Restoration Design in IP over Reconfigurable All-Optical Networks

Angela L. Chiu[1], Gagan Choudhury[1], Robert Doverspike[1], and Guangzhi Li[2]

[1] AT&T Labs Research, 200 Laurel Avenue, Middletown, NJ 07748, USA
[2] AT&T Labs Research, 180 Park Avenue, Florham Park, NJ 07932, USA
{chiu,rdd,gli}@research.att.com, gchoudhury@att.com

Abstract. Large IP backbone networks today are mostly deployed directly over sequences of point-to-point DWDM systems or chains of newer ROADM-based ultra long haul systems, interconnected by OEO regenerators. The next generation core optical network is moving toward an all-optical network architecture that is based on multi-degree ROADMs to reduce OEO regeneration cost as well as enabling automatic reconfigurability and dynamic restoration. In this paper, we study the restoration design in this new IP over reconfigurable all-optical network architecture to satisfy the resilience requirements for both IP and wavelength services. We propose two novel restoration schemes: 2-Phase Fast Reroute mechanism with optimized Traffic Engineering algorithm for restoring IP services and shared mesh restoration with standbys for restoring wavelength services. They both meet the requirement of sub-second restoration time and also maximize sharing among different failures with the objective of minimizing either overall capacity or overall cost. To further reduce the required restoration capacity in both IP layer and optical layer and address failures in both layers efficiently, we also propose an integrated IP-over-optical layer restoration strategy that enables sharing of restoration capacity among non-simultaneous failures across both IP and optical layers. Simulation results demonstrate significant improvements using our proposed schemes comparing with existing ones.

Keywords: IP-over-Optical, ULH, ROADM, reconfigurable all-optical network, IP service, wavelength service, fast reroute, traffic engineering, shared mesh restoration, restoration overbuild.

1 Introduction

After many years of research and industry efforts, ultra long haul (ULH) technologies for DWDM transport are maturing and carriers are deploying them for high capacity and capital savings [1, 2]. A first-generation ULH network typically consists of a set of point-to-point linear systems. Each linear system has two terminals. Between the two terminals, there are one or multiple reconfigurable optical add-drop multiplexers (ROADMs), where traffic can be added/dropped or expressed through optically (such an OADM is also called a degree-2 ROADM). With such a linear ULH system, a wavelength connection (also called a lightpath) is able to travel a long distance

(typically 1500km or longer) without requiring optical-electronic-optical (OEO) regeneration. This distance limit is called "ULH-reach". An OEO regenerator is needed when a connection length is longer than the ULH-reach or when it has to travel through two linear ULH systems even if its length is within the ULH-reach. Since the OEO regenerators are expensive devices, such a first generation ULH network is not inexpensive. They also complicate dynamic reconfiguration and restoration in the optical layer (OL).

In order to reduce the cost of OEO regeneration and enable automatic reconfigurability and dynamic restoration, next generation core optical networks are moving toward all-optical mesh networks from point-to-point linear ULH systems. This can be done via (1) converting the terminals and degree-2 ROADMs to higher-degree ROADMs to switch and route wavelengths optically [3, 4], also called as photonic cross connect (PXC), and (2) strategically placing OEO regenerators to reduce cost.

The next generation core IP network is the principal overlay network that is transported over the ROADM network, as shown in figure 1. All the middle SONET and Digital Cross-Connect System (DCS) layers are eliminated. Traditional sub-wavelength TDM private line service will be transported over the IP network via pseudo-wire circuit emulation with guaranteed minimum latency and quality of service (QoS) [5, 6]. Besides providing direct links for the IP layer, the optical network also provides wavelength services via optical connections that consist of one or multiple wavelengths. Both IP and wavelength services have very stringent quality requirements for their high priority traffic. One main requirement is resiliency against network failures including two main features: 1)The ability to restore within sub-second for high priority traffic for any single link/node failure; 2) For a small percentage of the traffic that is mission critical, the ability to restore from any double failures.
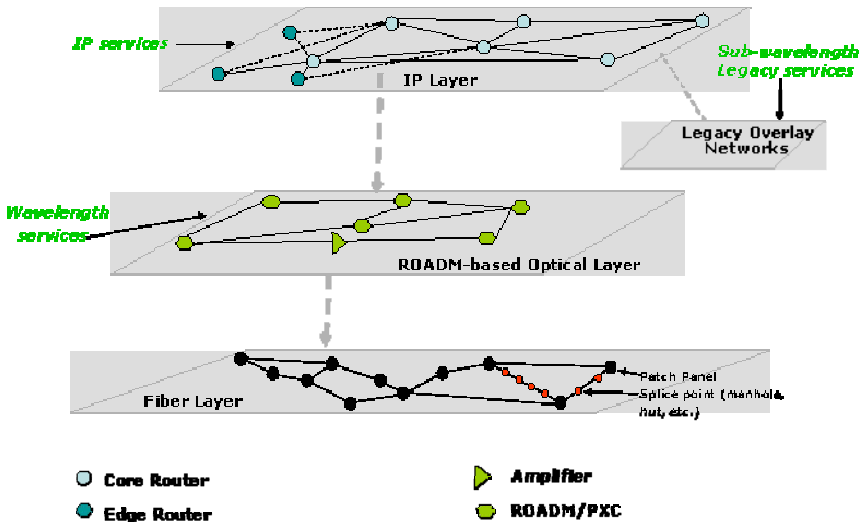

Fig. 1. IP over Reconfigurable All-Optical Network Architecture

The main challenge in reliable network design is to provide fast restoration while planning restoration capacity in a cost effective manner. After evaluating different IP-over-optical network architectures, we find that it is most cost effective to restore IP traffic in IP layer and restore wavelength traffic in optical layer since optical layer cannot restore failures that originate at the IP layer, such as router failures or router line card failures, unless some integrated method is used. However, the IP network is originally designed to support best-effort traffic and IP routing protocol re-convergence may take much longer than the sub-second restoration requirement. In this paper, we propose a novel 2-Phase Fast Reroute (FRR) mechanism with optimized Traffic Engineering algorithm to restore IP services. It meets the requirement of sub-second restoration time [7] and also maximizes sharing among single/double failures of links, routers and Shared Risk Groups (SRGs) and also among different phases of restoration with the objective of minimizing either overall capacity or overall cost. For wavelength service restoration, we propose to use shared mesh restoration (SMR) with standbys. Standbys are pre-cross-connected lightpaths providing connectivity between switching nodes (nodes with fiber link degree-2 or higher). It overcomes the problem of optical impairments for long connections, avoids today's wavelength power "balancing" delays, provides wavelength conversion for capacity efficiency and allows sharing of links across non-simultaneous failures.

Shared mesh restoration has been widely used in opaque optical networks such as AT&T's Intelligent Optical Network [8], where all restoration channels on each link are pre-installed and equipped with OEO regeneration at each end. They are thus optically isolated from other channels. When a failure is detected, a pre-planned restoration path will be dynamically established by cross-connecting unassigned channels. The whole process is automatic and rapid. In an all-optical ULH network, dynamically establishing a new restoration connection from scratch involves not only tuning the lasers and receivers to the appropriate frequencies and cross connecting the ROADMs/PXCs, but also triggering several feedback loop segments that are responsible for power equalization. This is because the new restoration wavelength(s) change the power profile on each link along the restoration path. The whole process can be slow and often unacceptable to large carriers. Furthermore, unlike an opaque optical network, an all-optical ULH channel does not necessarily have OEO regeneration at each end, thus extra OEO regenerators may be required in some nodes. Optimal regenerator placement becomes a critical problem in all-optical ULH network design. Because of these differences between electrical/opaque-optical networks and all-optical ULH networks, shared mesh restoration schemes [9, 10, 11] for opaque optical networks cannot be applied to all-optical ULH networks. Here, we propose using SMR with pre-configured (standby) lightpaths for restoring wavelength services in all-optical ULH networks.

To further reduce the required restoration capacity in both IP layer and optical layer and address failures in both IP and optical layers efficiently, we also propose to use an integrated IP-over-Optical restoration strategy that enables sharing of restoration capacity among non-simultaneous failures across layers. The basic idea is to install spare line cards on IP routers and automatically configure new IP links (or increase the capacity of existing IP links) via the optical layer after an IP-layer failure. Since component unavailability at the IP layer (e.g., failure/maintenance of a router line card or common equipment) and the optical layer (e.g., failure/maintenance of a

transponder or amplifier) are generally statistically independent (i.e., nonsimultaneous with high probability), this is a clever method to pool spare capacity in the optical layer to cover failures at both IP and optical layers.

The remainder of the paper is organized as follows. Section 2 briefly reviews the assumed next generation core optical network architecture. Section 3 describes the proposed schemes for restoring IP and wavelength services. Section 4 presents numerical studies demonstrating performance advantages of the proposed schemes comparing with existing methods. The comparisons are based on simulations using a representative network. We conclude the work in section 5.

2   Next Generation Core Optical Network Architecture

The next generation core optical network consists of a set of ROADM-based network nodes interconnected by multi-wavelength fiber links with optical amplifiers (OA, also known as in-line amplifiers ILA) in a mesh topology [12]. The nodes are usually multi-degree ROADMs with wavelength blocking and/or wavelength switching capability. It also provides network operators a way to dynamically drop and add wavelengths at network nodes without having to manually balance optical parameters every time when distances traveled by individual wavelengths are changed. Figure 2 illustrates a state-of-art ROADM architecture with a broadcast-and-select switch fabric [13, 14, 15, 4, 16]: splitters send copies of each entering multi-channel signal to each output port. Nx1 wavelength selective switches (WSS's) are used to choose which of the input channels at each frequency to put on the output fiber. These devices can simultaneously provide power equalization at the channel level.
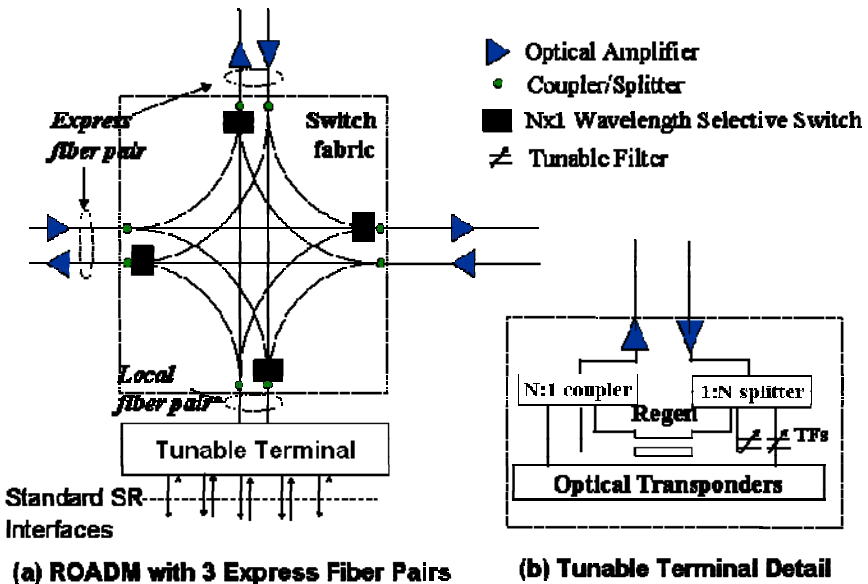

Fig. 2. A Broadcast-and-Select ROADM Architecture

From the perspective of the switch fabric, the terminal used to add and drop channels is just another output port. It contains tunable lasers and filters that allow any wavelength at the terminal/switch fabric interface to be routed to any terminal transmit/receive port. It also contains regenerators that can do wavelength translation. Each wavelength can carry a 10Gbps or 40Gbps optical signal, and is moving towards 100Gbps rate in laboratory experiments. The optical signal is able to travel a long distance (typically 1500km or longer), i.e., the ULH-reach, without requiring OEO regeneration. Note that the ULH-reach is typically rate dependent. The main function of the reconfigurable all-optical network is to offer wavelength services as well as providing transport for links at the higher layers, mainly the IP layer, as shown in figure 1. In order to provide high resiliency for both IP and wavelength services, we examine the restoration design in the IP over reconfigurable all-optical network architecture.

The IP core network is built on top of the core optical network. We assume that the core IP routers are collocated at the same office as core optical ROADMs. IP links are wavelength connections in the optical network. Multiple IP links often share a common optical link. Thus a single optical layer failure would cause multiple simultaneous IP link failures. But there is a high probability that IP router/line card failures and optical layer failures are non-simultaneous. We take this observation and propose a novel integrated IP-over-optical layer restoration scheme.

3   Restoration Design in IP over Reconfigurable All-Optical Networks

3.1   IP Service Restoration: 2-Phase Fast Reroute with Optimized Traffic Engineering

A principal challenge for IP service restoration is to provide sub-second single-failure restoration for high-priority IP services while maintaining efficient bandwidth utilization. Two most prevalent methods for IP-layer restoration today are IP reroute and MPLS Fast Reroute (FRR) [17]. IP reroute is the default and the most common restoration method in large commercial IP networks today. It routes traffic along the shortest path using certain metric such as hop count, route mile. It uses Interior Gateway Protocol (IGP) protocols OSPF or IS-IS signaling for general topology discovery and updates and then re-computes paths upon a failure. Using default OSPF or IS-IS timer values, the re-convergence may take 10's of seconds. Currently, studies and experiments are being conducted to aggressively tuning down the timers so the re-convergence time can be reduced to a few seconds [18, 19]. MPLS Fast Reroute is an IETF standardized protocol [17] where primary and backup (restoration) Label Switched Paths (LSPs) are established for next-hop or next-next-hop MPLS FRR. When a failure is detected at the upstream router from the failure, the MPLS forwarding label for the backup LSP is "pushed" on the MPLS shim header at the upstream router and "popped" at the downstream or next-next-hop router. These labels are pre-calculated and stored in the forwarding tables, so restoration is very fast. However, in this scheme, IP flows stay routed over the backup paths until the failure is restored. Because these paths are segmental "patches" to the primary paths,

the technique has poor capacity for restoring all traffic assuming that the backup paths follow the shortest paths. Algorithms for pure IP based Fast Reroute (without using MPLS signaling) have also been proposed and analyzed [18, 20].

For restoring IP services, we propose a novel 2-Phase Fast Reroute mechanism with optimized Traffic Engineering algorithm. It meets the requirement of sub-second restoration time and also maximizes sharing among single/double failures of links, routers and Shared Risk Groups and also among different phases of restoration with the objective of minimizing either overall capacity or overall cost. In phase 1, pre-computed next-hop or next-next-hop FRR tunnels would be used to restore only non-best effort traffic. Traffic switchover time has been measured in lab experiments at 50ms. Failed best effort traffic is not restored until the second phase. Note that best effort traffic that is on the backup path of the rerouted traffic is not guaranteed to have capacity, although for most failure states most flows would have enough capacity. However, this phase lasts only for a few seconds until OSPF/IS-IS or OSPF-TE/IS-IS-TE re-converges and provides a more optimal end-to-end path to be established for every primary LSP affected by the failure. During the second phase, where failed traffic is rerouted over an optimized end-to-end path, sufficient capacity is being planned for all traffic including all best effort traffic that has been indirectly affected by the failure. The failure information and Link Bandwidth availability information is obtained from IGP and its TE extension. Using this information, new end-to-end paths are established for all failed primary tunnels using a Constrained Shortest Path (CSPF) algorithm. After that all traffic flowing over the Next-Next-Hop FRR tunnels are switched over to the new and more efficient end-to-end primary tunnels. This phase lasts until the associated failure is repaired.

A key aspect of the 2-Phase FRR method is that an optimized traffic engineering algorithm is used during both phases of restoration that attempts to maximize sharing among all independent failure scenarios and attempts to minimize either the total capacity or total cost (including transport cost and IP port cost) or a linear combination of the two. Traffic Engineering algorithms have been proposed and analyzed in the past (e.g., see [21] and the references therein) but they have not been used in the past during two phases of Fast Reroute. Our proposed algorithm is distributed in nature. For each failure scenario, restoration traffic is routed using a Dijkstra shortest path algorithm where link weights along the path are chosen dynamically and depends on available and total capacity of the link, latency, and other cost measures such as IP port costs (during the operational phase this information is available through router provisioning and from OSPF-TE or IS-IS-TE Link State Advertisement packets). The algorithm reuses capacity already allocated for other independent failure scenarios as much as possible but also adds capacity, if needed, using a shortest path algorithm with dynamic link weights. Note that traditional IP routing based on OSPF/IS-IS uses static link weights and so is significantly less efficient than the proposed traffic engineering algorithm using dynamic link weights. The algorithm is also integrated with an algorithm for designing IP express links over the optical link topology with the objective of minimizing the sum of transport and IP port costs by choosing links along heavy-traffic paths while also minimizing the impact of Shared Risk Groups. The network design algorithm at first creates a large number of candidate IP express links (each one using one or more optical links) and ranks them according to the amount of working path traffic carried by them (in the

numerical results section we will assume that working path uses shortest distance routing so as to minimize working path capacity, but this assumption can be relaxed to further reduce the overall capacity and cost). At any stage of the algorithm the next candidate link would typically be parts of many SRGs where SRG for a given failure scenario is defined as the set of links that fail together during that failure scenario. The size of the SRG affects the amount of restoration capacity needed during the failure scenario. The algorithm adds express links based on both the working path traffic carried by the express link (higher the better) and the maximum size of SRG group of which it is a member (lower the better). After adding the express link the entire restoration algorithm is repeated to identify if the overall network cost is reduced. Typically, as we add successive express links (using the two criteria mentioned above), the cost keeps decreasing, reaches a minimum and then increases. The 2-Phase method also works well with multiple failure scenarios.

The proposed 2-Phase Fast Reroute with optimized traffic engineering also ensures that in a dynamic environment the blocking probability requirement is met for many service classes each with its distinct arrival rate, holding time and bandwidth requirement and novel sharing arrangements based on trunk reservation, upper limit and guaranteed minimum policies [22].

3.2   Wavelength Service Restoration: Shared Mesh Restoration with Standbys

In a reconfigurable all-optical network, connections at the optical layer come from two major sources: 1) wavelength services and 2) links of higher layer network, such as IP network. In today's commercial IP networks, failure is usually restored in IP layer, i.e., IP links are usually provisioned as un-restorable connections in optical layer. Alternative solution would be integration of IP layer and optical layer restoration that will be discussed in the subsection on integrated IP/OL restoration. For restoring wavelength services when a failure occurs in the optical layer, 1+1 Tail-end switch is the only commercial available restoration form in all-optical networks today and is only used for a few priority circuits. The drawback is that it requires 1-to-1 dedicated backup connections for single failure protection and 2-to-1 dedicated backup connections for double failure protection, which usually result in 150-200% restoration overbuild of transport resources due to significantly longer diverse backup paths. P-cycle based restoration is a popular restoration topic in academic [23]. However, typically these methods only demonstrate significant capacity advantages under single optical failure scenario and, as such, virtually all such proposals avoid the issue of multiple failure scenarios. Furthermore, if designed for single-failure only, such a method can have serious complications during unplanned multiple-failures, where contention (two or more connections trying to use the same restoration capacity) can occur and has to be resolved in real-time.

In order to provide fast and cost effective restoration, we propose to use shared mesh restoration with standbys extended from our previous work [4, 24]. The standbys provide connectivity between switching nodes (nodes with fiber link degree-2 or higher). All-optical switching is used at the intermediate nodes along a standby's path. As in [4, 24], the standbys can be pre-configured using OEO regenerators at their two ends which send test signals continuously. This overcomes the problem of optical impairments for long connections, avoids today's wavelength power

"balancing" delays, provides wavelength conversion for capacity efficiency and allows sharing of links across non-simultaneous failures. Variations of "standby" schemes, including "hot standby", "cold standby" and "no-standby" proposed and analyzed in [24] provide tradeoffs between restoration speed and restoration capacity overbuild. To illustrate the concept of standby restoration, we illustrate a simple example in figure 3.
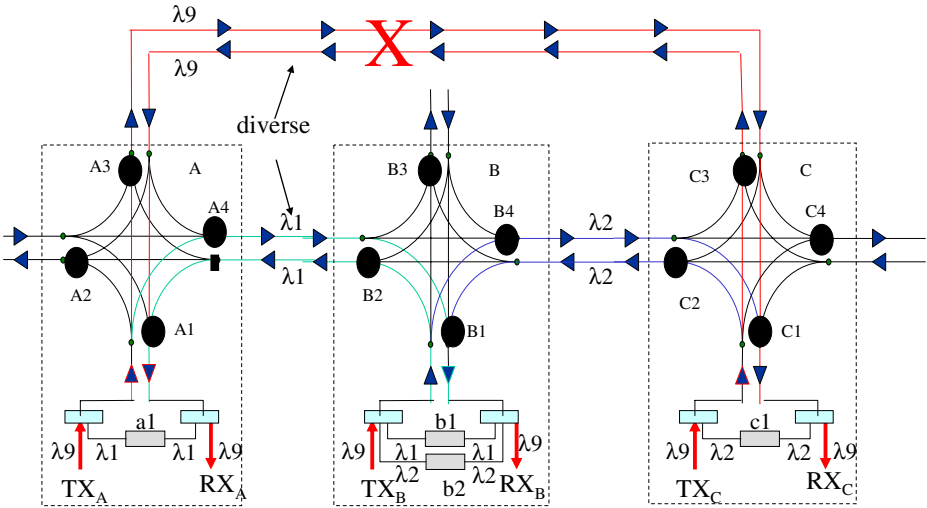

Fig. 3. An Illustration of Standbys

For an all-optical ROADM network, channels cannot be simply connected in series to create a restoration path due to wavelength continuity, ULH reach constraint, power equalization process. We propose to pre-place a few OEO regenerators strategically and pre-establish a set of "standby" optical connections between regenerators. During the restoration process, these "standbys" would be dynamically connected together in series to provide the desired restoration path. Since the "standby" connections are pre-established, i.e., the restoration wavelengths are actively present on the restoration paths, the optical transients associated with firing up their lasers would not be an issue at the point in time where restoration is needed. This will allow the restoration process to be done much more rapidly than it could otherwise be done. Those standbys are capable to be shared among multiple restoration paths via controlling the wavelength selective switches and tuning those regenerators. In figure 3, we assume that there is one service connection between node A and node C on wavelength $\lambda 9$. There are two standbys pre-established: one between node A and node B on wavelength $\lambda 1$ and the other between node B and node C on wavelength $\lambda 2$. Note that there could be other nodes between node A and node C, node A and node B, node B and node C. If a fibre-cut between node A and node C brings down the working path, a pre-computed disjoint restoration path through the overlay route A-B-C is selected. Different from opaque optical networks where restoration channels are pre-installed on each network link, here the restoration

channels are pre-reserved on pre-established standbys, which are restoration connections between OEO regenerators. In some sense, the standby restoration is performed on the standby overlay network bypassing the underlying transparent all-optical ULH network. Then all-optical network restoration time is equivalent to existing opaque optical network restoration time, which can achieve sub-second restoration on average using ROLEX-like signalling methodology [9].

To reduce network cost, the service providers would like to pre-establish as few restoration "standbys" as possible since standbys are reserved and not used to transfer traffic during network normal operation. To guarantee full restoration of any failure, the service providers need to pre-establish enough "standbys". Then the question is where and how many "standbys" the network should pre-establish, i.e., where and how many OEO regenerators the network should pre-placed in advance for restoration purposes.

There are many components in the restoration design of optical layer including service path selection algorithm, OEO regenerator placement, wavelength assignment [25], path based shared restoration with pre-cross-connected standbys, solution for dual failure restorable connections, as well as trap scenario avoidance for path selection. The detailed solutions are listed in following:

- Due to wavelength continuity constraint in all-optical networks, we select the service path to minimize the number of regenerators required with tie breaking of small number of hops. Once the number of regenerators is determined, we need to decide where to place the OEO regenerators considering the number of available wavelengths.

- Since the OEO regenerators separate the connection into multiple lightpaths, we can calculate the available number of wavelengths on each lightpath. One approach would be to maximize the sum of these available wavelengths over all separated lightpaths when all OEO placements are compared. However, if there is a choice of one OEO placement with 10 wavelengths available for either of two lightpaths and another OEO placement with 20 available wavelengths on one lightpath and only one available wavelength on the other lightpath, the maximum sum would vote for the asymmetric configuration. Obviously this is not a good idea to spend the last wavelength left on the second placement. So an improved approach should penalize a configuration which tries to spend the last few available wavelengths on their paths. Our proposed OEO placement solution would be to minimize the sum of the inverse of available wavelengths over all separated lightpaths, i.e., min $\Sigma 1/A(i)$, where $A(i)$ is the available wavelengths on lightpath i with OEO placement. This gives a high number for one of the lightpaths having only a few wavelengths still available. It tends to balance the number of available wavelengths on all lightpaths.

- Although there are many wavelength assignment schemes proposed in the literature, we choose first-fit wavelength assignment solution in this paper since it is simple and efficient. We order the wavelengths from 1 to W, where W is the maximum number of channels that the system supports, and select the smallest available wavelength along the lightpath. As we discussed before, with the help of regenerators, all optical connections are separated into lightpaths. Since

regenerator provides wavelength conversion capability, now wavelength assignment is on top of lightpaths instead of connections.

- With the concept of lightpath standbys and the help of regenerators, we can apply the path-based shared mesh restoration that was used in opaque optical networks [11] to all-optical networks. Specifically, the restoration channel sharing is on top of lightpath standbys instead of optical links only. We assume that each wavelength supports one OC-192 connection and each OC-192 demand is routed independently one at a time. To deal with optical signal reach constraints, we first create possible express links that are within the maximum optical signal reach distance on the network in two steps: (1) compute the potential path between any two nodes; (2) form an extended network by creating express links on these paths. Those express links are pre-established standbys candidates. To select the restoration path with minimum number of restoration wavelengths in the extended network for service path Ps, a matrix *failroute* is maintained where *failroute[i,j]* maintains how many standbys are needed on express link j if link i fails. Then $max_{i \in E}$ *failroute[i,j]* and $max_{i \in Ps}$ *failroute[i,j]* represent the total standby channels reserved on link j and the required standby channels required on link j if service path Ps fails, where E stands for the set of network original links.

  The difference, $S[j] = max_{i \in E}$ *failroute[i,j]* $- max_{i \in Ps}$ *failroute[i,j]*, would be the amount of sharable standby channels, and $max(0, 1-S[j])$ would be the additional required channels on link j if path Ps fails. Thus, after service path Ps is selected, we reset link weights on the extended network as additional required channels and select the shortest new link weight path as the restoration path Pr, which should be the smallest number of channels required path. After Ps and Pr are selected, then the matrix *failroute* is updated. After all connections are routed on the network, each link j would have $max_{i \in E}$ *failroute[i,j]* restoration channels and each restoration channel is a lightpath standby segment. Since each standby segment requires 2 unidirectional restoration regenerators, the total number of unidirectional restoration regenerators will be $H = 2 * \Sigma_j (max_{i \in E}$ *failroute[i,j]*), where j includes express links.

- To deal with double failures, we would select two failure disjoint restoration paths, one for 1:1 restoration and the other for shared mesh restoration. To make sure enough capacity is reserved on the second restoration path no matter what order failure happens, we first select service path Ps, and dedicated restoration path Pr1, then we select the second restoration path Pr2 using SMR on extended network as we discussed above. The only difference is that in formula of sharable standby channels, $S[j] = max_{i \in E}$ *failroute[i,j]* $- max_{i \in \{Ps, Pr1\}}$ *failroute[i,j]*, we use Ps+Pr1 instead of Ps. The reason is that we want to reserve enough capacity for both scenarios either Ps fails first and Pr1 fails second or Pr1 fails first and Ps fails second. In this case, no matter what failure order, there is enough capacity reserved on the third restoration path.

- For path based restoration, one has to deal with trap scenario during path selection, i.e., although there are two or more failure disjoint paths between two nodes, a simple shortest path as service path may fail to find a failure disjoint

restoration path. In this study, we use the following algorithm to avoid such cases: we first use maxflow-mincost algorithm to find 2 failure disjoint paths for single failure restorable connections and 3 failure disjoint paths for double failure restorable connections between the two ends of the connections via node splitting approach for node disjoint. Among the selected paths, we select the shortest path for service path. In this way, we guarantee the existence of single or double failure disjoint restoration paths.

3.3   Integrated IP-over-Optical Layer Restoration

As mentioned above in today's commercial IP networks, IP links are usually provisioned as un-restorable in the optical layer to satisfy minimum cost objectives and the restoration capacity is reserved in the IP layer. Again, this is mostly due to the presence of extra IP-layer capacity to restore failures that originate at the IP-layer plus the ability of the IP-layer to differentiate classes of service on a much finer scale than in the optical layer. However, as we explore the failure state space in more detail, we find that it is able to share some optical layer restoration capacity among non-simultaneous failures to restore or add some IP links via the optical layer. An innovative, yet practical method was developed by [26, 27, 28], which propose interaction between the IP and optical layers. By use of spare router line cards, this method optimizes network capacity by leveraging 1) the ability of the optical layer to make rapid OL connections and 2) the assumption that the optical layer requires extra restoration capacity (wavelengths) for restorable wavelength service and 3) except for extremely rare failure events, the non-simultaneity of failures between components in the IP and optical layers. Figure 4 shows a simple example to illustrate this concept, where dotted thick lines represent spare router line cards connecting to the ROADMs. Upon a router failure or router line card failure, those spare line cards will be used to restore failed IP links or add extra IP link capacity. A few key elements in this integrated method are described in the following:

- 1:N Interface Protection [27, 4]: The goal here is to provide rapid recovery from the failure of a router or ROADM line card or the connection between a router line card and the ROADM port. If one of the N working interfaces fails for whatever reason, the ROADM will rapidly switch the failed connection to the (spare) protection interface and (spare) line cards. This should be done locally at the node very rapidly and without involving either the remote router or any other ROADM. The benefits to the IP network of doing APS between the router and the ROADM are obvious when compared to the alternative, which is to have additional inter-router links. In addition, if done quickly enough the APS switch should hide the failure from the remote routers, thus eliminating disruptive IP layer convergence to a new forwarding table. A prototype demonstration that this is possible does exits using an opaque OXC [27]. It utilizes an extension of the OIF UNI together with enhancements to the router software. We will explore the usage of this fast protection in a transparent architecture.
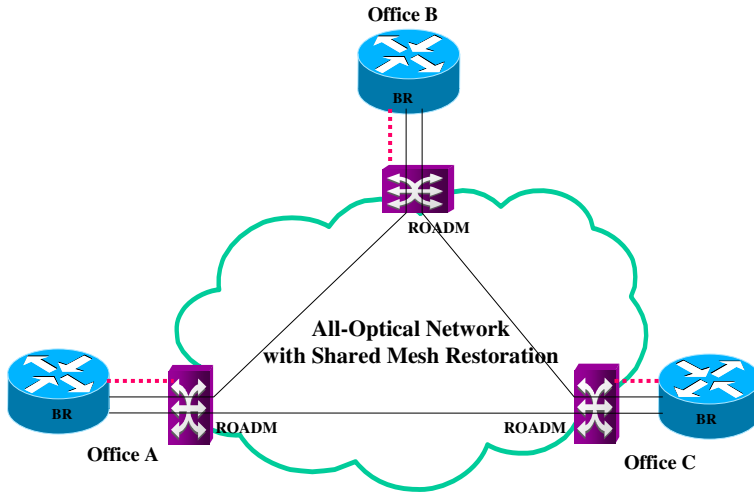
**Fig. 4.** Integrated IP-over-Optical Layer Restoration

- Restore from any single router failure or complete node failure [26, 28]: by use of spare router line cards, this method optimizes network capacity by leveraging 1) the ability of the optical layer to make rapid OL connections to add new IP link(s) or increase the capacity of existing IP links or virtual links (a logical link with 0 bandwidth). The latter will be faster since there is complex state information associated with each new IP router interface which would need to be dynamically associated with the backup interface.

- Restore IP links from OL failure(s) [28]: there are various options to restore IP links at the OL from OL failures. For bandwidth efficiency, composite IP link groups could recover with partial bandwidth to take advantage of the fact that there is no restoration requirement for significant part of the IP traffic – the best effort traffic. Another option is where IP links remain unprotected. Upon an OL failure, we reshape IP topology by making rapid OL connections to add new IP link(s) or by increasing the capacity of existing IP links or virtual links.

We propose to study the adaptation of theses schemes to the future IP over reconfigurable all-optical network and explore the cost advantages in future work.

4   Simulation Experiments and Results

4.1   Network and Traffic Models

We evaluate the performance of our proposed approaches via simulation on a hypothetical U.S. backbone network from [29] consisting of 28 nodes and 45 fiber links as shown in figure 5. The number over each link is the distance between two nodes in 10-mile unit. Each node consists of one router and one ROADM. Each fiber pair is assumed to support 100 wavelengths. There are two types of traffics generated
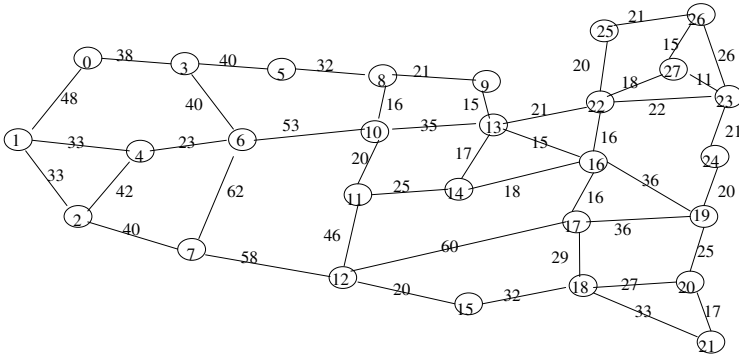
Fig. 5. A Hypothetical U.S. Backbone Network

at each node: IP traffic and wavelength traffic. All traffic demands are generated based on node traffic probability, which is proportional to the node degree. We randomly select a demand according to source node traffic probability and destination node traffic probability. For IP traffic, the bandwidth of each demand is uniformly generated from an interval (100, 1000) Mbps and there are 5000 such demands. Each IP demand flow has 40% chance to be best effort traffic and a small probability (up to 10%) requiring double failure restoration. For simplicity, the IP link bandwidths are assumed to be deployed in units of 10Gbps channels (the associated IP port may either be an OC192 port or a 10 Gigabit Ethernet port, the latter being usually cheaper). For each IP link the OSPF Administrative Weight metric is assumed to be distance plus 20 miles (this allows OSPF shortest path routing to be shortest distance routing as well in most cases but in the event of two nearly equal distance paths it chooses one with fewer hops). For wavelength traffic, we assume that each demand requires one wavelength and all demands need to be single failure protected at least. A small percentage of wavelength demands (up to 10%) needs to be double failure restorable. In our simulation, we assume wavelength demands arrive at the network one by one, and are never disconnected (or more realistically, have holding times in years), which is typical for commercial wavelength services. For both IP service and wavelength service restoration, we evaluate the performance of our proposals using restoration overbuild. It is defined as (total wavelength-mile for both service and restoration)/(wavelength-mile for service only) -1.

4.2   Results for IP Service Restoration

We will show results mainly for the 2-Phase Fast Reroute method with optimized Traffic Engineering. We will also provide comparison with pure shortest path routing (the most prevalent method in today's IP Networks) with and without Fast Reroute. In addition to single failure protection we also protect a subset of traffic under double failures (2 optical links or 1 node and 1 optical link, total of 2250 failure conditions). By pruning the failure state space, we reduced the number of cases significantly and we achieved further speed-up by using a very fast but less efficient pure shortest path routing algorithm for capacity allocation during multiple failure events. However,

majority of the traffic demand only needs single failure protection and for restoring those traffics we use slower but highly efficient Traffic Engineering algorithm that can efficiently reuse the capacity that is needed to be used for double failure protected demands. Each failure in the optical layer may fail multiple links in the IP layer and that is appropriately taken into account. Each node has an IP Router and there is IP traffic demand between every pair of IP Routers (28X27 = 756 demands). We route all IP working path demand along the shortest path and this requires each optical link to be used as IP link. In addition, we use an automated algorithm to add several IP Express Links in an optimal way in order to minimize total network cost (sum of transport cost and IP port cost) as explained in Section 3.1. The optimal choice of IP Express Links depend on the relative costs of IP ports and the underlying transport mechanism (note that 10 Gigabit Ethernet ports tend to be cheaper than OC192 ports and newer Ultra Long Haul transport technology is cheaper than older technology). To illustrate this point we consider two cases. In Case 1 (Table 1) 10Gbps IP port cost is assumed to be equivalent to 100 miles of transport cost and in Case 2 (Table 2) 10Gbps IP port cost is assumed to be equivalent to 300 miles of transport cost (transport cost is assumed to be proportional to Route Miles in both cases). Both tables show the result of incrementally adding Express Links (all results are based on 2-Phase Fast Reroute with optimized Traffic Engineering which tries to minimize a linear combination of cost and capacity). The tables show that as the number of Express links increases, at first the cost decreases, reaches a minimum and then start increasing. However, the relative cost difference among the alternatives is significantly higher in Table 2 (higher IP port cost) compared to Table 1. Also the

Table 1. (IP Port Cost = 100 Miles of Transport Cost)

Alternative	# of IP Logical Links	Capacity in 1000 10Gbps-Miles		Restoration Overbuild	# of IP 10Gbps Physical Links (W. + Rest.)	Normalized Total Cost
		Working Path	Working Path + Restoration			
1	45	138.57	202.6	0.46	735	104.2
2	62	147.16	208.7	0.42	627	102.4
3	81	150.93	212.9	0.41	568	101.8
4	**92**	**157.46**	**213.0**	**0.35**	**520**	**100**
5	103	164.05	217.4	0.33	510	101.3

Table 2. (IP Port Cost = 300 Miles of Transport Cost)

Alternative	# of IP Logical Links	Capacity in 1000 10Gbps-Miles		Restoration Overbuild	# of IP 10Gbps Physical Links (W. + Rest.)	Normalized Total Cost
		Working Path	Working Path + Restoration			
1	45	138.57	203.2	0.47	731	120.7
2	81	150.93	211.2	0.40	539	106.5
3	103	164.05	216.3	0.32	475	102.5
4	**122**	**169.41**	**217.5**	**0.28**	**442**	**100**
5	130	171.69	220.4	0.28	456	102.0

optimal number of IP logical links is higher in Table 2. It is important to note that minimizing either the total capacity (working path + restoration) or the restoration overbuild does not necessarily translate to minimum cost. For example, Table 1 shows that Alternative 1 has the minimum total capacity and Alternative 5 has the minimum restoration overbuild, but Alternative 4 has the minimum cost (we choose this alternative for the rest of the study).

Under no failure condition and under the second and more permanent phase of Single Failure Fast Reroute the maximum IP link utilization is assumed to be 95%. During the rarer double failures and during the initial transient phase of Fast reroute (a few seconds) the maximum IP Link utilization is assumed to be 100%.

Table 3 below shows the impact of increasing the percentage of traffic that requires double failure protection.

Table 3.

Percentage of Traffic Requiring Double Failure Protection	0%	2.5%	5%	7.5%	10%
Restoration Overbuild	0.353	0.354	0.358	0.365	0.380

Table 4 below compares three IP design alternatives in terms of restoration overbuild and cost. Alternative 1 uses 2-Phase FRR with optimized Traffic Engineering and is the one we propose. Alternative 2 uses IP reroute using OSPF shortest-path routing (without Fast Reroute) and is the most common routing mechanism in today's IP network. Alternative 3 uses simple MPLS FRR, i.e., no Traffic Engineering and instead it uses shortest path routing during the two phases. Clearly, our proposed mechanism performs significantly better than the other two.

Table 4.

IP Routing alternative	# of IP Logical Links	Capacity in 1000 Wavelength-Miles		Restoration overbuild	# of IP 10 Gbps Physical Links (W + Rest)	Normalized Total Cost
		Working Path	Work. Path + Rest.			
1	**92**	**157.46**	**213.0**	**0.353**	**520**	**100**
2	92	157.46	305.87	0.943	675	140.9
3	92	157.46	335.81	1.132	752	155.1

4.3   Results for Wavelength Service Restoration

We compare our wavelength service restoration scheme with existing commercialized 1+1 for single failure protection and 1+2 for double failure protection. For shared mesh restoration with standbys, we use our proposed algorithm to select restoration paths. For 1+1 or 1+2 protection, we always use shortest disjoint restoration paths. To evaluate the impact of traffic size, we first fix the double failure restoration ratio as
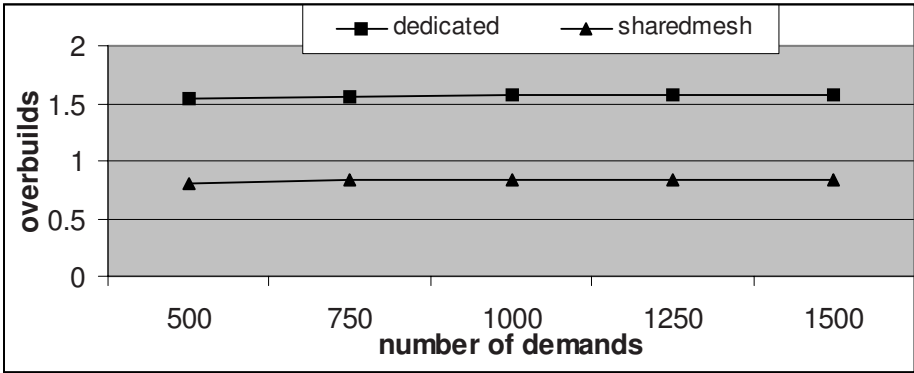
Fig. 6. Restoration Overbuild Comparison with a Fixed Double Failure Ratio (5%)
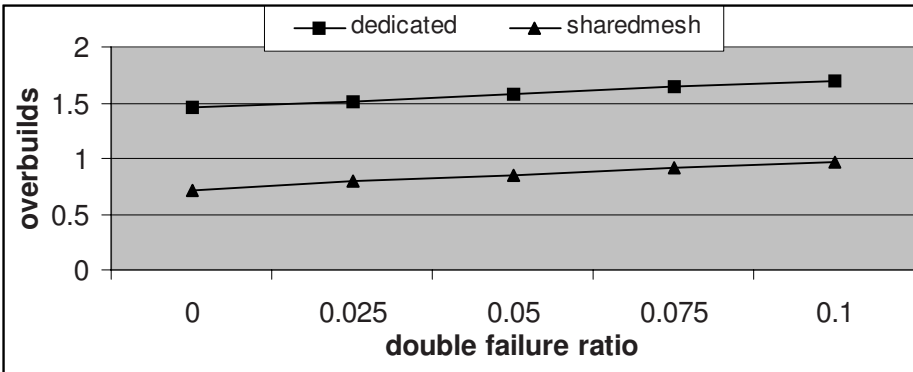

Fig. 7. Restoration Overbuild Comparison with a Fixed Number of Demands (1000)

5% and increase wavelength demands from 500 to 1500 with step size of 250. Then we fix the number of wavelength demands as 1000 and increase the double failure ratio from 0% to 10% in step size of 2.5% to evaluate the impact of demand double failure restoration requirement. In both cases, we compare restoration overbuilds of both schemes and leave the detail economic analysis for future study. Figure 6 and figure 7 show the simulation results.

Figure 6 shows that 1+1 protection results in very high restoration overbuild around 150% while our proposed shared mesh restoration with standbys demonstrates significant improvement by reducing the overbuild by 50% for all demand sizes studied.

Figure 7 shows that as we increase the double failure ratio from 0% to 10%, restoration overbuilds for both 1+1 protection and proposed shared mesh restoration with standbys increase steadily. Meanwhile, the improvements in restoration overbuild of our proposed scheme over 1+1 protection stay close to 50% for all double failure ratios.

We simulated both IP and wavelength service restoration solutions with different network topology and traffic patterns. On steady state, we observed similar levels of improvements comparing our proposed solutions and existing solutions. We omit to report them in this paper.

5   Summary

Large IP backbone networks today are mostly deployed directly over sequences of point-to-point DWDM systems or chains of newer ROADM-based ULH systems, interconnected by OEO regenerators. The next generation core optical network is moving toward an all-optical network architecture that is based on multi-degree ROADMs to reduce OEO regeneration cost as well as enabling automatic reconfigurability and dynamic restoration via wavelength switching and tuning. In this paper, we study the restoration design in this new IP-over-reconfigurable all-optical network architecture to satisfy the network reliability requirements for both IP and wavelength services. For restoring IP services, we propose a novel 2-Phase Fast Reroute mechanism with optimized Traffic Engineering algorithm. It meets the requirement of sub-second restoration and also maximizes sharing among single/double failures of links, routers and Shared Risk Groups and also among different phases of restoration with the objective of minimizing either overall capacity or overall cost. Simulation results have shown that the proposed restoration scheme provides significant savings comparing to existing IP reroute (e.g., OSPF or IS-IS) and MPLS Fast Reroute. In order to provide fast and cost effective restoration for wavelength services, we propose to use shared mesh restoration with pre-configured (standby) lightpaths. Simulation results demonstrate significant improvements in restoration overbuild using our proposed scheme when compared to existing 1+1 protection. To further reduce the required restoration capacity in both IP layer and optical layer and address failures in both layers efficiently, we also propose to use an integrated IP-over-optical layer restoration strategy that enables sharing of re storation capacity among non-simultaneous failures across both IP and optical layers. Detailed economical studies of the integrated method would be conducted in future work.

References

1. Wallace, A.F.: Ultra long-haul DWDM: network economics. In: OFC, vol. 2, TuT1-1 - TuT1-2, Anaheim (March 2001)
2. MCI Picks Ciena and Siemens for ULH, http://www.lightreading.com/document.asp?doc_id=48055.
3. Chiu, A., Yu, C.: Economic Benefits of Transparent OXC Networks as Compared to Long Systems with OADM's. In: OFC, Atlanta, pp. 411–412 (March 2003)
4. Strand, J., Chiu, A.: Realizing the Advantages of Optical Reconfigurability and Restoration with Integrated Optical Cross-Connects. Journal of Lightwave Technology 21(11) (November 2003)

5. Afferton, T., Doverspike, R., Kalmanek, C., Ramakrishnan, K.K.: Packet-Aware Transport for Metro Networks. IEEE Communications Magazine, 120–127 (March 2004)
6. Wei, J., Ramakrishnan, K.K., Doverspike, R., Pastor, J., Aguirre-Torres, L., Kalmanek, C., Afferton, T.: Convergence through Packet-Aware Transport. The Journal of Optical Networking, Special Issue on Convergence 5(4), 221–245 (2006)
7. Choudhury, G.: Models for IP/MPLS Routing Performance: Convergence, Fast Reroute, and QoS Impact. In: Keynote Speech, ITCOM's Conference on Performance, Quality of Service and Control of Next Generation Communication Networks II, Philadelphia, Pennsylvania, USA, pp. 1–12 (2004)
8. Cortez, B.: The Emerging Intelligent Optical Network: Now a Reality. In: OFC 2002 (2002)
9. Doverspike, R., Sahin, G., Strand, J.L., Tkach, R.W.: Fast Restoration in a Mesh Network of Optical Cross-connects. In: Proceedings of Optical Fiber Conference (OFC '99), San Diego, CA, February 1999 (1999)
10. Doshi, B., Dravida, S., Harshavardhana, P., Hauser, O., Wang, Y.: Optical Network Design and Restoration. Bell Labs Technical Journal, 58–84 (1999)
11. Li, G., Wang, D., Kalmanek, C., Doverspike, R.: Efficient Distributed Path Selection for Shared Restoration Connections. IEEE Conference on Computer Communication (INFOCOM) 2002, New York and Special Issue of IEEE Transactions on Networking 11(5), 761–771 (2003)
12. Simmons, J.: Network Design in Realistic All-Optical Backbone Networks. IEEE Communication Magazine 44(11) (November 2006)
13. Ranalli, E., et al.: Liquid-crystal based wavelength selectable crossconnect. In: ECOC, pp. 68–69 (1999)
14. Boskovic, A., et al.: Broadcast and Select OADM nodes: Application and Performance Trade-offs. In: OFC, TuX2, pp. 158–159 (March 2002)
15. Vasilyev, M., et al.: Broadcast and select OADM in 80 /spl times/ 10.7 Gb/s ultra-longhaul network. Photonics Technology Letters 15(2), 332–334 (2003)
16. Woodward, S.L., Feuer, M.D., Calvitti, J., Falta, K., Verdiell, J.: A High-Degree Photonic Cross-Connect for Transparent Networking, Flexible Provisioning & Capacity Growth. In: ECOC 2006, paper Th1.2.2 (2006)
17. Pan, P., Swallow, G., Atlas, A.: Fast Reroute Extensions to RSVP-TE for LSP Tunnels. In: IETF RFC 4090 (May 2005)
18. Choudhury, G.L.: IP Fast-Reroute: An Analysis of Applicability to a Core Network. In: NANOG 33 Conference, January 31 - February 1, 2005, Las Vegas, Nevada, USA (2005)
19. Francois, P., Filsfils, C., Evans, J., Bonaventure, O.: Achieving Sub-Second IGP Convergence in Large IP Networks. ACM SIGCOMM Computer Communications review 35(2), 35–44 (2005)
20. Ward, D., Atlas, A.: IP FRR: Overview and Things We are Struggling to Solve. In: NANOG 33 Meeting, January 31-February 1, 2005, Las Vegas, Nevada, USA (2005)
21. Kandula, S., Katabi, D., Davie, B., Charny, A.: Walking the Tightrope: Responsive yet Stable Traffic Engineering. In: SIGCOMM 2005, August 21-26, 2005, Philadelphia, PA, USA, pp. 253–264 (2005)
22. Choudhury, G.L., Leung, K., Whitt, W.: An Algorithm To Compute Blocking Probabilities In Multi-Rate, Multi-Class, Multi-Resource Loss Models. Advances in Applied Probability 27(4), 1104–1143 (1995)
23. Grover, W., Stamatelakis, D.: Bridging the ring-mesh dichotomy with p-cycles. In: DRCN 2000 (2000)

24. Li, G., Chiu, A., Strand, J.: Resilience Design in All-optical ULH Networks. Journal of Optical Networks (2006)
25. Zang, H., Jue, J.P., Mukherjee, B.: A Review of Routing and Wavelength Assignment Approaches for Wavelength-Routed Optical WDM Networks. Optical Network Magazine 1 (January 2000)
26. Chiu, A., Strand, J.: Joint IP/Optical Layer Restoration after A Router Failure. In: OFC2001, Anaheim, CA (March 2001)
27. Sebos, P., Yates, J., Li, G., Wang, D.: Ultra-fast IP link and interface provisioning with applications to IP restoration. In: IEEE OFC 2003 (2003)
28. Li, G., Wang, D., Yates, J., Doverspike, R., Kalmanek, C.: IP over Optical Cross-Connect Architectures. In: Optical Fiber Communications Conference (OFC 2004), Communication Magazine, vol. 45(2), pp. 34–39 (2007)
29. Xiong, Y., Mason, L.: Restoration strategies and spare capacity requirements in self-healing ATM networks., IEEE/ACM ToN, 7(1) (February 1999)