Distance-Adaptive Routing and Spectrum Assignment of Deadline-Driven Requests in Reconfigurable Elastic Optical Networks

Jared Morell and Gokhan Sahin
Electrical and Computer Engineering Department
Miami University
Oxford, OH, U.S.A.
e-mail: morellja@muohio.edu, sahing@muohio.edu

***Abstract*—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 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. We consider the DDR-provisioning problem in a reconfigurable elastic optical network that supports such bandwidth readjustments through minimal reconfiguration in the network, and develop a distance-adaptive routing, spectrum assignment, and reconfiguration algorithm for this purpose. Our results show major improvement in performance due to bandwidth reallocation of DDRs in elastic networks over a range of reconfiguration delay parameters.**

Keywords - deadline-driven traffic; elastic optical network; optical OFDM; routing modulation level and spectrum assignment; network reconfiguration.

I. INTRODUCTION

In high capacity optical transport networks, wavelength division multiplexing (WDM) has allowed for the spectrum division of channels into wavelengths of smaller bandwidth on a fixed-grid. Its rigidity and coarse granularity pose drawbacks, however, for clients and service providers. These wavelength-routed networks require full allocation of a wavelength even when the transmission across the channel is not sufficient to fill the wavelength's capacity. Similarly, if a connection needs more than a wavelength's worth of spectrum, it must occupy at least another wavelength to be accepted. This results in poor spectrum utilization.

Over the past few years, a substantial amount of effort has been put forth in designing and developing elastic optical networks based on spectrum slicing [1]. This network type has become known as spectrum-sliced elastic optical path network, or SLICE. Spectrum slicing is a means to divide the spectrum on smaller levels and "slice-off" just the amount necessary for an end-to-end connection request. Technological advances such as the bandwidth-variable (BV) optical cross-connect (OXC), BV transponder, optical multilevel modulation, and optical orthogonal frequency-division multiplexing (OFDM) have enabled an optical fiber to be transformed into a much more manageable and effective data transport [1], [2]. Fig. 1 shows the spectrum usage difference in WDM and OFDM networks.

In optical OFDM networks, data is taken and spread over the necessary number of overlapping, low data rate subcarriers. The subcarriers are able to overlap within a connection and be fully recovered at the receiver due to their orthogonal nature. Being able to pack the subcarriers into a much smaller area than would be possible with WDM enhances the spectrum utilization greatly [3].

There are a number of different traffic types that optical networks support, e.g., best-effort and minimum bandwidth that require various quality-of-service guarantees. An emerging class of applications that need on-demand and flexible bandwidth allocation are deadline-driven applications [4]. 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 [5]. Because these fields utilize a distributed network of computers to perform a task, various parts are needed by certain times for everything to run smoothly. They therefore could benefit from deadline-aware service. Other systems that do not need immediate updates could make use of a deadline-driven service as well.

In this paper, we address the issue of routing, modulating, and dynamically allocating spectrum to deadline-driven requests (DDR) in reconfigurable elastic optical networks. The capability of elastic optical networks to dynamically adjust the spectrum allocated to connections makes them a suitable candidate for carrying DDRs due to their flexible bandwidth requirement. Although various aspects of routing and spectrum

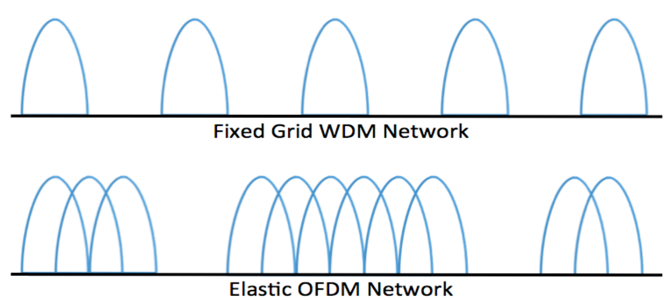

Figure 1. Comparing the spectrum of a WDM network with its fixed grid and that of an OFDM network with overlapping subcarriers.

assignment in elastic optical networks have been considered in recent work (see for instance [1-3, 6-8, 10]), we are not aware of prior research that addresses deadline-driven traffic in these networks. The problem that we consider is also different from earlier DDR work in WDM networks [4] due to the fundamental architectural differences between WDM-based and elastic optical networks. We explicitly consider various aspects of the problem that are specific to elastic optical networks, such as the need to allocate a contiguous band of subcarriers to a connection; the ability to choose a distance-adaptive modulation format for each sub-carrier, and the delay associated with reconfiguring the BV transponders and the BV OXCs in order to adjust the rates of an existing connection.

The rest of the paper is organized as follows. Section II formally presents the problem we are concerned with and the model we will be using. In Section III, we describe the method by which we accomplish the issue. Section IV delivers the results of our simulation. Finally, Section V concludes the paper and notes possible future work.

II. Problem Definition and Model

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. In addition, in order to make use of the orthogonality of the subcarriers and maintain other architecturally desirable characteristics, they should be assigned in a contiguous manner within the link. Subcarriers from different connections should never overlap on a link. Fig. 2(a) illustrates this.

Each OFDM subcarrier can be modulated using, e.g., binary phase-shift keying (BPSK, 1 bit per symbol), quadrature phase-shift keying (QPSK, 2 bits per symbol), quadrature amplitude modulation (8QAM, 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.

Additionally, because DDRs are allowed flexible bandwidth, spectrum allocation is not restricted to its availability at the time of arrival. Instead, reallocation may occur between the time of arrival and the deadline. Decreasing the bandwidth of an ongoing connection to accommodate an incoming request was first proposed in [4] for use in WDM networks. In this paper we will expand upon their ‘Changing-Rates’ method and apply it to an elastic network model. In our model, transmission of data can resume only after a delay following the bandwidth adjustment of an existing connection in order to accommodate the necessary reconfigurations, whereas an instantaneous bandwidth adjustment was assumed in [4].

The model that we are using for the elastic spectrum is a common one used for optical OFDM networks [2], [6], [7], [8]. 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

$$T = MC, \tag{1}$$

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). Fig. 2 (b) shows how the frequency slots relate to the actual overlapping subcarriers on the spectrum. To avoid interference, a guardband, G, of a certain number of frequency slots is required between connections.

Once a request arrives at the network, we are tasked with finding a path from its source to destination node with at least enough contiguous subcarriers to transfer its total file size by the deadline. If such a path is not immediately available, we attempt to perform a reallocation algorithm which would adjust the bandwidth of the smallest number of ongoing connections to make room for the incoming request. The next section describes the algorithm by which this is done.

III. Provisioning Deadline-Driven Requests in Reconfigurable Elastic Optical Networks

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

$$R = (\sigma, \delta, S, D), \tag{2}$$

where σ is the source node, δ is the destination node, S is the size of the file to be transferred (Gb), and D is the deadline of the request. Requests arrive uniformly across the network with arrival rate, λ requests per second. Upon arrival, the K shortest paths are computed for the request, without regards to spectrum availability at the time. We then iterate over the K shortest paths, searching for a void of frequency slots. To do so, we must first characterize each link in the network by a subcarrier availability vector, of size *maxSub*, equal to the number of subcarriers on a link [2].

$$u_l = [u_{li}] = (u_{l1}, u_{l2}, \ldots, u_{lmaxSub}). \tag{3}$$

The variable l designates the particular link. The value of each u_{li} is equal to 1 if that subcarrier 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 $l \in p$. This is what is used to first search for a sequence of unoccupied subcarriers on each path. Applying the subcarrier contiguity constraint, we search for the smallest spectrum void (free sequence of frequency slots) that can occupy the incoming request.

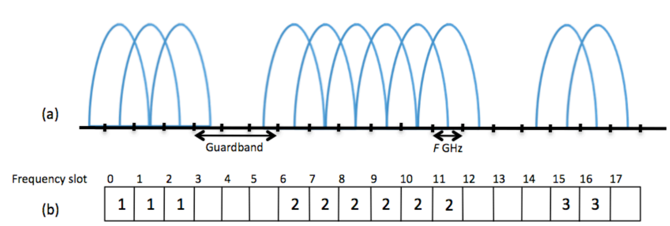

Figure 2. Relationship between the overlapping subcarriers (a) of several connections and an equivalent frequency slot model (b).

To figure out how many frequency slots are necessary, we must also determine the highest modulation level that is acceptable to apply to each subcarrier. It is a common simplifying assumption for such studies that the sole quality of transmission factor is the distance traversed [2]. 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 [9]. Thus, each path has its own M_p modulation level multiplier, and as a result, its own frequency slot transmission capacity, T_p, and necessary number of subcarriers used to transmit data,

$$X_{p\_min} = \text{ceil}[\ S / (\ (D - A)T_p\ )\ ], \tag{4}$$

where A is the arrival time of the request. However, not only is it necessary to find X_{p_min}, but there must be a guardband between neighboring connections. Therefore, the total minimum number of frequency slots required by incoming request, R, is

$$Y_{p\_min} = X_{p\_min} + G. \tag{5}$$

Iterating from the shortest to the longest of the K paths, we search for the smallest void, *minVoid*, of frequency slots in U_p that is greater than or equal to Y_{p_min}. If found on p, we are able to accommodate the incoming request. We then set the total frequency slots occupied by this request, $Y_R = minVoid$, note its starting index, compute its departure time,

$$dep_R = S / (\ (Y_R - G)T_p\ ) + t, \tag{6}$$

with t being the current time in the system, update the u_l vectors, and add the connection to the list of those currently in the network. If no void large enough was found on any path, we go to the reallocation phase.

In reallocation, we attempt to shrink the bandwidth of ongoing connections in order to accommodate the incoming request, while still being able to complete data transfer by their deadlines. We attempt reallocation with the shortest path first, and then move to each successive path from there. While attempting reallocation, we are always looking for a section of frequency slots that will affect the smallest number of ongoing connections.

For path, p, we first determine a new X_{p_min} and Y_{p_min}. These are determined by including a time penalty in the calculation in Equation 4, as follows.

$$X_{p\_min} = \text{ceil}[\ S / (\ (D - A - tPen)T_p\ )\ ], \tag{7}$$

where *tPen* is a constant penalty time to account for reconfiguration. Y_{p_min} is computed in the same way as in Equation 5. We then locate any connections that interfere on at least one link of the incoming request's path, and add them to a list, *List*. These are the connections that will be checked for reallocation at each step in the process.

We iterate over U_p, starting from index 0, up to $maxSub - 1$, examining each frequency slot, s, and the possibility of reallocating connections around it. Void sizes are kept track of in the same manner as the original search for free frequency slots. If the frequency slot we are examining is:

A. The start of a new void ($U_{ps} = 1$ and $U_{ps-1} = 0$)

Simply set the void counter, *void*, to 1.

B. Still in the void ($U_{ps} = 1$ and $U_{ps-1} = 1$)

Increase *void*.

If $s = maxSub - 1$, we iterate over *List*, checking what connections lie in the range of ($s - Y_{p\_min}$, s). Those that do would need to be reallocated by enough subcarriers so that they no longer are in that range. If each connection in *List* is either able to be reallocated or does not fall in that range of subcarriers, we have arrived at a possible reallocation scenario and make note of the connections that need adjusted and by how much.

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

If:

1) *This is the first frequency slot ($s = 0$)*

Perform a similar attempt to reallocate as in (*B.*), but check in the range, [0, Y_{p_min}).

2) *This is the end of a void ($U_{ps-1} = 1$)*

Iterate over a block of frequency slots encompassing the void in an attempt to reallocate connections within that block. The block is of size Y_{p_min}, and is initially placed with its starting index at $s - Y_{p\_min}$, and its last index at $s - 1$. Within that block, connections are attempted to be reallocated in a similar means as above. Then the block is moved forward one frequency slot and the process starts over. This continues until the block's last index reaches $s + Y_{p\_min} - void$.

3) *Otherwise*

Attempt to reallocate connections as done in (*B.*), but in the range, [s, $s + Y_{p\_min}$).

Once the frequency slot iteration has completed, we check if any possible connection reallocations were found. If they were, the section with the minimum number of connections being affected was selected. From there, those connections would be reduced and their parameters changed as follows,

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

$$A_R = A_R + tPen, \tag{9}$$

$$Y_R = Y_R - red, \tag{10}$$

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

where *red* is the amount by which that connection needs reduced. S_R now becomes the file size left to transfer, and A_R becomes the time at which connection resumes. Also, if any connections were reduced from their front, their initial indices would need shifted as well. The u_l vectors are updated as well, indicating free slots where the connections were reduced.

Now the incoming request can be added to the network where other connections have made space available. Its selected path and starting index are noted, departure time set, as in Equation 11, and Y_R set to Y_{p_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, in order to make room for the incoming connection, the request is blocked.

IV. RESULTS

Two different networks were used for this simulation. They were the US NSFNET topology, consisting of 14 nodes and 21 links, and the pan-European COST 266 network, made up of 28 nodes and 41 links [10]. Using $F = 5$ GHz, $C = 2.5$ Gbps, $G = 2$, $tPen = 0.05$ s, $maxSub = 600$, we evaluated the performance of our algorithm based on the fraction of blocked requests, while varying λ, *D* and *S*. On each simulation run, we held a different value of λ constant for use on all nodes in the network and uniformly distributed the values of *D* and *S* between set limits. Fig. 3 shows the algorithm's performance on both optical networks when compared to the scenario when no reallocation is considered. Here, the ranges of *D* and *S* are 3 – 100 s and 30 – 500 GB, respectively.

We also examined the case of having similar file sizes, but tighter deadlines, in the range of 2 – 5 s. This can be seen in Fig. 4.

In both cases, it is shown that the reallocation algorithm allows the network to service a far greater percentage of connection requests than if no reallocation were considered. There is a 20 – 35% difference in blocking probability between using and not using reallocation, 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 placed into equivalent spectrum space.

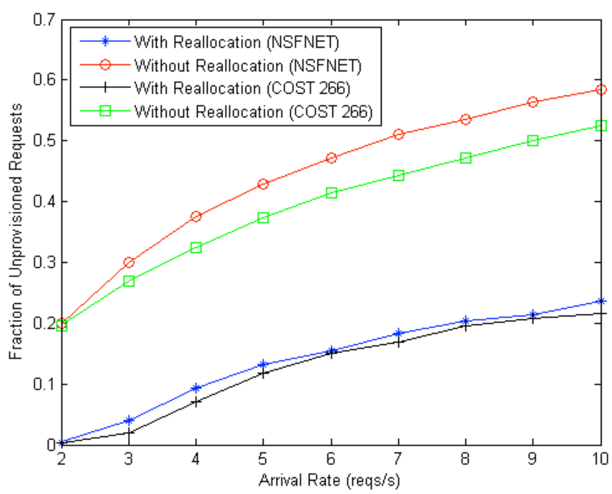

Figure 3. Comparing the routing and spectrum assignment algorithm with and without the reallocation phase

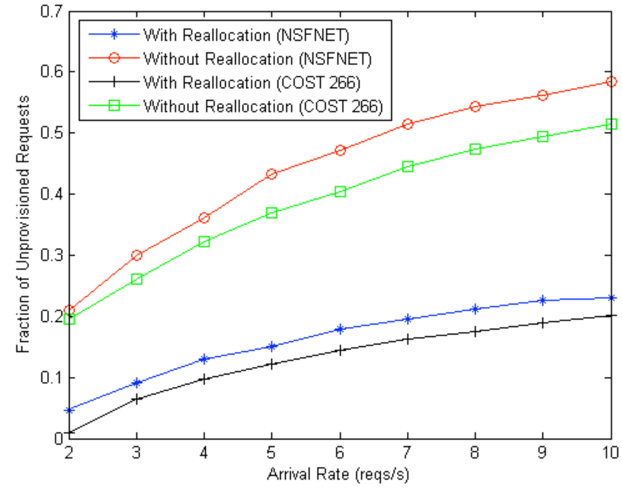

Figure 4. Comparing the routing and spectrum assignment algorithm for tight deadlines with and without the reallocation phase

V. CONCLUSION AND FUTURE WORK

We have developed an algorithm that allows a reconfigurable elastic optical network to accommodate deadline-driven requests with possibility of changing data rates. The method for reallocation allows the network to reconfigure ongoing connections in order to make room for a new request. In the future, it would be worth looking into reallocating and/or shifting connections at times other than request arrivals. Also, the means by which a void is chosen could be altered and different network and traffic scenarios could be analyzed.

REFERENCES

[1] B. Kozicki, H. Takara, Y. Tsukishima, T. Yoshimatsu, K. Yonenaga, and M. Jinno, "Experimental demonstration of spectrum-sliced elastic optical path network (SLICE)," Optics Express, vol. 18, no. 2, pp. 22105-22118, October 2010.

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

[3] J. L. Vizcaino, Y. Ye, and I. T. Monroy, "Energy efficiency analysis for flexible-grid OFDM-based optical networks," Computer Networks, vol. 56, no. 10, pp. 2400-2419, July 2012.

[4] D. Andrei, M. Tornatore, M. Batayneh, C. U. Martel, and B. Mukherjee, "Provisioning of deadline-driven requests with flexible transmission rates in WDM mesh networks," IEEE/ACM Transactions on Networking, vol. 18, no. 2, pp. 353-365, April 2010.

[5] I. Foster, M. Fidler, A. Roy, V. Sander, and L. Winkler, "End-to-end quality of service for high-end applications," Computer Communications, vol. 27, no. 14, pp. 1375-1388, September 2004.

[6] M. Jinno et al, "Distance-adaptive spectrum resource allocation in spectrum-sliced elastic optical path network," IEEE Communications Magazine, vol. 48, no. 9, pp. 138-145, August 2010.

[7] G. Zhang, M. De Leenheer, and B. Mukherjee, "Optical grooming in OFDM-based elastic optical networks," Proc. OFC/NFOEC, Paper OTh1A, March 2012.

[8] T. Takagi et al, “Dynamic routing and frequency slot assignment for elastic optical path networks that adopt distance adaptive modulation,” Proc. OFC/NFOEC, Paper OTuI, March 2011.

[9] A. Bocoi, M. Schuster, F. Rambach, M. Kiese, C. Bunge, and B. Spinnler, “Reach-dependent capacity in optical networks enabled by OFDM,” Proc. OFC/NFOEC, Paper OMQ4, March 2009.

[10] A. Betker et al, “Reference transport network scenarios,” MultiTeraNet Report, July 2003, retrieved: September, 2012, http://wall.ikr.uni-stuttgart.de/en/Content/Publications/Archive/Ko_ITGPhotNetze04_36322.pdf.