Optical Networking and Real-Time Provisioning: An Integrated Vision for the Next-Generation Internet

Chadi Assi, Abdallah Shami, and Mohamed A. Ali,
City College of the City University of New York
Russ Kurtz, REALTECH Systems Corporation
Dan Guo, Sorrento Networks

Abstract
This article considers the problem of real-time provisioning of optical channels in hybrid IP-centric DWDM-based networks. First, we present an overview of the emerging architectural alternatives for IP over optical networks, namely, the overlay, the peer, and the augmented models. Then lightpath provisioning issues are detailed for route selection, with a particular focus on the "routing and wavelength assignment" (RWA) problem. In particular, a broad overview is presented, with methodologies and associated algorithms for dynamic lightpath computation being outlined. Additionally, two broad constraint-based RWA algorithms for dynamic provisioning of the optical channels are presented and evaluated. Finally, the implications of implementing the proposed RWA schemes for the lightpath provisioning aspects for each of the three emerging IP-over-optical network interconnection models are examined.

Recently, there has been a dramatic increase in data traffic, driven primarily by the explosive growth of the Internet as well as the proliferation of virtual private networks (VPNs). At the same time, the rise of optical networking, first with wavelength-division multiplexing (WDM) transmission technology and more recently with optical multiplexers and optical cross-connect (OXC) devices, is moving us toward the vision of creating an "all-optical" Internet. In particular, these technologies yield the ability to add, drop, and in effect construct wavelength-routed networks, heralding a new era in which bandwidth is relatively abundant and inexpensive. To some, a key realization of this vision will occur when lightpaths (wavelengths) can be provisioned automatically to create bandwidth between end users, with timescales on the order of minutes or seconds. Dynamic wavelength provisioning is the main focus of this article, and will help open up a whole new world of responsive, customer-driven bandwidth services.

To better understand and appreciate the provisioning issue, we need to look into how circuits are currently provisioned in a typical network. Provisioning a cross-country SONET service today requires several steps. First, connectivity from the customer premise to the carrier's POP must be established for each end of the circuit. Second, a physical path must be mapped out between the many physical hubs in a carrier's network between the two points. Each path must be checked for fiber/ring bandwidth availability. Terminating equipment must be ordered and installed on each end of each fiber path, and each interconnect point must have capacity on the optical cross-connect system. All of the cross-connects and physical interconnects must then be made and each segment documented and tested. This process is extremely manual and generally takes several months to accomplish. DWDM complicates this process even further because tens and soon hundreds of wavelengths are supported on individual fiber strands. Clearly, an automated optical routing layer will facilitate much faster provisioning.

Before this vision can be realized, however, networks need to slim down. Today's core network architecture model has four layers: IP and other content-bearing traffic; over ATM for traffic-engineering; over SONET for transport; and over WDM for fiber capacity. This approach has significant functional overlap among its layers and typically suffers from the lowest common denominator effect where any one layer can limit the scalability of the entire network. When first conceived, this layering made sense, but as IP and DWDM evolve, a more efficient interworking is called for, i.e., one that exploits the complementary features of each domain. In effect, high-performance routers plus a smart optical transport layer equipped with a new breed of photonic networking components and subsystems together are setting the foundation for the next-generation networking paradigm.

This work was supported in part by DARPA/Air Force as a part of the NGI initiative under cooperative agreement number F30602-00-2-0591.
The solution, many believe, is to layer IP directly over the optical substrate [2]. If IP can be mapped directly onto the WDM layer, some of the functional overlap can be eliminated, potentially collapsing today's vertically layered network architecture into a horizontal model where all network elements work as peers to dynamically establish optical paths. To bring the IP and WDM layers together, however, new capabilities must be added to both layers. A framing standard is needed for carrying packets directly over lambda. Signaling standards are needed so that IP devices can control optical resources [3]. More importantly, with the conventional multi-layered architecture out of the way, automated provisioning systems will gain direct access to WDM resources, and dynamic lightpath provisioning will become easier and more practical to implement.

Once the view about network topology has changed, one will have to re-think routing as well. For example, initially there was fixed routing over fixed circuits (PSTN), and next came dynamic routing over fixed circuits (IP). Subsequently there was a move toward dynamic routing over virtual circuits (i.e., IP over ATM). Now, with recent advances in multi-protocol label switching (MPLS), we have label switching over virtual circuits [4]. Furthermore, industry organizations such as the Optical Internetworking Forum (OIF) and the Internet Engineering Task Force (IETF) are now extending the MPLS-framework (Generalized-MPLS, also referred to as multi-protocol lambda switching (MPAS)) to support not only devices that perform packet switching (routers), but also those that perform switching in time (SONET), wavelength (OXC's), and space. Therefore most likely the next evolution will be label swapping over dynamic circuits or lightpaths (see [3, 5-11]).

The Internet Drafts cited above describe several dynamic routing possibilities [7-11]. The simplest is to treat the optical layer as completely separate from the IP layer. In this "overlay" model, optical transport offers only higher capacity and higher reliability. A more ambitious "integrated" model links the routing decisions at the IP layer with the dynamic reconfiguration capabilities of optical crossconnects (MPAS) [8]. The main goal of these initiatives is to provide a framework for real-time provisioning of optical channels, through combining recent advances in MPLS traffic-engineering control planes with emerging optical switching technologies in a hybrid IP-centric optical network [2].

This article considers the problem of real-time provisioning of optical channels in a hybrid IP-centric DWDM-based networking model. Provisioning in this work implies that an optical channel is successfully routed if both an active path (working) and another alternate node/link-disjoint path (backup) are set up at the same time. Provisioning of connections requires algorithms for route selection, and signaling mechanisms to request and establish connectivity within the network along a chosen route. In particular, the problem of route selection in such wavelength-routed networks is referred to as the "routing and wavelength assignment (RWA) problem" [13]. Here, we present a review of RWA schemes and also compare the performance of two different constraint-based routing/RWA algorithms for dynamic provisioning of the optical channels. Specifically, the RWA schemes are used to compute end-to-end dedicated and shared backup paths to protect against single link/node failures. These algorithms are examples of approaches that might be used to simplify the complex problem of dynamic lightpath computation. Methodologies and associated algorithms for dynamic lightpath computation are outlined. We present an overview of the emerging architectural alternatives of the two-layer model, referred to in the literature as "the interconnection models," for IP-over-optical networks, namely, the overlay, the peer, and the augmented models [8]. Finally, we examine the implications of implementing the proposed RWA schemes for the lightpath provisioning aspects for each of the three emerging interconnection models.

The remainder of this article is organized as follows. We present an overview of the emerging architectural alternatives of the two-layer model. The proposed dynamic RWA algorithm is also presented. We present an overview of fault-tolerant routing, and we present the simulation results. The implementation of real-time provisioning at the optical layer is then presented. Finally we offer a summary and a conclusion.

**IP-over-optical Network Architectural Alternatives (Two-Layer Model)**

In the network model considered here, clients (e.g., IP/MPLS routers) are attached to an optical core network, and connected to their peers over dynamically switched optical paths (lightpaths) spanning potentially multiple OXC's. The interaction between the client and the optical core is over a well-defined signaling and routing interface, referred to as the user-network interface (UNI). Meanwhile, the optical core network consists of multiple OXC's interconnected by optical links in a general mesh topology. This network may be multi-vendor, where individual vendor OXC's constitute sub-networks. Each sub-network itself is assumed to be mesh-connected. The interaction between the sub-networks is over a well-defined signaling and routing interface, referred to as the network-network interface (NNI) (see Fig. 1).

Each OXC is assumed to be capable of switching a data stream from a given input port to a given output port. This switching function is controlled by appropriately configuring a cross-connect table. A lightpath is a fixed bandwidth connection between two network elements such as IP/MPLS routers established via the OXCs. Two IP/MPLS routers are logically connected to each other by a single-hop channel. This logical channel is the so-called lightpath. A continuous lightpath is a path that uses the same wavelength along all links along the entire route form source to destination.
Interconnection Models

One approach for transporting IP traffic over WDM networks is to use a multi-layered architecture comprised of an IP/MPLS layer over ATM over SONET over WDM. If an appropriate interface is designed to provide access to the optical network, multiple higher-layer protocols can request lightpaths to peers connected across the optical network. This architecture has four management layers. Another approach is to use a packet over SONET approach, doing away with the ATM layer, by putting IP/PPP/HDLC into SONET framing. This architecture has three management layers. The fact that both approaches support multiple protocols increases complexity for IP-WDM integration because of various edge interworkings required to route, map, and protect client signals across WDM sub-networks.

The two-layer model, which aims at a tighter integration between IP and optical layers, offers a series of advantages over the current multi-layer architecture model. MPLS [4] and its extension, GMPLS [9], have been proposed as the integrating structure between IP and optical layers. Nevertheless, routing in non-optical and optical parts of hybrid IP networks needs to be coordinated. To examine the architectural alternatives for the two-layer model (IP-over-optical network), it is important to distinguish between the data plane and control planes over the user-network interface (UNI). The IP-over-optical network architecture is classified according to the organization of the control plane, i.e., whether there is a single integrated or separate independent monolithic routing and signaling protocol spanning the IP and optical domains. Several models have been proposed, including overlay, augmented, and peer-to-peer models [8].

The Overlay Model — Under the overlay model, the IP domain is more or less independent of the optical domain, that is, the IP domain acts as a client to the optical domain. The IP/MPLS routing and signaling protocols are independent of the routing and signaling protocols of the optical layer. Thus the topology distribution, path computation, and signaling protocols would have to be defined for the optical domain. In this model, the client routers request high-bandwidth connections (lightpaths) from the optical network through the UNI. The client routers are provided with no knowledge of the optical network topology or resources. In this scenario, the optical network provides point-to-point connection to the IP domain. The overlay model may be statically provisioned using a network management system or may be dynamically provisioned.

The Peer Model — In the peer model, the two layers are collapsed into a single integrated layer managed and traffic engineered in a unified manner. In this regard, the OXCs are treated just like any other router (IP/MPLS routers and OXCs act as peers) and there is only a single instance of a routing protocol spanning an administrative domain consisting of the core optical network and the surrounding edge devices (IP/MPLS routers, ATM switches). Thus, from a routing and signaling point of view, there is no distinction between the UNI, the NNI (network-network-interface), and any other router-to-router interface. This allows the IP edge devices to have full access to the topology of the core network. A common IGP like OSPF or IS-IS may be used to exchange topology information. The assumption in this model is that all the optical switches and the routers have a common addressing scheme.

The Augmented Model — In the augmented model, the IP and optical domains can be functionally separated, each running its own routing protocol, but exchanging full reachability information across the UNI using a standard protocol. For example, IP addresses could be assigned to optical network elements and carried by optical routing protocols to allow reachability information to be shared with the IP domain to support some degree of automated discovery. This model combines the best of the peer and overlay interconnection models; it is relatively easy to deploy compared to the peer model in the near term. Also, this is a convenient solution, since it allows implementation of both provisioning and restoration procedures for optical sub-networks independent of the client network routing. In addition, this approach supports the common scenario in which the optical network and client networks are administered by different entities.

The Central issue in this model is how the routing information is exchanged at the IP-optical UNI: There are two possibilities for this. The first is to consider the interdomain IP routing protocol, BGP, which may be adapted for exchanging routing information between IP and optical domains. The second is to consider the use of OSPF areas (OSPF supports a two-level hierarchical routing scheme through the use of OSPF areas) to exchange routing information across the two domains. On the other hand, running a protocol like BGP across the UNI may be considered too involved, at least for initial implementations of the UNI. A simpler approach would be to limit the reachability information passed through the optical network.

Dynamic RWA

Provisioning of connections requires algorithms for route selection, and signaling mechanisms to request and establish connectivity within the network along a chosen route. The problem of route selection in such wavelength-routed networks is referred to as the "routing and wavelength assignment (RWA) problem," and consists of two sub-problems. The first is the routing problem, which determines the path along which the connection can be established. The second problem is to assign a wavelength (or a set of wavelengths) to each link along the selected path (wavelength assignment problem). Real-time provisioning implies that both the path and wavelength should be chosen/assigned dynamically (dynamic RWA), depending on the network state. In general, all networking models described above, regardless, require route/wavelength computation/assignment to provision a lightpath, i.e., dynamic RWA engine.

Overview of the RWA Problem

Given a set of connections, the problem of setting up lightpaths by routing and assigning a wavelength to each connection is known as the "routing and wavelength assignment (RWA) problem" [12]. Typically, connection requests may be of three types: static, incremental, and dynamic [14]. With static traffic, the entire set of connections is known in advance, and the problem is then to set up lightpaths in a global fashion while minimizing network resources such as the number of wavelengths or the number of fibers in the network. Here, the RWA problem for static traffic is known as static lightpath establishment (SLE) and can be formulated as a mixed-integer linear program [15]. In the incremental-traffic case, connection requests arrive sequentially, a lightpath is established for each connection, and the lightpath remains in the network indefinitely.

For the case of dynamic traffic, a lightpath is set up for each connection request as it arrives, and the lightpath is released after some finite amount of time. The objective in the incremental and dynamic traffic cases is to set up lightpaths and assign wavelengths in a manner that minimizes the...
amount of connection blocking [14]. This problem is referred to as the dynamic lightpath establishment (DLE). Generally, the DLE is more difficult to solve, and therefore heuristics methods are generally employed. Heuristics exist for both the routing sub-problem and the wavelength assignment sub-problem.

For the routing sub-problem, there are three basic approaches that can be found in the literature: fixed routing, fixed-alternate routing, and adaptive routing [14]. Fixed routing is one variant of "static routing" in which routing decisions do not vary with time. Moreover in fixed routing the same fixed route for a given source-destination pair is always selected. The fixed alternate routing approach considers multiple routes between a source-destination pair, and in each node in the network maintains an ordered list of a number of fixed routes to each destination node. When a connection request arrives, the source node attempts to establish the connection on each of the routes from the list in sequence, until a route with a valid wavelength assignment is found. Conversely, in adaptive routing [14, 16--18] the route from a source node to a destination node is chosen dynamically, depending on the network state. Adaptive routing requires extensive support from the control and management protocols to continuously update the routing table at the node. An advantage of adaptive routing is that it results in lower connection blocking than fixed and fixed-alternate routing.

Meanwhile for the wavelength assignment sub-problem, a number of heuristics have been proposed [19--21]. These heuristics are Random Wavelength Assignment, First-Fit, Least-Used, Most-Used, Min-Product, Least-Loaded, MAX-SUM, Relative Capacity Loss, Wavelength Reservation, and Protecting Threshold. In [17] the authors propose an adaptive unconstrained routing (AUR), which incorporates network state information into route computation and channel allocation.

Currently, the algorithms that offer the best performance are Relative Capacity Loss (RCL) [21], and Distributed Relative Capacity Loss (DRCL) [14]. RCL calculates the Relative Capacity Loss for each path on each available wavelength and then chooses the wavelength that minimizes the sum of the relative capacity loss on all the paths. DRCL is proposed in [14] and is based on RCL but it is more efficient in a distributed environment. For a tutorial review of the RWA problem, refer the reader to [14].

Optical networks can also pose added wavelength continuity constraints [13], and these may require the use of wavelength conversion (also referred to as wavelength translation or wavelength changing). A wavelength converter is a device that takes at its input a data channel modulated onto an optical carrier with a wavelength \( \lambda_{in} \), and produces at its output the same data channel modulated onto an optical carrier with a different wavelength \( \lambda_{out} \). If wavelength converters are included in the OXCs in WDM networks, connections can be established without the need to find an unoccupied wavelength, which is the same on all the links making the route. This means that networks with wavelength converters are equivalent to traditional circuit-switched networks. Wavelength converters thus result in improvements in network performance. On the other hand, it has been shown that a careful wavelength assignment in wavelength-continuous networks can lead to improved performance, thus reducing the benefits of wavelength converters [22]. In [22] the authors investigate the benefits of limited wavelength conversion for ring and mesh-torus topologies with fixed shortest path routing. The authors of [23] used a hypercube network to study limited conversion with fixed shortest path routing and a First-Fit wavelength selection algorithm. It is shown that limited wavelength conversion (25 percent) achieves the same performance improvement as full wavelength conversion [22, 23]. Additionally, many other constraints can also serve to complicate the RWA process, especially in all-optical networks. Specifically, besides wavelength continuity requirements, these include analog attenuation effects and power limitations. For example, adequate signal-to-noise ratios (SNRs), crosstalk, and dispersion effects caused by subsystem components and fiber links can be computed along candidate paths. This information can be incorporated into route resolution strategies by defining new cost functions [24]. In [24] the authors have extended the routing and wavelength assignment problem to account for the power degradation of a routed signal due to non-ideal behavior of optical components such as multiplexers, demultiplexers, taps, and fiber links.

The Proposed Dynamic RWA Algorithms

Some combined RWA algorithms are now presented. Specifically, these algorithms integrate and collapse both the routing and wavelength assignment sub-problems into a single dynamic constraint-based routing problem. Thus, the emphasis here is on the adaptive routing problem, rather than focusing on the wavelength assignment problem. It has been shown that the routing scheme has much more of an impact on the overall network performance than the wavelength-assignment scheme [14, 16]. Moreover, both algorithms are also shown to be capable of supporting fault-tolerant adaptive routing and are amenable to fully distributed implementations.

The network is viewed as a multi-layered graph, each corresponding to a specific wavelength. For a connection request and on a given wavelength, Dijkstra's shortest path algorithm, which is suitably modified for WDM networks, is used for computing a constraint path. This is achieved by associating each link in the network with a specific weight function that incorporates WDM-specific information such as the number of available wavelengths and the total wavelengths [18]. This means that the algorithm might compute (on-line) W Paths, each corresponding to one of the W wavelengths. Then one of these paths is selected according to a global selection criterion. Thus the problem of wavelength assignment is totally mitigated and both the routing and wavelength assignment sub-problems are now integrated and collapsed into a single dynamic constraint-based routing problem. This is in contrast to the work reported in [18], in which a single path is first calculated, then a wavelength is assigned to the path by propagating a wavelength request to all the routers along the path. In this way such an algorithm avoids the overhead associated with such a wavelength request (probe message). The emphasis here is on the adaptive routing problem, rather than focusing on the wavelength assignment problem. It has been shown that the routing scheme has much more of an impact on the overall network performance than the wavelength assignment scheme [14--16].

The algorithm is implemented as per the following:

1. First consider a multi-fiber IP-centric DWDM-based network whose physical topology consists of multiple OXCs interconnected via point-to-point WDM links in an arbitrary mesh topology.
2. Assume that none of the OXCs has wavelength conversion capability. Hence, to meet a connection request, a lightpath, which uses the same wavelength on all the links along the entire route form source to destination, has to be set up.
3. Both algorithms are based on a fully distributed implementation in which all nodes maintain a synchronized and identical topology and link state information (traffic engineering database, TED).
4. Assuming that \( W \) is the number of wavelengths per fiber,
the network is represented by \( W \) identical graphs, each conforming to the physical topology and a particular wavelength. Hence the network can be viewed as \( W \) identical wavelength graphs, each representing a wavelength. In view of this multi-graph model, each physical link is now represented by \( W \) virtual links (channels), each corresponding to one of the wavelength graphs. Figure 2 illustrates the concept of the multi-graph approach for a simple network with four nodes, four physical links, and \( W = 2 \).

5. For a given connection request, a constraint route is calculated, for each of the wavelength graphs, throughout the entire network from source to destination, typically using a shortest path algorithm but with the link weights adjusted to attain some sort of local resource optimization. Clearly there are at most \( W \) paths that can be calculated, each corresponding to a given wavelength, provided that each path can meet the given routing constraint. As a result we get the vector \( V = \langle \text{Path}_i, \text{Wavelength}_i, \rangle , i = 1, ..., W \), where the number of entries stored in \( V \) may vary from no entries at all (request is blocked) to a possible maximum of \( W \) entries. Finally, to globally optimize the network resources, provided that the number of entries stored in \( V \) is more than one, an entry \( \langle \text{Path}_i, \text{Wavelength}_i, \rangle \) has to be selected, out of all the other possible entries. Thus by virtually separating wavelengths, both the routing and wavelength assignment sub-problems are now reduced into a single dynamic constraint-based routing problem.

Algorithm I: Full Adaptive Routing — The implementation of this algorithm is as follows:

1. For a given wavelength graph \( \lambda_i \), each virtual link, in each of the \( W \) wavelength graphs, is assigned a cost. Basically, the cost of a link at a given wavelength \( \lambda_i \) is defined here as the inverse of the number of available channels over that particular link. Hence initially the cost of a given link throughout the entire network is set = \((1/F)\), where \( F \) is the number of fibers (per link) connecting two adjacent OXCs. In general, the cost of link \( L_j \) at wavelength \( \lambda_i \), \( C(L_j^{\lambda_i}) \) is given by:

\[
C(L_j^{\lambda_i}) = \frac{1}{F - N(L_j^{\lambda_i})} \quad \text{if } |N(L_j^{\lambda_i})| < |F|
\]
\[
= \infty \quad \text{if } |N(L_j^{\lambda_i})| = |F|
\]

Where \( N(L_j^{\lambda_i}) \) is the number of occupied (unavailable) \( \lambda_i \)'s on link \( L_j \).

2. For a given wavelength \( \lambda_i \), we associate each path throughout the entire network with a total cost, \( C_{sd}^{\lambda_i} \), which is defined here as the summation of the costs of all individual links spanning the entire path from source to destination.

\[
C_{sd}^{\lambda_i} = \sum_{j=1}^{n} C(L_j^{\lambda_i}) \tag{2}
\]

where \( \langle L_1, L_2, ..., L_n \rangle \) is the set of \( n \) links that comprise the path.

3. For a given connection request, run Dijkstra's algorithm on the first wavelength graph \( \lambda_1 \) to find the shortest path (the path with minimum \( C_{sd}^{\lambda_1} \)). Store the calculated path along with its corresponding wavelength \( \lambda_1 \) as the first entry of the vector \( V \). Note that the calculated local path, for a given wavelength graph \( \lambda_1 \), is not necessarily the path with the minimum number of hops.

4. Repeat Step 3 for each of the remaining \( W-1 \) wavelength graphs. Note that the vector \( V \) might now have up to \( W \) entries.

5. Examine the contents of the vector \( V \) and perform one of the following instructions:

a) If the vector \( V \) has no entries at all, reject the connection request; otherwise go to step b.

b) If the vector \( V \) has only one entry, select this entry as the combination \( \langle \text{Path}_1, \text{Wavelength}_1 \rangle \) that satisfies the connection request. After assigning the path, update the weights associated with all links along the entire path (only on the corresponding wavelength graph \( \lambda_1 \)) by basically decremented the number of the available \( \lambda_1 \)'s (channels) on every link along the selected path by one; otherwise go step c.

c) If the vector \( V \) has more than one entry, select one of those entry combinations that satisfy one of the following global path selection schemes:

Total Cost-Based Selection — In this scheme a total cost, \( C_{sd}^{\lambda_i} \), is associated with each computed path within the vector \( V \)</div>
However, that the total cost of the path with wavelength \( \lambda_j \) (1) is higher than that with \( \lambda_i \) (11/12).

**Future Cost-Based Selection** — This scheme uses the same total cost \( C_{\text{tot}} \) of Eq. (3), but with the individual link cost \( C_{ij} \) of Eq. (1) redefined as:

\[
C_{ij}(\lambda, \text{Future}) = \frac{1}{F - 1 - N} \left( \frac{L_j}{L_j} \right) \quad \text{if } N \left( \frac{L_j}{L_j} \right) < [F]
\]

Thus the total future cost of this scheme is given by:

\[
C_{\text{tot}}^{\lambda, \text{Future}} = \sum_{j=1}^{n} C_{ij}(\lambda, \text{Future})
\]

The path with the minimum future cost is selected and assigned to the connection. If all paths have the same future cost of infinity (i.e., along a path at least one link has one link channel available), select the path with the least cost using the original definition of individual link cost of Eq. (1). Note that this selection criterion strongly favors assigning paths with the least current/future utilized resources. This is illustrated in Fig. 3c, where the path with infinite future cost is not selected.

Two comments are in order here. First, for any of the three selection schemes described above, after a path is assigned the weights associated with all links along the selected path should be updated as indicated in step 5b above. Second, all link state updates must be propagated and advertised to all other core nodes throughout the network. Note that the frequency of the link state update per unit time is proportional to the number of accepted/released calls. In other words, the link state update is triggered once for every accepted/released call. As a result, the signaling overhead associated with a high volume of calls may be excessive.

**Algorithm II: Semi-Adaptive Routing** — This algorithm adopts the same implementation procedures developed for the full adaptive algorithm described above, except for the following fundamental differences:

1. A shortest path algorithm (Dijkstra's algorithm) is initially run off-line to calculate the shortest path (only minimum number of hops) between every source-destination node pair (routing tables) throughout the entire network. These off-line computed routing tables are stored at each node in every wavelength graph. Thus the initial routing tables are identical for all \( W \) wavelength graphs.

2. For an initial connection request the ingress node at every wavelength graph consults its own routing table for the shortest path. As a result, similar to algorithm I, we may get as much as \( W \) paths, where one of them can then be selected according to the selection schemes described above.

3. For all the consecutive connection requests, the routing tables remain unchanged, so that Step 2 is repeated until the cost of a link \( L_j \) in a given wavelength graph \( \lambda_i \) goes to infinity (no more available \( \lambda_j \)'s). In this case link \( L_j \) is removed from wavelength graph \( \lambda_i \) and the routing tables are calculated for each node again. Note that Dijkstra's algorithm is run this time online to find the shortest path (the path with minimum \( C_{ij} \)).

Note that the signaling overhead associated with the link state updates for this algorithm is considerably less than that of algorithm I (link state updates are only triggered when the cost of the link goes to infinity). In addition the time associated with computing a path for the semi-adaptive algorithm is less than that of the full adaptive algorithm, since the path is directly read off the routing table.

Finally, in the case where the lightpath is wavelength-continuous, as this work has assumed, optical non-linearities, chromatic dispersion, amplifier spontaneous emission, and other factors together limit the scalability of an all-optical network. Routing in such networks will then have to take into account noise accumulation and dispersion to ensure that lightpaths are established with adequate signal qualities. This work assumes that the all-optical (sub-)network considered is geographically constrained so that all routes will have adequate signal quality, and physical layer attributes can be ignored during routing and wavelength assignment. However, the policies and mechanisms proposed here can be extended to account for physical layer characteristics, and requires future work.

**Fault-Tolerant Routing**

Given the wide range of services envisioned for future IP networks, network survivability is a crucial concern. Survivability schemes can be classified into two forms: protection, which refers to pre-provisioned failure recovery; and restoration, which refers to more dynamic signaled recovery [2]. A common approach to protection is to set up two physically link-disjoint paths for every connection request. One path, called the primary, is used to transmit data, while the other path is reserved as a backup in the event that a link in the primary path fails. To further protect against node failures, the primary and backup paths may also be node-disjoint [25, 26].

Fixed-alternate routing provides a straightforward approach to handling protection [17]. On the other hand, in adaptive routing, a protection scheme may be implemented in which the backup path is set up immediately after the primary path has been established [14]. The same routing protocol may be used to determine the backup path, with the exception that a link cost is set to infinity if that link is being used by the primary path. The resulting route will then be link-disjoint from the primary path. Since these schemes require backup path routing at setup time, they must be more closely incorporated with the primary lightpath RWA algorithms. On the other hand, channel restoration does not rely on pre-computed backup routes, and instead dynamically re-computes a new path for a broken channel [26]. This has the advantage of low overhead in the absence of failures. However, this does not guarantee successful recovery, since the attempt to establish a new path may fail...
due to resource shortage at the time of failure recovery. Additionally, recovery timescales are usually longer [2].

By making use of the WDM channel routing capabilities, a variety of lightpath protection schemes can be designed. For example, dedicated backup channels can be provisioned for users requiring high availability. Here a pre-computed link-disjoint backup channel is reserved for each primary channel at setup time, and in case of a fault condition on the primary path, a channel switch-over is performed. The dedicated backup reservation method has the advantage of shorter restoration time since the resources are reserved for the backup path when establishing the primary path itself. However, this method reserves excessive resources. For better resource utilization, multiplexing techniques [27] can be employed. If two primary lightpaths do not fail simultaneously, their backup lightpaths can share a wavelength channel. However, in case of primary link failure, the backup capability of the other is no longer preserved. Therefore, although channel-blocking rates will be less (than the dedicated case), channel recovery probability will also be lower.

In addition, recently the concept of a shared risk link group (SRLG) definition has also been proposed to help identify risk associations between various entities (see [28]). This concept is used to ensure that the primary and the backup path are not affected by the same failure. By using this concept, adequate resource "disjointness" can be introduced into the constraint-based path computation phase, thereby reducing the probability of simultaneous lightpath failures (e.g., between working and protection paths). Further details are beyond the scope of this article, and interested readers are referred to [8, 28].

Overall, the proposed adaptive RWA scheme can be extended to ensure diversity in routes. This can be achieved by coordinating each diversely routed lightpath group by a single network entity. To create a diversely routed lightpath group, a user registers with a coordinator and receives the group identifier. For groups originating through the same client router, this router would typically act as the coordinator. To ensure diversity in routes, N SRLG and node disjoint routes through the network are selected, where N represents the number of diverse routes required.

Simulation Results

The performance of the proposed dynamic RWA algorithms is evaluated via simulation of the mesh-based NSFNET shown in Fig. 4. The NSFNET consists of 14 nodes and 21 physical links. Each adjacent node pair is connected through a bi-directional physical link that consists of N fibers, where each fiber is assumed to have the same number of wavelengths (W). The simulation results calculated in this section assume that N = 2 and W = 4. We use a dynamic traffic model in which call requests arrive at each node according to a Poisson process with a network arrival rate λ. An arrival session is equally likely to be destined to any node in the network. The session holding time is assumed to be exponentially distributed with mean 1/μ. The blocking probability is the metric used to evaluate the network performance. In each simulation run, a large number of requests are generated one after the other, and the results are averaged over many simulation runs. If at any time a connection request cannot be satisfied according to the algorithms developed above, the connection request is dropped.

Figure 5 shows the simulated blocking probability vs. the call arrival rate for both algorithms when the total cost-based path selection criterion (scheme I) is used. The simulated blocking probability is also shown in the figure for the conventional static RWA scheme used in most algorithms. As expected, it can be seen from the figure that the performance of both dynamic algorithms is significantly better than that of the static algorithm. Note, however, that the performance of the full-adaptive algorithm is slightly better than that of the semi-adaptive algorithm. These results always hold, independent of which path selection scheme described above (I, II, or III) is used.

Figure 6 shows the simulated blocking probability vs. the call arrival rate of the full-adaptive algorithm for all three path selection schemes described above. As can be seen from Fig. 6, the path selection process based on future cost performs the best, followed by the path selection process based on the balanced cost, then by the path selection process based on total cost. Also included in the figure is the adaptive First Fit (FF) algorithm [14]. Note that the performance of the future cost algorithm is significantly better than that of the FF. However, the performance of the balanced cost algorithm is almost the same as that of the FF algorithm (same results were also obtained in [14] with the DRCL algorithm), and both of them outperform the total cost scheme.

Figure 7 shows the required number of wavelengths vs. the call arrival rate for three different cases:

- A system that uses end-to-end dedicated backup paths to provide 100 percent protection against single link/node failures.
- A system that uses end-to-end shared backup paths to provide 100 percent protection against single link/node failures.
- A system that provides no protection at all. As expected, the number of wavelengths required to provide shared protection is considerably less than that required for the dedicated case (also reported in [26]).
IEEE Network

node has a representation of the full physical network topology given threshold of the number of available wavelengths per fiber, below which the updates can be triggered. Once each link state updates can be triggered, for instance, based on updating and advertising all of the above link attributes. The have defined and assigned implementing a simplified link state advertisement algorithm to capture optical link parameters and any constraints specific to model extended OSPF. This algorithm is capable of periodically reservable link bandwidth, current bandwidth reservation, current bandwidth usage, and link coloring [1291]. These extensions for different elements of the physical plant hierarchy, i.e., by component is to define a naming and addressing convention for different elements of the physical plant hierarchy [11]. Here we have defined and assigned a naming and addressing convention for different elements of the physical plant hierarchy, i.e., by implementing a simplified link state advertisement algorithm to model extended OSPF. This algorithm is capable of periodically updating and advertising all of the above link attributes. The link state updates can be triggered, for instance, based on a given threshold of the number of available wavelengths per fiber, below which the updates can be triggered. Once each node has a representation of the full physical network topology and the available resources on each link, a path selection algorithm is required, i.e., dynamic RWA.

A path selection process: Uses the information distributed by the dynamic link state advertisement algorithm to select an explicit route that meets the specific requirements of the traffic flow. This process can be performed either off-line or online using a constraint-based routing calculation. The source router (pce model) or the border OXC/central management node (augmented/overlay models) are basically responsible for computing the complete path all the way to the destination through the optical domain, and then initiating path setup using the signaling protocol (e.g., CR-LDP or RSVP). The route may be specified either as a series of nodes (routers/OXCs) or in terms of the specific links used (as long as IP addresses are associated with these links).

Route Selection Using the Proposed RWA
Numerous policies can be used to route lightpaths through the network, such as the constraint-based routing algorithms proposed here. This scheme can be used directly for computing the route and assigning the wavelength for both the overlay and the augmented models. In this case a connection request is initiated by a client IP/MPLS router (border router, that is, a router directly connected to the optical network) and sent to an ingress optical node (border OXC, that is, the OXC connected to the border router) using UNI signaling. Such provisioning requests may specify the desired destination client router. Note that the source end-point is implicit in this case. The ingress optical node processes the request and computes an appropriate route along with a wavelength through the network (using topology and state information that has been propagated using OSPF link state advertisements). Note that the request may also be received by an ingress OXC from a central management node, specifying the source and destination end-points.

The routing within the optical and IP domains in the case of the augmented model may be separated, with a standard routing protocol running between domains [7, 8]. This is similar to the IP interdomain routing model, where the central issue is how the routing information is exchanged at the IP-optical UNI. There are two possibilities for this. The first is to consider the interdomain IP routing protocol, BGP, which may be adopted for exchanging routing information between IP and optical domains. The second is to consider the use of OSPF areas (OSPF supports a two-level hierarchical routing

Real-Time Provisioning at the Optical Layer
Provisioning end-to-end circuits is an endless source of struggle for service providers and a frustration for end-users. Provisioning of connections requires algorithms for route selection, and signaling mechanisms to request and establish connectivity within the network along a chosen route. In this section we examine the problem of route selection in the context of applying/adapting the RWA algorithm presented above to each of the three interconnection models described earlier. The implications on both the route selection and signaling mechanism components will also be outlined for each of the three interconnection models.

Dynamic Lightpath Computation
Dynamic computation of a lightpath involves the implementation of two traffic engineering components: an information distribution mechanism that provide knowledge of the relevant attributes of available network resources, and a path selection process that uses the information distributed by the dynamic link state advertisement algorithm to select a path that meets the specific requirements of the traffic flow. In a fully distributed IP-over-optical network implementation, these are:

An information distribution mechanism: Provides knowledge of the network's topology and the available resources. This component is implemented by defining relatively simple extensions to the interior gateway protocol (IGP), e.g., open shortest path first (OSPF) so that link attributes are included as part of each router's link state advertisement. Some of the traffic-engineering extensions that need to be added to the IGP link state advertisement include maximum link bandwidth, maximum reservable link bandwidth, current bandwidth reservation, current bandwidth usage, and link coloring [29]. These extensions capture optical link parameters and any constraints specific to optical networks. Such topology and link state information is then flooded to all nodes via updates. Another important component is to define a naming and addressing convention for different elements of the physical plant hierarchy [11]. Here we have defined and assigned a naming and addressing convention for different elements of the physical plant hierarchy, i.e., by implementing a simplified link state advertisement algorithm to model extended OSPF. This algorithm is capable of periodically updating and advertising all of the above link attributes. The link state updates can be triggered, for instance, based on a given threshold of the number of available wavelengths per fiber, below which the updates can be triggered. Once each node has a representation of the full physical network topology and the available resources on each link, a path selection algorithm is required, i.e., dynamic RWA.

A path selection process: Uses the information distributed by the dynamic link state advertisement algorithm to select an explicit route that meets the specific requirements of the traffic flow. This process can be performed either off-line or online using a constraint-based routing calculation. The source router (pce model) or the border OXC/central management node (augmented/overlay models) are basically responsible for computing the complete path all the way to the destination through the optical domain, and then initiating path setup using the signaling protocol (e.g., CR-LDP or RSVP). The route may be specified either as a series of nodes (routers/OXCs) or in terms of the specific links used (as long as IP addresses are associated with these links).

Route Selection Using the Proposed RWA
Numerous policies can be used to route lightpaths through the network, such as the constraint-based routing algorithms proposed here. This scheme can be used directly for computing the route and assigning the wavelength for both the overlay and the augmented models. In this case a connection request is initiated by a client IP/MPLS router (border router, that is, a router directly connected to the optical network) and sent to an ingress optical node (border OXC, that is, the OXC connected to the border router) using UNI signaling. Such provisioning requests may specify the desired destination client router. Note that the source end-point is implicit in this case. The ingress optical node processes the request and computes an appropriate route along with a wavelength through the network (using topology and state information that has been propagated using OSPF link state advertisements). Note that the request may also be received by an ingress OXC from a central management node, specifying the source and destination end-points.

The routing within the optical and IP domains in the case of the augmented model may be separated, with a standard routing protocol running between domains [7, 8]. This is similar to the IP interdomain routing model, where the central issue is how the routing information is exchanged at the IP-optical UNI. There are two possibilities for this. The first is to consider the interdomain IP routing protocol, BGP, which may be adopted for exchanging routing information between IP and optical domains. The second is to consider the use of OSPF areas (OSPF supports a two-level hierarchical routing
scheme through the use of OSPF areas) to exchange routing information across the two domains [7, 8].

However, in the case of the peer model additional extensions need to be added to the routing protocol (OSPF) so that the segment of the entire route that crosses the optical core (between the ingress and egress OXCs) must be treated as a virtual link of fixed capacity and advertised as such in further OSPF updates. The routing in this case is referred to as a "flat" routing organization [7-8]. Under this approach there is only one instance of the routing protocol running in the IP and optical domains. An IGP such as OSPF or IS-IS with suitable optical extensions is used to exchange topology information. These optical extensions will capture the unique optical link parameters. The OXCs and the routers maintain the same link state database. The routers can then compute end-to-end paths to other routers across the OXCs. This lightpath is always a tunnel across the optical network between edge routers. Once created such lightpaths are treated as virtual links and are used in traffic engineering and route computation. As and when forwarding adjacencies (FAs) are introduced in the link state, corresponding links over the IP optical interface are removed from the link state advertisements. Finally, the details of the optical network are completely replaced by the FAs advertised in the link state [7, 8].

RWA Implications on Signaling Mechanisms

Once a lightpath request from a source is received by the ingress node, it computes the complete path all the way to the destination through the optical domain using the proposed RWA algorithm. The output of this calculation is an explicit route consisting of a sequence of hops that provides the shortest path through the network that meets the constraints. This explicit route is then passed to the signaling component that initiates path setup (to reserve resources) using the signaling protocol, e.g., CR-LDP or RSVP-TE [9]. Note that the implications of using the proposed RWA scheme on the signaling is that the conventional overhead associated with the wavelength request (probe message) is no longer needed, since the RWA scheme selects the route and assigns the wavelength simultaneously.

Conclusion

This article has considered the problem of real-time provisioning of optical channels in a hybrid IP-centric DWDM-based networking model. Provisioning implies that an optical channel is successfully routed if both an active path (working) and another alternate link-disjoint path (backup) are set up at the same time. Specifically, the work presented here has addressed the implementation issues of the path selection component of the traffic engineering problem in such a network. Methodologies and associated algorithms for dynamic lightpath computation were outlined.

We have presented and compared the performance of two different constraint-based routing and wavelength assignment (RWA) algorithms for dynamic provisioning of the optical channels. Specifically, the RWA scheme is used to compute end-to-end dedicated and shared backup paths to protect against single link/node failures. Three path selection schemes have also been proposed for each algorithm. Both algorithms are based on a fully distributed implementation. The performance of both algorithms was then compared with that of the conventional static RWA algorithm. It was shown that the dynamic full-adaptive algorithm outperforms the semi-adaptive algorithm, and both algorithms significantly outperform the conventional static algorithm. It was also shown that the Future Cost-Based Selection scheme outperforms both the total-based and the balanced selection schemes.

Acknowledgment

The authors would like to thank Nasir Ghani and Debanjan Saha for their helpful comments and discussions.

References


Additional Reading

Biographies

CHADI M. ASSL (assl@ee1.so10.engr.cuny.edu) received the B.E. degree in electrical and electronics engineering from the Lebanese University, Beirut, in 1997, and the M.S. degree in electrical engineering from the City University of New York in 1999. He is currently pursuing the Ph.D. degree in electrical engineering at the City University of New York. In the summer of 2000 he was an intern at Sorrento Networks, San Diego, where he developed software for the RWA and wavelength conversion in optical networks. His current research interests are in the area of optical networking, network architectures, IP over WDM, and specifically protection/restoration and real-time provisioning.

ABDALLAH A. SHAMI (sham@ee1.so10.engr.cuny.edu) received the B.E. degree in electrical and electronics engineering from the Lebanese University, Beirut, in 1997, the M.S. degree in electrical engineering from the Saint Joseph University in Lebanon in 1998. He is currently pursuing the Ph.D. degree in electrical engineering at the City University of New York. In the summer of 2000 he was an intern at Sorrento Networks, San Diego, where he developed software for the RWA and wavelength conversion in optical networks. His current research interests are in the area of optical networking, IP over WDM, and specifically traffic engineering issues and real-time provisioning.

MOHAMED A. ALI (eeali@ee1.so10.engr.cuny.edu) received his Ph.D. degree in electrical engineering from the City College of the City University of New York in 1999. He is a faculty member at Sorrento Networks, San Diego, CA, where he is currently a professor. His research interests cover the fields of optical device technology, high-speed optical communication systems and architectures, optical switches and high-performance IP routers, and WDM transmission and networking. His major interests are in the areas of computer simulation and modeling of high-speed communication systems. He has published more than 60 papers in professional journals and international conferences. Dr. Ali is a recipient of the NSF Faculty Career Development Award.

RUSS KURTZ (rkurtz@realtech.com) is CTO and VP of REALabs at REALTECH Systems Corporation. In this capacity, he is responsible for an organization that provides premium consulting work with leading network equipment vendors and carriers. His group is also responsible for assessing network architecture and service configurations. Kurtz received a BSEE from Purdue University in 1977 and a Doctorate in relativistic physics in 1981 from the University of Texas at Austin. He joined Bell Labs in 1982. In 1990 he assumed responsibility for evolution planning for AT&T's transport network, including office cross-connect and multiplexing equipment and fiber optic line technologies. This work resulted in AT&T's current program to comprehensively deploy SONET-based WDM line systems in ring and point-to-point arrangements. In partnership with then AT&T Network Cable Systems, the work helped define what is now known as Lucent's Trust Wave dispersion compensating fiber product. In 1995 he took over a group responsible for developing an ATM infrastructure for AT&T's SONET and packet-based data services. Kurtz joined REALTECH Systems Corporation in 1999. He has authored and co-authored public papers on electronic hardware designs and telecommunications network design.

DAN GUO (dguo@turinetworks.com) is a senior staff engineer at Turin Networks, Petaluma, CA, where he helped create the Traverse(tm) Multiservice Optical Transport Platform. Prior to joining Turin Networks, he was a manager and network architect at Sorrento Networks, San Diego, CA, where he designed Sorrento's optical channel protection protocol engine and wavelength routing software. He previously held senior engineering positions at Cascade Communications Inc., Waltham, MA, and at GTE Laboratories, Waltham, MA, working on IP and ATM network design and traffic engineering. Prior to his career in industry, he was a member of the research staff in the Center for Telecommunication Research (CTR) at Columbia University, and a research investigator in the Optical Network Technology Consortium (ONTC) funded by DARPA. He received a B.S. degree from the University of Science and Technology of China in 1987, and M.Phil and Ph.D. degrees in computer science from Columbia University, New York, NY, in 1993 and 1995, respectively. He has published more than 20 papers in research journals and international conferences.

Errata

In the article entitled “A Survey on TCP-Friendly Congestion Control,” published in the May/June issue of IEEE Network Magazine, the following table was reproduced incorrectly. The correct version of the table is published here.

<table>
<thead>
<tr>
<th>Protocol</th>
<th>Unicast/multicast</th>
<th>Congestion control mechanism</th>
<th>Network support</th>
<th>Protocol complexity</th>
<th>Smoothness of the rate</th>
<th>Bias against high RTTs</th>
<th>TCP friendliness</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAP</td>
<td>Unicast</td>
<td>Rate</td>
<td>End-to-end</td>
<td>Low</td>
<td>Sawtooth</td>
<td>Yes</td>
<td>Limited</td>
</tr>
<tr>
<td>LDA (+)</td>
<td>Unicast</td>
<td>Rate</td>
<td>End-to-end</td>
<td>High</td>
<td>Sawtooth</td>
<td>Yes</td>
<td>Acceptable</td>
</tr>
<tr>
<td>TCRC</td>
<td>Multicast</td>
<td>Rate</td>
<td>End-to-end</td>
<td>Medium</td>
<td>Smooth</td>
<td>Yes</td>
<td>Good</td>
</tr>
<tr>
<td>TEAR</td>
<td>Multicast</td>
<td>Rate</td>
<td>End-to-end</td>
<td>Low</td>
<td>Smooth</td>
<td>Yes</td>
<td>Good</td>
</tr>
<tr>
<td>MLA &amp; LPR</td>
<td>Multicast</td>
<td>Window</td>
<td>End-to-end</td>
<td>Low</td>
<td>Sawtooth</td>
<td>Yes</td>
<td>Good</td>
</tr>
<tr>
<td>MTCP</td>
<td>Multicast</td>
<td>Required</td>
<td>Window</td>
<td>Low</td>
<td>Sawtooth</td>
<td>Yes</td>
<td>Good</td>
</tr>
<tr>
<td>NCA</td>
<td>Multicast</td>
<td>Required</td>
<td>Window</td>
<td>Medium</td>
<td>Sawtooth</td>
<td>Yes</td>
<td>Good</td>
</tr>
<tr>
<td>PASMCC</td>
<td>Multicast</td>
<td>Optional</td>
<td>Window</td>
<td>Medium</td>
<td>Sawtooth</td>
<td>Yes</td>
<td>Good</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Protocol</th>
<th>Unicast/multicast</th>
<th>Congestion control mechanism</th>
<th>Network support</th>
<th>Protocol complexity</th>
<th>Smoothness of the rate</th>
<th>Bias against high RTTs</th>
<th>TCP friendliness</th>
</tr>
</thead>
<tbody>
<tr>
<td>RLC</td>
<td>Multicast</td>
<td>Rate</td>
<td>End-to-end</td>
<td>Medium</td>
<td>Layer-dependent</td>
<td>No</td>
<td>Acceptable</td>
</tr>
<tr>
<td>FUD-DL</td>
<td>Multicast</td>
<td>Rate</td>
<td>End-to-end</td>
<td>High</td>
<td>Layer-dependent</td>
<td>No</td>
<td>Acceptable</td>
</tr>
<tr>
<td>LTS</td>
<td>Multicast</td>
<td>Rate</td>
<td>End-to-end</td>
<td>Medium</td>
<td>Layer-dependent</td>
<td>Yes</td>
<td>Acceptable</td>
</tr>
<tr>
<td>TRFR</td>
<td>Multicast</td>
<td>Rate</td>
<td>End-to-end</td>
<td>Medium</td>
<td>Layer-dependent</td>
<td>Yes</td>
<td>Acceptable</td>
</tr>
<tr>
<td>MDIA</td>
<td>Multicast</td>
<td>Rate</td>
<td>End-to-end</td>
<td>Medium</td>
<td>Layer-dependent</td>
<td>Yes</td>
<td>Acceptable</td>
</tr>
<tr>
<td>Rainbow</td>
<td>Multicast</td>
<td>Window</td>
<td>Required</td>
<td>Low</td>
<td>Sawtooth</td>
<td>Yes</td>
<td>Good</td>
</tr>
</tbody>
</table>

Table 1. Characteristics of the presented congestion control protocols.