ABSTRACT

ADAPTIVE RESIZING OF DEADLINE-DRIVEN REQUESTS FOR PROVISIONING TRAFFIC IN ELASTIC OPTICAL NETWORKS

by Jared Anthony Morell

Spectrum-sliced elastic optical networks, enabled by technological advances such as CO-OFDM, bandwidth-variable transponders, bandwidth-variable optical cross-connects, and optical multi-level modulation, provide a means to divide the spectrum of light transmitted over optical fibers on a finer granularity than WDM and to slice-off just the adequate amount for each connection. It is envisioned that these networks will carry various types of traffic with different service level guarantees, including Deadline-Driven Requests (DDRs) that require the data to be transferred by a given deadline without imposing a specific constant bandwidth requirement. As a result, DDRs can be provisioned with variable transmission rates between their arrival times and deadlines. For this thesis, the connection-request-provisioning problem is considered in a reconfigurable elastic optical network that supports such bandwidth readjustments on DDRs through minimal reconfigurations within the network. Various distance-adaptive routing, spectrum assignment and reconfiguration algorithms are developed for this purpose.
ADAPTIVE RESIZING OF DEADLINE-DRIVEN REQUESTS FOR PROVISIONING TRAFFIC IN ELASTIC OPTICAL NETWORKS

A Thesis

Submitted to the
Faculty of Miami University
in partial fulfillment of
the requirements for the degree of
Master of Science in Computational Science and Engineering
Department of Electrical & Computer Engineering
by
Jared Anthony Morell
Miami University
Oxford, Ohio
2013

Advisor________________________
Dr. Gokhan Sahin

Co-Advisor________________________
Dr. Donald Ucci

Reader________________________
Dr. Chi-Hao Cheng
# Table of Contents

List of Figures .................................................................................. iii  
List of Tables ................................................................................... iv  
Acknowledgments ............................................................................. v  
1 Introduction
   1.1 Fiber-Optic Communications .................................................. 1  
   1.2 Goals ................................................................................... 2  
   1.3 Overview ............................................................................. 2  
2 Fixed-Grid and Elastic Optical Networks ..................................... 4  
   2.1 Overview of WDM Networks ............................................... 4  
   2.2 Hardware Developments .................................................... 6  
   2.3 Overview of OFDM ............................................................ 7  
   2.3 Elastic CO-OFDM Networks .............................................. 8  
3 Provisioning Traffic Requests ..................................................... 9  
   3.1 Deadline-Driven Requests ................................................ 9  
   3.2 Provisioning Requests in WDM Networks ......................... 9  
   3.3 Provisioning Requests in SLICE Networks ....................... 10  
4 Heuristic Routing and Spectrum Assignment Algorithm ............ 13  
   4.1 Routing and First Attempt at Spectrum Assignment ............ 13  
   4.2 Connection Reallocation .................................................. 15  
5 Numerical Results ......................................................................... 19  
   5.1 Simulation Setup .................................................................. 19  
   5.2 Minimal vs. Greedy with and without Reallocation .......... 20  
   5.3 Limiting Reallocations ...................................................... 22  
   5.4 Varying Number of Path Options ..................................... 24  
   5.5 Mixed Traffic ................................................................. 25  
6 Conclusions and Potential Future Work .................................... 27  
7 Works Cited ................................................................................. 28
List of Figures

2-1 Optical Cross-Connect 5
2-2 Bandwidth-Variable Optical Cross-Connect 7
2-3 Fixed-Grid WDM vs. Elastic CO-OFDM Network 8
3-1 Subcarriers and Frequency Slots 12
4-1 Reallocation Diagram 17
5-1 COST 266 Network Topology 20
5-2 US NSFNET Topology 20
5-3 RSA Algorithm with and without Reallocation Results 21
5-4 Minimal vs. Greedy Method Results 22
5-5 Limiting Maximum Number of Reallocated Connections Results 23
5-6 Varying Number of Available Paths Results 25
5-7 Mixed Traffic Results 26
List of Tables

4-1 List of Symbols Used 18
Acknowledgments

This thesis would not have been possible without the gracious help and support of a number of people. I would first like to thank my advisor, Dr. Gokhan Sahin, who has provided me with countless hours of guidance, encouragement, and assistance. Dr. Sahin was thoughtful enough to deliver me a multitude of ideas and resources to help point me towards a particular area of optical networks that I have found both interesting and cutting-edge. Whenever my work seemed stuck or I experienced drawbacks in my progress, he would always do whatever he could to keep me motivated, on track, and moving forward. I would not have had nearly the success I did in developing the new methods presented in this thesis without the innovative and thought-provoking discussions I have had with Dr. Sahin. Even the birth of his child and being across the Atlantic Ocean didn’t prevent Dr. Sahin from giving me his full support.

I would like to thank Dr. Donald Ucci for serving as my thesis co-advisor. Dr. Ucci provided me with invaluable feedback through the process of writing my thesis. Working as his teaching assistant for a number of courses also gave me the opportunity to learn and model my work after such an enthusiastic and hard-working man. Our numerous discussions around the office have expanded my outlook and attitude toward my future.

The third member of my committee, Dr. Chi-Hao Cheng, has been gracious enough to read and evaluate this thesis. His comments and insight have furthered my knowledge and understanding of the subject matter and enabled the development of my work.

Lastly, I would not be where I am today without my loving parents, Joe and Loretta Morell. They instilled in me the values I try to carry with me through life and have patiently put up with me through these 24 years. They have always believed in me and encouraged me to be the best person I can be.
Chapter 1

Introduction

1.1 Fiber-Optic Communications

Fiber-optic cables provide the ability to develop high-speed, high-capacity telecommunications networks that can span entire continents. Fiber-optic communication was developed in the latter part of the 20th century. With the invention of the laser in 1960s and the development of low power-loss glass fibers in the 1970s, optical fibers became a feasible means to transmit data [1]. By the 1990s, as bandwidth-intensive applications grew and the Internet rapidly expanded, optical fiber links developed to handle heavier traffic loads and became the backbone of telecommunication companies worldwide [1].

Signals are transmitted by pulsing light through thin, flexible transparent glass or plastic fibers. The near-infrared spectrum (approximately 770 – 1675 nm wavelength) is used for optical communication because this is a low-loss region for silica glass fibers [1]. Using various multiplexing schemes, each fiber has the capacity to carry many independent channels by which signals can be transmitted using a different part of the light spectrum.

Fiber-optic communication is superior to electrical systems in many ways, especially over long distances [1]. For one, carrier frequencies used for optical communication are in the $10^{13}$ to $10^{14}$ Hz range, whereas coaxial cables use frequencies between $10^6$ and $10^9$ Hz [1]. With optical frequencies being several orders of magnitude higher than those used in electrical communication systems, information can be transmitted across fibers at a much higher rate. Second, light propagates through the fiber with very little attenuation, allowing far fewer repeaters to be needed than in electrical cables when spanning great lengths [1]. As a result, system cost and complexity can be reduced. Fibers are also immune to electromagnetic interference, so no crosstalk occurs between fibers or fibers and other cables, and no
environmental noise can degrade the signal [1]. Additionally, fibers save space and resources when used instead of electrical cables. Being small in size and very low weight for the information carried within them, optical fibers are better suited than bulky electrical cables to be placed in any number of locations [1]. Finally, fibers are far more secure than electrical systems transmitting equivalent data. The optical signal is very well contained within the fiber and an opaque coating on its outside eliminates signal emissions. As a result, they are much more difficult to tap and have signals intercepted than are wires in electrical systems [1].

Optical networks are used to provide a transmission service for a wide array of applications, ranging from High-Definition Television (HDTV) to large-scale grid computing to database queries. With the ability to bundle a large number of high-speed, high-capacity fibers, optical networks can be used to simultaneously connect various forms of traffic, such as Internet Protocol (IP), Synchronous Optical Network (SONET), and Asynchronous Transfer Mode (ATM), to locations around the world.

1.2 Goals

The purpose of this research is to develop a number of methods by which an optical network can accommodate various types of traffic requests while optimizing network characteristics such as spectrum efficiency and fraction of provisioned requests and bytes. To accomplish this, a simulation model will first be developed for a promising new, bandwidth-elastic optical network scheme on a number of different network topologies. Then, new methods by which certain traffic can be dynamically assigned spectrum over time will be introduced onto this new elastic network. The research is intended to show that these particular methods are beneficial to network performance by allowing more network connection requests to be serviced than would otherwise be possible. An extensive number of scenarios will be simulated in order to evaluate the effectiveness of the algorithms under differing network conditions.

1.3 Overview

This thesis will demonstrate, analyze, and compare several algorithms that solve the problem of provisioning various types of traffic including deadline-driven traffic in elastic optical networks under a number of different network and traffic scenarios. Chapter 2 describes how most optical
backbone networks operate today as well as the recent hardware developments that will push optical networks into a new era that will allow much greater flexibility and spectrum utilization. Chapter 3 focuses on the issue of provisioning connection requests in both fixed-grid and elastic optical networks. Chapter 4 details the algorithms used to provision connection requests and reallocate spectrum of ongoing deadline-driven connections in reconfigurable, elastic optical networks. Chapter 5 analyzes results for the algorithms run for a number of different simulations. Finally, Chapter 6 concludes the paper.
Chapter 2

Fixed-Grid and Elastic Optical Networks

2.1 Overview of WDM Networks

In high-capacity, optical backbone networks, Wavelength-Division Multiplexing (WDM) has traditionally been used to divide the spectrum of the optical fiber into a number of parallel channels, called wavelengths. Cumulatively, the wavelengths are capable of being transmitted on a single signal [2]. These wavelengths are of equivalent bandwidth, because WDM makes use of rigid, wideband signals distributed on a fixed-grid optical spectrum as defined by ITU standards [ITU-T G.694.2]. WDM describes the same concept as Frequency-Division Multiplexing (FDM) but is the term used for optical networks, since the carrier signal is typically described by its wavelength, and not its frequency, as in a radio carrier.

In WDM, the system will multiplex a number of different carrier signals, each transmitting at a different frequency, onto a single optical fiber. The different wavelengths of laser light are used as independent channels. To transmit data, each carrier signal is modulated using a particular technique for modifying some aspect of the signal, such as its phase or amplitude, so that a particular number of bits can be represented. Each carrier of differing wavelength is sent through the fiber and can be uniquely extracted at the receiver. A wavelength corresponds to the bandwidth surrounding the carrier frequency that is required for successful recovery of the data signal at the receiver after demodulation. Carrier frequencies must be placed far enough apart so that no interference takes place between wavelengths [2]. These wavelengths are located on the same section of spectrum for all links in a network, based on the ITU standards [ITU-T G.694.2]. Similarly, there are a predefined set of carriers located at the same frequency for all links in a network, creating a “fixed-grid”. For a connection to be established within a traditional WDM network, any wavelengths that the connection requires must occupy the same spectral location across the network. This is because switching a signal’s wavelength between links (known as
wavelength conversion) is not typically used in optical transport networks. Wavelength conversion at the optical level is still deemed too costly for practical purposes. Performing optical-electrical-optical conversion to undertake wavelength conversion is also too complex and time consuming to be used in optical transport networks.

Several devices make WDM in fiber-optic communication possible. First, it is necessary to convert incoming electrical signals into optical signals at a particular wavelength for transmission. At the receiver, the system must be able to do the opposite to translate optical signals into electrical ones. Optical transponders are used for this very purpose. An Optical Cross Connect (OXC) is used to switch optical signal paths. These OXCs utilize Wavelength Selective Switches (WSS) to access individual wavelengths in the fiber’s signal and either send them to another link on that same wavelength or interface them with the electrical router to be terminated or initialized on the optical network. A diagram of an OXC can be seen in Fig. 2-1. Current systems typically allow for optical bypass [3]. This means that signals that are passing through a node, but not terminating in it, remain in the optical domain so as to avoid costly optical-electrical-optical conversion.

![Diagram of an OXC](image)

**Figure 2-1:** An optical cross-connect that is able to route individual wavelengths and add or drop signals to and from the electrical switch.
Using these technologies, WDM-based optical networks can transmit vast amounts of data across countries and continents. However, with the ever-expanding bandwidth needs of countless applications, more must be done to accommodate heavier and more varying traffic types.

### 2.2 Hardware Developments

The spectrum used by end-to-end paths in WDM, called lightpaths, is required to be on a fixed-grid. This means that lightpaths need to be set up so that they occupy entire wavelengths of network-designated size at the wavelengths’ specific intervals. If multiple wavelengths are required to carry the desired connection, the guardbands between those wavelengths are wasted simply to comply with the existing fixed-grid structure. Furthermore, if less than an entire wavelength’s worth of bandwidth is required for a connection to be made, the client is still required to reserve a wavelength along the path. With emerging technologies, it is possible to “slice-off” just the required bandwidth, at a much finer granularity and use only the necessary amount of network resources. A substantial amount of effort has been put forth in designing and developing elastic optical networks based on spectrum slicing [4]. This network type has become known as Spectrum-Sliced Elastic optical path network, or SLICE. A number of newly developed components and technologies make SLICE a viable means to enhance optical networks.

Over the past few years, several advancements have been made in high-capacity, optical hardware. Bandwidth-Variable (BV) transponders provide variable granularity in the spectral domain by enabling the dynamic adjustment of input and output wavelength transmission bandwidth. The term, granularity, is used to describe the increments of spectral width that connections can occupy within the same fiber. BV transponders transmit signals of different bandwidth on the same fiber. Additionally, BV-OXCs enable switching to be made at variable-sized bandwidths so that end-to-end paths can be set up with just the required bandwidth. BV-OXCs select exactly the bandwidth of an incoming signal within a fiber and route it to an outgoing fiber or an electrical network. Figure 2-2 illustrates the concept behind a BV-OXC. Developments in optical multilevel modulation have enabled an adaptive modulation scheme to be applied to different wavelength signals. This means that not all wavelengths in a fiber must be modulated in the same way. By incorporating all of these technologies into optical networks, finer granularity is achieved and spectrum can be better utilized [5].
2.4 Overview of OFDM

Orthogonal Frequency-Division Multiplexing (OFDM) has made SLICE a viable means to achieve more efficient spectrum usage. By using low-rate, orthogonally modulated subcarriers, variable data rate signals can be managed with finer granularity. The orthogonal nature of the subcarriers allows them to overlap during transmission and to be individually recovered at the receiver. By making use of these overlapping subcarriers, data can be transmitted in parallel within a connection. Doing so allows more of the spectrum to be utilized than would be possible by transmitting data on a single wideband channel. OFDM also mitigates the effects of Inter-Symbol Interference (ISI). The subcarriers of a signal are set to low rates and, as such, the symbol duration is longer than if a single-carrier system was used with an equivalent total data rate. Additionally, OFDM can be efficiently implemented using the Fast Fourier Transform (FFT) [6]. The FFT converts the time signal into an equivalent frequency representation by computing the Discrete Fourier Transform (DFT). Computing the DFT directly is too slow for practical communications purposes. Evaluating the DFT of $N$ points takes $O(N^2)$ operations. Using an FFT on those same $N$ points takes only $O(N \log N)$ operations. For large $N$, the difference in speed between using the DFT and FFT is immense.

Coherent detection in optical OFDM networks has provided the ability to not only encode and decode the light by amplification, as in direct detection, but by the phase of the light as well.
This is what is known as Coherent Optical OFDM, or CO-OFDM. CO-OFDM expands upon the methods by which the light can be modulated. SLICE, coupled with the ability to use CO-OFDM, has transformed optical fibers into a much more manageable and effective data transport [4, 7].

2.5 Elastic CO-OFDM Networks

In CO-OFDM networks, data can be taken and spread over the necessary number of overlapping, low data-rate subcarriers. The subcarriers are able to overlap within a connection and to be fully recovered at the receiver due to their orthogonal nature. Each individual subcarrier can be modulated with a particular modulation scheme, as in WDM networks. Being able to pack the subcarriers into a much smaller area than would be possible with WDM greatly enhances the spectrum utilization [5]. Additionally, using low-rate modulation on a number of narrowband subcarriers produces an improved Signal-to-Noise Ratio (SNR) when compared to rapidly modulating a wideband signal and provides higher flexibility in bandwidth adjustment [7]. A spectrum comparison of OFDM and WDM is shown in Fig. 2-3.

Since CO-OFDM is capable of managing signals with varying data rate and bandwidth, its use will enable optical networks to be built to meet the growing and changing requirements of future traffic [8].

![Figure 2-3: Comparing the spectrum of a WDM network on the fixed grid and that of a CO-OFDM network that utilizes sets of overlapping subcarriers placed at varying intervals for different connections.](image)
Chapter 3

Provisioning Traffic Requests

3.1 Deadline-Driven Requests

Optical networks support a number of different traffic types. Two examples are Best-Effort and Minimum Bandwidth traffic that require various Quality-of-Service (QoS) guarantees. An emerging class of applications that need on-demand and flexible bandwidth allocation are deadline-driven applications [9]. As the name suggests, these are applications that require data be transferred by a given deadline. Because these applications do not require a strict, specified bandwidth, variable transmission rates can be used to accommodate the requests. Traffic such as this arises in such cases as eScience and grid-computing [8]. Since these fields utilize a distributed network of computers to perform a task, various components are needed at specific times for smooth operation. They therefore could benefit from deadline-aware service. Other systems that do not need immediate updates could benefit from using a deadline-driven service as well.

3.2 Provisioning Requests in WDM Networks

In order to accommodate any sort of request in WDM networks, several constraints need to be met. First, a connection lightpath must occupy the same wavelength on all links of its path from start to end. This is because most OXCs in use do not allow for wavelength conversion. This is a difficult task to perform at the optical level and it is, in general, costly in time to conduct any optical-electrical-optical conversions. It should be noted, however, that a connection might make use of more than one wavelength and lightpath if the bandwidth provided by one wavelength would not suffice. The second core issue when attempting to establish a connection is addressing the needs of the client. As stated in Section 3.1, these might include a minimum
bandwidth requirement at all times or, in the case of DDRs, completing the transmission by a given deadline.

Because DDRs are allowed flexible bandwidth, spectrum allocation in reconfigurable networks is not restricted to its availability at the time of arrival. Instead, to make use of its flexibility, a connection’s spectrum may be reallocated between the time of arrival and the deadline. Doing so would be beneficial to network performance in terms of its ability to serve more connections or bandwidth. Decreasing the bandwidth of an ongoing deadline-driven connection to accommodate an incoming DDR was first proposed in [9] for use in WDM networks. If a lightpath with sufficient bandwidth is not available for the DDR when it arrives, other deadline-specified connections’ transmission rates could be reduced in order to make spectrum available to the new connection. Whenever ongoing connections use more than a needed amount of bandwidth to transmit their data on time, it could be possible to reduce their spectrum usage by a sufficient level to provide the new connection request with its needed bandwidth. This would make it possible for the ongoing connections (as well as the new request) to complete data transfers by their corresponding deadlines. Reallocating DDR bandwidth over time would improve both spectral efficiency and reduce the number of blocked requests in the network [9].

3.3 Provisioning Requests in SLICE Networks

Establishing a connection in an OFDM-based network is substantially different from that in WDM networks. What was a wavelength continuity concern between links in a WDM network becomes an issue of subcarrier continuity in an OFDM network. Additionally, in order to make use of the orthogonality of the subcarriers and maintain other architecturally desirable characteristics, all of a connection’s subcarriers should be assigned in a contiguous manner within the link. Subcarriers from different connections should never overlap on a link. Figure 3-1(a) illustrates the spectrum of three different connections in a SLICE network fiber.

Each OFDM subcarrier can be modulated using one of several modulation techniques in a SLICE network that makes use of CO-OFDM. For example, the system may employ binary phase-shift keying (BPSK or 1 bit per symbol), quadrature phase-shift keying (QPSK or 2 bits per symbol), quadrature amplitude modulation (8 QAM or 3 bits per symbol), and so on, to determine the data rate that each subcarrier can deliver. It is desirable to select the highest modulation level while
still maintaining acceptable quality of transmission. Doing so will provide a connection with the highest data rate possible while staying within the capabilities of the network architecture and hardware.

In elastic networks, the goal will be to expand upon the method by which DDRs were dynamically allocated bandwidth. We note that adjusting the transmission rate of a connection involves a coordinated reconfiguration of all the bandwidth-variable components (transponders and the intermediate optical cross-connects) on its path. As a result, the reallocation of bandwidth is anticipated to take a non-negligible amount of time, even though there are efforts to achieve hitless reconfiguration [10]. In the model for this thesis, transmission of data can resume only after a delay following the bandwidth adjustment of any existing connections in order to accommodate the necessary changes. This reallocation penalty is based on the time it takes to reconfigure the appropriate optical network components.

The model that will be used for the elastic spectrum is a common one used for optical OFDM networks [4, 11-13]. Rather than looking at the spectrum as an open resource where subcarriers can be placed at any frequency, it can be thought of as divided into frequency slots, of width, $F$ GHz, able to transmit at capacity, $T$, given by

$$T = MC,$$

where $T$ is in Gbps, $M$ is the modulation multiplier (equal to 1 for BPSK modulation, 2 for QPSK modulation, etc.), and $C$ is the base capacity of a subcarrier using BPSK modulation (in Gbps). Using the frequency slot model makes network management and optical node architecture far less complicated. Figure 3-1 shows how the frequency slots relate to the actual overlapping subcarriers on the spectrum. To avoid interference, a guardband, $G$, equal to a certain number of frequency slots is required between connections. The guardband allows the BV-OXCs to select individual connections at the optical level in an effective manner and either route them to other links or translate them to electrical signals.

Requests are serviced on a first-come, first-served basis. Once a request arrives at the network, be it a DDR or any other traffic type, it is necessary to find a path from its source to destination node that meets all criteria for both its traffic type and the elastic network. If such a path is not immediately available to the request, an attempt is made to perform a reallocation algorithm.
which would adjust the bandwidth of a number of ongoing connections to make spectrum available for the incoming request. The next section details the algorithm by which this can be accomplished.

Figure 3-1: Relationship between the overlapping subcarriers (a) of several connections and an equivalent frequency slot model (b).
Chapter 4

Heuristic Routing and Spectrum Assignment Algorithm

4.1 Routing and First Attempt at Spectrum Assignment

A request, $R$, arrives at the network and is defined by the following parameters:

$$R = (\sigma, \delta, S, D),$$

where $\sigma$ is the source node, $\delta$ is the destination node, $S$ is the size of the file to be transferred (Gb), and $D$ is the deadline of the request (a coherent listing of symbols used can be found on page 18).

Requests arrive across the network based on a Poisson distribution, with an arrival rate parameter of $\lambda$ requests per second [14]. Mixed-traffic scenarios are also simulated, where requests are either deadline-driven or have a fixed-bandwidth requirement. For fixed-bandwidth connections, the data rates and durations are based on finding the minimum bandwidth required to transfer the data by the given deadline [See (4)].

Upon arrival, the $K$ shortest paths are computed between $\sigma$ and $\delta$ for the request, without regards to spectrum availability at the time. Please note that these paths could also be pre-computed for all source-destination pairs ahead of time and stored for future retrieval, if desired. The $K$ shortest paths are then iterated from shortest to longest, searching for a segment of contiguous, available frequency slots that occupy the same section of spectrum on all links in the path. To do so, each link, $l$, in the network must first be characterized by a frequency slot availability vector of length, $\text{maxSub}$ [4]. The variable, $\text{maxSub}$, defines the link capacity as well, since it quantifies the number of frequency slots on each link in the network. The frequency slot availability vector of link $l$ is

$$u_l = [u_{li}] = (u_{l1}, u_{l2}, \ldots, u_{l\text{maxSub}}).$$

(3)
The value of each \( u_{li} \) is equal to 1 if that frequency slot is available on that link, and 0 if it is not. It is possible to compute the path, \( p \), subcarrier availability vector, \( U_p \), by using the Boolean AND operation over all \( l \) in \( p \). Using \( U_p \), it is then possible to search for a sequence of unoccupied frequency slots on each of the \( K \) paths. Applying the subcarrier continuity constraint, and available spectrum void (free sequence of frequency slots) for which the incoming request can occupy is sought out.

To determine the number of frequency slots that are necessary for \( R \) to transmit all of \( S \) by \( D \), it is first necessary to calculate the highest modulation level able to be applied to all subcarriers on the path. It is a common simplifying assumption for such studies that the most dominant quality of transmission factor is the distance traversed [15]. For distances up to 6000 km, BPSK modulation can be used. For every halving of the transmission distance, the signal quality improves enough to increase the modulation level by 1 bit per symbol [16]. Thus, each path has its own modulation level multiplier, \( M_p \), and as a result, its own frequency slot transmission capacity, \( T_p \), and its own minimum necessary number of frequency slots required to transmit all of \( S \) by \( D \),

\[
X_{p,\text{min}} = \text{ceil}[\frac{S}{(D - A)T_p}],
\]

where \( A \) is the arrival time of the request. However, because of the guardband, the total minimum number of frequency slots required by incoming request, \( R \), becomes

\[
Y_{p,\text{min}} = X_{p,\text{min}} + G.
\]

There are two methods by which the incoming request may occupy spectrum:

- **Minimal** approach: search \( U_p \) from shortest to longest path for the smallest sequence of available frequency slots, \( \text{minVoid} \), that is greater than or equal to \( Y_{p,\text{min}} \). If \( \text{minVoid} \) is found on \( p \), it is possible to immediately accommodate the incoming request and the total number of frequency slots occupied by the request, \( Y_R \), is set to \( \text{minVoid} \).

- **Greedy** approach: search \( U_p \) from shortest to longest path for the largest sequence of frequency slots, \( \text{maxVoid} \). If \( \text{maxVoid} \) is found on \( p \), it is possible to immediately accommodate the incoming request and \( Y_R \) is set to \( \text{maxVoid} \).

If the method selected was able to find a place for the incoming request, the starting frequency slot index that the connection will occupy is set; its departure time from the network is then computed as,
\[ dep_R = \frac{S}{(Y_R - G)T_p} + t \]  \hspace{1cm} (6)

(with \( t \) being the current time in the system), the \( u_i \) vectors are updated, and the new connection is added to the list of those currently in the network. If Greedy or Minimal was unable to find a set of frequency slots that could immediately accommodate the incoming request, the algorithm proceeds to the Reallocation phase.

### 4.2 Connection Reallocation

In Reallocation, an attempt is made to reduce the bandwidth of ongoing deadline-driven connections in order to make room for the incoming request. The algorithm requires that connections must maintain transmission on at least one frequency slot plus the guardband (i.e., connections cannot be reduced to zero frequency slots and then resume transmission at a later time). It also must be made certain that any ongoing connections that are reallocated are still able to complete data transfer by their deadlines. The Reallocation algorithm will stop after it finds a path on which there is at least one section of spectrum that is able to reduce connections by the necessary number of frequency slots to accommodate \( R \). If more than one available section is found on \( p \), the section that affects the least number of ongoing connections is selected.

For path \( p \), a new \( X_{p_{\text{min}}} \) and \( Y_{p_{\text{min}}} \) are first determined, with an included time penalty, \( t_{\text{Pen}} \), incurred for network reconfiguration as follows:

\[ X_{p_{\text{min}}} = \text{ceil}[ \frac{S}{(D - A - t_{\text{Pen}})T_p} ] , \]  \hspace{1cm} (7)

\( Y_{p_{\text{min}}} \) is computed in the same way as (5).

Any connection that occupies any link \( l \) in \( p \) is added to a list of interfering connections, List. These connections will be checked for reallocation along each step of the Reallocation process for \( p \).

The vector \( U_p \) is iterated from index 0 up to \( \text{maxSub} - 1 \), while examining each frequency slot, \( s \), and the possibility of reallocating connections surrounding \( s \) to make room for \( R \). The system keeps track of sequences of available frequency slots with a counter, \( \text{void} \). If the frequency slot we are examining is:
A. The beginning of a new sequence of frequency slots ($U_{ps} = 1$ and $U_{ps-1} = 0$):

set void to 1.

B. Still in a sequence of available frequency slots ($U_{ps} = 1$ and $U_{ps-1} = 1$):

increase void.

If at the end of the $U_p$ vector ($s = \text{maxSub} - 1$), $List$ is examined to determine which connections are occupying frequency slots in the range $(s - Y_{p_{min}}, s)$. Those that do would need to be reallocated by enough frequency slots so that they no longer fall in that range. If every connection in $List$ is either able to be reduced by the required number of frequency slots (meaning it first must be deadline-driven), or does not interfere in that range, a potential reallocation scenario has been found. The connections that need adjustment and the amount by which they need adjustment are recorded.

C. Occupied ($U_{ps} = 0$):

and if:

1) This is the first frequency slot or the previous slot was occupied ($s = 0$ or $U_{ps-1} = 0$):

a similar attempt to reallocate as in (B.) is conducted, but in the range, $[s, s + Y_{p_{min}}]$.

2) This is the end of a sequence of available frequency slots ($U_{ps-1} = 1$):

the system blocks off a section of frequency slots. The block has length of $Y_{p_{min}}$; it is initially placed so that its starting index is located at the frequency slot whose index is $s - Y_{p_{min}}$. In that block’s range, the algorithm attempts to reallocate connections in a similar means as above. The block is then shifted forward by one frequency slot and another attempt is made. This continues until the block’s last index reaches the frequency slot with index equal to $s + Y_{p_{min}} – \text{void}$.

Figure 4-1 illustrates each Reallocation scenario.
After the frequency slot iteration has completed on $p$, it is determined whether any potential reallocation scenarios had been found. If so, the spectrum section that affects the least number of connections is selected. Those connections are then reduced by their pre-determined amounts and have their defining parameters changed as follows,

$$S_R = S_R - (Y_R - G)T_p (t - A_R),$$  \hspace{1cm} (8)

$$A_R = A_R + t_{Pen},$$  \hspace{1cm} (9)

$$Y_R = Y_R - red,$$  \hspace{1cm} (10)

$$dep_R = S / ( (Y_R - G)T_p ) + A_R,$$  \hspace{1cm} (11)

where $red$ is the amount by which that particular connection needs to be reduced. $S_R$ now becomes the file size left to transfer, and $A_R$ becomes the time at which connection resumes after network reconfiguration. Additionally, if any connections were reduced from their lower
indexed side, their starting indices would need to be shifted as well. All affected $u_l$ vectors are updated to reflect the frequency slots freed by the reduced connections.

Now the incoming request can be added to the network in the section of frequency slots on all links of $p$ that ongoing connections made available. Its selected path and starting index are noted, $dep_R$ set as in (11), and $Y_R$ set to $Y_{p, \text{min}}$. The $u_l$ vectors are once again adjusted to account for the incoming connection.

If it is not possible to reallocate connections on any path by enough frequency slots as to make room for the incoming connection request, the request must be blocked.

<table>
<thead>
<tr>
<th>$R$</th>
<th>Connection request currently being serviced by the network</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma$</td>
<td>Source node of connection</td>
</tr>
<tr>
<td>$\delta$</td>
<td>Destination node of connection</td>
</tr>
<tr>
<td>$S$</td>
<td>Size of the file to be transferred (Gb)</td>
</tr>
<tr>
<td>$D$</td>
<td>Deadline specified by the request</td>
</tr>
<tr>
<td>$A$</td>
<td>Arrival time of request</td>
</tr>
<tr>
<td>$t$</td>
<td>Time in the system</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>Arrival rate of requests to the network (requests/s)</td>
</tr>
<tr>
<td>$K$</td>
<td>Number of shortest paths to work with</td>
</tr>
<tr>
<td>$G$</td>
<td>Guardband – number of frequency slots required between neighboring connections</td>
</tr>
<tr>
<td>$l$</td>
<td>Link number in the network</td>
</tr>
<tr>
<td>$s$</td>
<td>Frequency slot index on a link</td>
</tr>
<tr>
<td>$u_{ls}$</td>
<td>Availability of frequency slot $s$ on link $l$ (0 – unavailable, 1 – available)</td>
</tr>
<tr>
<td>$p$</td>
<td>Path index out of the $K$ paths for a connection</td>
</tr>
<tr>
<td>$U_p$</td>
<td>Frequency slot availability vector of path $p$</td>
</tr>
<tr>
<td>$C$</td>
<td>Transmission capacity of a frequency slot using BPSK modulation (Gbps)</td>
</tr>
<tr>
<td>$F$</td>
<td>Spectral width occupied by a frequency slot (GHz)</td>
</tr>
<tr>
<td>$M_p$</td>
<td>Modulation level multiplier for path $p$</td>
</tr>
<tr>
<td>$T_p$</td>
<td>Transmission capacity of a frequency slot on path $p$ using maximum modulation level allowed (Gbps)</td>
</tr>
<tr>
<td>$X_{p, \text{min}}$</td>
<td>Minimum number of frequency slots needed to transmit all of $S$ by $D$ on path $p$</td>
</tr>
<tr>
<td>$Y$</td>
<td>Number of contiguous frequency slots occupied by a connection including $G$</td>
</tr>
<tr>
<td>$dep$</td>
<td>Time at which the connection will depart the network and free its resources</td>
</tr>
<tr>
<td>$t_{Pen}$</td>
<td>Time it takes for network resources to be reconfigured (s)</td>
</tr>
</tbody>
</table>

Table 4-1: Listing of symbols used and their definitions.
Chapter 5

Numerical Results

5.1 Simulation Setup

Network models and the routing and spectrum assignment methods were all developed in C++. Simulations were run on each individual arrival rate for at least 40,000 connection arrivals/departures, assuming a Poisson arrival process. Every 40,000 runs, the fraction of unprovisioned (blocked) requests was computed. Simulation terminated when the difference between consecutive fractions of unprovisioned requests was less than 0.01. This process was done twice more and averaged over the three simulations to determine the average fraction of unprovisioned requests for a given network scenario under the stress of a particular arrival rate.

Two different networks were used for this simulation. They were the pan-European COST 266 network, composed of 28 nodes and 41 links, and the US NSFNET topology, consisting of 14 nodes and 21 links [17]. The two networks are depicted in Figs. 5-1 and 5-2.

Using $F = 5$ GHz, $C = 2.5$ Gbps, $G = 2$, $t_{Pen} = 0.05$ s, $maxSub = 600$, the performance of the algorithm was evaluated based on the fraction of unprovisioned requests. This was always observed, relative to a domain of $\lambda$ values. The particular values of $F$, $C$, $G$, $t_{Pen}$, and $maxSub$ are based on work done in similar elastic optical network studies [4, 11]. The value of $S$ for each connection request was chosen from a uniform distribution of discrete values between 20 and 500 GB, spaced at 60 GB intervals. The $D$ for each connection request was selected from one of two different continuous uniform distributions, depending on the simulation scenario (2 – 5 s for tight deadline cases and 3 – 100 s for relaxed deadlines).
5.2 *Minimal vs. Greedy* with and without *Reallocation*

The first performance evaluation tested the performance of the algorithm with and without the *Reallocation* phase. Figure 5-3 shows this performance using *Minimal* initial spectrum selection on both the NSFNET and COST 266 optical network topologies for $K = 3$, when compared to the scenario when no deadline-driven connection reallocation was considered. The leftmost plot in
Fig. 5-3 depicts the case for relaxed deadlines, while the rightmost plot shows the results from the tight deadline requirement.

![Graph showing comparison of blocking probabilities](image)

**Figure 5-3:** Comparing the Minimal routing and spectrum assignment algorithm with and without the reallocation phase for requests subject to relaxed deadlines (left) and tight deadlines (right) on the NSFNET topology with $K = 3$.

In both simulation scenarios, it was observed that the reallocation algorithm allows the network to service a far greater percentage of connection requests than if no reallocation of ongoing connections were considered. There is a $20 - 35\%$ difference in blocking probability for the Minimal algorithm between using and not using the Reallocation phase, over the simulated arrival rates for the relaxed deadline scenario, and a $16 - 31\%$ difference for the tight deadline case. This clearly demonstrates better spectrum utilization, as more connections can be accommodated in equivalent spectrum space.

It was then desirable to test the effectiveness of the Minimal initial spectrum search method when compared to the Greedy one, while also observing the effect of Reallocation. To do so, both Greedy and Minimal methods were run on the COST 266 network with and without the Reallocation phase for $K = 3$ and under the case of relaxed deadlines. In addition to observing the effect that the algorithms had on the fraction of unprovisioned requests, the fraction of unprovisioned bytes was also recorded. Figure 5-4 shows the results of this simulation.
Figure 5-4: Comparing the *Minimal* and *Greedy* choice methods for the initial spectrum search with and without the *Reallocation* phase by the fraction of unprovisioned requests (left) and fraction of unprovisioned bytes (right). Run on the COST 266 topology with $K = 3$ and relaxed deadlines.

Utilizing spectral reallocation of connections shows vast improvement over the cases where *Reallocation* was omitted. It can also be seen that the *Greedy* method outperforms the *Minimal* choice both with and without *Reallocation* under the two performance metrics. For instance, with *Reallocation*, *Greedy* blocked 9% and 10% fewer of the total requests received than *Minimal* did at 5 and 9 requests per second, respectively.

### 5.3 Limiting Reallocations

Allowing unlimited connection reallocations may put a strain on network resources. As a result, it was decided to observe how limiting the maximum number of connection reallocations per connection request would affect network performance. Figure 5-5 depicts a number of different scenarios in which the maximum number of reallocations per connection request was varied between 1, 3, and an unlimited amount. The left column of plots show the results of varying the maximum number of reallocations when using the *Greedy* method of initial spectrum selection on the NSFNET topology for relaxed deadlines, while the right column shows results from using the *Greedy* method on the COST 266 network for tight deadlines. In Fig. 5-5., the first row of plots uses a $K$ of 1, the second row uses $K$ of 3, and the last row uses a $K$ of 5.
Figure 5-5: Observing the effect of limiting the maximum number of reallocated connections per new connection request. Simulations were run for the Greedy method on the NSFNET topology with relaxed deadlines for $K = 1$ (upper-left), $K = 3$ (middle-left), and $K = 5$ (lower-left); as well as for the COST 266 topology with tight deadlines for $K = 1$ (upper-right), $K = 3$ (middle-right), and $K = 5$ (lower-right).
The simulations show a very slight improvement (averaging at most 0.01 decrease in overall request blocking probability for the range of arrival rates) between limiting the reallocations to 1 and limiting reallocations to either 3 or an unlimited amount. There is essentially no difference between limiting reallocations to 3 and providing unlimited reallocations. Similar results were observed for tight deadlines in NSFNET and relaxed deadlines in the COST 266 topology that are not presented here. A consequence of this study reveals that by allowing just a single connection reallocation per new request in a network, an optical network can benefit substantially by using Reallocation with ongoing deadline-driven connections to boost network performance.

5.4 Varying Number of Path Options

The next situation that was observed was the effect of varying $K$ between 1, 3, and 5 paths, while holding all else constant. Figure 5-6 shows various scenarios in which network performance was evaluated based on the number of available shortest paths. The upper two plots derive from simulating the Greedy method on the NSFNET topology for tight deadlines. The upper-left simulation restricted connection reallocations to 1 per request and the upper-right did not apply a reallocation restriction. The lower-left and -right plots of Fig. 5-6 are the results of simulating the Greedy method on the COST 266 topology for relaxed deadlines and restricting connection reallocations to 3 and an unlimited amount, respectively.

In the tight deadline case, it is shown that network performance, based on blocked requests, improves as the number of available paths increases. There are, however, diminishing gains from 3 to 5 paths when compared to increasing available paths from 1 to 3. As for the relaxed deadline scenario, a large gain in network performance is still seen from 1 to 3 paths. However, there is virtually no difference in network performance between using 3 and 5 available paths. For large scale networks, it could be worth considering running the algorithm for a limited number of checked paths to ensure better efficiency in time.
5.5 Mixed Traffic

The final performance evaluation was to incorporate mixed traffic into the system. Fixed-bandwidth traffic was used in addition to deadline-driven traffic and the percentage of DDRs in the network was varied between 25% and 100%. Fixed-bandwidth traffic would not allow for reallocation of its spectrum once the connection had been made. Thus, as more of the network traffic became deadline-driven, performance increased. Figure 5-7 illustrates four mixed-traffic cases. In each case, the probability of a request being deadline-driven or fixed-bandwidth is adjusted to four different ratios for each arrival rate. The upper set of simulations in Fig. 5-7 make use of the Greedy method on the COST 266 network with maximum reallocations set to 3...
for $K = 3$ and tight deadlines (upper-left), and for $K = 5$ and relaxed deadlines (upper-right). The lower set of simulations show the effect of mixed traffic on the NSFNET topology with maximum reallocations set to 3 for $K = 3$ and relaxed deadlines (lower-left), and for $K = 5$ and tight deadlines (lower-right). All plots demonstrate the effectiveness of utilizing DDRs in a network as well as the benefit that adaptive reallocation has on network performance.

Figure 5-7: The effect that the ratio of DDRs to fixed-bandwidth traffic has on network efficiency. Simulations were run for the Greedy method on the COST 266 topology with a maximum of 3 connection reallocations per request for tight deadlines and $K = 3$ (upper-left) and for $K = 5$ and relaxed deadlines (upper-right); as well as for the NSFNET topology for $K = 3$ and relaxed deadlines (lower-left) and for $K = 5$ and tight deadlines (lower-right).
Chapter 6

Conclusions and Potential Future Work

In this thesis, algorithms have been proposed for efficient provisioning of requests in elastic CO-OFDM networks with the ability to resize the bandwidth of deadline-driven connections. Numerical results demonstrate that reallocating connections enable the network to accommodate requests that otherwise would have been rejected. Even by allowing just one reallocation per connection request, spectrum utilization and network performance see vast improvement. By taking advantage of the presence of applications that make use of deadline-driven traffic, networks can efficiently manage their spectrum to achieve far greater performance. Furthermore, it was observed that a greedy approach that allowed an arriving connection to use as much spectrum as possible resulted in substantially better performance than a conservative approach that occupied the minimum spectrum needed to meet the connection deadline. We also observed that considering alternate paths for routing a connection yielded diminishing benefits, with the highest gain observed from 1 to 3 paths.

In the future, other methods could be developed that may improve upon results seen here, in one or more different network or traffic scenarios. It may be worth investigating reallocating and/or shifting connections in the network at times other than request arrivals. For instance, a deadline-driven connection might be allowed an increase in spectral resources when a neighboring connection departs the system. Recent work has been made to design an elastic network defragmentation scheme [18], so this and other new methods could be considered. Additionally, it might be possible to develop an accurate mixed integer programming model of the algorithm to compare the algorithm’s results with optimal case results. Finally, the actual signaling and control mechanisms for coordinated bandwidth resizing of ongoing connections could be investigated from both speed and robustness perspectives.
Chapter 7

Works Cited


