency ($\eta = 1$) in order to extend the TR, allowing the signal to travel a longer distance without regeneration. Therefore, the spectrum required for each demand is large. In addition, when there are insufficient RNs the demands are forced to choose paths that require fewer regenerators (the shorter paths), restricting routing choices, which in turn gen-

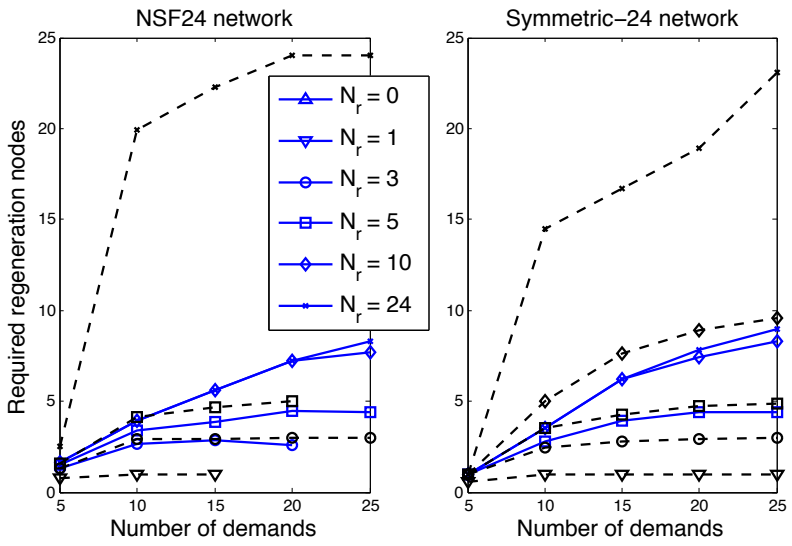

Figure 6.6: Required RN for PB and LB MILP formulations with limited regeneration resources, with optimized spectral efficiency $\eta = 1$ bit/symbol, with TR constraints (dash lines represent LB MILP)

erates congestion and increases the spectrum needed. Nevertheless, the LB MILP remains more flexible than the PB method. When all nodes are allowed to regenerate anywhere ($N_r = 24$), the required spectrum is minimized. Figure 6.6 shows that the LB MILP uses the maximum number of RN allowed in an effort to decrease the spectral usage, while the PB MILP is less efficient, often using less than the maximum number.

6.5 Summary

We compare the current PB MILP formulation to an LB MILP formulation for impairment-constrained elastic optical networks. The PB MILP formulation is limited in its ability to optimize the RSA because regeneration resources and modulation selection are predetermined. In general the LB MILP formulation requires less spectrum and fewer regeneration resources. The price paid for this savings is a significant increase in computation time compared with the PB method.

Chapter 7

GRWA for Dynamic WDM Networks using a Time-slotted MILP

In this chapter we return to WDM systems. Translucent fiber-optic networks have been carefully planned to achieve high capacity utilization efficiency as required by society's ever-increasing traffic demand. Existing research treats the problem of resource placement largely as a static design problem, which is solved with linear programming (LP) to find the optimal solution. The dynamic operational problem (grooming, regeneration, routing, and wavelength assignment) is approached using heuristic methods with the goal of improving the overall network performance given an existing network infrastructure. Our work combines these two approaches and solves a real-time dynamic traffic scenario with integer linear programming (ILP), seeking to maximize the overall network throughput. The traffic is served in a time-slotted fashion so that the network throughput is optimized at each time slot given the existing network state. The solution is compared with results from existing heuristic

methods. We incorporate physical impairment limitations into our network model, and consider several grooming options.

7.1 Introduction

Translucent fiber-optic networks need to be carefully planned at many different stages. Not only the capital investment should be minimized during network design and network upgrade, but also the service capability should be maximized once the equipment is deployed and real-time traffic demands arrive.

As we mentioned in previous chapters, many works have applied the powerful mixed-integer linear programming (MILP) tool to solve optical networking issues: [1], [43],[44], [45], [46],[47]. However, in these and other previous work when researchers use LP to address network problems, they only consider a static traffic model. The dynamically fluctuating nature of real-time traffic cannot be properly represented with this kind of model.

An effective alternative to using an LP method is to employ a heuristic method, which can handle dynamic environments. Heuristic methods usually have the advantage of a simple implementation, short calculation time, and good scalability. The drawback of heuristic methods is that they often do not provide optimal solutions. This is partly due to the fact they do not consider all possible solutions, and often pick one based on the network designer's understanding of the current network state.

In order to solve the network operation problem for dynamic real-time traffic and find a solution closer to the optimal solution, we propose to solve an ILP in a time-slotted fashion: for each time slot the network throughput is maximized while preserving as much of the resources as possible for future traffic. This requires a different objective function than previous LP work such as in Chapter 5 and 6. The dynamic traffic model differs from the static traffic model in that each call request

has a arrival time and a call duration, and the network capacity is updated continuously. Our approach bundles incoming traffic into time slots and solves the grooming, routing and wavelength assignment (GRWA) problem for those incoming calls optimally given the existing network state. We incorporate physical impairments (PIs) into our solution by imposing all-optical reach constraints. We also allow for multiple line-rates across the network. Another aspect that differentiates our approach from others is that we incorporate grooming into our solution in real time. This operation requires unique ILP constraints so that various grooming rules are enforced.

7.2 Implementing Traffic Grooming in MILP

Recall traffic grooming can currently only be done in the electrical domain, which implies that we can only perform grooming either at the source of a demand or at an OEO regeneration site of a demand, which are the only two places where the signal reverts to its electrical form. This restriction requires us to carefully track the wavelength availability of channels for grooming opportunities. A wavelength must be considered as unavailable to groom onto along an all-optical segment, where it remains in the optical domain. To address this constraint, a novel approach must be applied. It is essential in our formulation to clearly mark all-optical segments that include more than one physical link so that OEO conversion is forbidden within the segment. Our approach is to create a virtual link between the two ends of a segment whenever it is formed, while marking the original physical resource unavailable to all new traffic demands. The virtual link holds only the wavelength used in the segment and its free capacity is adjusted based on the channel usage on the segment. New arrivals only see the one wavelength on these virtual links, and it will appear as a single hop to prevent a call from attempting to groom mid-way. When the routing and wavelength assignment ILP uses this virtual link, it considers only solutions that

include utilizing that wavelength from the segment source to the segment end, but not partially in between. This ensures that grooming is only performed at points where signals undergo OEO conversion. This modification only lasts until the termination of the call that forms the segment.

7.3 Mathematical Model

To the author's knowledge, this is the first published work that solves the grooming problem for real-time traffic using an ILP. Using an ILP allows us to fully exploit the versatility and routing potential of the network by considering all possible solutions, as compared to heuristic methods that consider only a small set of candidate paths. Our approach also differs from ILP approaches that also restrict the number of candidate paths for each node pair, such as [59].

7.3.1 Time-Slotted Approach

The novelty of our approach stems partly from viewing the dynamic problem as a succession of small static problems with strong initial conditions (the current network state). We solve the real-time dynamic traffic situation by using a time-slotted approach, applying the ILP to each time-slot successively. Similar to the dynamic heuristic method, the network state is updated continuously and forms a base for the ILP execution for next time slot.

The pooling of traffic arriving within a time slot provides the algorithm flexibility to find an optimal GRWA solution for those calls, i.e., optimized for that time slot. In order to take the nature of the dynamic traffic into consideration, we split the time axis during which traffic arrives into fixed time slots, which creates a discrete-time approximate model of the real-time traffic, as shown in Fig. 7.1. In this approximation the demands is always serviced starting at the beginning of a time slot and

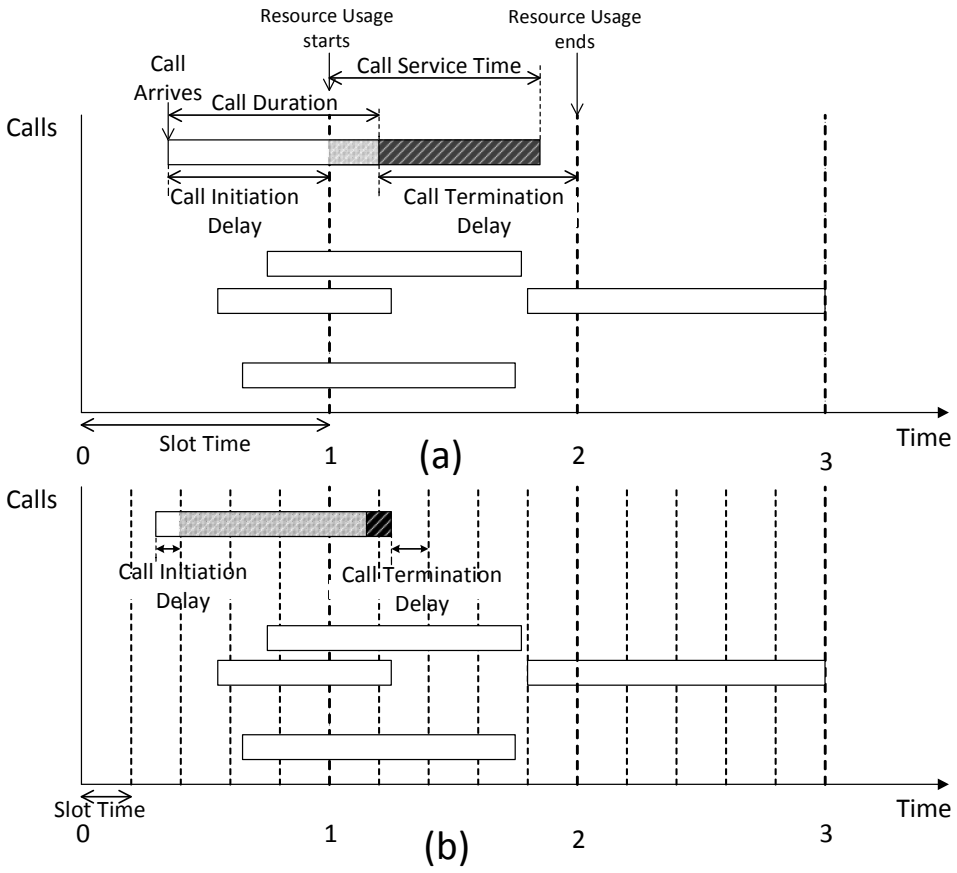

Figure 7.1: Illustration of time-slotted traffic model. Time slot duration of (a) 1 time unit and (b) 0.2 time units.

terminates at the end of a time slot. This can be thought of as introducing a certain amount of delay (no longer than the duration of the time slot) for the demand to be serviced, called the initiation delay. The result is also an over-allocation of resources for demands after their termination (termination delay), again no longer than the duration of the time slot. By doing this, we create an opportunity to find an optimal solution for all the calls that arrive within the same time slot. Fig. 7.1 illustrates the call initialization delay and call termination delay suffered by calls, which depends on the time slot duration. The time slot duration also affects the number of calls to be calculated per ILP solution. In this example, when the time slot duration is set to 1 time unit, Fig. 7.1(a), 5 calls arrive within 2 time slots (around 2.5 calls per time slot); when the time slot duration is set to 0.2 time units, Fig. 7.1(b), 5 calls arrive

within 8 time slots, which makes it on average 0.625 calls per time slot. The delay introduced in servicing calls can be considered a trade-off for system performance improvement.

At the beginning of each time slot, the calls that have arrived in the previous time slot are used as demand requests and the solution is found based on the current network state (wavelength usage, etc). The network state is then updated based on the ILP's GRWA solution. The available capacity of each wavelength is updated. The ILP enforces the restriction of no wavelength conversion for all-optical segments. Then, before the next time slot starts, call terminations are performed, where some resources are released and the network state is again updated. The above steps are performed iteratively until all traffic samples terminate. The overall goodput (fraction of the data rate request accepted) is calculated by dividing the sum data-rate of traffic accepted by the system by the sum data-rate of the traffic requested.

7.3.2 ILP Model

In this section we introduce the ILP model to optimize the network goodput by optimally aggregating sub-wavelength data traffic into high data-rate streams. The notation in Tables 7.1, 7.2, and 7.3 is used in our mathematical formulation. In particular we use the symbols i and j to index the head and tail of a physical link, and symbols s and d to index the source and destination of a demand.

Objective function:

$$\begin{aligned} \max & \sum_{sd \in D} Accept_{sd} \times Datarate_{sd} \\ + & \sum_{ij \in L, ld \in Ld} ULLd_{ij,ld} \times w_{ld} * ld + \\ & \sum_{ij \in L, ld \in Ld, sd \in D} U_{ij,ld,sd} \times w \end{aligned} \tag{7.1}$$

Table 7.1: Sets used by ILP

N^{sd}	Set of network nodes
nbN	Number of nodes in the network
N^{r}	Set of nodes that are able to regenerate signal
N^{nr}	Set of nodes that are not capable of regeneration or those with the capability but not selected for regeneration
L	Set of network links
D	Set of network demands between node pairs. These demands often ask for data rate lower than line rate of a wavelength
Ld	Set of available wavelengths on each fiber link.

Table 7.2: Parameters used by ILP

$A_{ij,ld}$	how much of channel capacity is available of each link ij and ld pair, $ij \in L, ld \in Ld$
R_n	the number of regenerators allocated at node n
Tr_{ld}	Transmission reach for each wavelength, this is based on the channel line rate of that wavelength
$Datarate_{sd}$	Data rate that is requested by demand sd in D
$Length_{ij}$	Link length of link ij in kms
C_{ld}	Channel capacity (i.e.: channel line rate) of each wavelength
$S_{n,sd}$	$S_{n,sd} = -1$ if node n is demand sd's source node; $S_{n,sd} = 1$ if node n is demand sd's destination node; otherwise, $S_{n,sd} = 0$.
w	small negative weight parameter ≈ -0.01 to be put on each link-lambda pair usage; therefore, the shortest path routing is followed when there exist multiple equivalent solutions (equal throughput).
w_{ld}	small negative weight parameter ≈ -0.001 put on wavelength ld. Different values are used to enforce preferences for particular wavelengths; for example, a higher w_{ld} is assigned to wavelengths with low channel capacity to allocate the wavelength with the minimum sufficient capacity to a demand first.

The objective function is designed to maximize the demands serviced while using as few network resources as possible. The network resource usage is used as a tie breaker when multiple solutions result in the same goodput.

Constraints:

Table 7.3: Variables used by ILP

$Accept_{sd}$	$Accept_{sd} = 1$ if demand sd is accepted; otherwise $Accept_{sd} = 0$
$U_{ij,ld,sd}$	$U_{ij,ld,sd} = 1$ if wavelength ld on link ij is assigned to demand sd
$UL_{ij,sd}$	$UL_{ij,sd} = 1$ if link ij is assigned to demand sd; otherwise $UL_{ij,sd} = 0$
$ULLd_{ij,ld}$	number of demands that share wavelength ld on link ij.
$\delta_{ij,sd,ld}$	physical distance from the head of link ij to the head of the segment for demand sd and wavelength ld. $\delta_{ij,sd,ld} = 0$ if wavelength ld on link ij is not assigned to demand sd.
$Y_{n,sd,ld}$	physical distance from node n to the head of the segment for demand sd and wavelength ld. $Y_{n,sd,ld} = 0$ if node n is not on the lightpath selected for demand sd.

Conservation flow constraint:

$$\sum_{ij \in L, j=n} UL_{ij,sd} - \sum_{ij \in L, i=n} UL_{ij,sd} = S_{n,sd} \times Accept_{sd}, \quad \forall n \in N, sd \in D \tag{7.2}$$

Link capacity constraint:

$$\sum_{sd \in D} U_{ij,ld,sd} \times Datarate_{sd} \leq C_{ld} \times A_{ij,ld}, \forall ij \in L, ld \in Ld \tag{7.3}$$

No call splitting:

$$\sum_{ld \in Ld} U_{ij,ld,sd} \leq 1, \forall sd \in D, ij \in L \tag{7.4}$$

Wavelength continuity constraint:

$$\sum_{ij \in L, j \neq s, j \neq d} U_{ij,ld,sd} = \sum_{ij \in L, i \neq s, i \neq d} U_{ij,ld,sd}, \quad \forall sd \in D, ld \in Ld \tag{7.5}$$

Transmission reach constraints, $\forall sd \in D, ij \in L, ld \in Ld$:

$$\begin{aligned}
\delta_{ij,sd,ld} &\leq U_{ij,ld,sd} \times Tr_{ld}, \\
\delta_{ij,sd,ld} &\leq Y_{i,sd,ld} \\
Y_{i,sd,ld} - \delta_{ij,sd,ld} &\leq Tr_{ld} \times (1 - U_{ij,ld,sd}) \\
Y_{n,sd,ld} &= 0, \forall n \in N^r \text{ or } n = s \\
Y_{n,sd,ld} &= \sum_{ij \in L: j=n} \delta_{ij,sd,ld} + Length_{ij} \times U_{ij,ld,sd}, \\
&\qquad \forall n \in N^{nr} \text{ and } n \neq s
\end{aligned} \tag{7.6}$$

Regenerators resource constraint:

$$\sum_{sd \in D, nj \in L, ld \in Ld} U_{nj,ld,sd} \leq R_n, \forall n \in N^r \tag{7.7}$$

Link-wavelength usage count:

$$ULLD_{ij,ld} = \sum_{sd \in D} U_{ij,ld,sd}, \forall ij \in L, ld \in Ld \tag{7.8}$$

Link usage with link-wavelength usage:

$$UL_{ij,sd} = \sum_{ld \in Ld} U_{ij,ld,sd}, \forall ij \in L, sd \in D \tag{7.9}$$

No grooming constraint (optional):

$$\sum_{sd \in D} U_{ij,ld,sd} \leq 1, \forall ij \in L, ld \in Ld \tag{7.10}$$

7.3.3 Explanation of Constraints

Equation (7.2) ensures that for transient nodes (neither the source nor the destination for the demand considered) the traffic entering the node equals the traffic exiting the node. For end nodes, the traffic that exits the source node and enters into the destination node equals the demand's datarate request. Equation (7.3) ensures that, for the wavelength (lambda) considered, the traffic volume assigned to new demands is less than the available capacity. Note that a link-lambda pair could be used for one demand and later also assigned to another demand if there is capacity left over; so $A_{ij,ld}$ ranges from 0 to 1. Equation (7.4) ensures that each demand request can only be satisfied by at most one lambda on any link. This assumption is solely for the sake of model simplicity, and could be removed to consider an expanded version of the problem. Equation (7.5) ensures that for each segment the optical signal maintains the same lambda. The constraint is placed on the nodes that do not have regeneration capability (also on nodes with regeneration capability if the regeneration function is not used). For these nodes the traffic that enters and the traffic that exits the node for a demand use the same lambda.

Equations (7.6) ensure that the optical signal will not travel further than the transmission reach without regeneration: if a link is assigned to a demand, its head node should be within the transmission reach from the segment's head node. The distance from the segment head node to a link ij should be no greater than the distance to the node i. If a link is assigned to a demand ($U_{ij,ld,sd} = 1$), $Y_{i,sd,ld}$ (abbreviated Y) and $\delta_{ij,sd,ld}$ should be equal. If it is not assigned to the demand, Y should be within the transmission reach. For the source node of the demand or a node performing a regeneration operation for that demand, Y is 0. This is how the Y's are derived from the source of the demand. If a link is assigned to the demand, its end node Y is calculated by adding the physical link distance to the Y value of its

head node. For any node along the lightpath for a demand, the Y should be within the transmission reach.

Equation (7.7) enforces that the total number of OEO regenerators at any given regeneration node at any given time slot is no greater than the number of regenerator circuit the node has. Equations (7.8) and (7.9) show the relation between the three variables we use in our programming. Equation (7.10), which enforces how many demands can share a wavelength ld on a link ij, is optional and is used to compare the grooming case versus the no-grooming case.

7.4 Numerical Results

Numerical simulation experiments are conducted on a 14-node bidirectional-link NSF nationwide network, shown in Fig. 2.2. Each link has 8 wavelengths, and each wavelength supports either a 10, 40 or 100 Gbps line-rate and has unlimited, 2500 and 2000 kms TR, respectively. The ILP is solved using the optimizer CPLEX [57]. The demand requests arrive at the system following a Poisson process and are uniformly distributed among node pairs. The call duration follows an exponential distribution with a mean value of 1 (arbitrary time units). For Figs. 7.2-7.7, the time slot duration is 0.1 time units. The data-rate requests for each demand follow a uniform distribution ranging from 1 to 30 Gbps.

Since in real-time scenarios the processing time is critical, the ILP solve-time is limited to 100 seconds (wall-time) on a desktop computer; when the calculation time exceeds this limit, the best feasible solution at that time is selected.[1] The computation time is therefore significant when the call-durations are short compared to this delay. With optimized processing the delay may be reduced significantly, yet the algorithm remains complex for large networks. The proposed technique, as stated, is suitable

[1]We tested with different simulation time constraints and noticed that while it is important to have a solve-time sufficiently long to find at least one feasible solution, finding the optimal solution does not improve the system performance significantly.

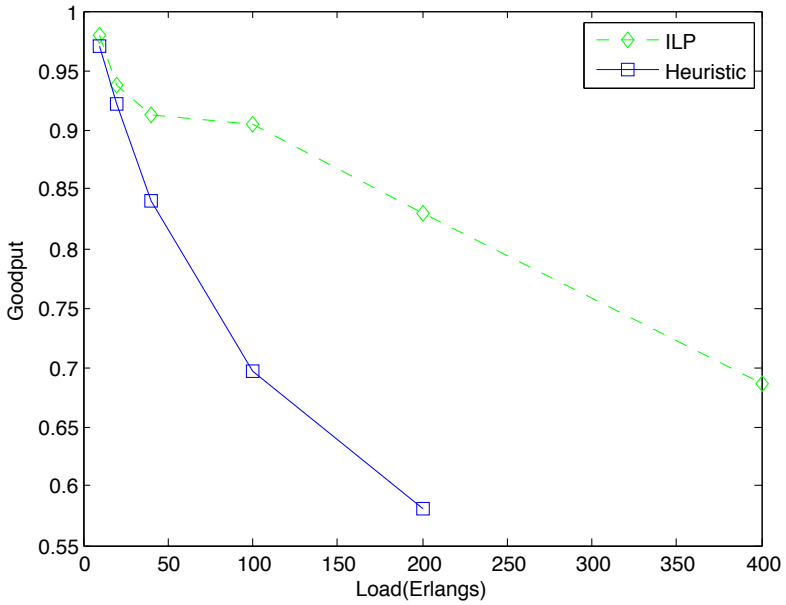

Figure 7.2: Goodput using ILP and heuristic method versus load

to calls with long service time requirements, such as large file transfers and other e-science applications.

We compare the solution from solving the ILP with the solution from a previously proposed heuristic method in Chapter 3. For a high load case, we can see significant improvements in terms of both network goodput, Fig. 7.2, and call blocking probability, Fig. 7.3. This is because the ILP considers all possible lightpaths given by the network structure while the heuristic method often considers the shortest path or just a few alternatives before it rejects the request. It is obvious that the network operator benefits from the more comprehensive search within the solution space. From the plots we can see that for low loads the ILP method provides little advantage since the low traffic rarely causes congestion.

The grooming case outperforms the no-grooming case as the load increase, as seen in Figs. 7.4 and 7.5. In the no-grooming case, a channel is set as unavailable once assigned to a demand, no matter how much capacity is still available. If the call cannot groom with others, there is also a smaller chance for it to find a feasible

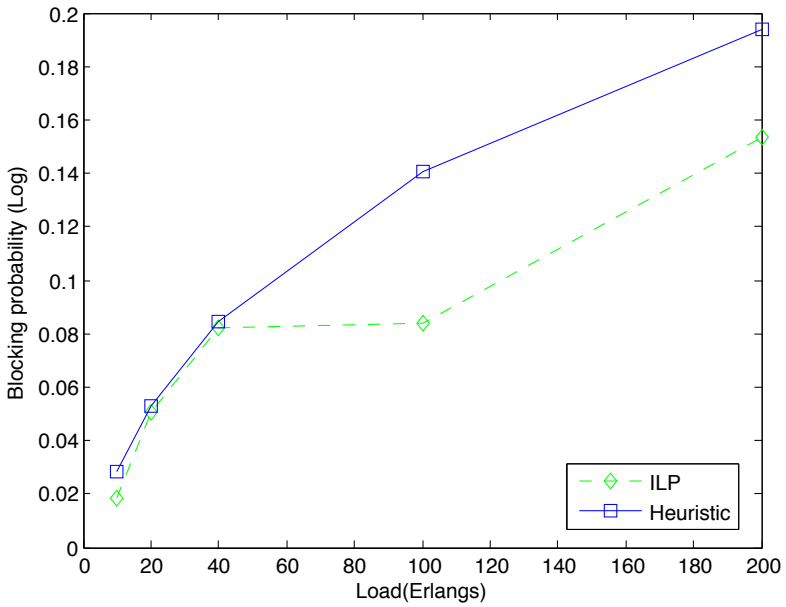

Figure 7.3: Blocking probability using ILP and heuristic method versus load

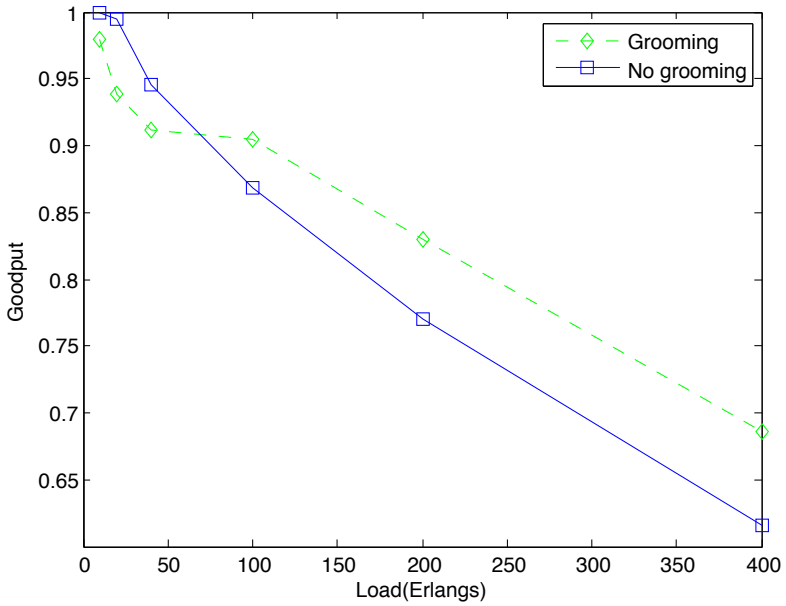

Figure 7.4: Goodput using grooming and no-grooming methods versus load

lightpath when the network is heavily loaded. That is why we can see the grooming case outperforms the no-grooming case in our ILP solution.

For some cases (for low load) the performance improvement is not obvious or sometimes reversed. This is because the time-slotted ILP solution is based on an

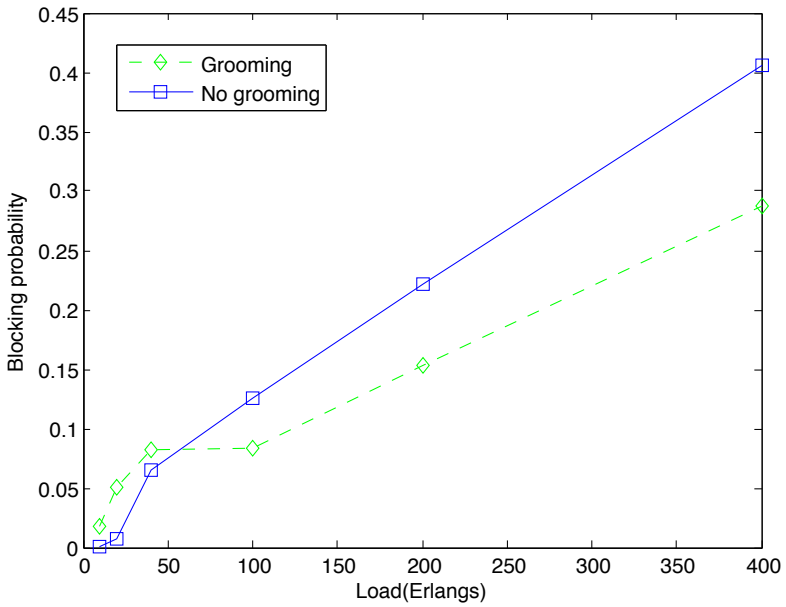

Figure 7.5: Blocking probability using grooming and no-grooming methods versus load

objective function to maximize the single time slot throughput, rather than the whole simulation throughput. In the future, we will consider improvement to address this issue.

In Figs. 7.6 and 7.7, we compare the difference when regeneration is allowed versus not allowed, i.e., between a transparent and a translucent network. If no regeneration is possible, the physical impairments put a restriction on the signal transmission reach, which then makes certain node pairs unreachable for high line-rates. The regeneration operation basically extends the transmission reach for each line-rate lightpath, thereby introducing more candidate solutions. For node pairs that are too far apart in terms of physical distance, it is possible to create a lightpath for some line-rates with the signal being regenerated somewhere along the lightpath. For heavily loaded networks, the direct connection (the shortest path) between a node pair may be unavailable, and the demand may be required to be rerouted over a longer lightpath, where the signal quality degradation could then become an issue. Regeneration makes rerouting possible.

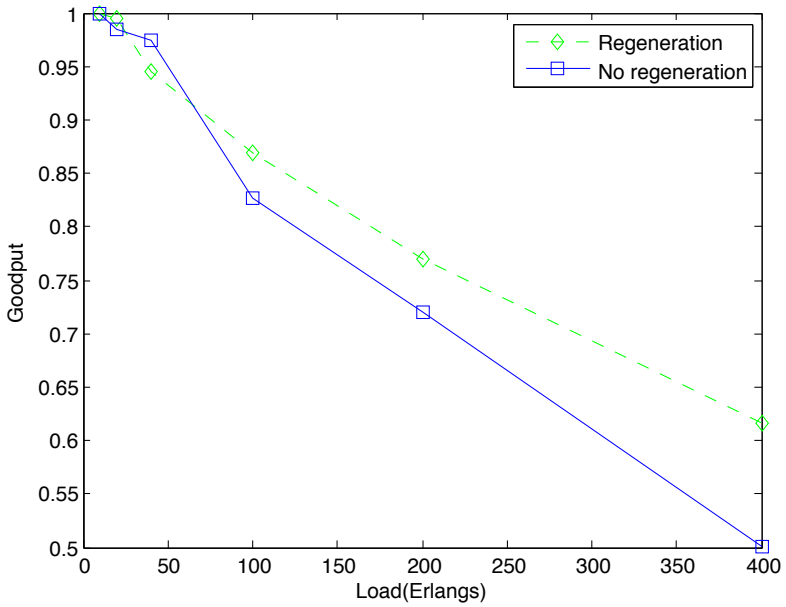

Figure 7.6: Goodput using regeneration and no-regeneration methods versus load

Another important impact on performance caused by regeneration is that it provides opportunities for grooming. When a signal undergoes OEO conversion, it can then be multiplexed with other streams to become a higher data-rate stream, which improves the channel capacity usage efficiency. That is why researchers propose to intentionally control TR of some line-rate signals to force them through OEO conversions, to take advantage of these grooming opportunities.

Note that although obvious differences in performance exist between different methods, the difference is not large. Since the ILP considers all options, the network is highly efficient and can handle large volumes of traffic.

In Table 7.4 we compare the system performance for a traffic model with load of 100 Erlang using different time slot durations. The value of the time slot duration can be considered as the maximum delay of each call before it can be processed. From the ILP's point of view, it affects the number of simultaneous demands for service. On average, the number of calls per time slot is equal to the load (Erlang) times the slot duration divided by the call duration (arbitrary units of time). As the slot size

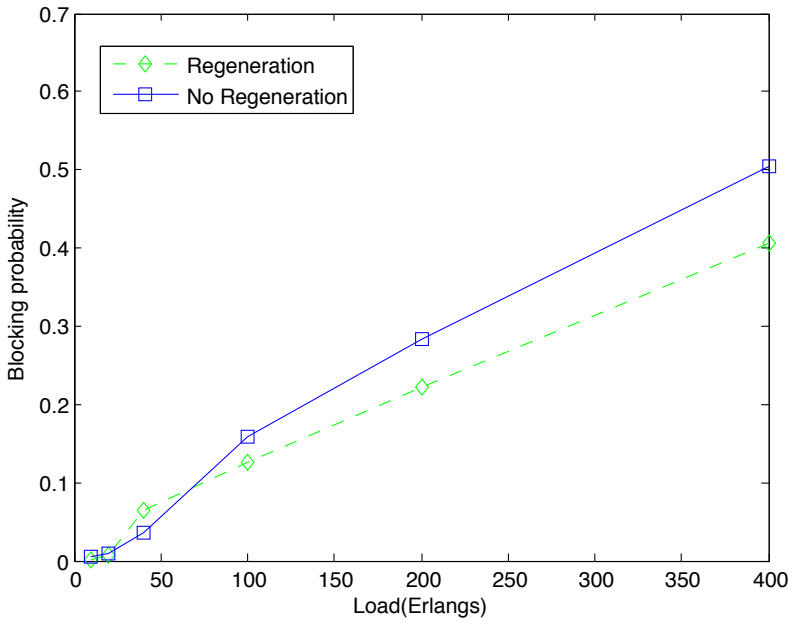

Figure 7.7: Blocking probability using regeneration and no-regeneration methods versus load

increases, the number of calls to allocate per time slot increases. The ILP solution calculated with more calls results in a better solution, since the optimization only considers new calls (not calls established in previous slots). However, the increased time slot duration wastes resources since the termination of calls is delayed until the end of each time slot. Based on these two reasons, we can explain why the performance is similar with different time slot durations, as seen in the table.

Table 7.4: Effects of Time Slot Duration, Load of 100 Erlang

Time slot duration	0.2	0.1	0.05	0.03
Goodput	0.92	0.91	0.88	0.91
Blocking probability	0.09	0.08	0.05	0.09

7.5 Summary

We propose an approach to solve the dynamic real-time traffic GRWA problem using an ILP in a time-slotted fashion. The traffic only endures a short delay before

being processed. Meanwhile, the system efficiency is increased compared to heuristic methods, both in goodput and call blocking probability. Grooming and regeneration effects on the network performance are also discussed.

Using an ILP in a dynamic system may become too computationally burdensome to be scalable as the network size increases. In future work we will consider relaxation techniques for obtaining faster near-optimal solutions using this approach.

Chapter 8

Estimation of Network Performance Through Analytical Modeling

8.1 Introduction

In previous chapters, we propose several heuristic algorithms and MILP formulations for solving the *routing and wavelength assignment* (RWA) for wavelength division multiplexed (WDM) networks and *routing and resource allocation* (RSA) problems for elastic optical networks (EON). In particular, we are interested in the impact of technologies such as signal regeneration, wavelength conversion, and modulation conversion on network performance metrics such as the amount of spectrum needed. It is desirable for network designers to begin with a fast coarse estimate of network performance implementing such technologies without requiring computationally burdensome and detailed algorithms. In this chapter, we create an analytical model to provide such an estimate.

Analytical models have been proposed for network design problems to show insight on the relationship between network parameters and to generate preliminary approximations of network performance much more quickly than can be obtained by simulation. In [60] the authors analyze the impacts of wavelength conversion on a translucent optical network. [61] develops an approximation for the traffic grooming problem in optical networks. Korotky first introduced the *network global expectation model* in [62], and derives a formula to represent network cost. In [63] he analyzes the impact on network capacity using mixed-bit-rate transmission with this model. The authors in [64] use the network global expectation model to estimate the number of channels used given a certain traffic load, then use the number of channels to calculate the impact of physical impairments and analyze the quality of transmission issue. [65] uses a different approach: they develop an analytical model for dynamic networks by implementing a finite state machine.

In this chapter we propose an analytical model to estimate the network performance as measured by the average link spectrum usage, highest required spectrum, and network capacity for physically impaired EONs, when the aforementioned technologies (signal regeneration, wavelength and modulation conversion) are implemented. To the best of our knowledge, there is no published analytical model that considers these effects on optical networks. We therefore compare the results of our analytical model with results from simulation. Our analytical models closely approximate results obtained via simulation, with modeling accuracy depending on the network topology and how much information is known about the routing protocol. In comparison with our one-shot probabilistic approach, over 100 random trials are needed for the simulations to converge, resulting in over 30× longer calculation time. We first apply our analytical model on a symmetric 36-node mesh network as shown in Fig. 2.4, then we apply the model on a real world network, the NSF-24 network shown in Fig. 5.2.

8.2 Analytical Model

We model a network with sparse signal regeneration by assuming that a node is capable of signal regeneration with probability q independently of the other nodes. We also assume that all nodes are capable of converting the wavelength to avoid any wavelength contention, i.e, the wavelength continuity constraint is not enforced. The case where wavelength continuity is required is beyond the scope of this work. The traffic demands on the network are from every node to every other node, and the goal is to estimate the minimum spectrum needed when the bit rate of demands are fixed and to maximize the throughput the network can handle when bit rate is not fixed but there is a limit of bandwidth for the transponders.

8.2.1 Expected Value of Longest Segment Length

Figure 8.1 illustrated a lightpath that consists of two transparent segments that represent portions of the lightpath where the signal remains in the optical domain. Optical signals are regenerated at node 4. Recall from Section 2.3 in Chapter 2 that both the transmission reach (TR) and the bandwidth used depend on the bit rate and spectral efficiency of the traffic. Therefore, when the bit rate and spectral efficiency are fixed (no modulation conversion), the reach is also fixed. In order to satisfy the quality of transmission (QoT) for the lightpath, the TR constraint has to be satisfied for all transparent segments, and thus the TR has to be longer than the longest segment. Therefore, when modulation conversion is not available (i.e., spectral efficiency is fixed), the spectral efficiency is limited by the longest transparent segment of the two (in this case segment #1).

In order to find the longest segment, we first derive the expected value of the length of the first transparent segment. Let the number of hops on a path be denoted as the random variable H. The probability that the node at hop k on a path p is

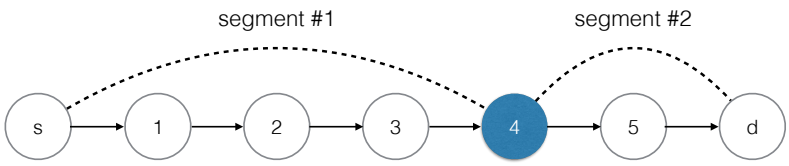

Figure 8.1: A lightpath that consists of two transparent segments with node 4 as regeneration node

assigned as the first regeneration node for this path is $(1-q)^{k-1}q$ for $k < H$. Since the lightpath has a limited length, the expected value of the first transparent segment length (in number of hops), S_1, is then

$$E[S_1 \mid H = h] = \sum_{k=0}^{h-2} k(1-q)^k q + (h-1)(1-q)^{h-1} + 1 \tag{8.1}$$

The first term includes cases when any node from node 1 to nodes $h-1$ is a regeneration node; the second term is when there is no regeneration node; the constant 1 at the end is because k in this equation represents the number of non-regeneration nodes before the first regeneration node, and starts at $k = 0$.

The geometric distribution model can only generate the expectation length of the first transparent segment of the path given its hop count, not the longest path. In our example in Fig. 8.1, we have to take into account the probability that the second transparent segment (segment #2) is longer than transparent segment #1. In order to find the expected value of the length of the longest transparent segment, we derive a new model based on enumeration of the possible outcomes of the random regeneration assignment. We follow the stars and bars approach[1] [66] and derive the formula for a path of h hops as follows. Define the longest transparent segment as a random variable $S^{(\ell)}$, and let the number of regeneration nodes on the path be labeled

[1]The stars and bars method stems from a combinatorial mathematics theorem to solve simple counting problem such as how many ways there are to put N indistinguishable balls into K distinguishable bins.

$N^{(r)}$. Then,

$$\Pr\{S^{(\ell)} = k \mid H = h\} = \sum_{i=0}^{h-1} \Pr\{S^{(\ell)} = k \mid N^{(r)} = i, H = h\} \Pr\{N^{(r)} = i \mid H = h\} \tag{8.2}$$

$$\Pr\{S^{(\ell)} \leq k-1 \mid N^{(r)} = i, H = h\} = \frac{\sum_{j=0}^{i+1} (-1)^j \binom{i+1}{j} \binom{h-1+j(k-1)}{i}}{\binom{h-1+i}{i}} \tag{8.3}$$

$$\Pr\{N^{(r)} = i \mid H = h\} = q^i (1-q)^{(h-1-i)} \binom{h-1}{i} \tag{8.4}$$

In (8.3), we show the conditional probability that the longest transparent segment has at most $k-1$ hops given there are i regeneration node on a h hop path. Similarly we can express the probability that the longest transparent segment has no more than k hops. Then, we can find the probability that the longest transparent segment has exact k hops by subtracting the two. Combining (8.2), (8.3), and (8.4), we derive the expected value of the length of the longest transparent segment for any path given the probability of regeneration node assignment q.

$$E[S^{(\ell)} \mid H = h] = \sum_{k=1}^{h} k \Pr\{S^{(\ell)} = k \mid H = h\} \tag{8.5}$$

For example, for a path length of $H = 10$, the expected value of the longest transparent segment expressed as the number of hops for different regeneration probabilities q is given in Fig. 8.2. In this figure, we also show the results from simulation, averaged over different numbers of trials. Notice the necessity of having sufficient simulation trials in order for the results to converge.

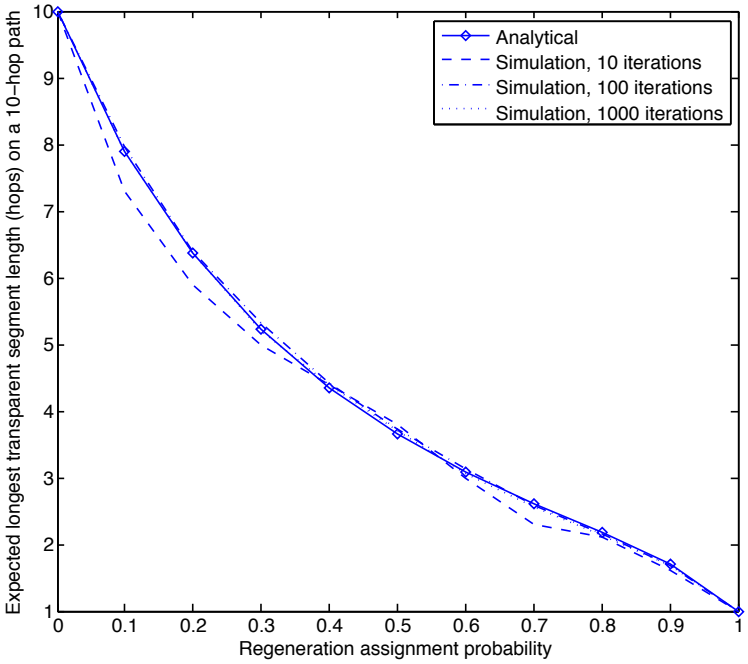

Figure 8.2: Expected longest transparent segment length for different regeneration assignment probabilities q for a path of 10 hops

8.2.2 Expected Number of Segments of Length k

When modulation conversion is available at a regeneration node, the spectral efficiency for each demand on any link doesn't depend on the longest segment on the path anymore. Instead, it depends on the length of the transparent segment that the link belongs to. Compared to the no-modulation conversion case, the problem is more complex in that we need not only the segment length, but also its distribution.

Let us assume there are $N^{(r)} = i$ regeneration nodes on a h-hop path, which implies there are $i + 1$ transparent segments. Then let $Y^{(k)}$ be the number of segments that have k hops. In order to find $E[Y^{(k)}|H, N^{(r)} = i]$, we define the indicator variables $X_m^{(k)}$ so that $X_m^{(k)}$ is 1 if the mth segment has k hops and 0 otherwise. Then we have $Y^{(k)} = X_1^{(k)} + X_2^{(k)} + ... + X_{i+1}^{(k)}$. By linearity of expectation, $E[Y^{(k)}|H, N^{(r)} = i] = E[X_1^{(k)}|H, N^{(r)} = i] + E[X_2^{(k)}|H, N^{(r)} = i] + ... + E[X_{i+1}^{(k)}|H, N^{(r)} = i]$. So now the

problem reduces to calculating the $E[X_m^{(k)}|H, N^{(r)}]$ for each m,

$$\begin{aligned} E[X_m^{(k)}|H = h, N^{(r)} = i] &= \Pr\{X_m^{(k)} = 1|H = h, N^{(r)} = i\} \\ &= \Pr\{\text{segment m has k hops}|H = h, N^{(r)} = i\} \\ &= \binom{h-1-i}{k-1} \left(\frac{1}{i+1}\right)^{k-1} \left(1 - \frac{1}{i+1}\right)^{h-i-k} \end{aligned} \tag{8.6}$$

Therefore,

$$\begin{aligned} E[Y^{(k)}|H = h, N^{(r)} = i] &= (i+1) \Pr\{\text{segment m has k hops}|H = h, N^{(r)} = i\} \\ &= \binom{h-1-i}{k-1} \left(\frac{1}{i+1}\right)^{k} \left(1 - \frac{1}{i+1}\right)^{h-i-k} \end{aligned} \tag{8.7}$$

Knowing the probability that a h-hop path has i regeneration nodes, (8.4), we get the expected value of the number of segments that has k hops for a h-hop path given the regeneration assignment probability q. For each segment length, assuming all links have the same length, we can calculate the spectral efficiency according to the reach and bit rate of the demand using the transmission reach formula (2.1). Once we find the expected number of segments for all k's, we can derive the spectrum usage for the network.

8.2.3 Path Hop Density Function

Since in our analytical model, we analyze the network performance on a path basis, we want to distinguish the paths according to their length (number of hops). The hop density function $f_H(h)$ represents the fraction of paths in the network of each path length. Note that for a particular topology, we can easily find the hop density function of the number of hops per path for a certain routing algorithm (e.g., Dijkstra's shortest path routing algorithm for static shortest path routing).

We derive the expected number of transparent segments with length k on a h hop path in the section above. Given the path hop density, the expected number of transparent segments of length k over the whole network is then simply

$$E[Y^{(k)}] = \sum_{h} f_H(h) \sum_{i=0}^{h-1} E[Y^{(k)}|H = h, N^{(r)} = i] \Pr\{N^{(r)} = i|H = h\} \tag{8.8}$$

8.3 Performance Measures

In this section we derive performance metrics to estimate the resources needed and the capacity of the network. In this section we test our expressions on the 36-node symmetric network shown in Fig. 2.4.

8.3.1 Spectrum Usage for One Demand

In Section 8.2, we derive two important expected values: the expected value of the length of the longest transparent segment and the expected number of segments of length k. When modulation conversion is not available, the transmission reach T_r is the length of the longest transparent segment. If modulation conversion is available, the transmission reach enforced on each link is the length of the segment it belongs to.

We use the transmission reach formula for demand d given in Chapter 2, (2.1). According to the relationship between bandwidth B, bit rate R_d and spectral efficiency, $\eta_d = R_b/B$, the transmission reach formula can be re-written as

$$T_r = \frac{\alpha + \beta B}{R_d} + \gamma \tag{8.9}$$

i.e.,

$$B = \frac{R_d(T_r - \gamma) - \alpha}{\beta} \tag{8.10}$$

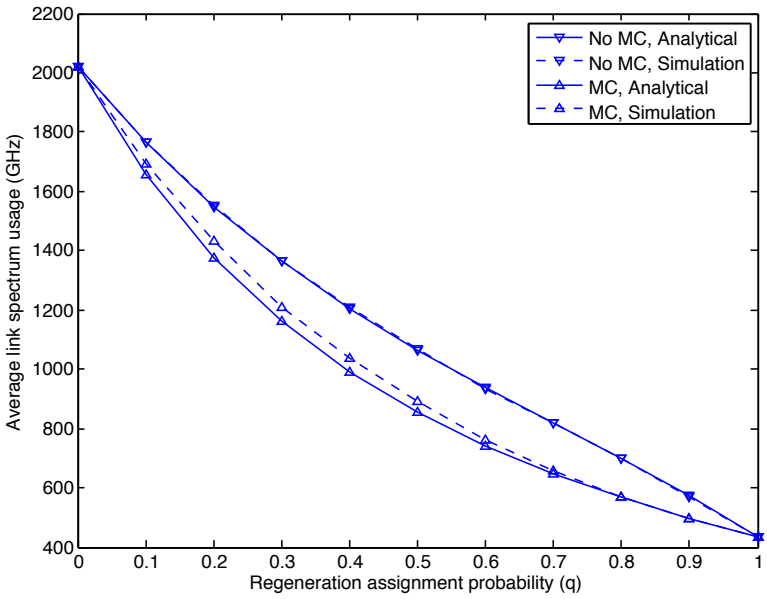

Figure 8.3: Comparison of average link spectrum usage between simulation and analytical model with and without modulation conversion

Using as the transmission reach either the longest segment for the path (if no modulation conversion is used) or each segment length on the path separately (if modulation conversion is used), this equation gives us the spectrum used by demand d. Summing up the spectrum usage of each demand and then dividing by the number of links, we can get the average link spectrum usage of the network shown in Fig. 8.3.

8.3.2 Highest Spectrum Required on a Link

Again assuming we know the routing table and the traffic demand matrix, we can find the highest required spectrum on any link by identifying the link that is shared by the most number of paths assuming each demand uses the same amount of bandwidth. We define $F^{(max)}$ as the highest spectrum usage on any link of the network; call that link $l^{(max)}$. If no modulation conversion is used, the spectrum used by the path

depends on the longest segment of the path and

$$F^{(max)} = \sum_{p:l^{(max)} \in p} \frac{R_d(E[S^{(\ell)}|H = h_p]L - \gamma) - \alpha}{\beta}. \tag{8.11}$$

where h_p is the number of hops of path p, and L is the link length and assumed to be the same for all links.

If modulation conversion is used, since we do not know the length of the segments that share the link, the best we can do is to use the average segment length. First we calculate the total spectrum usage of a path p with length $H = h_p$ as

$$B_p = \sum_{k=1}^{h_p} \sum_{i=0}^{h_p-1} \binom{h_p - 1 - i}{k-1} \left(\frac{1}{i+1}\right)^k \left(1 - \frac{1}{i+1}\right)^{h_p-i-k} \frac{R_d(kL - \gamma) - \alpha}{\beta}. \tag{8.12}$$

Then,

$$F^{(max)} = \sum_{p:l^{(max)} \in p} \frac{B_p}{h_p} \tag{8.13}$$

This is based on the assumption that all paths use about the same spectrum, which may or may not be accurate depending on the load balance.

In Fig. 8.4 we show a comparison between the analytical model and simulation results when no modulation conversion is available. In Fig 8.5 we show results for the modulation conversion case. When there is modulation conversion, the spectrum usage on any link not only depends on the paths that share the link, but also depends on the segments within those paths that share the link. Since with our probabilistic model we cannot identify which segment of each link shares the link (actually what we are interested in is the segment length), we use the average segment length for calculating the spectrum usage. However, in reality the link that requires the most spectrum may change due to the assignment of segment lengths. Such variation cannot be shown using our model. The inaccuracy is greater compared to the no modulation conversion case. As expected, the more information one has about the

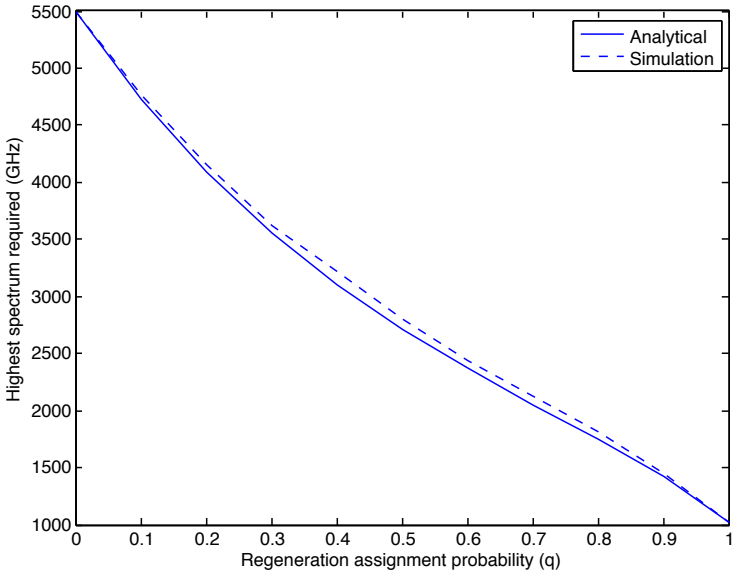

Figure 8.4: Comparison of highest spectrum required between simulation and analytical model without modulation conversion

network, the more accurate the prediction. The highest required spectrum (i.e., when $q = 0$, so the length of the segments that share the link all equal the path length) can be seen as the upper bound for the spectrum required on the network. This is essentially a transparent network. The lowest required spectrum (i.e., when $q = 1$, so the length of the segments that share the link all equal one hop) can be seen as the lower bound for spectrum requirement. This is essentially an opaque network.

8.3.3 Capacity of the Network

Assuming each optical transponder has a limit on the spectrum it can use to transmit signals, we can use our model to calculate the total capacity of the network. Results in this section assume each demand can use no more than 50 GHz.

Following (2.1) in Chapter 2, we have

$$R_d = \frac{\alpha + \beta B}{T_r - \gamma} \tag{8.14}$$

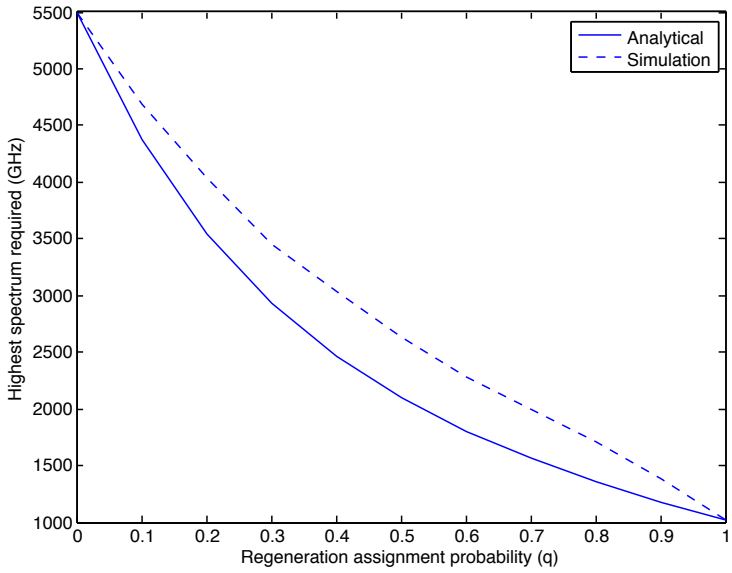

Figure 8.5: Comparison of highest spectrum required between simulation and analytical model with modulation conversion

We know the relationship between T_r and the segment length, and thus can find a limit on R_d, which gives us the network capacity as $C = \sum_d R_d$, where the summation is over all demands in the demand matrix.

For the case where modulation conversion is not used,

$$C = \sum_{d} \sum_{k=1}^{h_d} \Pr\{S^{(\ell)} = k \mid H = h_d\} \frac{\alpha + \beta B}{kL - \gamma} \tag{8.15}$$

where h_d is the number of hops on the path used by demand d. For the case with modulation conversion, the network capacity is the same, since the bit rate of the demand is limited by its longest transparent segment.

In Fig. 8.6 we show how the capacity of the network changes with the regeneration assignment probability of the nodes. Notice that the two curves align closely. This is because when all links are with the same length the estimation of the length of the longest segment is very accurate as shown in Fig. 8.2. In the following section, we discuss our analytical model on a real topology NSF-24 network.

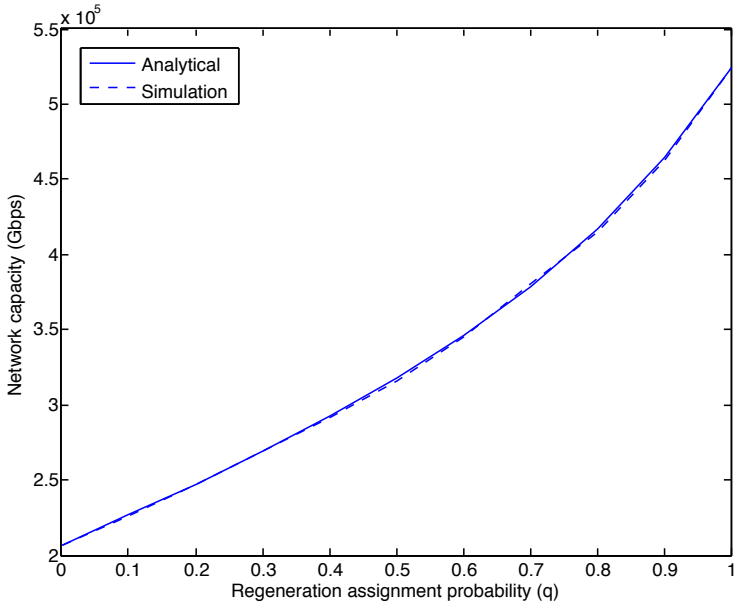

Figure 8.6: Comparison between simulation and analytical model of the network capacity assuming a per-demand bandwidth limit of 50 GHz

8.4 Real Topology

Consider a real topology such as the NSF-24 network shown in Fig. 5.2. Also assume shortest path routing and first fit (FF) wavelength selection. We can first find the hop density function of the paths $f_H(h)$, and then, following (8.8) above, calculate the expected number of segments with given length h. And finally, we can calculate the average link spectrum usage using (8.10) or (8.12), the highest spectrum required using (8.11) or (8.13) depending on whether modulation conversion is implemented or not, and the network capacity using (8.15).

8.4.1 Average Link Spectrum Usage for the NSF-24 Network

In Figs. 8.7 and 8.8, we compare the average link spectrum usage obtained using the analytical model and simulation for the NSF-24 network. Since we don't know exactly the links contained in each segment, we can only approximate the segment length. We tested two cases: (1) we use the average link length over all links in the

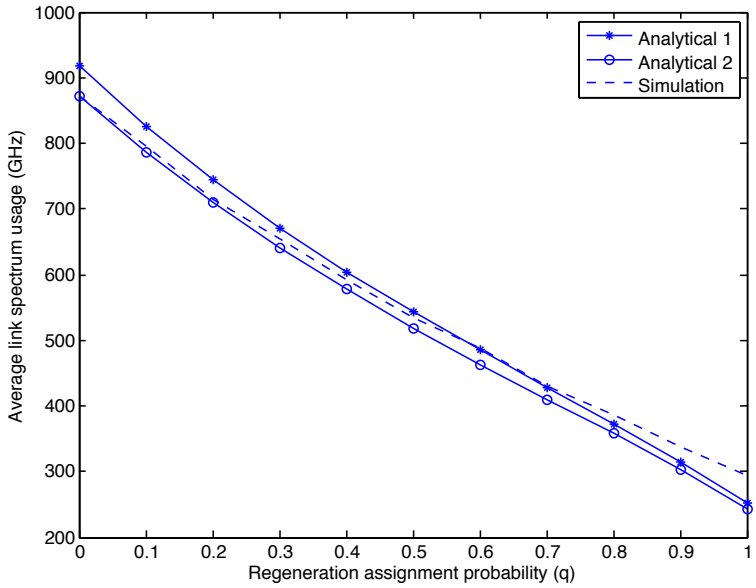

Figure 8.7: Average link spectrum usage comparison using average link length over all links (analytical 1), average link length over single path (analytical 2), and simulation results, without modulation conversion

network (simulating the scenario when links length assignment is not known) and (2) we average the link length over the links in the specific path (simulating the scenario when links length assignment is known). Results show that when we approximate the link length using method (1), we overestimate the average link spectrum when q is small, and underestimate the average link spectrum usage when q is large. When we approximate the link length using method (2), the results between the analytical model and simulation is close when q is small, and gradually increases when q reaches 1. Note that since we use the average link length as the length of all links, when no modulation conversion is allowed, we always underestimate the length of the longest segment; therefore, in Fig. 8.7, the analytical model generally yields a smaller average link spectrum usage than simulation results.

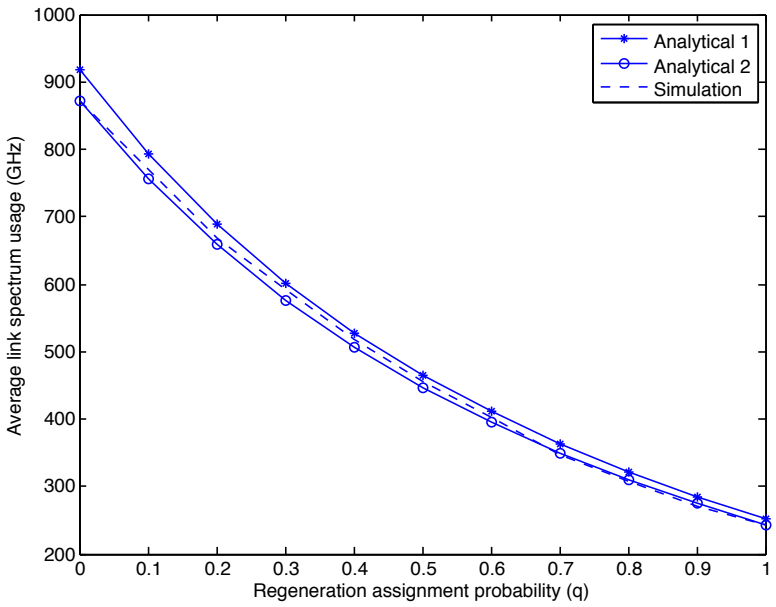

Figure 8.8: Average link spectrum usage comparison using average link length over all links (analytical 1), average link length over single path (analytical 2), and simulation results, with modulation conversion

8.4.2 Highest Spectrum Required for a Link for the NSF-24 Network

In Figs. 8.9, we compare the highest spectrum required between the two analytical cases and simulation with modulation conversion. Results show that the difference between estimation and simulation is greater using method (2) than method (1). In method (2), the inaccuracy in the link length approximation accumulates over all the demands that share the link and therefore, causes a greater gap. In method (1), however, the transparent segment length is underestimated for some demands and overestimated for others. Therefore the analytical results are closer to simulation results.

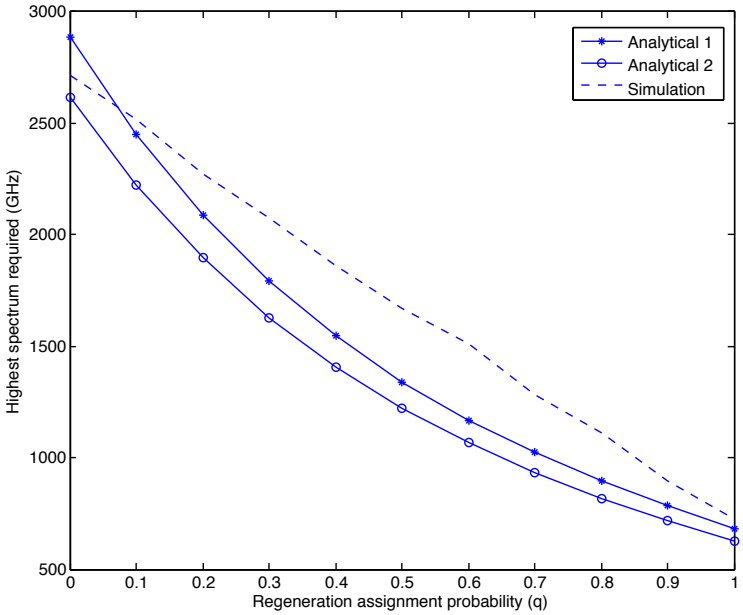

Figure 8.9: Highest required spectrum comparison with modulation conversion using average link length over all links (analytical 1), average link length over single path (analytical 2) and simulation results

8.4.3 Network Capacity on a Link for the NSF-24 Network

In Fig. 8.10, we compare the network capacity between the two analytical cases and simulation. Results show that when we approximate the link length using method (1), we underestimate the network capacity when q is small, and overestimate the network capacity when q is large. When we approximate the link length using method (2), the results between the analytical model and simulation is close when q is small, and gradually increases when q reaches 1. When $q = 0$, there is only one segment for each path, therefore, the estimation of method (2) is the same as simulation. When $q = 1$, the longest segment on the path depends is the longest link on that path, therefore, using average link length as the length of all links in method (2) underestimates the longest segment length and overestimate the network capacity. Method (1) on the other hand, underestimates the longest segment length for paths with longer links and overestimate it for paths with shorter links.

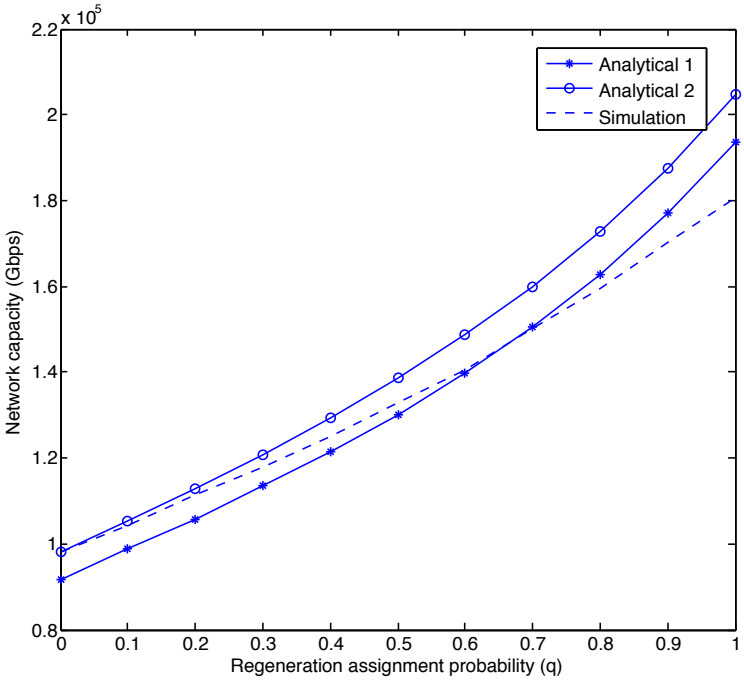

Figure 8.10: Network capacity comparison with modulation conversion using average link length over all links (analytical 1), average link length over single path (analytical 2), and simulation results

8.5 Complexity and Accuracy

In order to show the trade-off between computational complexity and accuracy, we compare our analytical model with simulations using different numbers of trials in Figs. 8.11 and 8.12. Results from the analytical model are consistent with the simulation results based on 100 random trials. The simulation converges after about 10 trials (the difference between 50 and 100 trials is not obvious). For simulations using 50 and 100 trials, the running times are approximately 8 and 16 seconds[2], respectively, and appear to increase with higher q. The running time for the analytical model, however, is less than 1 second and remains low with increased q. We conclude that the analytical model is at least 8 times faster for our problem, with similar accuracy. As the size of the topology increases, the running time is of great concern

[2]The simulations were run in Matlab R2013a within OS X on a Macbook pro laptop with 2.3 GHz Intel Core i7 processor and 8 GB of memory.

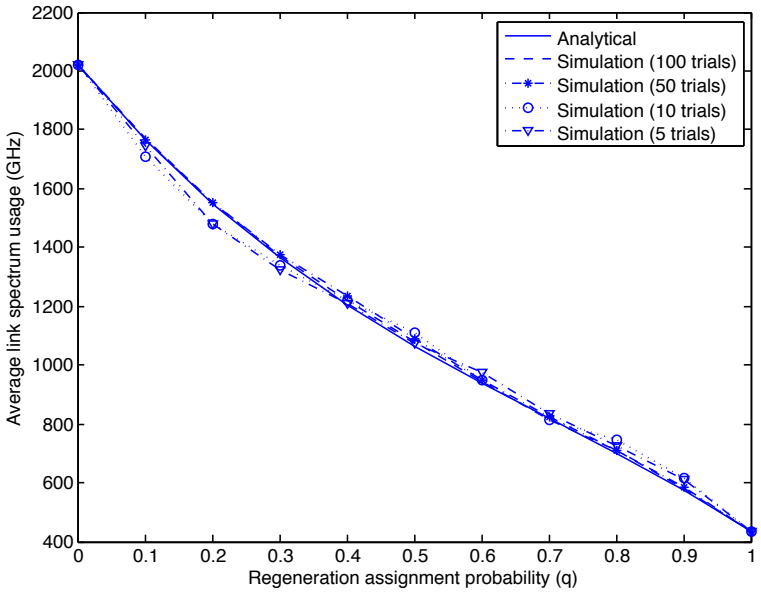

Figure 8.11: Average link spectrum usage of regenerated network without modulation conversion for 36-mesh network

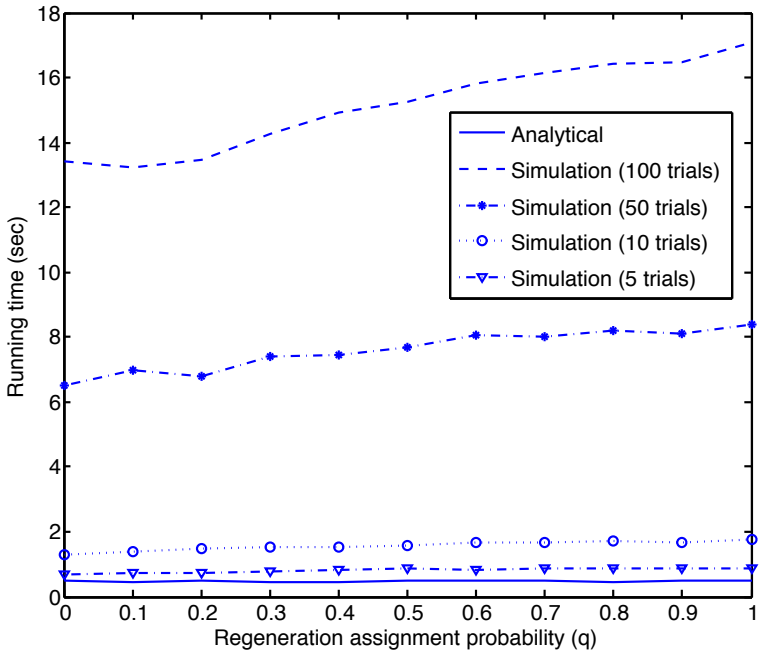

Figure 8.12: Running time for Fig. 8.11

and therefore, the analytical model is preferable for obtaining an initial estimate of the average link spectrum needed.

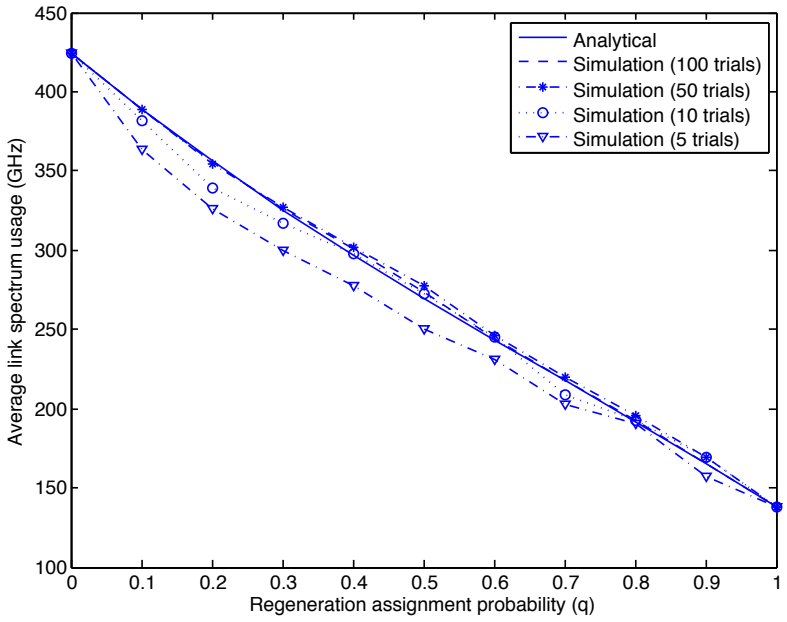

Figure 8.13: Average link spectrum usage of regenerated network without modulation conversion for 16-mesh network

The convergence of simulation results depend on the size of the network. Comparing the result of average link spectrum usage for a smaller 16-node symmetric network, a 100-node symmetric network, and the 36-mesh network shown above, Figs. 8.13 and 8.14 show that larger size networks requires fewer trials of simulation to converge to a result. These figures also show that the analytical model is consistent with simulation results. The running time for simulation also increases drastically with the network size: we observe that the average runtime (over all q's) for 100 trials for the 16-mesh network is 2 seconds, yet for the 100-mesh network it is 165 seconds. The runtime for our analytical model does not scale that much when size of the network increases. It runs less than 0.1 seconds for the 16-mesh network and around 4 seconds for the 100-mesh network. The speed up is 34 and 42 times, respectively.

We also implemented our analytical model on a ring network and found similar outcomes, except that the analytical model's runtime is in this case similar to that of a simulation of 10 trials. This is due to the fact that the range of path lengths

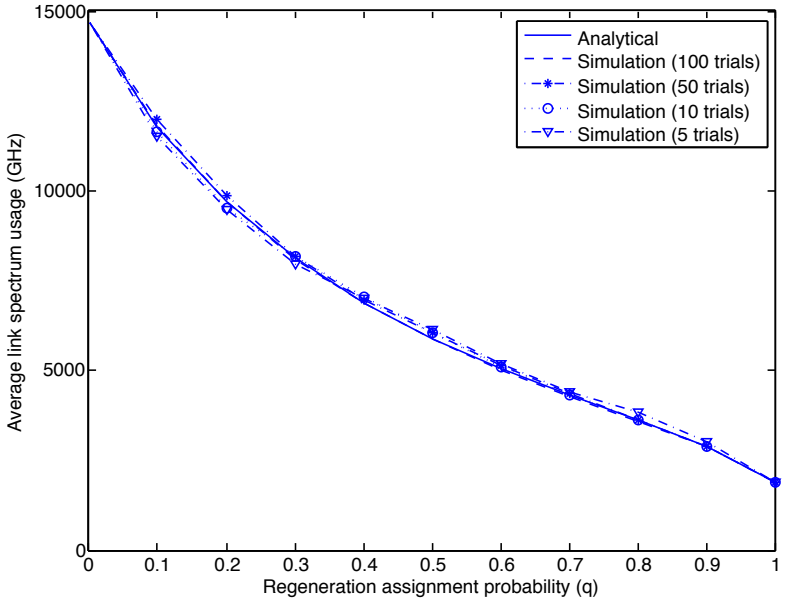

Figure 8.14: Average link spectrum usage of regenerated network without modulation conversion for 100-mesh network

is larger compared to a mesh network with same number of nodes. Therefore the amount of calculation also increases for the analytical model.

Once the routing is known, the analytical model can be applied to any network and the results appear to be consistent with simulation. We tested the analytical model on the 36-mesh network with modified routing (each demand randomly selects from four alternate shortest paths). The results show similar trend as those shown above.

8.6 Summary

In this article, we propose an analytical model to investigate the impact of signal regeneration and modulation conversion on network performance as measured by average link spectrum usage, highest spectrum required, and network capacity. We show the trade-off between the number of regeneration nodes and spectrum usage. We compare our model with simulation results using 100 trials and show good consistency.

We also show that our model can reduce the calculation complexity as measured by a speed-up in the runtime by at least one order of magnitude, depending on the topology.

Chapter 9

Summary and Conclusions

In this thesis, we investigate the impact of physical impairments and resource allocation approaches on the network performance for both static and dynamic fiber optic networks. Heuristic algorithms and mixed-integer linear programming (MILP) formulations are developed according to the assumed constraints and objectives. An analytical model is derived to provide a coarse estimation of the network performance and give network designers insights into the relationship between network parameters.

9.1 Summary

The first part of the work involve several heuristic algorithms for wavelength division multiplexed (WDM) networks. The centralized heuristic grooming, routing, and wavelength assignment (GRWA) algorithm presented in Chapter 3 and published in [67] is essentially a weighted shortest path routing algorithm that bases the weights on signal regeneration and traffic grooming opportunities. Its goal is to balance the saving in spectrum usage from traffic grooming with the additional congestion resulting from rerouting on a path longer than the shortest path. We break down the algorithm into a sequential process that considers the three different categories of traffic grooming opportunities: destination grooming, source grooming, and intermediate segment

grooming. The algorithm also guarantees the established lightpath have acceptable quality of transmission (QoT) by limiting the length of transparent segments. As per reviewers suggestions when this paper was first published, we developed another centralized heuristic GRWA algorithm for dynamic translucent networks, described in Chapter 4 and published in [68], called the *grooming adaptive shortest path* (GASP). Instead of valuing traffic grooming, signal regeneration, wavelength availability separately, we combine them into one step.

A distributed GRWA algorithm that implements ant colony optimization (ACO) presented in Chapter 4 and published in [68] proposes a distributed control and management mechanism for network operation. It reduces the information sharing down to only among relevant neighbors. It also localizes the GRWA decision making to the source node of the traffic demand, which shortens the call setup delay. The ACO algorithm is then optimized for our specific network performance goal, which includes QoT constraints. A series of algorithm modifications are explored to improve the general ACO model to better fit our problem. The network performance and network control overhead between distributed and centralized GRWA algorithms are compared.

A part of the work addresses elastic optical networks (EONs). We implement a MILP approach to solve the routing and spectrum assignment (RSA) problem for EONs, described in Chapter 5 and submitted for publication [56]. The programming not only uses flexible spectrum assignment for EONs but also includes variables and constraints for wavelength and modulation conversion. The complexity of the formulation is discussed. We then develop a recursive model that breaks down the problem into subproblems with subsets of traffic demands and solve them sequentially. This recursive model can be easily applied to dynamic network and network expansion with existing infrastructure as well as ongoing traffic. In Chapter 6, also submitted for publication [69], we investigate the merit of this link-based MILP in contrast

with using a path-based (PB) MILP formulation to solve the network design problem where only a subset of the RSA solutions are considered. We study the drawbacks of pre-determining the regeneration resources allocation as in previous publications and the impacts of the number of candidate paths in the solution pool.

The MILP formulation is modified in Chapter 7 (published in [70]) to apply to dynamic WDM networks with an objective function that minimizes the occurrence of traffic demands being blocked. The formulation is formulated specifically in a time-slotted fashion that matches the traffic arrival and termination processes. The resource allocation optimization is performed for each time-slot successively, with the existing network state as initial constraints. This approach is the first to explore the possibility of using MILP to solve dynamic RWA problems.

The goal of analytical models is to mathematically explain the relationship between network parameters. We present a model that also gives a preliminary estimate of network performance given certain assumptions, such as uniform traffic, uniform link length, etc. In Chapter 8 we express the RSA problem for EON in analytical form. The analytical model is able to estimate key network performance such as total spectrum usage, highest spectrum requirement and network capacity, while considering the QoT of the network.

9.1.1 Traffic Grooming in WDM Networks

When implemented in the heuristic methods in Chapters 3 and 4, traffic grooming is encouraged with caution: on one hand we want to encourage new traffic to fill the gap left by existing traffic, on the other hand we want to avoid excessive spectrum usage when grooming requires rerouting further from the shortest path. Therefore, we always give priority to grooming opportunities that don't require or require very little rerouting, such as source-to-destination grooming. We then quantify the benefits of grooming by using weights within the routing algorithm so that we can combine

the two merits into one. In Chapter 4 the traffic grooming information is not just collected but also affects the ants behavior in exploring the network. We develop a novel approach to implement traffic grooming in MILP in Chapter 7 so that traffic grooming is enforced at either the source or regeneration node where all groomed traffic undergoes OEO conversion.

9.1.2 Regeneration Node Placement

In Chapters 3 and 4 we follow the heuristic method for regeneration node placement proposed in [21] that values the connectivity of the topology. In Chapter 5 we solve the regenerator placement problem within the mixed-integer linear program by using variables that represent regenerator placement decisions as part of the objective function. In this way, we can optimize the problem for our topology of interest.

9.1.3 Regeneration Circuit Allocation

We implement several schemes for regeneration circuit allocation heuristically: in Chapter 3, a regenerator is assigned to call demands as needed; in Chapter 4, a regenerator is assigned to a call demand whenever possible. We also show in Chapter 7 that for a network with a similar scale as the NSF-24 network, the number of regenerators needed is reasonably low. We then optimize regenerator usage by including it as part of the objective function in our MILP formulation in Chapter 5.

9.2 Conclusions

In Chapter 3 we show that with traffic grooming, the blocking probability of dynamic WDM networks is reduced and the throughput is increased. We use the ongoing traffic trends to show that at any moment in the operation of the dynamic network, the network with traffic grooming implemented serves more traffic. We also show the

effects of having a different number of regeneration nodes in the network, and point out that while having more regeneration nodes implies more traffic grooming opportunities, therefore leading to a higher spectrum usage efficiency, after the number of regeneration nodes reaches certain point (for example, 10 for the NSF-24 network), the marginal improvement of adding new regeneration nodes decreases notably. This is partly because the network is not symmetric and some paths are less frequently used than others.

In Chapter 4 we introduce a distributed GRWA algorithm using ant colony optimization (ACO) for dynamic WDN networks. We acknowledge the performance disadvantage of using a distributed algorithm compared to a centralized one by presenting the ACO and GASP algorithm performance side-by-side. This results in a trade-off between blocking probability and distributed management (reduced control overhead). The ACO has been modified by many previous works for better performance, yet we show that for our specific question of GRWA, it is only meaningful to optimize the algorithm with network performance as the criteria. By adjusting the parameters of the algorithm the network performance in terms of blocking probability can be improved.

In Chapter 3 we point out that the improvement to network throughput flattens out as the number of regeneration nodes increases. This is because the topology is not ideally symmetric and certain paths are not used as often as others. This conclusion also implies that the translucent optical network design is both effective in terms of increasing network throughput and economically reasonable by reducing the need for expensive electronic devices. We support this conclusion again in Chapter 5 and 6 with MILP formulations.

In Chapter 5 we investigate the implementation of technologies such as wavelength and modulation conversion on different topologies (namely, connectivity, node degree, etc., but with the same scale in terrestrial coverage and order of topology) and show

that the effects vary based on the topology. A more symmetric topology often requires fewer conversion nodes than a realistic topology since the majority of lightpaths share common intermediate nodes. A more symmetric topology also requires less spectrum because there are fewer bottlenecks than networks such as the NSF-24 network; also more alternative paths are available.

MILP often yields solutions much better compared to the heuristics, but due to its complexity it can only be applied to small scale networks. In Chapter 5 we show the difficulty of solving an MILP even for a limited number of demands for the NSF-24 topology. In order to solve this issue, we introduce the recursive MILP formulation that makes the execution time much smaller and predictable (i.e., a higher chance of finding a solution within a fixed time limit). The trade-off is the solution quality.

The popular accepted path-based (PB) MILP formulation is compared with our link-based (LB) MILP formulation in Chapter 6. We show that due to the pre-determination of regeneration resources for PB formulation, the regeneration resources usage is not optimized for the whole network, yielding less efficient assignment (the PB cannot use all regeneration resources due to pre-assignment even when they are available) and more required spectrum for the same traffic matrix. The advantage of using PB formulations is its reduced computation complexity and therefore faster calculation time. For PB formulations, adding more candidate paths in the solution pool results in better network performance, yet this depends on the network topology (connectivity, scale, etc.) as shown in comparing between the NSF-24 network and a symmetric 24-node network. The size of the solution pool needed is found to depend on many factors. The LB formulation, however, provides an optimal solution regardless of the variation in topology parameters.

To implement MILP on dynamic networks, it is essential to break down the continuous traffic into a sequence of sets of calls according to their arrival times, and optimize the RWA for each set separately, as we propose in Chapter 7. The size of

the set of calls determines the setup delay and extra assignment of resources. We notice that the overall blocking probability of the network remains rather invariant to the size of time granularity of the traffic demand partitioning.

Analytical model runs much faster compared to even heuristic algorithms for finding network performance objectives. Our analytical model described in 8 shows orders of magnitide of speed-up compared to heuristic simulation. It is also highly consistent with simulation runs for a large number of trials. The accuracy of the analytical model varies with the topology of the network.

9.3 Future work

The network resource placement and assignment problem has been approached in many ways, from heuristic algorithms to linear programming optimization. The report herein suggests further study in particular areas of interest.

In particular, the resource assignment mechanism could be further improved if additional information about future traffic needs were included. In terms of traffic grooming, we encourage the grooming of new traffic with other new traffic or existing traffic. But since the demand termination time is unknown, it becomes possible that one of the groomed traffic quickly terminates after grooming starts, leaving the spectrum inefficiently utilized. It is expected that with certain traffic reservation/planning, the algorithm can be improved so that the uptime of traffic grooming is high. In Chapter 5, we show that with the recursive MILP that can be used to mimic dynamic networks, employing modulation conversion does not always improve the performance of the network. This is also because optimization is only made for the current traffic while not taking future traffic into consideration. We plan to include traffic scheduling into our design in the future.

In some of our work we use relaxation of integer parameters to make the solutions more computationally tractable. For example, the spectrum assignment in our MILP is an arbitrary value, while in reality, arbitrary spectrum assignment is not practical (using OFDM technology, the spectrum assignment can only be made in units of the size of subcarrier). Similarly for spectral efficiency and modulation we assume complete flexibility when such an assumption is overstated. We plan to further investigate the impact of these assumptions in future work.

The analytical model presented in Chapter 8 can also be extended. We could consider scenarios where regeneration resources are not placed randomly but clustered in a certain order. The model may also be able to handle dynamic traffic instead of a static traffic matrix.

In our later work, we start to consider the effects that network topology has on network performance. The dependence of traffic on network resource usage and physical impairments is expected to be highly topology dependent. Although real networks such as the NSF-24 network are often used in the literature to implement designs, it is critical to develop algorithms that take network topology into consideration. We plan to introduce topology related knowledge, such as through graph theory, into research on algorithm design for fiber optic networks.

Bibliography

[1] M. Batayneh, D. Schupke, M. Hoffmann, A. Kirstaedter, and B. Mukherjee, "On routing and transmission-range determination of multi-bit-rate signals over mixed-line-rate WDM optical networks for carrier ethernet," *Networking, IEEE/ACM Transactions on*, vol. 19, no. 5, pp. 1304–1316, 2011.

[2] G. P. Agrawal, *Fiber-optic communication systems*, 1997, vol. 1.

[3] N. S. Bergano and C. Davidson, "Wavelength division multiplexing in long-haul transmission systems," *Lightwave Technology, Journal of*, vol. 14, no. 6, pp. 1299–1308, Jun 1996.

[4] J. Hayes, "Copper, fiber or wireless?" http://www.ecmag.com/section/systems/copper-fiber-or-wireless.

[5] H. Zhu, H. Zang, K. Zhu, and B. Mukherjee, "A novel generic graph model for traffic grooming in heterogeneous WDM mesh networks," *Networking, IEEE/ACM Transactions on*, vol. 11, no. 2, pp. 285–299, 2003.

[6] O. Awwad, A. Al-Fuqaha, and M. Guizani, "Genetic approach for traffic grooming, routing, and wavelength assignment in WDM optical networks with sparse grooming resources," in *Communications, 2006. ICC '06. IEEE International Conference on*, vol. 6, 2006, pp. 2447–2452.

[7] B. Chen, G. Rouskas, and R. Dutta, "On hierarchical traffic grooming in WDM networks," *Networking, IEEE/ACM Transactions on*, vol. 16, no. 5, pp. 1226–1238, 2008.

[8] S. Balasubramanian and A. Somani, "On path-level traffic grooming strategies in WDM metro optical networks," *Communications Magazine, IEEE*, vol. 46, no. 11, pp. 91–97, 2008.

[9] M. Jinno, H. Takara, B. Kozicki, Y. Tsukishima, Y. Sone, and S. Matsuoka, "Spectrum-efficient and scalable elastic optical path network: architecture, benefits, and enabling technologies," *Communications Magazine, IEEE*, vol. 47, no. 11, pp. 66–73, November 2009.

[10] M. Klinkowski and K. Walkowiak, "Routing and spectrum assignment in spectrum sliced elastic optical path network," *Communications Letters, IEEE*, vol. 15, no. 8, pp. 884–886, August 2011.

[11] S. Talebi, F. Alam, I. Katib, M. Khamis, R. Salama, and G. N. Rouskas, "Spectrum management techniques for elastic optical networks: A survey," *Optical Switching and Networking*, vol. 13, no. 0, pp. 34 – 48, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1573427714000253

[12] J. Kahn and K.-P. Ho, "Spectral efficiency limits and modulation/detection techniques for dwdm systems," *Selected Topics in Quantum Electronics, IEEE Journal of*, vol. 10, no. 2, pp. 259–272, March 2004.

[13] P. Winzer and R.-J. Essiambre, "Advanced modulation formats for high-capacity optical transport networks," *Lightwave Technology, Journal of*, vol. 24, no. 12, pp. 4711–4728, Dec 2006.

[14] K. Christodoulopoulos, I. Tomkos, and E. Varvarigos, "Elastic bandwidth allocation in flexible OFDM-based optical networks," *Lightwave Technology, Journal of*, vol. 29, no. 9, pp. 1354–1366, May 2011.

[15] G. Zhang, M. De Leenheer, and B. Mukherjee, "Optical traffic grooming in OFDM-based elastic optical networks [invited]," *Optical Communications and Networking, IEEE/OSA Journal of*, vol. 4, no. 11, pp. B17–B25, Nov 2012.

[16] A. Nag and M. Tornatore, "Transparent optical network design with mixed line rates," in *2008 2nd International Symposium on Advanced Networks and Telecommunication Systems, ANTS 2008*, 2008.

[17] A. Nag, M. Tornatore, and B. Mukherjee, "Optical network design with mixed line rates and multiple modulation formats," *Lightwave Technology, Journal of*, vol. 28, no. 4, pp. 466–475, 2010.

[18] A. Klekamp, R. Dischler, and F. Buchali, "Limits of spectral efficiency and transmission reach of optical-OFDM superchannels for adaptive networks," *Photonics Technology Letters, IEEE*, vol. 23, no. 20, pp. 1526–1528, Oct 2011.

[19] S. Zhang, C. Martel, and B. Mukherjee, "Dynamic traffic grooming in elastic optical networks," *Selected Areas in Communications, IEEE Journal on*, vol. 31, no. 1, pp. 4–12, January 2013.

[20] K. Zhu and B. Mukherjee, "Traffic grooming in an optical WDM mesh network," *Selected Areas in Communications, IEEE Journal on*, vol. 20, no. 1, jan 2002.

[21] X. Yang and B. Ramamurthy, "Sparse regeneration in translucent wavelength-routed optical networks: architecture, network design and wavelength routing," *Photonic Network Communications*, vol. 10, no. 1, pp. 39 – 53, 2005.

[22] D. Chaves, C. Ayres, R. Carvalho, H. Pereira, C. Bastos-Filho, and J. Martins-Filho, "Multiobjective sparse regeneration placement algorithm in optical networks considering network performance and CAPEX," in *Transparent Optical Networks (ICTON), 2010 12th International Conference on*, July 2010.

[23] A. Colorni, M. Dorigo, V. Maniezzo *et al.*, "Distributed optimization by ant colonies," in *Proceedings of the first European conference on artificial life*, vol. 142. Paris, France, 1991, pp. 134–142.

[24] G. Di Caro and M. Dorigo, "Mobile agents for adaptive routing," in *System Sciences, 1998., Proceedings of the Thirty-First Hawaii International Conference on*, vol. 7, 1998, pp. 74–83 vol.7.

[25] J. Pedro, J. Pires, and J. Carvalho, "Distributed routing path optimization for OBS networks based on ant colony optimization," in *Global Telecommunications Conference, 2009. GLOBECOM 2009. IEEE*, 2009, pp. 1–7.

[26] R. M. Garlick and R. S. Barr, "Dynamic wavelength routing in WDM networks via ant colony optimization," in *Ant Algorithms : Third International Workshop, ANTS 2002, volume 2463 / 2002 of Lecture Notes in Computer Science*. Springer-Verlag, 2002, pp. 250–255.

[27] S.-H. Ngo, X. Jiang, and S. Horiguchi, "Adaptive routing and wavelength assignment using ant-based algorithm," in *Networks, 2004. (ICON 2004). Proceedings. 12th IEEE International Conference on*, vol. 2, 2004, pp. 482–486 vol.2.

[28] K. Bhaskaran, J. Triay, and V. Vokkarane, "Dynamic anycast routing and wavelength assignment in WDM networks using ant colony optimization (ACO)," in *Communications (ICC), 2011 IEEE International Conference on*, 2011, pp. 1–6.

[29] Y.-M. Kim, E.-J. Lee, and H.-S. Park, "Ant colony optimization based self-organizing QoS framework in IP networks," *Communications Letters, IEEE*, vol. 14, no. 11, pp. 1074–1076, 2010.

[30] J. Mapisse, P. Cardoso, and J. Monteiro, "Ant colony optimization routing mechanisms with bandwidth sensing," in *EUROCON - International Conference on Computer as a Tool (EUROCON), 2011 IEEE*, 2011, pp. 1–4.

[31] Y. Wang, J. Zhang, Y. Zhao, J. Wang, and W. Gu, "ACO-based routing and spectrum allocation in flexible bandwidth networks," *Photonic Network Communications*, vol. 25, no. 3, pp. 135–143, 2013. [Online]. Available: http://dx.doi.org/10.1007/s11107-013-0397-z

[32] G. Pavani and H. Waldman, "Traffic engineering and restoration in optical packet switching networks by means of ant colony optimization," in *Broadband Communications, Networks and Systems, 2006. BROADNETS 2006. 3rd International Conference on*, 2006, pp. 1–10.

[33] X. Li, Y. Aneja, and F. Baki, "Ant colony optimization metaheuristic for the traffic grooming in WDM networks," in *Proceedings of the 2nd international conference on Combinatorial Optimization and Applications*, ser. COCOA 2008. Berlin, Heidelberg: Springer-Verlag, 2008, pp. 235–245.

[34] J. Triay and C. Cervello-Pastor, "An ant-based algorithm for distributed routing and wavelength assignment in dynamic optical networks," *Selected Areas in Communications, IEEE Journal on*, vol. 28, no. 4, pp. 542–552, 2010.

[35] E. Dijkstra, "A note on two problems in connexion with graphs," *Numerische Mathematik*, vol. 1, no. 1, pp. 269–271, 1959. [Online]. Available: http://dx.doi.org/10.1007/BF01386390

[36] A. Mokhtar and M. Azizoglu, "Adaptive wavelength routing in all-optical networks," *Networking, IEEE/ACM Transactions on*, vol. 6, no. 2, pp. 197–206, Apr 1998.

[37] M. L. Fredman and R. E. Tarjan, "Fibonacci heaps and their uses in improved network optimization algorithms," *J. ACM*, vol. 34, no. 3, pp. 596–615, Jul. 1987.

[38] T. Erlebach and S. Stefanakos, "On shortest-path all-optical networks without wavelength conversion requirements," in *STACS 2003*, ser. Lecture Notes in Computer Science, H. Alt and M. Habib, Eds. Springer Berlin Heidelberg, 2003, vol. 2607, pp. 133–144.

[39] N. Garcia, P. Lenkiewicz, M. Freire, and P. Monteiro, "On the performance of shortest path routing algorithms for modeling and simulation of static source routed networks – an extension to the Dijkstra algorithm," in *Systems and Networks Communications, 2007. ICSNC 2007. Second International Conference on*, Aug 2007, pp. 60–60.

[40] A. Zyane, Z. Guennoun, and O. Taous, "Performance evaluation of shortest path routing algorithms in wide all-optical WDM networks," in *Multimedia Computing and Systems (ICMCS), 2014 International Conference on*, April 2014, pp. 831–836.

[41] M. Klinkowski and K. Walkowiak, "Offline RSA algorithms for elastic optical networks with dedicated path protection consideration," in *Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), 2012 4th International Congress on*, Oct 2012, pp. 670–676.

[42] A. Pages, J. Perello, S. Spadaro, and G. Junyent, "Split spectrum-enabled route and spectrum assignment in elastic optical networks," *Optical Switching and Networking*, vol. 13, no. 0, pp. 148 – 157, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1573427714000332

[43] A. Jarray and B. Jaumard, "Exact ILP solution for the grooming problem in WDM ring networks," in *Communications, 2005. ICC 2005. 2005 IEEE International Conference on*, vol. 3, May 2005, pp. 1708–1712 Vol. 3.

[44] S. Bandyopadhyay, Q. Rahman, S. Banerjee, S. Murthy, and A. Sen, "Dynamic lightpath allocation in translucent WDM optical networks," in *Communications, 2009. ICC '09. IEEE International Conference on*, June 2009, pp. 1–6.

[45] Y. Chen, A. Bari, and A. Jaekel, "Optimal regenerator assignment and resource allocation strategies for translucent optical networks," *Photonic Network Communications*, vol. 23, no. 1, pp. 16–24, 2012. [Online]. Available: http://dx.doi.org/10.1007/s11107-011-0331-1

[46] H. Takada and A. Anzaloni, "Integer linear programming models and performance evaluation on wavelength rearrangement in a mesh-restored all-optical network," *Communications Letters, IEEE*, vol. 10, no. 2, pp. 111–113, Feb 2006.

[47] G. Shen and R. Tucker, "Sparse traffic grooming in translucent optical networks," *Lightwave Technology, Journal of*, vol. 27, no. 20, pp. 4471–4479, Oct 2009.

[48] X. Wan, N. Hua, and X. Zheng, "Dynamic routing and spectrum assignment in spectrum-flexible transparent optical networks," *Optical Communications and Networking, IEEE/OSA Journal of*, vol. 4, no. 8, pp. 603–613, Aug 2012.

[49] X. Liu, L. Gong, and Z. Zhu, "Design integrated RSA for multicast in elastic optical networks with a layered approach," in *Global Communications Conference (GLOBECOM), 2013 IEEE*, Dec 2013, pp. 2346–2351.

[50] Z. Zhu, W. Lu, L. Zhang, and N. Ansari, "Dynamic service provisioning in elastic optical networks with hybrid single-/multi-path routing," *Lightwave Technology, Journal of*, vol. 31, no. 1, pp. 15–22, Jan 2013.

[51] K. Kuang, X. Wang, S. Wang, S. Xu, H. Liu, and G. Liu, "Dynamic routing and spectrum allocation in elastic optical networks with mixed line rates," in *High Performance Switching and Routing (HPSR), 2014 IEEE 15th International Conference on*, July 2014, pp. 1–6.

[52] K. Christodoulopoulos, I. Tomkos, and E. Varvarigos, "Routing and spectrum allocation in OFDM-based optical networks with elastic bandwidth allocation," in *Global Telecommunications Conference (GLOBECOM 2010), 2010 IEEE*, Dec 2010, pp. 1–6.

[53] H. Wang and G. Rouskas, "Traffic grooming in optical networks: Decomposition and partial linear programming (LP) relaxation," *Optical Communications and Networking, IEEE/OSA Journal of*, vol. 5, no. 8, pp. 825–535, Aug 2013.

[54] Z. Liu and G. Rouskas, "Link selection algorithms for link-based ILPs and applications to RWA in mesh networks," in *Optical Network Design and Modeling (ONDM), 2013 17th International Conference on*, April 2013, pp. 59–64.

[55] K. Christodoulopoulos, P. Soumplis, and E. Varvarigos, "Planning flexible optical networks under physical layer constraints," *Optical Communications and Networking, IEEE/OSA Journal of*, vol. 5, no. 11, pp. 1296–1312, Nov 2013.

[56] X. Wang, M. Brandt-Pearce, and S. Subramaniam, "Impact of wavelength and modulation conversion on transluscent elastic optical networks using MILP," *submitted to IEEE/OSA Journal on Optical Communications and Networks*, 2014.

[57] *IBM ILOG CPLEX studio version 12.51.*

[58] J. Löfberg, "YALMIP: a toolbox for modeling and optimization in MATLAB," in *Proceedings of the CACSD Conference*, Taipei, Taiwan, 2004. [Online]. Available: http://users.isy.liu.se/johanl/yalmip

[59] A. Al-Fuqaha, G. Chaudhry, M. Guizani, and G. Brahim, "Routing in all-optical dwdm networks with sparse wavelength conversion capabilities," in *Global Telecommunications Conference, 2003. GLOBECOM '03. IEEE*, vol. 5, Dec 2003, pp. 2569–2574 vol.5.

[60] S. Subramaniam, M. Azizoglu, and A. Somani, "All-optical networks with sparse wavelength conversion," *Networking, IEEE/ACM Transactions on*, vol. 4, no. 4, pp. 544–557, Aug 1996.

[61] S. Antonakopoulos and L. Zhang, "Approximation algorithms for grooming in optical network design," in *INFOCOM 2009, IEEE*, April 2009, pp. 1548–1556.

[62] S. Korotky, "Network global expectation model: a statistical formalism for quickly quantifying network needs and costs," *Lightwave Technology, Journal of*, vol. 22, no. 3, pp. 703–722, March 2004.

[63] S. K. Korotky, R.-J. Essiambre, and R. W. Tkach, "Expectations of optical network traffic gain afforded by bit rate adaptive transmission," *Bell Labs Technical Journal*, vol. 14, no. 4, pp. 285–295, Winter 2010.

[64] L. Velasco, A. Jirattigalachote, P. Monti, L. Wosinska, S. Spadaro, and G. Junyent, "Probabilistic-based approach for fast impairments-aware RWA in all-optical networks," in *Optical Fiber Communication (OFC), collocated National Fiber Optic Engineers Conference, 2010 Conference on (OFC/NFOEC)*, March 2010, pp. 1–3.

[65] Y. Yu, J. Zhang, Y. Zhao, H. Li, Y. Ji, and W. Gu, "Exact performance analytical model for spectrum allocation in flexible grid optical networks," *Optical Fiber Technology*, vol. 20, no. 2, pp. 75 – 83, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1068520013001764

[66] W. Feller, "An introduction to probability theory and its applications," *An introduction to probability theory and its applications*, vol. 1, 2, 1950.

[67] X. Wang, M. Brandt-Pearce, and S. Subramaniam, "Grooming and RWA in translucent dynamic mixed-line-rate WDM fiber optics networks subject to physical impairments," *OFC conference*, Mar. 2012.

[68] ——, "Dynamic grooming, routing, and wavelength assignment for real-time optical networks using ant colony optimization," in *IEEE GLOBECOM 2013*, 2013.

[69] ——, "Path-based vs. link-based resource allocation for physically impaired elastic optical networks," *submitted to IEEE/OSA Photonics Technilogy Letters*, 2014.

[70] ——, "Dynamic grooming and RWA in translucent optical networks using a time-slotted ILP," in *Global Communications Conference (GLOBECOM), 2012 IEEE*, Dec 2012, pp. 2996–3001.