An Analysis of Path Recovery Schemes in GMPLS Optical Networks with Various Levels of Pre-Provisioning

David Griffith, Richard Rouil, Stephan Klink, and Kotikalapudi Sriram
National Institute of Standards and Technology (NIST)
100 Bureau Drive, Stop 8920
Gaithersburg, MD 20899-8920
Email: david.griffith@nist.gov

ABSTRACT

The amount of resource provisioning prior to failure events to support optical path recovery has a significant impact on network performance, and so designing recovery mechanisms for any large network necessitates balancing multiple (in some cases, competing) requirements. The Common Control and Measurement Plane (CCAMP) working group's Protection and Restoration Design Team conducted a qualitative analysis of path protection schemes using different degrees of backup resource pre-provisioning. The resulting analysis grid highlights some of the trade-offs between the different approaches. In this paper, we describe the results of a simulation study that we conducted using the NIST GMPLS/Lightwave Agile Switching Simulator (GLASS) simulation tool. By measuring network performance with respect to the metrics used by the design team, we were able to produce quantitative results that confirm the design team's qualitative analysis and provide additional information for carriers and service providers who are designing optical networks.

1. INTRODUCTION

Among the various services that carrier networks must support, rapid recovery of data connections from failures is among the most essential. Because of the convergence of voice, video, and data services that has taken place in the last decade, service providers' networks now carry vast quantities of critical, delay-sensitive, loss-sensitive traffic. Failures that result from a disruption of the physical network can have serious economic consequences, particularly if the affected traffic cannot be recovered for a long period of time. Legacy carrier networks employed a variety of failure recovery mechanisms, such as redundant connections in SS7 networks and Automatic Protection Switching (APS) in SONET/SDH networks. The past five years have seen the advent of automatically switched optical transport networks and a concerted effort in multiple standards development organizations to create a single control plane to govern both new and legacy optical systems as well as the higher layer networks that reside above them.

The new control plane is built around a framework that was originally created for Multi-Protocol Label Switching (MPLS). It uses Internet routing mechanisms such as OSPF, IS-IS, and signaling based on RSVP, with extensions to each protocol to support traffic engineering. The bulk of the control plane development work is being done by the Internet Engineering Task Force (IETF), which created the framework for MPLS. At present, the Common Control and Measurement Plane (CCAMP) working group is responsible for defining the core functions of this new control plane. During the past year, the working group has been developing a set of four Internet Drafts[1−4] that define additional capabilities for the control plane in order to support protection and restoration at both the span and path levels. One of these four drafts is an analysis document[2] that describes some of the design issues for protection mechanisms, such as designing escalation strategies and determining what type of recovery mechanisms is appropriate for protected traffic. The analysis document also examines the design trade-offs associated with the degree to which backup resources are reserved and/or committed prior to a failure. These trade-offs are captured in a table that qualitatively shows how different levels of resource pre-selection perform with respect to a set of metrics. While such a table is useful, it does not provide carriers

This research was partially supported by the National Institute of Standards and Technology (NIST), the Advanced R&E Activity (ARDA), the Laboratory for Telecommunications Sciences (LTS) MENTER project, the Defense Advanced Research Projects Agency (DARPA) Fault Tolerant Networks (FTN) program, and the National Communications System (NCS).

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **2003**	2. REPORT TYPE	3. DATES COVERED **00-00-2003 to 00-00-2003**

4. TITLE AND SUBTITLE **An Analysis of Path Recovery Schemes in GMPLS Optical Networks with Various Levels of Pre-Provisioning**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **National Institute of Standards and Technology,100 Bureau Drive, Stop 8920,Gaithersburg,MD,20899**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
Proc. of SPIE Vol. 5285 OptiComm 2003: Optical Networking and Communications, edited by Arun K. Somani, Zhensheng Zhang, (SPIE, Bellingham, WA, 2003), pp. 197-208

14. ABSTRACT
see report

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**12**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

and service providers with quantitative performance information that can be used to make decisions about what kind of protection to employ in their networks.

In this paper, we describe a set of experiments that we performed using an optical network simulation tool. The goals of these experiments were to verify the Protection and Restoration Design Team's analysis grid and to develop a set of design guidelines that could be used by network managers to determine what degree of backup resource reservation is appropriate. Other groups have used simulations to examine other issues associated with developing recovery mechanisms for GMPLS optical networks. In one study,[5] the authors use simulations to compare the performance of span, subpath, and path recovery schemes with respect to a set of performance metrics that are analogous to the ones that we use in this paper. They do not consider pre-provisioning issues, however. In another study,[6] the authors use simulations to examine design choices for optical cross-connects and restoration signaling protocols. In another study,[7] the authors report on experiments that were used to verify previous simulation results that demonstrate that GMPLS signaling can be used to achieve rapid restoration of connections in optical mesh networks.

The rest of this paper is organized as follows. In Section 2, we describe the work that the CCAMP Protection and Restoration Design Team has done in putting together the path recovery analysis grid. We describe the different types of path protection schemes that the design team considered and the metrics that they used to evaluate the schemes' effectiveness for recovery. In Section 3, we describe the GLASS simulation tool and the experiments that we performed to enhance and quantify the IETF analysis grid. We describe our results and discuss some of the design trade-offs that follow from our analysis in Section 4. We summarize our work and discuss future work in Section 5.

2. THE CCAMP PROTECTION AND RESTORATION DESIGN METHODOLOGY

The CCAMP working group is developing a set of mutually independent signaling and measurement protocols that can support many types of tunneling technologies in core networks. Examples include opaque and transparent optical switches, ATM and Frame Relay switches, and MPLS label switching routers. The working group has already generated Requests for Comments (RFCs) describing enhancements to the Resource ReserVation Protocol with Traffic Engineering (RSVP-TE)[8] and the MPLS Label Distribution Protocol (LDP),[9] both of which can be used as complementary signaling protocols for the common control plane. They have also defined a Link Management Protocol (LMP)[10] that can be used to manage multiple data and control channels between switching entities in core optical networks. A set of extensions to LMP for Wavelength Division Multiplexed (WDM) networks[11] allow communication between optical cross-connects and optical line systems. Other tasks include developing mechanisms to distribute network resource state information using routing protocols such as OSPF.[12,13] The group is developing a tunnel route tracing protocol[14] and defining signaling mechanisms to carry out link and path protection and fast restoration.

Protection and restoration are both types of recovery schemes, but with some important differences. In protection schemes, the network operator creates dedicated backup resources for protected traffic. In restoration schemes, the network operator establishes new connections or activates reserved resources for displaced traffic when a fault occurs. In order to support protection and restoration, the GMPLS signaling protocols must be able to support the creation of backup label switched paths (LSPs) and rapid switching operations to move protected traffic that is affected by failures to backup LSPs. A protection and restoration design team within CCAMP is developing a set of four documents that specify how the GMPLS control plane can be used to recover from network failures.[15] The terminology document[1] standardizes the vocabulary associated with GMPLS recovery, and is used as input for the other three documents. The analysis document[2] evaluates tradeoffs between different types of recovery schemes and identifies issues that operators must consider when deciding how traffic from failed connections will be recovered. The functional specification[3] describes different types of restoration mechanisms and the signaling procedures and message types that are used to implement them. The protocol specification[4] describes in greater detail the signaling messages and objects that are used to carry out different types of failure recovery.

2.1. Types of Restoration Schemes

There are four distinct phases of LSP creation: routing, signaling, resource selection, and resource allocation. The degree of pre-planning associated with a given LSP recovery scheme depends on which of the above four phases take place prior to a failure event, and which occur when a failure happens. In the case of protection switching, all four phases are executed at the time that the LSP is created. The resulting protection LSP can be used either for 1+1 or for dedicated 1:1 recovery. In the case of pre-signaled and pre-selected recovery LSPs, the fourth set up phase is carried out after a failure occurs. Thus, specific backup resources on each link are assigned to the recovery LSP, but the cross connects at the intermediate nodes are not configured to establish an actual light path. In the case when a recovery LSP is pre-signaled only, it has reserved bandwidth on each of the links that compose its path, but it has not selected a particular resource (e.g., wavelength, label) from the restoration resource pool on each link. In the case when only the recovery LSP routing path is pre-planned, a route for the recovery LSP has been computed, but no signaling and reservation of resources has taken place. Finally, dynamic rerouting recovery schemes perform all four of the LSP set up phases above only after a failure has occurred.

In Fig. 1, we show the four stages of LSP setup and the levels of pre-provisioning that are associated with halting the setup process at a particular point and deferring the remainder of the process until the occurrence of a failure event. The analysis draft and the simulations that we performed consider four levels of pre-provisioning, as indicated in Fig. 1. These four cases cover the full range of possible path restoration schemes. At one extreme (Case 1) the network computes the route and assigns specific resources to the backup LSP, although the cross-connections in the intermediate OXCs are not set until a failure occurs. At the other extreme (Case 4) the backup path setup process begins only when a failure is detected and reported to at least one edge node associated with the affected LSP.

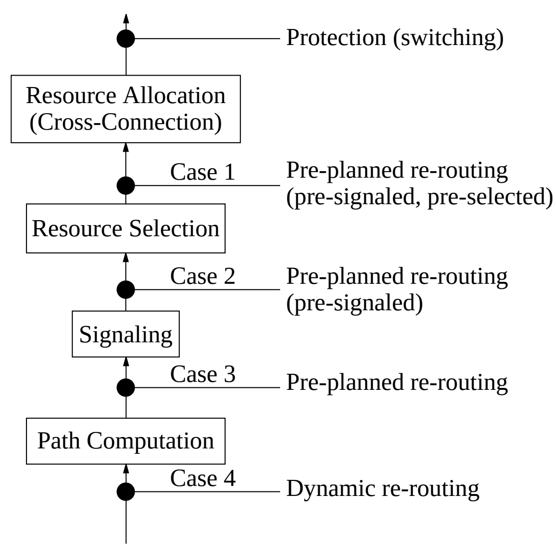

Figure 1. Phases of recovery LSP creation with different levels of pre-provisioning indicated. Pre-signaled LSPs have computed routes and reserved resources prior to failure but have neither selected nor allocated resources in advance.

2.2. Performance Metrics

The analysis draft[2] uses a set of four metrics to produce a comparative evaluation of the four types of path recovery schemes that are shown in Fig. 1. Together, these metrics define a 4-dimensional vector space over which each recovery scheme's behavior can be plotted for a given network topology. The goal is to choose the scheme that exhibits the best performance overall, or that exhibits optimal performance with respect to a subset of these metrics that are deemed to be critical by the network operator. The metrics listed in the draft are the following: fast convergence (performance), efficiency (scalability), robustness (availability), and resource optimization (optimality).

Convergence speed, which is determined by recovery time, and efficiency/scalability, which is determined by switching time, together determine the time from the detection of a failure to the completion of recovery operations. The recovery time consists of the following components: detection time, correlation time, and hold-off time. The detection time is the time that is required for the optical layer to discover a failure and to generate alarms to alert other network components that a problem exists. Correlation time is the time required to aggregate multiple alarms in order to report on as a batch or as a single message indicating the failure of a larger logical entity (e.g., an SRLG). Correlating failures is optional, but the price for not doing this is the greater control plane overhead that results from the generation of many failure messages if a large number of LSPs are lost. If the network supports escalation schemes that coordinate recovery operations between multiple layers, then higher layers may use hold-off timers to allow the lower layers an opportunity to recover from a fault. Using effective failure detection and correlation procedures can increase convergence speed; ideally it should scale well with respect to the number of failures in the network. The switching time depends on the degree of pre-provisioning of recovery resources on the backup LSP. With more pre-provisioning, the recovery is faster because less computational effort is required when a failure occurs. The recovery time is also impacted by the length of the connection (i.e., the number of hops); longer connections have greater switching times because of the greater distance over which switchover signaling messages must propagate. Ideally, switching time should be independent of the number of spans or LSPs that are affected by a failure event.

If the recovery scheme is robust, it will function well in a wide range of failure scenarios. All recovery schemes will perform well if an isolated or small-scale failure occurs. The ability of a recovery scheme to perform well when multiple failures occur is an indication of its robustness. If the recovery scheme is robust, then the availability of the LSPs will be high. Availability is a function of the amount of time that LSPs are unavailable due to failure events; availability should not be affected by the physical network topology.

The degree of resource optimization is a function of the resources required to implement a recovery scheme vs. the resources required to support the protected working LSPs. It is desirable to minimize this quantity. In situations where 1+1 protection is used, the network recovers quickly from failures but its resources are not optimized, because twice the total working path bandwidth is required to support this type of protection scheme. Other recovery schemes can achieve greater degree of resource optimization. For instance, 1:1 recovery uses network resources more efficiently than 1+1 protection if the backup LSP is used to carry low-priority traffic during those periods when the working LSP is not in a failed state. In shared mesh restoration schemes, increasing the number of protection LSPs that share a given set of resources on a given link will reduce the protection overhead further, but at a cost of reduced backup LSP availability if the respective working LSPs simultaneously fail and contend for the shared resource.

2.3. Analysis Grid

In Table 1, we show the analysis grid[2] for path recovery schemes. The grid classifies recovery schemes based on whether they implement routing computation and path selection dynamically or before the reported failure (pre-planned). The grid further classifies each subgroup of schemes by the starting point for backup LSP setup once a failure occurs. If path computation and selection happens after a failure indication is received, further path setup cannot begin with the resource selection or allocation phases, so only the bottom box in the dynamic column, corresponding to Case 4 in Fig. 1, applies.

The three boxes in the first column likewise correspond to Cases 1, 2, and 3 in Fig. 1. Within each box in the grid,[2] the authors report the relative behavior of the respective pre-provisioning level with respect to the four metrics that we discussed in Section 2.2. Some general behavioral trends are immediately clear from an examination of Table 1. First, the recovery speed decreases as more pre-provisioning is used. This is obvious, since fewer computations need to be executed on-the-fly when the backup LSP is pre-planned. For Case 1, responding to a failure requires only the propagation of Switchover and Switchover Acknowledgment messages on the forward and return paths on the backup LSP, respectively. Flexibility and robustness also decrease with increasing levels of pre-provisioning. By committing to using a particular set of resources for resource pools, a pre-planned backup LSP becomes vulnerable to failures of other working paths with which it may be sharing the selected recovery resources or failures of those resources themselves. This vulnerability can be reduced if new resources can be selected on the fly, but in some failure scenarios this may not be possible. Finally, LSP recovery schemes that use no pre-provisioning consume the fewest resources. In these schemes, there's no need to designate or manage reserve bandwidth on any links. The other schemes used more resources, although the level of resource consumption can be reduced if recovery resources are shared among multiple working LSPs. In general, therefore, the network will do well with respect to three of four metrics if dynamic path recovery is used. Unfortunately, there is a penalty to be paid in the form of increased recovery time. Recovery time constraints will therefore tend to determine which path recovery scheme is ultimately chosen by the network operator.

Table 1. Analysis grid,[2] in which four levels of pre-provisioning for path recovery are compared with respect to a set of four performance metrics.

		Path Search (computation and selection)	
		pre-planned	dynamic
first path setup phase after failure	allocation	**Case 1** faster recovery less flexible less robust most resource consuming	does not apply
	selection	**Case 2** relatively fast recovery relatively flexible relatively robust resource consumption depends on sharing degree	does not apply
	signaling	**Case 3** relatively fast recovery more flexible relatively robust less resource consuming (depends on sharing degree)	**Case 4** less fast (computation requirement) most flexible most robust least resource consuming

3. SIMULATION SETUP AND EXPERIMENT DESIGN

In order to enhance the analysis grid that we described in the previous section, we simulated the behavior of path recovery schemes using each of the four levels of pre-provisioning that were described in Section 2.1. We conducted our simulation experiments using an enhancement package that we developed as an extension for the Scalable Simulation Framework (SSF) network simulation tool.[16] This package, known as the GMPLS/Lightwave Agile Switching Simulator (GLASS),[17] consists of a set of objects that are interoperable with SSF and that allow the user to model optical cross connects, optical network interface cards (ONICs), and optical fiber bundles consisting of an arbitrary number of fibers, with an arbitrary number of wavelengths on each fiber. In this section, we describe the experiments that we performed, including the values that we assigned to various global parameters and other design assumptions that we made. We also describe the performance metrics that we used and discuss the methods that we used to compute them.

We simulated two large-scale backbone network topologies. In both topologies, each bidirectional link is composed of 8 unidirectional fibers (4 in each direction). There are 16 wavelengths on each unidirectional fiber; 15 carry data, and one is used solely to carry control messages. Each wavelength supports a bandwidth of 2.5 Gbps. In the NSFNet topology, shown in Fig. 2 (a), there are nineteen nodes with an average degree of 3.368. In the GEANT topology, shown in Fig. 2 (b), there are seventeen nodes with an average degree of 2.882. We assume that no switches in either network are capable of wavelength conversion. To determine the point when the network is stable, we ran the simulations for 1 simulation year. We found that the network load stabilized at around $t = 800000$ seconds. We ran all the simulations for 2000000 seconds, which is around 23 days.

3.1. Scenarios

To examine the behavior of each of the four pre-provisioning levels under a variety of operational conditions, we simulated two different types of failures in both network topologies over a range of offered loads. In order

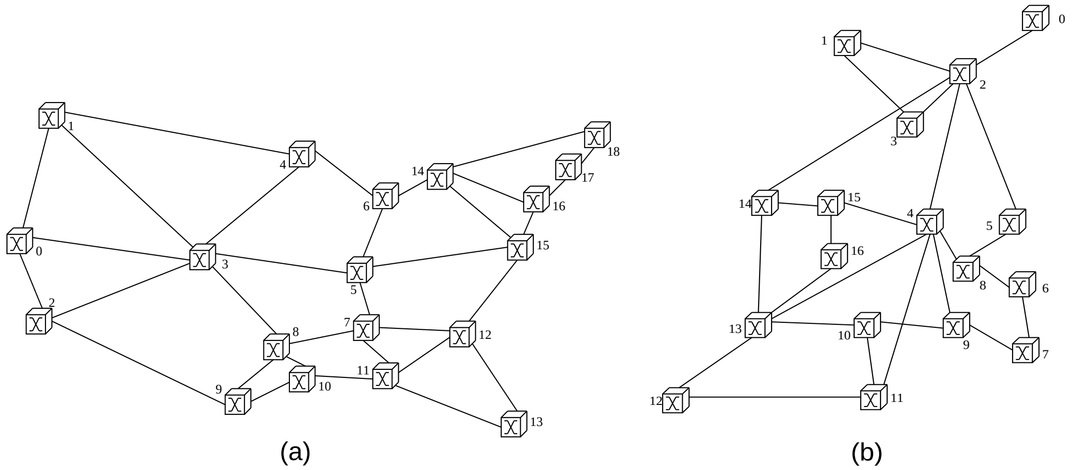

Figure 2. (a) The NSFNet network topology and (b) the GEANT network topology.

to produce a particular load level, we generated a Poisson stream of connection requests, where the source and destination node IDs were uniformly distributed over the entire node ID space, with the restriction that calls could not originate and terminate at the same node. All connections were protected using the same path recovery scheme; the same level of pre-provisioning was used for all connections. The connection duration was exponentially distributed with a mean value of $1/\mu = 100$ hours. The average arrival rate of the connection requests and the duration of the connections can be used to calculate the load offered to the network. For example, if $\lambda = 0.01$ requests/hour (corresponding to an average gap of approximately four days between requests), then the network sees a load of one Erlang. For the link failure scenario we considered loads of 50, 100, 500, and 1000 Erlangs, which respectively correspond to mean interarrival times of two hours, one hour, twelve minutes, and six minutes. For the node failure scenario we considered loads from 100 to 1000 Erlangs, at intervals of 100 Erlangs.

When the network receives a new connection request, it must generate two edge-disjoint paths between the designated source and destination nodes. To do this, we implemented a modified version of the K-Shortest-Path algorithm described by Bhandari.[18] We used the algorithm to compute a maximum of 5 edge-disjoint paths. In some cases, it was not possible to obtain 5 edge-disjoint paths; the network simply used what was available when this occurred. The links were weighted by distance, since it is desirable to have the shortest possible optical path across the network. Starting with the best path, the network controller attempted to find a wavelength that could be assigned to the path. It used the FirstFit algorithm to do this; if a wavelength could be found, then the path was designated as the working LSP. If the wavelength assignment algorithm was unsuccessful, the network controller would move on to the next best path. If the supply of candidate paths was exhausted, the connection request was considered failed. If a working path could be found, the network controller would begin to set up the backup path. The controller does this by examining the remaining members of the set of candidate paths and using a modified wavelength assignment algorithm, called FirstFitBackup. This algorithm is similar to FirstFit, but it allows multiple backup LSPs to share a single wavelength.

We considered the following two types of failure scenarios for each network topology: a link failure scenario and a node failure scenario. For the link failure scenario, we failed all eight unidirectional fibers associated with a given randomly chosen link at a fixed time $t = 1500000$ seconds (17.3 days). The link failure is automatically repaired after two days (48 hours). We repeated the scenario 20 times and formed an ensemble average for each of the four metrics of interest. We use a similar methodology for the scenarios involving node failures. To implement a node failure, we simultaneously caused all the unidirectional fibers originating from the affected node to fail. Each node's failure was simulated three times. We formed a network-wide ensemble average by repeating the experiment for each node in the network and averaging the final value of each metric over the full set of runs.

3.2. Signaling

The four types of recovery schemes described in Section 2.1 use GMPLS signaling protocols, with extensions as described in the protocol specification.[4] Responding to a failure requires a three-way handshake. First the node that detects the failure sends a Failure Indication message to each affected LSP's network ingress point. Next the ingress computes a backup path if dynamic recovery is being used and sends a Switchover Request message down stream along the backup path to configure the intermediate switches to support the displaced traffic. When this message is received by the egress node, that node will send a Switchover Response message back to the ingress node to complete the backup LSP setup. Typically, these two messages are respectively contained in Path and Resv RSVP-TE messages.

In the simulations, we have implemented the GMPLS signaling protocol with some limitations; not all types of failure scenarios are permitted. For example, we do not consider multiple failure events. This allows us to simplify the recovery protocol implementation. Another condition was that no connection setup or tear down attempt could fail due to failures in the GMPLS signaling protocol, such as lost messages. This condition allows us to have relatively simple management of the network resources. A network controller generates new connection requests and computes working and backup LSPs. If both paths can be computed, the network controller notifes the nodes along the path about the new connection so that they can be configured to support it. Similarly, at the end of a connection, each node will be notified about the end of the connection. Activating the backup LSP, however, is automatically accomplished using signaling messages.

3.3. Node Parameters

All switches in each network are transparent optical cross-connects, meaning that there is no conversion from photons to electrons and back for traffic that transits the switch. We also assume that there are no wavelength converters. In addition, the setup delay is 7 ms (for connect/disconnect lambdas). The delay for OE (Optical-Electrical) or EO (Electrical-Optical) is 0.05 ms, and happens at the Add/Drop Lambda (ADL) ports. The switch does its resource allocations for connection setups sequentially, using a FIFO queue to store configuration commands. There is no latency in the ONIC (Optical Network Interface Card). Thus, only the size of the message and the line rate determine the transmission delay for signaling messages. Each node's failure manager contains 2 queues; one is used for signaling messages and failure notifications, and the other one is used for recovery timeouts. The timeout queue has a higher priority than the message queue, and the timeout value, which is the maximum time allotted for a failed connection to be recovered before it is declared lost, is 1 sec. The processing time of each signaling message is 1 msec, and the queue size is infinite. We defined additional delays for resource selection (0.1 msec) and route computation (0.1 msec).

3.4. Simulation Metrics

In our simulations, we use four performance metrics that map to the qualitative measures presented in the analysis grid. They are (1) new connection blocking probability, which corresponds to resource optimization, (2) restorability, which corresponds to the success rate of connection recovery, (3) mean recovery time, which corresponds to efficiency (or switching time), and (4) availability, which corresponds to the down time of LSPs. We describe each of these metrics in more detail.

The new connection blocking probability, $Pr\{\text{blocking}\}$, is a measure of the number of new connection requests that cannot be satisfied because the required resources are not available. A connection request fails if it is not possible to successfully complete both the working LSP setup and the backup LSP pre-provisioning setup. We maintain two counters that are updated upon the arrival of each new connection request. One counter keeps track of the total number of new connection requests, and is incremented upon the arrival of each new request, while the other tracks the number of failed requests and is incremented only when a request fails. The estimated new connection blocking probability is the ratio of the second counter to the first.

The restorability, R, is the probability, expressed as a percentage, that a failed connection can be successfully recovered. For the link failure scenarios, this is just the ratio of the number of restored connections to the total number of affected connections. R is computed in this case by taking the ratio of the number of successfully restored connections to the total number of failed connections. When we compute this metric for the node failure scenarios, we exclude connections that terminate at the failed node, as their being restored is impossible.

The expected restoration time, $E[T]$, is the average value of all successfully restored connections' recovery times. The recovery time for each connection is the time from the onset of failure to the successful initiation of traffic on a backup LSP whose resources have been reserved, selected, and allocated.

The availability, A, is a measure, expressed as a percentage, of the average fraction of an affected connection's lifetime that is spent in an active state. For this set of connections that are successfully restored, $A = 1 - \mu E[T]$. For each connection that cannot be restored, if t_{start} and t_{end} are its respective beginning and ending times and t_f and t_r are, respectively, the node or link failure and repair times, with $t = 0$ being the time when the simulation starts, then for that connection

$$A = \begin{cases} \frac{t_f - t_{start}}{D}, & t_{end} < t_r \\ \frac{D - t_r + t_f}{D}, & t_{end} \geq t_r, \end{cases}$$

where $D = t_{end} - t_{start}$ is the connection's duration.

4. SIMULATION RESULTS AND ANALYSIS

In this section, we present the results of the simulations. The results associated with the link failure scenarios for the GEANT and NSFNet topologies appear in Fig. 3 and Fig. 4, respectively. The results from the node failure scenarios appear respectively in Fig. 5 and Fig. 6 for the GEANT and NSFNet topologies. The same set of trends can be observed in all four sets of plots, although the network's performance is more severely degraded in the case of node failures. When the network is lightly loaded, the performance of the four levels of pre-provisioning is similar with respect to all four metrics, although there are some important differences. When the load increases beyond 200 Erlangs, which corresponds roughly to an average link load of 35% and 15% for Cases 1 and 2 and Cases 3 and 4 respectively, we begin to see significant differences between the various recovery schemes. We show network-wide average link loads for the four recovery schemes versus offered load in Fig. 7. The connection request blocking probability increases significantly for Cases 1 and 2 while remaining low for Cases 3 and 4 for loads up to 500 Erlangs (below 1%) in the link failure scenarios and up to 600 Erlangs (below 3%). Beyond

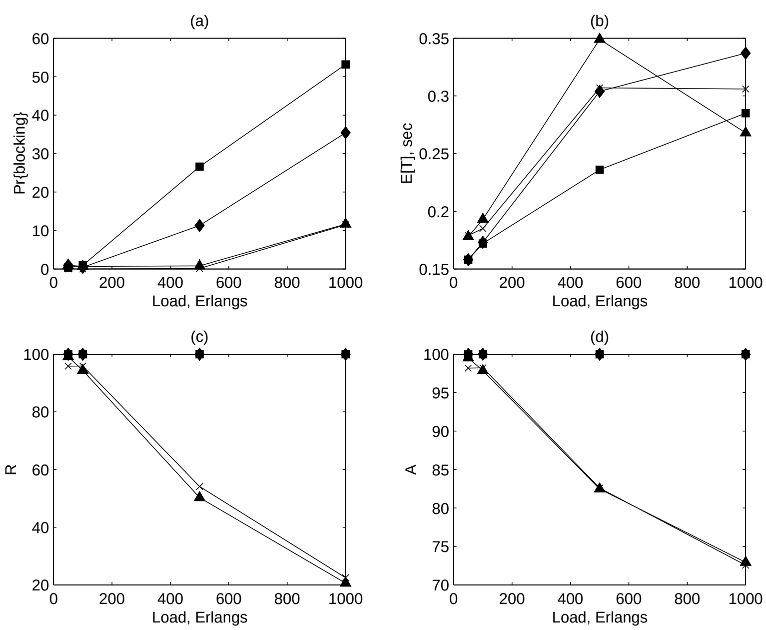

Figure 3. Performance of path recovery schemes in the GEANT network for the link failure scenario. (a): New connection blocking probability (%) vs. load. (b): Average path recovery time vs. load. (c): Restorability (%) vs. load. (d): Availability (%) vs. load. In all plots, the recovery schemes are denoted as follows: ■: Case 1, ⧫: Case 2 ▲: Case 3, ×: Case 4

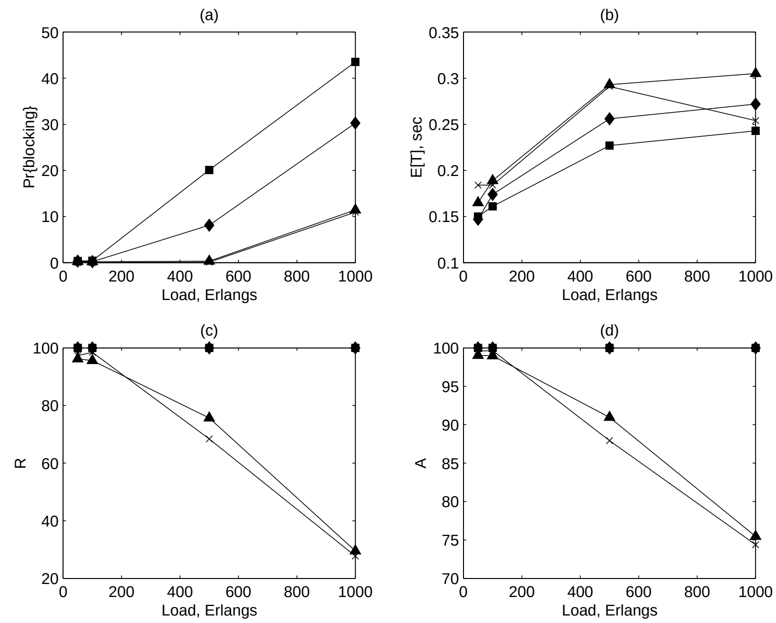

Figure 4. Performance of path recovery schemes in the NSFNet network for the link failure scenario. (a): New connection blocking probability (%) vs. load. (b): Average path recovery time vs. load. (c): Restorability (%) vs. load. (d): Availability (%) vs. load. In all plots, the recovery schemes are denoted as follows: ■: Case 1, ◆: Case 2 ▲: Case 3, ×: Case 4

these loads, the connection request blocking probability increases at a lower rate for Cases 3 and 4 (essentially the same rate for both cases) than for Cases 1 and 2. This effect can be seen in Fig. 7 as well. Case 1 is by far the worst performer with respect to this metric, as indicated it all four figures.

Because of the parameters we used, we observed little difference between the four pre-provisioning levels with respect to recovery time. In all four figures, we see that recovery time for Cases 3 and 4 peaks at 500 Erlangs and declines at higher loads. For Cases 1 and 2, the rate of increase of the mean recovery time with respect to the load is less over the range $[500, 1000]$ than it is over the range $[100, 500]$. This is because connections that span many hops are more likely not to be set up successfully in high-load environments, so that short connections (with fewer hops) are more abundant. In such a situation, the relative absence of long connections skews the recovery time distribution and produces a lower average value. If we were to use a route computation and selection time on the order of seconds, which is more common in large networks, we would see a dramatic gap between the restoration time curves for Cases 1 and 2 and Cases 3 and 4.

With respect to the restorability and availability measures, we obtained the surprising results that Cases 1 and 2 outperformed Cases 3 and 4 by a margin that increased dramatically with the offered load. For the link failure scenarios, we saw availability values of 100% in the first two cases; this is expected due to pre-selection of resources. Moreover, Cases 3 and 4 had availability values above 98% for loads below 100 Erlangs in the single link failure case. However, for increasing load, we observed that the values of R fell for Cases 3 and 4 to around 30% in the NSFNet topology and 20% in the GEANT in the link failure scenario. When we simulated node failures, we saw that R fell to 10% for NSFNet and 5% for GEANT. The availability measure, A, exhibited similar behavior. Cases 1 and 2 also performed poorly in the node failure scenarios relative to their values of 100% that were obtained when only single link failures occurred.

There are several reasons for the set of behaviors observed above. New connection requests are not accepted unless both the working and backup LSPs can be created (if pre-provisioning levels 1 or 2 are being used). Also, only working LSPs that do not have any links in common can share backup resources. The result of this is that any link failure in both topologies can be recovered from, but Cases 3 and 4 require multiple affected working paths (of which there are more, because the new connection blocking probability is lower than for Cases 1 and

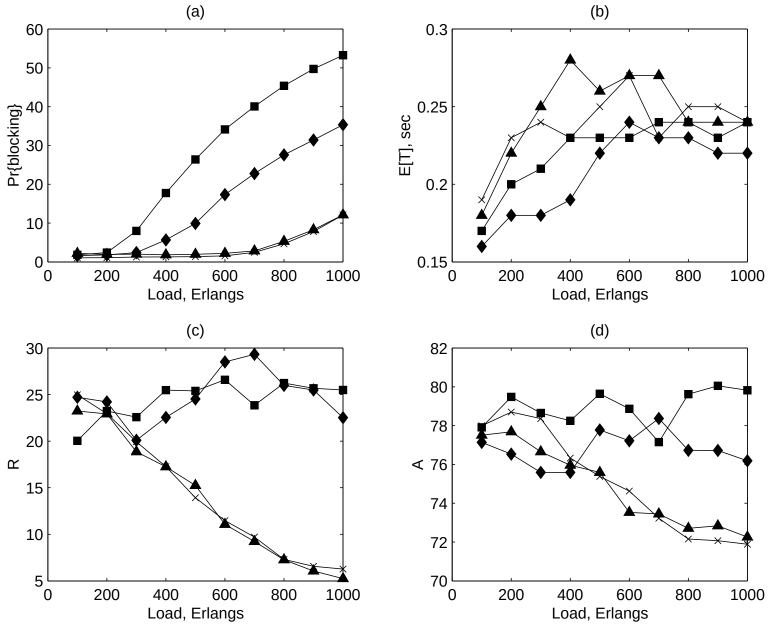

Figure 5. Performance of path recovery schemes in the GEANT network for the node failure scenario. (a): New connection blocking probability (%) vs. load. (b): Average path recovery time vs. load. (c): Restorability (%) vs. load. (d): Availability (%) vs. load. In all plots, the recovery schemes are denoted as follows: ■: Case 1, ♦: Case 2 ▲: Case 3, ×: Case 4

2) to contend for limited backup resources, resulting in lower restorability and less availability. Thus, at a load of 1000 Erlangs, only about half of new connection requests are honored if Case 1 is used, but every failed LSP can be recovered if a link failure occurs. If dynamic recovery (Case 4) is used, 90% of new connection requests are honored but only 20% of the LSPs that are lost because of a link failure can be recovered. If we consider the values of R and A when node failures occur, as shown in Fig. 5(c-d) and Fig. 6(c-d), we see that the performance difference between Cases 1 and 2 and Cases 3 and 4 is less pronounced than it is for link failures, but it is still noticeable. The degradation of performance in the pre-planned recovery schemes relative to their performance in the link failures scenarios is severe; it can be partly attributed to the simultaneous loss of working and backup LSPs when a node fails. It can be further attributed to the contention for backup resources that arises from the failures of multiple working LSPs that share recovery (backup path) resources.

5. CONCLUSIONS AND FUTURE WORK

In this paper, we described the work that the CCAMP working group in the IETF is doing to create extensions to GMPLS to support enhanced protection and restoration capabilities in an enhanced control plane for optical networks. We described the work that the Protection and Restoration Design Team has done to list the various design trade-offs associated with using a particular recovery scheme, particularly their development of an analysis grid that discusses the effect of degree of pre-provisioning on the performance of path recovery schemes. We simulated the behavior of the four types of path recovery schemes that appear in the design team's analysis grid, and we verified that the general rules that appear in the grid are valid, with some caveats that our simulations revealed. We showed that there is an important trade-off between the network's ability to recover failed connections and its ability to accept new connections. A network operator can guarantee greater restorability by using recovery schemes with a lot of pre-planning, at the expense of a greater rejection rate for new connections. Conversely, a network can support more traffic if recovery schemes with less pre-provisioning are used, but the number of connections that can be successfully restored when a failure occurs will be much lower. If the network is slightly loaded, then it is clearly the interest of the network operator to pre-signal and pre-select recovery resources. At higher loads, using pre-signaling only or a mixture of pre-signaling and dynamic recovery schemes,

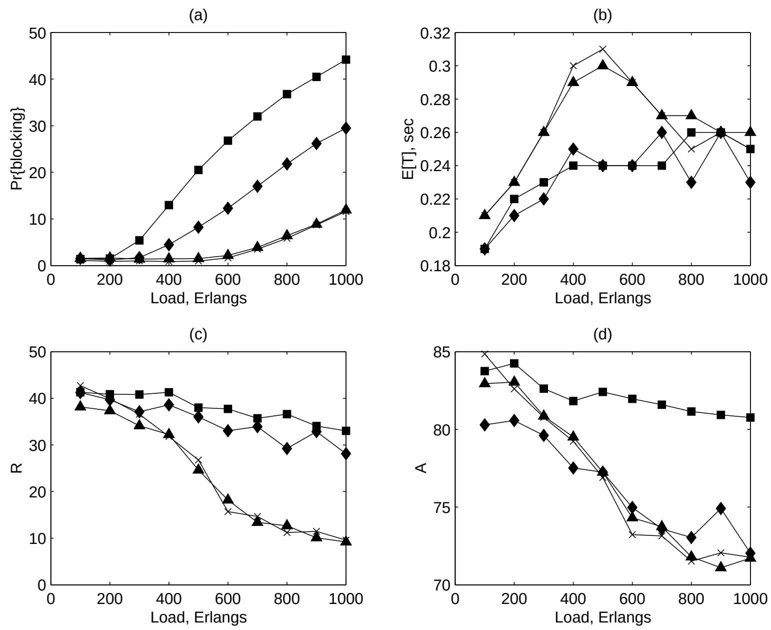

Figure 6. Performance of path recovery schemes in the NSFNet network for the node failure scenario. (a): New connection blocking probability (%) vs. load. (b): Average path recovery time vs. load. (c): Restorability (%) vs. load. (d): Availability (%) vs. load. In all plots, the recovery schemes are denoted as follows: ■: Case 1, ♦: Case 2 ▲: Case 3, ×: Case 4

one can obtain better overall network performance. Our next task is to use our simulation tool to determine what the optimal mixture of pre-provisioning levels should be as a function of network load.

Acknowledgements

The authors are grateful to S.K. Lee, N. Golmie, and D. Su of NIST, H. Choi, S. Subramaniam, G. Sahin, and Y. Kim of George Washington University, and O. Borchert for their valuable contributions.

REFERENCES

1. "Recovery (Protection and Restoration) Terminology for GMPLS," E. Mannie and D. Papadimitriou, Eds., IETF Internet draft (work in progress).
2. "Analysis of Generalized MPLS-based Recovery Mechanisms (including Protection and Restoration)," D. Papadimitriou and E. Mannie, Eds., IETF Internet draft (work in progress).
3. "Generalized MPLS Recovery Functional Specification," J. Lang and B. Rajagopalan, Eds., IETF Internet draft (work in progress).
4. "RSVP-TE Extensions in support of End-to-End GMPLS-based Recovery," J. Lang and Y. Rekhter, Eds., IETF Internet draft (work in progress).
5. J. Wang, L. Sahasrabuddhe, and B. Mukherjee, "Path vs. Subpath vs. Link Restoration for Fault Management in IP-over-WDM Networks: Performance Comparisons Using GMPLS Control Signaling," *IEEE Communications Magazine*, vol. 40, no. 11, pp. 80–87, November 2002.
6. M. Goyal, G. Li, and J. Yates, "Shared mesh restoration: a simulation study," *Proceedings of the Optical Fiber Communication Conference and Exhibit, 2002*, pp. 489–490.
7. G. Li, J. Yates, R. Doverspike, and D. Wang, "Experiments in fast restoration using GMPLS in optical/electronic mesh networks," *Proceedings of the Optical Fiber Communication Conference and Exhibit, 2001*, vol. 4, pp. PD34-(1–3).

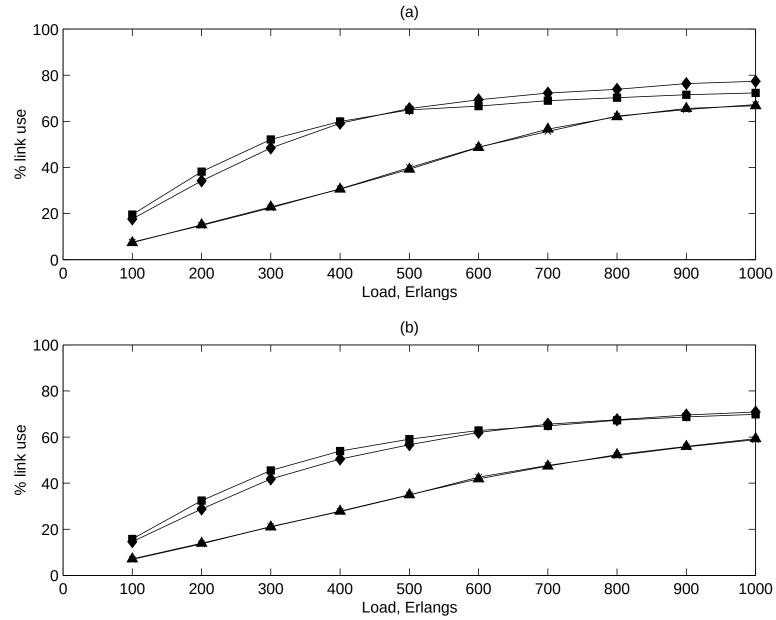

Figure 7. Average link usage vs. offered load. (a): Link usage for the GEANT topology. (b): Link usage for the NSFNet topology. In all plots, the recovery schemes are denoted as follows: ■: Case 1, ♦: Case 2 ▲: Case 3, ×: Case 4

8. "Generalized Multi-Protocol Label Switching (GMPLS) Signaling: Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions," L. Berger, Ed., RFC 3473, January 2003.
9. "Generalized Multi-Protocol Label Switching (GMPLS) Signaling: Constraint-based Routed Label Distribution Protocol (CR-LDP) Extensions," P. Ashwood-Smith, Ed., RFC 3472, January 2003.
10. "Link Management Protocol (LMP)," J. Lang, Ed., IETF Internet draft (work in progress).
11. "Link Management Protocol (LMP) for Dense Wavelength Division Multiplexing (DWDM) Optical Line Systems," A. Fredette, Ed., IETF Internet draft (work in progress).
12. "Routing Extensions in Support of Generalized MPLS," K. Kompella, Ed., IETF Internet draft (work in progress).
13. "OSPF Extensions in Support of Generalized MPLS," K. Kompella, Ed., IETF Internet draft (work in progress).
14. "Tracing Requirements for Generic Tunnels," J. Lang and D. Papadimitriou, IETF Internet draft (work in progress).
15. D. Griffith, "IETF Work on Protection and Restoration for Optical Networks," *Optical Networks Magazine*, vol. 4, no. 4, July/August 2003.
16. *Scalable Simulation Framework API Reference Manual, Version 1.0*, James H. Cowie, Ed., www.ssfnet.org/SSFdocs/ssfapiManual.pdf.
17. Oliver Borchert, Richard Rouil, "The GMPLS Lightwave Agile Switching Simulator - An overview," www.antd.nist.gov/glass
18. R. Bhandari, "Optimal physical diversity algorithms and survivable networks," *Proceedings of the Second IEEE Symposium on Computers and Communications, 1997*, Alexandria, Egypt, 1-3 July, 1997, pp. 433-441.