Reconfiguration and Dynamic Load Balancing in Broadcast WDM Networks

Ilia Baldine†
MCNC, RTP, NC 27709, USA

George N. Rouskas
Department of Computer Science, North Carolina State University, Raleigh, NC 27695-7534, USA

Received September 10, 1998; Revised November 20, 1998

Abstract. In optical WDM networks, an assignment of transceivers to channels implies an allocation of the bandwidth to the various network nodes. Intuition suggests, and our recent study has confirmed, that if the traffic load is not well balanced across the available channels, the result is poor network performance. Hence, the time-varying conditions expected in this type of environment call for mechanisms that periodically adjust the bandwidth allocation to ensure that each channel carries an almost equal share of the corresponding offered load. In this paper we study the problem of dynamic load balancing in broadcast WDM networks by retuning a subset of transceivers in response to changes in the overall traffic pattern. Assuming an existing wavelength assignment and some information regarding the new traffic demands, we present two approaches to obtaining a new wavelength assignment such that (a) the new traffic load is balanced across the channels, and (b) the number of transceivers that need to be retuned is minimized. The latter objective is motivated by the fact that tunable transceivers take a non-negligible amount of time to switch between wavelengths during which parts of the network are unavailable for normal operation. Furthermore, this variation in traffic is expected to take place over larger time scales (i.e., retuning will be a relatively infrequent event), making slowly tunable devices a cost effective solution. Our main contribution is a new approximation algorithm for the load balancing problem that provides for tradeoff selection, using a single parameter, between two conflicting goals, namely, the degree of load balancing and the number of transceivers that need to be retuned. This algorithm leads to a scalable approach to reconfiguring the network since, in addition to providing guarantees in terms of load balancing, the expected number of retunings scales with the number of channels, not the number of nodes in the network.

Keywords: broadcast optical networks, wavelength division multiplexing (WDM), reconfiguration, dynamic load balancing

1 Introduction

Single-hop lightwave networks have been proposed for Local and Metropolitan Area Networks (LANs and MANs) [1,2]. The single-hop architecture employs Wavelength Division Multiplexing (WDM) to provide connectivity among the network nodes. The WDM channels are dynamically shared by the attached nodes, and the logical connections change on a packet-by-packet basis creating all-optical paths between sources and destinations. Single-hop networks require the use of rapidly tunable optical lasers and/or filters that can switch between channels at high speeds. Such devices do exist today [3]; however, they have to be custom-built and they tend to be extremely expensive, accounting for a significant fraction of the overall cost of building a lightwave network. Consequently, media access protocols such as HiPeR-ℓ [4], FatMAC [5], DT-WDMA [6], and Rainbow [7] that require tunability only at one end have the potential of keeping the overall cost at reasonable levels, leading to network architectures that can be realized cost effectively.

When tunability only at one end, say, at the transmitters, is employed, each fixed receiver is permanently assigned to one of the wavelengths used for packet transmissions. In a typical near-term WDM environment, the number of channels that will be supported within the optical medium is expected to be smaller than the number of attached nodes. As a result, each channel will have to be shared by multiple receivers, and the problem of assigning receive...
wavelengths arises. Intuitively, this assignment must be somehow based on the prevailing traffic conditions. But with fixed receivers, the assignment of receive wavelengths is permanent and cannot be updated in response to changes in the traffic pattern.

Alternatively, one can use slowly tunable, rather than fixed, receivers. We will say that an optical laser or filter is rapidly tunable if the time it takes to switch between channels is comparable to a packet transmission time at Gigabit per second rates. Slowly tunable devices can be significantly less expensive than rapidly tunable ones, but their tuning times can also be significantly longer (up to several orders of magnitude). As a result, these devices cannot be assumed ‘tunable’ at the media access level (i.e., for the purposes of scheduling packet transmissions), as this requires fast tunability. However, use of slowly tunable receivers makes it possible to modify the assignment of receive wavelengths over time to accommodate varying traffic demands.

The issues that arise in reconfiguring a lightwave network by retuning a set of slowly tunable transmitters or receivers have been studied in the context of multihop WDM networks in [8,9]. The work in [9] considered the problem of constructing a sequence of branch-exchange operations of minimum length to take the network from an initial to a target connection diagram. The focus in [8] was on the design of dynamic policies for determining when and how to reconfigure the network. A comprehensive evaluation of reconfiguration policies and retuning strategies for single-hop networks has been conducted by the authors in [10], where we demonstrated the benefits of reconfiguration through both analytical and simulation results.

In this paper we consider the problem of reconfiguring a single-hop network by retuning a subset of the slowly tunable receivers in response to changing network traffic conditions. Our objective is to ensure that the traffic load remains balanced across the various channels, while minimizing the number of receivers that need to be retuned. We show that employing well-known load balancing algorithms leads to an approach that does not scale well with the size of the network. We then present a new approximation algorithm for the load balancing problem that provides for tradeoff selection, using a single parameter, between the two conflicting goals. Our algorithm is simple, fast, scalable, and tends to select the least utilized receivers for retuning, hence minimizing the impact of the reconfiguration phase on the carried traffic. Although our work is motivated by a problem in optical networks, our solution techniques are applicable to a generalized version of the classical multiprocessor scheduling problem [11], whereby it takes a non-negligible amount of time to transfer tasks among processors.

The next section introduces the network model, and discusses the issues arising during the reconfiguration phase. In Section 3 we describe two approaches for dynamically load balancing by retuning the slowly tunable receivers. In Section 4 we present some numerical results to compare the two approaches, and we conclude the paper in Section 5.

2 System Model

2.1 Network Model and Operation

We consider a packet-switched single-hop lightwave network with $N$ nodes, and one transmitter-receiver pair per node. The nodes are physically connected to a passive broadcast optical medium that supports $C < N$ wavelengths, $\lambda_1, \ldots, \lambda_C$, as shown in Fig. 1. Both the transmitter and the receiver at each node are tunable over the entire range of available wavelengths. However, the transmitters are rapidly tunable, while the receivers are slowly tunable. We will refer to this tunability configuration as rapidly tunable transmitter, slowly tunable receiver (RTT-STR). (We note that all our results can be easily adapted to the dual configuration, STT-RTR.)

Let $\Delta_r(\Delta_t)$ denote the normalized tuning latency of transmitters (receivers), expressed in units of packet transmission time. In the RTT-STR system under consideration, we have that $\Delta_r > \Delta_t \geq 1$, where $\Delta_r$ is a small integer, while $\Delta_t$ takes values that may be significantly greater than $\Delta_r$. The main motivation for employing slowly tunable receivers vs. fast tunable ones is the significant savings in cost that can be realized.

We distinguish two levels of network operation, differing mainly in the time scales at which they take place. At the packet scheduling level, connectivity among the network nodes is provided by reservation protocol such as HiPeR-$\ell$ [4] that requires tunability only at the transmitting end. The protocol schedules packets for transmission by employing scheduling algorithms that can effectively mask the tuning latency of tunable transmitters [12–16]. Since the