Network Design and Architectures for Highly Dynamic Next-Generation IP-Over-Optical Long Distance Networks

Angela L. Chiu, *Senior Member, IEEE*, Gagan Choudhury, *Fellow, IEEE*, George Clapp, Robert Doverspike, *Senior Member, IEEE*, Joel W. Gannett, *Senior Member, IEEE*, John G. Klincewicz, Guangzhi Li, *Member, IEEE*, Ronald A. Skoog, *Member, IEEE*, John Strand, *Senior Member, IEEE*, Ann Von Lehmen, *Fellow, IEEE*, and Dahai Xu, *Member, IEEE*

(Invited Paper)

***Abstract*—The DARPA CORONET project seeks to develop the target network architectures and technologies needed to build next-generation long-distance IP-over-Optical-Layer (IP/OL) networks. These next-generation networks are expected to scale 10–100 times larger than today's largest commercial IP/OL network. Furthermore, DARPA has established advanced objectives for very rapid provisioning of new IP or private line connections, very rapid restoration against up to three simultaneous network failures, and future dynamic "wavelength" services ranging from speeds of 40–800 Gigabits per second. Besides these ambitious goals, the CORONET project seeks to establish a commercially-viable network architecture that supports both commercial and government services. In this paper, we describe the CORONET program requirements, and present our initial architectures and analysis of the early phases of this long-term project. We propose a novel *2-Phase Fast Reroute* restoration method that achieves 50–100 ms restoration in the IP-Layer in a cost-effective manner, and a commercially viable OL restoration method that can meet the rapid CORONET requirements. We also estimate the magnitude of the extra capacity needed to provide dynamic wavelength services compared to that of static services, and show that the extra capacity to restore a small percentage of high priority traffic against multiple failures requires a small amount of extra capacity compared to that of single failures.**

***Index Terms*—Network operating systems, networks.**

Manuscript received February 10, 2009; revised April 06, 2009. Current version published June 12, 2009. This work was supported by the DARPA CORONET Program, Contract N00173-08-C-2011.

A. Chiu, G. Choudhury, R. Doverspike, J. Klincewicz, and J. Strand are with AT&T Labs–Research, Middletown, NJ 07748 USA (e-mail: chiu@research.att.com; gchoudhury@att.com; rdd@research.att.com; klincewicz@att.com; jls@research.att.com).

G. Li and D. Xu are with AT&T Labs–Research, Florham Park, NJ 07932 USA (e-mail: gli@research.att.com; dahaixu@research.att.com).

G. Clapp, J. Gannett, R. Skoog, and A. Von Lehmen are with Telcordia, Red Bank, NJ 07701 USA (e-mail: clapp@research.telcordia.com; jgannett@research.telcordia.com; rskoog@research.telcordia.com; avl2@research.telcordia.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JLT.2009.2021564

I. Introduction

The past several years have seen significant activity in developing and demonstrating capabilities for "on-demand" optical and data services. The grid and high-performance computing and e-Science research communities have been especially prolific in advancing capabilities for such services (e.g., [1]–[5]) while in the commercial world, two large carriers have announced rapid provisioning of SONET services using control plane technology. The recently launched DARPA CORONET program [6] moves well beyond the current state of the art in long-distance networks and envisions a high capacity, highly dynamic next-generation IP-over-Optical Layer (IP/OL) network architecture with greatly improved operation, performance, survivability, and security that would be attractive for transition to commercial networks.

The CORONET program envisages evolution of today's commercial network architectures to a more generic and simplified network architecture. Figs. 1 and 2 illustrate this evolution. Fig. 1 depicts today's complex layering of "legacy" technology layers that one would find in a large long-distance carrier, such as SONET/SDH rings, Digital Cross-Connect Systems (DCSs), Frame Relay Switches, ATM switches, and circuit switches. These layers were deployed over many years of technology evolution. The CORONET network vision in Fig. 2 simplifies this architecture to just three layers: IP Layer, Optical Layer, and Fiber Layer. The Optical Layer consists of *All-Optical Switches* (AOSs) and the links (or spans) between them. The AOS is an evolution of today's multi-degree Reconfigurable Optical Add Drop Multiplexer (ROADM) with tunable and steer-able add/drop terminals [7]. A ROADM uses Dense Wavelength Division Multiplexing (DWDM) to multiplex many high rate digital signals over a single fiber. These figures also show how today's services would migrate to this future architecture. For example, today's private line services at or below the rate of 10 Gigabits-per-second (Gbps) are envisioned to be carried by the IP layer, either by constant bit rate (CBR) circuit emulation or as packet-based Virtual Private Line (e.g., Ethernet Private Line). Private line services

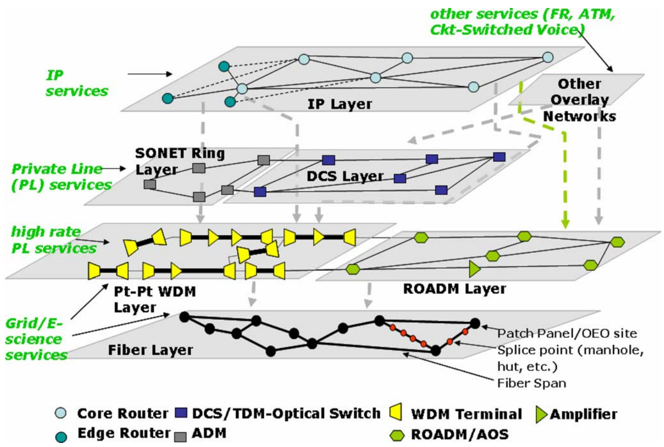

Fig. 1. Today's network layers.

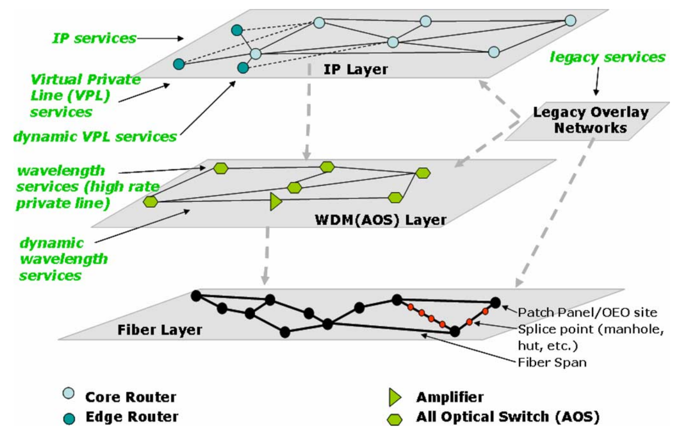

Fig. 2. CORONET network layers.

at 40 Gbps or higher rate (called "wavelength services") are carried directly on the Optical Layer.

In addition to advanced network architectures, the CORONET program envisages advanced service features and capabilities, such as the following.

- CORONET examines a potential future network significantly larger than today's core IP/OL network. The nominal-size network is characterized by combined traffic of 20 Terabits-per-second (Tbps), split across both IP-Layer services (75%) and wavelength services (25%), and AOS channels (wavelengths) that support 40 Gbps transmission rates. CORONET [6] specifies other values and combinations of these parameters, for example, up to 100 Tbps traffic, 100 Gbps wavelengths, and different mixes of IP and Wavelength services. If we consider the full range of these different network sizes, we estimate they fall in the range of 10 to 100 times the scale of current networks. However, this paper reports results on only the nominal-size network.
- Significantly more dynamic services: This can be characterized by service "connection" requests that can occur randomly among a large set of nodes scattered across the world and have widely varied holding times (anywhere from seconds to years).
- Significant variation in connection rates: connection requests carried by the IP layer can be requested at rates up to a quarter of the wavelength rate. Wavelength services are directly transported by the Optical Layer and can be requested in units of one, two, four, or eight times the wavelength rate. CORONET establishes various scale benchmarks for the network.
- High quality of service (QoS): blocking for connection requests is required to be less than 10^{-3}.
- Rapid Provisioning: A premium class of wavelength services is expected to be provisioned in 100 ms or less (slightly longer for inter-continental demands).
- Rapid restoration: All services classes, except IP-Layer *Best Effort* (BE) are expected to be restorable within 100 ms after a single network failure (slightly longer for inter-continental demands). In addition, a small subset of premium service classes must be restorable after any double failure and a smaller class must be restorable after any triple network failure.

These are extremely challenging service objectives. To gauge this challenge, let us more carefully contrast the networks of today vs. tomorrow, as depicted in Figs. 1 and 2. The CORONET vision is enabled by a future AOS with significantly increased optical reach, capacity, and flexibility. For example, CORONET assumes the nominal-size has an un-regenerated optical reach of up to 2000 km, 100 wavelengths per system, and rapid recovery from optical impairments, such as transients. While CORONET projects 100 ms provisioning for some priority class of services, note that today's high rate private line services are generally provided in a much slower and sometimes manual provisioning process that can take months. Some advanced "Bandwidth-on-Demand" services, such as AT&T's Optical Mesh Service [8] can be provisioned in seconds, but generally such services are not yet available for rates of 2.5 Gbps and higher. One of the most challenging aspect of the CORONET vision is the ability to provide connection requests for wavelength services with extremely short holding times (on the order of seconds).

We see the difficult challenge of providing CORONET's vision of network restoration when we note that in today's large carrier-based networks there are few, if any, restoration capabilities besides limited deployments of simple (but costly) electronic-based 1+1 protection at the DWDM-Layer. Virtually all restoration today is provided by the upper layers, such as reconfiguration in IP networks, as provided by Internal Gateway Protocols (IGPs)—such as Open Shortest Path Forwarding (OSPF) or Intermediate System-Intermediate System (IS-IS)—or various forms of restoration provided by SONET rings and DCSs. Clearly, the CORONET restoration requirements will require the development of advanced Optical-Layer restoration mechanisms. Furthermore, besides the establishment of advanced and rapid restoration methods, CORONET requires that both IP and Optical-Layer restoration methods be implemented in a cost-efficient manner. For example, CORONET constrains the network design such that the extra IP-Layer and Optical-Layer capacities needed for restoration do not exceed 75% of the working capacity (i.e., the amount of capacity to carry the working connections in the non-failure

Fig. 3. Global CORONET network.

Fig. 4. CORONET network in the continental USA.

state). This ratio is called the *spare capacity ratio* in CORONET terminology. Also, the stringent restoration requirements for services to be restorable against multiple failures require a highly connected, worldwide Optical-Layer topology.

This paper provides our preliminary networking architectures and design procedures to meet the CORONET vision. Furthermore, we present a preliminary assessment of the network impacts of CORONET advanced services capabilities on network capacity constraints and targets. There is a wide body of work in the literature that addresses IP-over-Optical-Layer architectures, interworking, and technologies, including many contributions by the present authors. For example see [9]–[14]. What principally differentiates our present paper from that of previous approaches is that we are further constrained by limitations on capacity (cost), network feasibility and operations required to achieve commercial viability, the long-term goal of the CORONET project. A summary of some of our earlier results can also be found in [15].

Section II discusses the Optical-Layer topology that consists of the AOSs and the links between them. Section III describes the traffic model that we formulated, as well as the characteristics of the traffic matrix. Section IV describes IP-Layer architecture, restoration methods, and network design procedures and tradeoffs. Section V describes the Optical-Layer architecture, provisioning process, restoration process, network design procedures and, finally, preliminary CORONET network capacity impacts and tradeoffs. Section VI concludes the paper.

II. Optical-Layer Topology

The topology of the CORONET Optical Layer is loosely modeled after existing networks of commercial carriers. The CORONET requirements referred to in [6] describe an optical network of 100 nodes: 75 in the continental USA and 25 in Europe and Asia. It is a hierarchical network with a mix of large and small nodes, distinguished principally by their degree and the volume and type of the services that they terminate. Furthermore, it also specifies some high-level connectivity and traffic requirements. Figs. 3 and 4 depict the global and continental USA topologies.

The network scale is characterized by the graphs in Figs. 5 and 6. Fig. 5 shows the distribution of the path lengths between all node pairs. The path length is measured in *routed kilometers* which is the sum of the link distances of the shortest path between a node pair. Each link distance is the aerial distance between the end nodes of the link as measured on a great circle, inflated by 20% to represent terrestrial or oceanic routing. Most node pairs lie within the continental USA and have an average path length of 2600 km, but there are many very long transoceanic paths that bring the overall average length to more than 7100 km. Fig. 6 depicts the nodal hop count of the shortest path between all node pairs. The average number of hops in the continental USA is 6.9, while the overall average is 7.9. The variation in hop count is not as dramatic as the variation in path length because Asia and Europe have relatively few nodes (12 and 13, respectively) and a very small average hop count within each continent (less than three hops). Furthermore, 37 of the 136 links lie outside the continental USA. The longest 20 of these, which are mainly transoceanic links, have an average link distance of 5869 km. These add only one hop, but a considerable distance when they are included in a path.

The volume of traffic in the CORONET network ranges from 20 Terabits per second (Tbps) (the nominal-sized network) up

TABLE I
CORONET SERVICES

Service Type	Granularity		Setup Type		Restoration Type		Packet Quality of Service		Total
IP Best Effort									1
IP Guaranteed Bandwidth	3	Fine Medium Coarse	3	Fast Scheduled Semi-permanent	3	1-failure 2-failure 3-failure	2	Latency-sensitive Loss-sensitive	54
Wavelength	4	1-λ 2-λ 4-λ 8-λ	4	Very fast Fast Scheduled Semi-permanent	3	1-failure 2-failure 3-failure			40
Total Number of CORONET Services									**95**

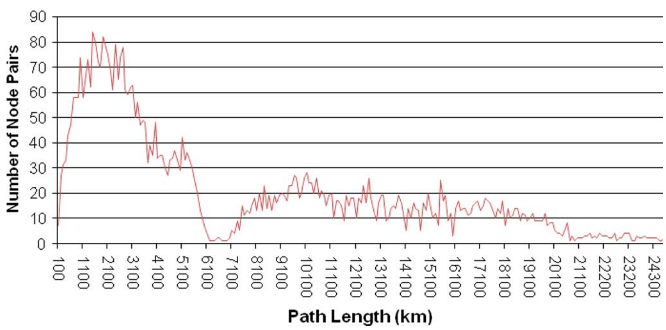

Fig. 5. Distribution of path distances.

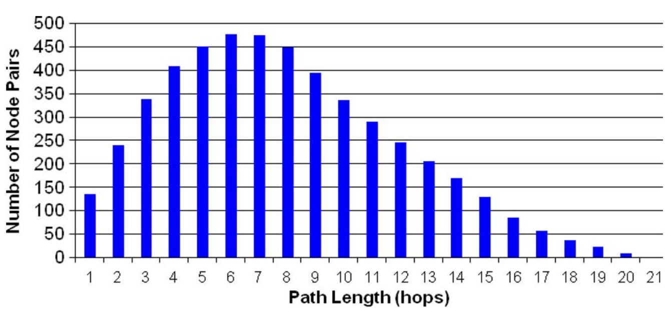

Fig. 6. Distribution of hop count for the shortest paths between all node pairs.

to 100 Tbps. Hence, a primary concern in designing the Optical Layer is to place links so there is sufficient network capacity between node pairs. This placement is subject to two stringent CORONET constraints for the nominal-sized network: 1) exactly one AOS-AOS system (a single fiber pair) per link can be placed on all but 10% of the links, in which case two systems are allowed; and 2) the ratio of the restoration capacity to the working capacity must be less than 75%.

The network restoration requirements mentioned above (i.e., that the network must be capable of restoring a subset of services against as many as three failures) means that the network must be richly interconnected. In the continental USA, for example, there are four separate paths between the east coast and the west coast. All the links are carefully selected to ensure the greatest possible path diversity between node pairs. In general, the CORONET topology is constructed so that *every* node pair with minimum nodal degree of k is k-diverse, thus providing the greatest possible support for restorable services.[1] For example, if a node pair is composed of a 3-degree node and a 4-degree node, the minimum degree is 3. A node pair is 3-diverse if there are three diverse paths between them. This characteristic is achieved even though nodal degree is tightly constrained: 50% of the nodes are degree two, 35% are degree three, 15% are degree four or higher, and the average nodal degree is 2.8. The algorithm to verify network link diversity is a standard max-flow algorithm with appropriately capacitated edges, such as can be found in [16].

III. TRAFFIC MODEL

The CORONET traffic model consists of 95 different classes of service (CoS), encompassing IP and wavelength services with different bandwidth granularities, quality of service (QoS) requirements, setup time requirements, and restoration classes. Table I shows the different service characteristics and the possible combinations.

The bandwidth granularities for IP range from 0.4 Gbps to a quarter of the rate of a wavelength. The wavelength services range from one wavelength to bundles of two, four, and eight wavelengths. Dynamic services include a "very fast" wavelength service setup class with 100 ms setup time and less than one minute holding time. In addition, there are both IP and wavelength services with setup times of two seconds ("fast") and 10 seconds ("scheduled"). All services (except IP best effort) are restorable against at least one failure, with up to 20% of the traffic restorable against two failures and up to 5% restorable against three failures. Wavelength classes of service with very fast setup are restorable against only one failure, instead of three as with the other classes.

As we mentioned in the previous section, the total volume of traffic is quite large compared to today's networks. There are four traffic scenarios, the smallest of which has 20 Tbps (the nominal-sized network) and the largest 100 Tbps. In addition to the sheer volume, a particularly challenging aspect of the traffic is the extreme dynamic arrival and departure process of the wavelength services. In carriers today, CORONET's "wavelength services" are most closely approximated by the highest speed category of private-line services that are supported by their transport networks. Carriers presently plan their transport

[1]In CORONET, we require traffic among nodes in the continental USA to stay within the continent, i.e., the traffic cannot traverse either working or restoration paths that leave the continent. Therefore, we do not include transoceanic links when calculating k-diversity for the continental USA. We do not apply this requirement to nodes outside the continental USA.

networks based on service forecasts that span one to five years, and the services are assumed to be very long-lived (typically months to years). A large carrier typically experiences a low level of churn in established private-line services during a single month. In CORONET, however, the very fast wavelength service lasts less than one minute, and in the network scenario with 100 Tbps of total traffic, there would be over ten million connections requests in a single month! Designing an efficient network that can support these services is indeed a very challenging task, and it is compounded by the bandwidth of some of the connection requests, e.g., a 3-restorable eight-wavelength connection with maximum connection-request blocking of 10^{-3}. Note that design of the IP-Layer topology creates links that are transported as connections in the Optical Layer. By definition of "wavelength services" we do not include upper-layer IP links. This combination of IP-Layer links and wavelength services creates the totality of connections that must be transported by the Optical Layer. In this paper, we will focus on the network architecture and design for the nominal-sized network (20 Tbps), partitioned into 75% IP-Layer traffic and 25% wavelength services. We note that the relative amount of Optical-Layer capacity required to meet the QoS and restoration objectives of each layer deviates quite significantly from these traffic percentages. Section V describes this surprising observation.

We define each traffic class by its layer, rate, QoS, and restoration characteristics. We then define Erlang-type traffic arrival and departure stochastic processes and compute their steady-state rate parameters for each traffic class and point-pair to satisfy CORONET traffic requirements. These traffic class parameters are proportioned to meet various ratios stated in the CORONET requirements, such as the proportion of very fast services, triple-failure restorable services, etc. As another example, high-rate wavelength services only occur at a limited number of large-cities. We define this collection of rate parameters per traffic class and node pair as the *traffic matrix*. We note that while carriers generally implement a planning process that adjusts network capacity each planning period to accommodate a forecast of traffic over the upcoming period, CORONET does not specify the concept of planning for multi-period network growth. Therefore, we should loosely think of the CORONET traffic matrix as the cumulative traffic forecast of a single future planning period.

To use this traffic model in the development of our provisioning, restoration, and network design processes, as well as our performance analysis tasks, we have developed a simulation methodology. Conceptually, there are about 130,000 independent traffic generators needed to model the connection requests that can arise between the pairs of nodes in the CORONET network. For each pair of nodes and allowable class of service between that pair of nodes, a traffic stream is generated based on the expected load (from the traffic matrix) and the traffic characteristics of that class of service. A class of service is differentiated by network layer, setup time, restoration level, and QoS, as well as statistical distributions for bandwidth, holding time, and inter-arrival time. Using discrete event simulation technology, we generate a stream of demands between their endpoint nodes throughout the network, creating 130,000 demand arrival processes that each separately evince the desired statistical properties.

IV. IP-Layer Architecture and Network Design

CORONET requirements make the simplifying assumption that IP-Layer routers are placed at each physical location (node). In contrast, in actual carrier networks, the majority of locations that have DWDM equipment do not contain routers. We use a two-layer hierarchical design consisting of a backbone layer and an access layer. A node that is incident to exactly two Optical-Layer links is designated to be an "access only" node, whereas a node incident to three or more Optical-Layer links is designated to be a "backbone + access" node. In creating the IP-Layer link topology, we first establish a "direct" IP link between any pair of routers that is directly connected by a single Optical-Layer link and have traffic between them. Then, "express links" between particular pairs of "backbone + access" nodes are added sequentially using a greedy heuristic algorithm. This algorithm judiciously adds express links one-by-one from the list of candidate express links. This selection algorithm balances two heuristic constraints: 1) choose links along high traffic volume paths and 2) minimize the impact of any single optical-link failure on the IP-Layer. Objective 2) is achieved by defining a metric that counts the number of IP-Layer links that route over each Optical-Layer link at each step of the design process. By maximizing this metric over the set of Optical-Layer links, we obtain a rough characterization of the impact of the worst single Optical-Layer failure on the IP-Layer. The algorithm favors the express link that minimizes this impact. The decision of when to stop adding express links is based on minimizing overall cost (Optical-Layer transport cost plus the cost of IP ports in the IP layer). We note that this notion of how higher-layer links share paths over lower-layer links is also referred to as Shared Risk Link Groups (SRLG).

All traffic between a pair of routers is carried over a primary *tunnel*, formally called an MPLS-TE LSP (MPLS-TE = Multi-Protocol Label Switching with Traffic Engineering and LSP = Label Switched Path), that is routed over one or more IP links ("direct" and/or "express"). There may be multiple primary tunnels between the same pair of routers to allow for differentiation among traffic types during both normal operation and restoration.

To provide restoration $\leq$100 ms and maintain efficiency of rerouting, restoration in the IP Layer is provided using a 2-phase Fast Reroute (FRR) technique [17], [18]. The first phase consists of a very fast switch-over to a pre-computed backup tunnel for priority traffic between the pair of routers that is adjacent to the failure. The backup tunnels may be of the Next-Hop type to handle link failures or of the Next-Next-Hop type to handle both router and link failures. A Next-Hop backup tunnel only bypasses the failed link and a Next-Next-Hop backup tunnel bypasses both the failed link and the neighboring router to which the link is connected [18]. The second phase consists of a slower reroute of the primary tunnel over an end-to-end path optimized for the current state of the network. In both phases, we achieve efficient routing of traffic using optimized MPLS-TE which maximizes sharing of bandwidth among all failure scenarios (including, in some cases, double/triple failures for a subset of traffic). We have to design the IP link capacities in such a way that both under the no-failure scenario and under all single, double and triple failure scenarios (including the two restoration phases of single failure) there is enough capacity for all traffic that we intend to restore. In other words, we have to design a network that satisfies the maximum of all capacity requirements under the various scenarios mentioned above. Our approach for designing optimized MPLS-TE networks uses a heuristic routing algorithm that tries to minimize the maximum of all capacity requirements. Details of this heuristic will be given in a future publication.

As indicated above, there are many types of IP traffic with different bandwidth/performance/restoration requirements that are differentiated during transport and routing using IP QoS. Also, traffic between any point-pair is dynamic, characterized by steady state average load (intensity) but allowing for shorter arrival/departure of individual connections. Sizing of each IP link is done by using a multi-rate Erlang loss model to compute the overall required blocking probability and to determine the upper limit of link utilization that would satisfy this requirement. Different blocking probability requirements are used during no-failure and failure conditions. CORONET requirements allow for higher blocking during network failures. Also, 40% of the traffic is designated as best effort and does not have to be restored because of network failure; however, BE traffic whose working path does not contain failed network components (which we term "unaffected BE" traffic) must still be transported and meet CORONET loss constraints.

Table II compares four different IP-layer link topologies which differ in the number of express links. Recall that IP traffic accounts for 75% of the network traffic in the scenario we report. For each design we list in the second column the total capacity in units of 1000 wavelength-km that would be required to restore the network during any single failure. This capacity in wavelength-km is computed by routing the IP-Layer links over the Optical Layer generally using shortest-path routing. This provides a measure of the transport cost. The third column lists the number of 40 Gbps links that are required. This provides a measure of IP port costs. Although initial CORONET requirements do not specify models for equipment costs, most carriers use capital costs for evaluating competitive network architectures and designs. We illustrate such design tradeoffs with a simple relative model where we assume that the IP port cost associated with a single 40 Gbps link is equivalent to 770 wavelength-km of transport cost. The sum of transport costs and IP port costs (normalized to hide actual costs) is shown in the last column. Clearly, the sum of transport cost and IP port cost, for our assumed cost ratio, is minimized (marginally) by our Case 3 design. This is better illustrated in Fig. 7 that graphs the columns just described. If one examines the rightmost graph of normalized cost, one interesting (and desirable) feature of the IP express link selection algorithm is that the "optimal cost" region is fairly flat (Case 2 and Case 4 designs differ from the Case 3 design only by 1% and 4.7%, respectively), allowing for a large number of alternate "near-optimal" designs. We see that

TABLE II
IP-LAYER NETWORK DESIGN (SINGLE FAILURE RESTORATION)

	Capacity in units of 1000 Wavelength-km.	Number of 40 Gbps IP Links	Normalized Total Cost
Case 1 (136 Direct Links + 0 Express Links)	1565	2700	113
Case 2 (136 Direct Links + 45 Express Links)	1713	1985	101
Case 3 (136 Direct Links + 73 Express Links)	1810	1839	100
Case 4 (136 Direct Links + 107 Express Links)	2005	1785	104.7

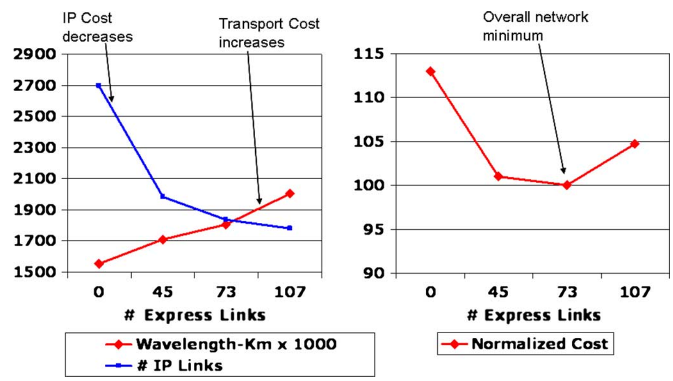

Fig. 7. IP network design (single failure restoration).

if we choose a design in this flat region and then perturb the traffic matrix or cost ratio by a small amount, then the design remains "near-optimal" under the new scenario as well.

Table III summarizes the impact on the network design of the requirements to restore certain traffic classes against double and triple network failures. Each row summarizes results for a particular design; Row 2 corresponds to the "Case 3" design in Table II. The first three columns are as in Table II. The fourth column reports the spare capacity ratio, defined as the percentage increase of total capacity-mileage required by a protection scheme over a service-only scheme. Our network design process, described above, makes very efficient use of excess restoration capacity. This is evidenced by the spare-capacity ratios (shown in the fourth column of Table III) which do not exceed 39%, which well exceed the CORONET capacity usage objectives [6]. This bound holds even under double and triple failure restoration scenarios due to the high degree of sharing of restoration capacity among the large number of failure scenarios. However, note that spare capacity can be shared over both applications: to accommodate dynamic virtual private line services and restore the network during network failures. The denominator of the spare-capacity ratio reflects the extra capacity needed to provide the blocking QoS for dynamic traffic in a network with no restoration.

Table IV shows the advantage of utilizing our optimized MPLS-TE scheme instead of standard OSPF routing. The columns are as in Table III. The first row corresponds to the "Case 3" design of Table II, which utilizes optimized

TABLE III
IP-LAYER RESTORATION SCENARIOS

	Capacity in units of 1000 Wavelength-km.	Number of 40 Gbps IP Links	Normalized Total Cost	Spare Capacity Ratio
No Failure Restoration	1335	1389	74.1	0
Single Failure Restoration	1810	1839	100	0.356
Single +Double Failure Restoration	1834	1882	101.7	0.374
Single + Double + Triple Failure Restoration	1853	1881	102.3	0.388

TABLE IV
CAPACITY COMPARISONS OF ROUTING PROTOCOLS (IP LAYER)

	Capacity in units of 1000 Wavelength-km.	Number of 40 Gbps IP Links	Normalized Total Cost	Spare Capacity Ratio
Optimized MPLS-TE	1810	1839	100	0.356
OSPF	2281	2322	126	0.684

TABLE V
IMPACT OF ALLOWING PACKET LOSS FOR BE TRAFFIC UNAFFECTED BY FAILURE (IP LAYER)

	Capacity in units of 1000 wavelength-km.	Spare Capacity ratio	Number of 40 Gbps IP Links	Normalized Total Cost
No packet loss to unaffected BE	1810	0.356	1839	100
Packet loss to unaffected BE during Fast Reroute	1684	0.239	1740	93.7
Packet loss to unaffected BE during all restorations	1565	0.172	1552	85.5

MPLS-TE. The second row reports the results with OSPF routing. Both rows assume single-failure restoration only. In the second row, OSPF weights are assigned to links that are proportional to route-km in order to minimize latency. Note that under the no-failure network state, both schemes use minimum-latency routing (a desirable feature for customers). Algorithms that successfully minimize network capacity in a restorable network exploit ways to share network restoration capacity across non-simultaneous network failures. This extra restoration capacity lies idle (unassigned) during no-failure conditions. These minimization algorithms accomplish this by seeking and sharing restoration capacity via flexible routing algorithms, which can be curtailed by minimum-latency routing constraints. The optimized MPLS-TE scheme achieves significant savings in spare capacity ratio (and cost) by relaxing the minimum-latency constraint only under failure conditions. As shown in Table IV, the spare capacity ratio is about 48% lower and the cost is about 21% lower compared to OSPF routing, which is the standard practice of most carriers.

Table V shows the advantage of allowing packet loss to BE traffic not directly affected by a failure. Only single-failure restoration is assumed in all cases. Column headings are as in the previous tables. The first row is the base case used in the previous tables and does not allow any packet loss to unaffected BE traffic. The second row allows the unaffected BE traffic to experience packet loss during the FRR phase only. This lowers the spare capacity ratio from 35.6% to 23.9%. The third row allows the unaffected BE traffic to experience packet loss during all failure conditions and lowers the spare capacity ratio further to 17.2%. We estimate the FRR phase would not exceed 10 seconds, whereas typical repair of a network failure may last several hours (perhaps longer for more complex failures); therefore, we feel most carriers would find it reasonable to operate under the assumptions of the second row, but not under the assumptions of the third row.

Table VI shows the impact of changing the fraction of traffic that needs to be restored because of double and triple failures. It is interesting to note that the spare capacity ratio stays below 57% even when 20% of the traffic has to be restored from double

TABLE VI
IMPACT ON OF CHANGING THE FRACTION OF DOUBLE/TRIPLE FAILURE RESTORABLE TRAFFIC (IP LAYER)

Fraction of Double-Failure Restorable traffic	Fraction of Triple-Failure Restorable traffic	Capacity in units of 1000 wavelength-km.	Spare Capacity Ratio
0.02	0.005	1847.1	0.384
0.06	0.015	1848.7	0.385
0.10	0.025	1853.1	0.388
0.15	0.0375	1960.1	0.458
0.20	0.05	2101.5	0.566

failures and 5% of the traffic has to be restored from triple failure. By contrast, if we use OSPF routing, the spare capacity ratio would be over 68% (Table IV) even without any double or triple failure restoration.

In summary, the key novel features of the IP network design algorithms and methodology are as follows:

- use of the 2-phase fast re-route (FRR) technique to achieve sub-100 ms restoration and maintain reroute efficiency;
- use of optimized MPLS traffic engineering (MPLS-TE) among single/double/triple failure restoration scenarios in order to greatly reduce the spare capacity ratio and total cost;
- handling dynamic traffic demands with pre-specified blocking probability requirements in conjunction with MPLS-TE routing using multi-rate Erlang loss models under stochastic traffic assumptions;
- a new automated algorithm for optimizing the design of an IP-Layer topology that prefers paths with higher traffic volume and lower probability of Optical-Layer (SRLG) impact.

V. OPTICAL-LAYER ARCHITECTURE AND NETWORK DESIGN

To meet the goals of the CORONET vision we break the analysis for the Optical Layer into the following tasks.

- Develop the process for real-time connection routing (provisioning) for wavelength services. This task consists of developing an efficient routing algorithm and set of protocols that meet the CORONET provisioning time requirements discussed in Section I.
- Develop the real-time restoration process for wavelength services. This task consists of developing an efficient re-routing algorithm and set of protocols that meet the CORONET restoration-time requirements for each traffic class, as discussed in Section III. Furthermore, there is interdependency between the provisioning and restoration processes that must be specified and coordinated.
- Develop algorithms for network design that determine the placement of network capacity, as well as provide routing guidance to the real-time provisioning and restoration processes. Furthermore, we use the network design process to give an initial assessment of the feasibility of the impact of the CORONET capacity constraints (such as spare capacity ratio and link fiber thresholds) and traffic and class-of-service assumptions (such as dynamic traffic and restoration requirements for IP vs. wavelength services).
- Finally, CORONET real-time provisioning, restoration and network design estimates must be appropriately analyzed via detailed network simulation. This last task occurs in the latter stages of the CORONET project and results will be reported in future publications.

A. *Optical-Layer Structure and Assumptions*

Let us review the critical Optical-Layer structural assumptions.

- The CORONET Optical Layer shown in Figs. 3 and 4 is assumed to have a single AOS at each node. As described in Section II, for the nominal-sized network (20 Tb/s) each AOS-AOS link is allowed to be connected by either one or two fiber pairs, each with 100 wavelengths; furthermore, two fiber pairs are allowed on no more than 10% of the links.
- As detailed in Section III, the connections to be carried on the Optical Layer are bidirectional with bandwidth $=$ $k \cdot \{\text{wavelength rate}\}$, where $k \in \{1, 2, 4, 8\}$.
- There are two types of connections: the static links of the IP network and dynamic wavelength services. In today's commercial IP networks, all network failures are restored in the IP-Layer alone. Thus, IP links are usually provisioned as un-restorable connections in the optical layer.
- Wavelength service connections are constrained to occur among 40 of the 100 nodes; these include all the degree 4/5 nodes and most of the degree 3 nodes. Demands requiring protection from more than one failure are terminated only at nodes with appropriate nodal degree (e.g., demands requiring restoration from three failures must be terminated at each end at nodes of degree ≥ 4).

Although many readers are familiar with optical-layer transmission equipment and terminology, we briefly review our connection model here. A point-to-point, Optical-Layer connection enters/leaves the Optical-Layer via bi-directional *optical transponders (OTs)* at each endpoint AOS. These OTs have a bidirectional drop-side port connecting directly or indirectly to the line-side ports of the IP-Layer routers or to metro/access equipment for wavelength services. On the network side, the generated inter-office signals from each OT traverse a path of AOS-AOS links and are assigned to one channel per link of the path. Each channel is assigned a specific wavelength. Channels are cross-connected at intermediate AOSs. If two channels are optically cross-connected at an intermediate AOS, then they must have the same wavelength; if two channels are cross-connected via an optical–electrical–optical (O-E-O) converter, called a *regenerator*, they may have different wavelengths. Regenerators are needed to: 1) satisfy reach constraints, or 2) do wavelength conversion. A sub-path of channels that are all optically cross-connected (and thus all have the same wavelength) is called a *lightpath*.

TABLE VII
RESTORATION SCENARIOS FOR DYNAMIC WAVELENGTH SERVICES

	Capacity in units of 1000 Wavelength-km.	Normalized Capacity
No Failure Restoration (See Note 1)	2009	100
Single Failure Restoration (See Note 2)	3268	163
Single +Double Failure Restoration (See Note 3)	3493	174
Single + Double + Triple Failure Restoration	3499	174

Note 1: All demands treated as requiring no restoration.
Note 2: Demands requiring double or triple failure restoration treated as requiring only single failure restoration.
Note 3: Demands requiring triple failure restoration treated as requiring only double failure restoration.

B. Real-Time Provisioning and Routing

Call admission, routing, and wavelength assignment can be done under software control and hence can be based on real-time state information, often called a *state-dependent* policy. Important constraints on our design of the wavelength service provisioning process are the following:

- the stringent set-up time for most classes of service (50 ms $+\ 2\times$round-trip fiber transmission delay time);
- restoration: since restorability constitutes part of the QoS of the particular class of service, the admission control policy must consider the ability of the connection to be restored against potential network failure. This decision involves service-path routing, restoration-path routing and availability of idle restoration capacity at the time of connection.
- the CORONET objective for commercial viability, which requires a careful approach to the utilization of complex and novel distributed software.

To address the first objective above, we have developed a novel process for distributed Optical-Layer connection provisioning called 3-*way handshake*. Because link state information quickly becomes stale with short holding time connections, we use the 3-way-handshake signaling protocol to obtain the most current information possible by probing multiple routes for available capacity at connection setup time. It makes three passes: from source to destination for probing, then backwards for cross-connects, and then forward again to tear down un-needed connections. Details can be found in [19]. However, to meet the CORONET capacity constraints, we must ensure that the provisioning and restoration processes are coordinated with the network design process. This is discussed further in the discussion of network design in Section V-D.

C. Optical-Layer Restoration Methodology

Although there are many possible OL restoration approaches from which to choose (for example, see [26], [27], and [28]), the most efficient methods are those where connections that have low probability of failing simultaneously can share the same restoration channels/wavelengths (i.e., wavelengths left unassigned to cover potential network failures) across different potential failures. We refer to such schemes as *shared-mesh restoration*. Numerous studies have demonstrated the capacity efficiency and appropriateness of shared-mesh restoration for optical networks. See for example [20], [21], [23], [24], [30], and [31]. Our goal for this task is to develop a restoration methodology that satisfies the following:

1) the stringent CORONET restoration-time requirements;
2) the stringent restoration network capacity requirements (e.g., the spare capacity ratio); and
3) the DARPA objective of long-term commercial viability.

Because of the richness of previous optical-layer mesh-restoration methods, including many from some of the present authors, we seek to extend existing work on shared mesh restoration to satisfy objectives 1 and 2 above. We are also investigating the potential use of p-cycles [32], [33] that will be reported in later publications. For the third objective, we can leverage our experience with restoration methods and signaling in large-scale commercial deployments of shared-mesh restoration in DCS networks [29].

Perhaps the most challenging aspect of the CORONET vision is emphasized by the fact that even after over a decade of DWDM technology deployment in long distance networks, there is no significant deployment of restoration in any large-scale commercial, DWDM-layer network today. Therefore, we feel a practical Optical-Layer restoration method must be developed to achieve the aforementioned objective of commercial viability.

The approach we suggest modifies and leverages two previous approaches. The first establishes an inventory of cross-connectable optical resources to support shared-mesh restoration [20], [21]. The second specifies a signaling and channel selection algorithm that operates in optical environments and rapidly enough to satisfy the CORONET restoration requirements [34]. Refs. [20] and [21] define a layer of *hot standby* lightpaths which sit idle until needed. The term "hot" implies that there is always an active optical signal between the ends of the lightpath even when it sits idle between failures. When

TABLE VIII
COMPARISON OF CAPACITY REQUIRED FOR DYNAMIC WAVELENGTH AND IP-LAYER SERVICES

	Wavelength Service	**IP Service**
Proportion of total demand	25%	75%
Normalized Capacity (Service Only)	150	100
Normalized Capacity (With Restoration)	262	138

a failure occurs, a connection reroutes over a path of standby lightpaths.

Using hot standbys serves two purposes: 1) it mitigates the effect of optical transients (e.g., see [25]) as many optical connections are rapidly re-configured after a failure and 2) it enables more rapid restoration by reducing the number of real-time routing decisions and intermediate cross-connections. In the CORONET vision optical amplifiers are assumed to be sufficiently advanced to obviate the need for item 1). Hence, the concept of "hot" standby path becomes unnecessary. Thus, we propose a method that modifies the approaches of [20] and [21] to define standby lightpaths as lightpaths with no end O-E-O devices connected. When a network failure occurs, an active connection is rerouted to a sequence of one or more standby lightpaths, $(L_1, \ldots, L_\mathrm{n})$, by tuning the end OTs of the original connection to the wavelength of L_1 at one end and the wavelength of L_n at the other end. Next, each L_k is interconnected to L_{k+1} by cross-connecting their channels via regenerators at their intersecting AOS. We also consider strategies where the cross-connection of L_k to L_{k+1} is optical and only uses a regenerator when needed. Thus, under CORONET assumptions, the major distinction becomes whether regenerators are dedicated to a standby lightpath or shared among multiple standby lightpaths, i.e., only connected when needed for restoration. We use the term "Regen Dedicated Standby Restoration" to designate the former and "Regen Shared Standby Restoration" to designate the latter. We recommend the "Regen Shared Standby Restoration" method because it results in a much lower number of regenerators compared to the former method (see [21]).

The signaling and channel selection algorithm is an extension of the Robust Optical Layer End-to-End X-connection (ROLEX) method, originally established in [34]. Our extension operates by calculating and storing a pre-calculated path of *standby edges* for each connection that is link or node-disjoint from the service path. A standby edge is defined to be a set of co-terminus, unassigned standbys. Recall that each standby may itself route over multiple AOS-AOS links. Note that in the degenerate case, all standbys could be exactly one hop long, in which case the standbys simply represent all individual restoration channels. The restoration paths are calculated and stored at provisioning time or soon thereafter (recall the restoration paths are only needed when failures occur, which generally have long inter-arrival). The restoration paths can also be re-optimized and updated periodically as the network evolves. When a failure occurs, for each failed connection, the method begins signaling from both ends simultaneously to choose specific unassigned standbys to cross-connect. The messages meet at an AOS in the middle and have protocols to resolve contention during channel selection. More details can be found in [35]. The method to choose these pre-calculated restoration paths is described in the next section.

D. Network Design and Capacity Placement

As stated previously, we must design a network that satisfies the CORONET capacity constraints: for example, the spare capacity ratio for the combined IP and Optical-Layer networks cannot exceed 75% and the number of fiber pairs per AOS-AOS link (nominal sized network) is constrained. Therefore, the network design objectives are to place the optical capacity efficiently and ensure that the provisioning and restoration processes use these them efficiently. In particular, in the network design process we have implemented a *routing engine* to determine how service and restoration paths use the optical resources, namely wavelength and OT or regenerator capacity. This same centralized routing engine is used by the provisioning process described earlier in a clever way so that rapid provisioning is achieved, yet complex restoration calculations can be executed and updated as network conditions change. An initial description of the provisioning is described in [19] and further implementation details will be described in future publications.

We have developed and analyzed various initial routing strategies for the Routing Engine. The preliminary routing method we use is a *static* path-selection algorithm for the service and restoration paths. This is a simple strategy that does not change routing based on network state. For each connection request with a class of service that is k-restorable ($k = 1, 2, 3$), the static path-selection algorithm generates paths consistent with the diversity conditions stated in Section II; namely, it selects the set of $(k+1)$-diversely routed paths that have the minimum total distance over all such sets. From this set, the shortest path is chosen as the working path and the second shortest path is chosen as the first restoration (backup) path. If $k \geq 2$, the third shortest path is chosen as the second restoration path to cover dual-failures. The forth shortest path is chosen as the third restoration path if $k = 3$. We are also investigating state-dependent routing algorithms which may provide more efficient use of capacity. Results of this work will be described in future publications.

The network design process uses the routing engine and the simulation stream of dynamic wavelength services described in Section III to compute network capacity requirements over the planning horizon (sometimes called *dimensioning* the network). It begins by estimating an upper bound on network capacity for each link using Erlang approximations. Then it uniformly reduces link capacity over multiple simulation runs until QoS thresholds are encountered. A more sophisticated stochastic optimization approach uses the simulation streams to gradually build a piecewise-linear envelope (lower bound approximation) of the expected QoS stochastic function. Then, more rapid link-dimensioning can be executed on the piecewise-linear envelope rather than repeating entire simulation streams for every link capacity scenario. However, this latter approach is still under development.

We applied our routing engine and capacity estimation algorithms to compare the impact on network capacity of restoration against multiple failures. This is illustrated in Table VII (analogous to Table III for the IP Layer). Compared to the "No Failure Restoration" case, restoration of all connections against a single node or link failure increases capacity requirements by 63%. Requiring 10% of wavelength service connections to belong to a class of service that is restorable against two failures (one of which can be a node failure) increases the capacity by 72% over the no-failure case but only an increase of 7% capacity over the single-failure case. If 2.5% of wavelength service connections must be restorable against three failures (one of which can be a node failure), there is a negligible increase in capacity over the two-failure case. The principal reason for these interesting results is because the number of single-failure-restorable connections dominates the restoration capacity requirements, yet they do not have to be restored during multiple failure events. Thus, during multiple-failure events the multiple-failure-restorable connections have first crack at the extra bandwidth placed for single-failure events.

Finally, we use the routing engine and capacity algorithm to contrast the capacity requirements between wavelength and IP services. Table VIII shows that although IP-Layer traffic is 75% of total network traffic bandwidth (measured by total expected load), it require only 40% of total capacity in the network without restoration capacity (in unites of wavelength-kilometers). It only requires 35% of total capacity in the network with restoration capacity.

Some of the reasons for this disproportional difference in required capacity between IP Layer and Optical Layer are as follows:

- the much larger granularity of the wavelength demands. Consider a simple queueing example: Suppose a single link has an expected load of 200 Gbps (5×40 Gbps wavelength). To achieve 10^{-4} blocking would require 32 wavelengths of capacity (16% utilization); however, if the same 200 Gbps load was divided into 10 Gbps connections (the size of the largest IP demand) less than half as much capacity would be required to achieve the same blocking;
- the small expected loads between individual node-pairs for wavelengths services, most of which are below <1 over the simulation period; IP best effort traffic, which is 40% of the IP bandwidth requirement, does not require restoration.

VI. Conclusion and Future Studies

A major element of future work in the CORONET program will be the simulation of all IP and OL provisioning and restoration algorithms on the worldwide network described in Section II at traffic levels of 20–100 Tbps. These simulations will be used to validate the performance of the algorithms and protocols outlined above against CORONET performance metrics, and to gain insight into a highly dynamic network environment.

Another important element of future work in dynamic networking is the investigation of cross-layer interactions, for example in traffic engineering or restoration. As an example of the latter, as we mentioned previously, in today's commercial IP networks, IP links are usually provisioned as un-restorable connections in the Optical Layer. However, as we explore the failure state space in more detail, we find opportunities to share Optical-Layer restoration capacity among non-simultaneous failures between the two layers. We propose to further investigate a suite of innovative, yet practical methods [22], [23], and [24] which exploit the ability of a dynamic Optical Layer to make rapid connections. We will perform cost studies to quantify possible advantages of such integrated restoration methods.

References

[1] L. Ong, J. Verrant, T. Lehman, J. Sobieski, and R. Summerhill, "Application of ASON multi-domain architecture to Internet2," presented at the Global Interoperability in Multi-Domain and Multi-Layer ASON/GMPLS Networks Workshop, ECOC 2007, Berlin, Germany, September 16, 2007.

[2] Dynamic Resource Allocation via GMPLS Optical Networks (DRAGON) Project. [Online]. Available: http://www.dragon.east.isi.edu/

[3] V. Maglaris, "What should come after GÉANT2," presented at the NORDUnet Conf., Helsinki, Finland, Apr. 11, 2008. [Online]. Available: http://www.geant2.net/upload/pdf/Vasilis_Maglaris.pdf

[4] I. Yamauchi, T. Miyake, Y. Mikamo, J. Shimada, K. Kobayashi, and H. Esaki, "JGN II (Japan Gigabit Network II): A research and development system for advanced broadband networks," in *Proc. Symp. Applications and the Internet Workshops (SAINT 2005 Workshops)*, 2005, pp. 34–37 [Online]. Available: http://www.jgn.nict.go.jp/jgn2plus/english/index.html

[5] N. Taesombut, F. Uyeda, L. Smarr, T. A. DeFanti, P. Papadopoulos, and J. Leigh *et al.*, "The OptIPuter: High-performance, QoS-guaranteed network service for emerging E-science applications," *IEEE Communications*, vol. 44, no. 5, pp. 38–45, May 2006.

[6] "Dynamic multi-terabit core optical networks: architecture, protocols, control and management (CORONET)," DARPA, BAA06-29 [Online]. Available: www.darpa.mil/STO/Solicitations/CORONET/index.htm

[7] M. D. Feuer, D. C. Kilper, and S. L. Woodward, "ROADMS and their system applications," in *Optical Fiber Telecommunications VB*, I. P. Kaminow, T. Li, and A. Willner, Eds. New York: Academic Press, 2008, ch. 8.

[8] R. Doverspike, "Practical aspects of bandwidth-on-demand in optical networks," presented at the Panel on Emerging Networks Service Provider Summit, Optical Fiber Communication Conf. (OFC), Anaheim, CA, Mar. 2007.

[9] A. Nucci, N. Taft, C. Barakat, and P. Thiran, "Controlled use of excess backbone bandwidth for providing new services in IP-over-WDM networks," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 9, pp. 1692–1707, Nov. 2004.

[10] J. Fang, M. Sivakumar, A. Somani, and K. Sivalingam, "On partial protection in groomed optical WDM mesh networks," in *Proc. IEEE Int. Conf. Dependable Systems and Networks*, Yokohama, Japan, Jun. 28–Jul. 1, 2005, pp. 228–237.

[11] K. Ratnam, M. Gurusamy, and L. Zhou, "Differentiated QoS routing of restorable sub-lambda connections in IP-over-WDM networks using a multi-layer protection approach," in *Proc. Int. Conf. Broadband Networks*, Oct. 2005, vol. 1, pp. 127–136.

[12] L. Song, J. Zhang, and B. Mukherjee, "Dynamic provisioning with availability guarantee for differentiated services in survivable mesh networks," *IEEE J. Sel. Areas Commun. (OCN Supplement)*, vol. 25, no. 4, pp. 32–44, Apr. 2007.

[13] K. Ratnam, M. Gurusamy, and L. Zhou, "Differentiated survivability framework and modeling for heterogeneous grooming optical networks," in *IEEE GLOBECOM 2007*, Washington, DC.

[14] K. Ratnam, L. Zhou, and M. Gurusamy, "Efficient multi-layer operational strategies for survivable IP-over-WDM networks," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 16–31, Aug. 2006, Part Supplement.

[15] A. Chiu, G. Choudhury, G. Clapp, R. Doverspike, J. Gannett, J. Klincewicz, G. Li, R. Skoog, J. Strand, A. Von Lehmen, and D. Xu, "Network design and architectures for highly dynamic next-generation IP-over-optical long distance networks," presented at the Optical Fiber Communication Conf. (OFC), San Diego, CA, Mar. 2009.

[16] R. Ahuja, T. Magnanti, and J. Orlin, *Network Flows: Theory, Algorithms, and Applications*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[17] A. Chiu, G. Choudhury, R. Doverspike, and G. Li, "Restoration design in IP over reconfigurable all-optical networks," in *Proc. IFIP Int. Conf. Network and Parallel Computing*, Dalian, China, Sep. 18–21, 2007, pp. 315–333.

[18] G. Choudhury, "Models for IP/MPLS routing performance: Convergence, fast reroute, and QoS impact," in *Keynote Presentation, SPIE Conf. Performance, QoS and Control of Next-Generation Communication Networks (ITCOM)*, Philadelphia, PA, Oct. 2004, vol. 5598.

[19] R. Skoog and A. Neidhardt, "A fast robust signaling protocol for enabling highly dynamic optical networks," presented at the Optical Fiber Communication Conf. (OFC), San Diego, CA, Mar. 2009.

[20] J. Strand and A. Chiu, "Realizing the advantages of optical reconfigurability and restoration with integrated optical cross-connects," *J. Lightw. Technol.*, vol. 21, no. 11, pp. 2871–2882, Nov. 2003.

[21] G. Li, A. Chiu, and J. Strand, "Resilience design in all-optical ultra long-haul networks," *J. Opt. Netw.*, vol. 5, no. 8, pp. 625–636, Jul. 2006.

[22] A. Chiu and J. Strand, "Joint IP/optical layer restoration after a router failure," presented at the Optical Fiber Communication Conf. (OFC), Anaheim, CA, Mar. 2001.

[23] P. Sebos, J. Yates, G. Li, D. Wang, A. Greenberg, M. Lazer, C. Kalmanek, and D. Rubenstein, "Demonstration of ultra-fast IP restoration using reconfigurable transport networks," presented at the Optical Fiber Communication Conf. (OFC), Atlanta, GA, Mar. 2003.

[24] G. Li, D. Wang, R. Doverspike, C. Kalmanek, and J. Yates, "Economic analysis of IP/optical network architectures," presented at the Optical Fiber Communication Conf. (OFC), Los Angeles, CA, Mar. 2004.

[25] J. Strand and A. Chiu, "Impairments and other constraints on optical layer routing," Internet Engineering Task Force, RFC 4054, May 2005.

[26] J. Strand, "Optical network architecture evolution," in *Optical Fiber Telecommunications IVB*, I. Kaminow and T. Li, Eds. New York: Academic Press, 2002.

[27] G. Ellinas *et al.*, "Routing and restoration architectures in mesh optical networks," *Optical Networks Mag.*, vol. 4, no. 1, pp. 91–106, Jan. 2003.

[28] W. Grover, *Mesh-Based Survivable Transport Networks: Options and Strategies for Optical, MPLS, SONET and ATM Networking*. Upper Saddle River, NJ: Prentice-Hall, 2003.

[29] B. Cortez, "The emerging intelligent optical network: Now a reality," presented at the Optical Fiber Communication Conf. (OFC), Anaheim, CA, Mar. 2002.

[30] G. Li, A. Chiu, and J. Strand, "Failure recovery in all-optical networks," presented at the 5th Int. Workshop on Design of Reliable Communication Networks (DRCN2005), Island of Ischia, Italy, Oct. 16–19, 2005.

[31] J. Simmons, "Cost vs capacity tradeoff with shared mesh protection in optical-bypass-enabled backbone networks," presented at the Optical Fiber Communication Conf. (OFC), Anaheim, CA, Mar. 2007.

[32] A. Kodian and W. D. Grover, "Failure-independent path protecting P-cycles; efficient and simple fully pre-connected optical path protection," *J. Lightw. Technol.*, vol. 23, no. 10, pp. 3241–3259, Oct. 2005.

[33] H. Luss and R. T. Wong, "Survivable telecommunications network design under different types of failures," *IEEE Trans. Syst., Man, Cybern. Part A, Syst. Humans*, vol. 34, no. 4, pp. 521–530, Jul. 2004.

[34] R. Doverspike, G. Sahin, J. Strand, and R. Tkach, "Fast restoration in a mesh network of optical cross-connects," presented at the Optical Fiber Communication Conf. (OFC), San Diego, CA, Feb. 1999.

[35] A. Chiu, R. Doverspike, G. Li, and J. Strand, "Restoration signaling protocol design for next-generation optical network," presented at the Optical Fiber Communication Conf. (OFC), San Diego, CA, Mar. 2009.

Angela L. Chiu (SM'07) received the S.M and Ph.D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, MA.

She is a Principal Member of Technical Staff with AT&T Labs–Research. Her research interests include architecture, design, and analysis of IP and optical networks. During her time at AT&T, she has led the IP QoS and MPLS architecture and design for AT&T IP backbone.

Gagan Choudhury (M'81–SM'05–F'09) received the Ph.D. degree in electrical engineering from the State University of New York at Stony Brook.

He is a Lead Member of Technical Staff with AT&T Labs–Research. His research interests include analysis and robust design of large IP/Optical networks, data plane and control plane modeling of unicast and multicast routing protocols, queueing theory, and algorithms for multidimensional transform inversion.

George Clapp is a Chief Scientist in Telcordia's research group in network architectures and the management of IP/optical networks. He leads research in optical network design and the interworking of IP/MPLS networks with dynamic optical services. He previously worked at Ameritech, where he initiated Ameritech's Internet service and led the creation of a Network Access Point for the National Science Foundation. He has also chaired IETF working groups that defined the transport of IP over public data services.

Robert Doverspike (M'92–SM'98) received the Ph.D. degree in mathematics from Rensselaer Polytechnic Institute, Troy, NY.

He is Executive Director of Network Evolution Research at AT&T Labs–Research. His research interests include optimization and restoration in multi-layered transmission and switching networks, packet transport in metro and long distance networks, and advanced interactions of transport and IP network architectures.

Joel W. Gannett (M'80–SM'91) received the Ph.D. degree in electrical engineering and computer science from the University of California at Berkeley.

He is a Senior Scientist in the Network Operations and Management Research Department at Telcordia Technologies, Red Bank, NJ. He has extensive experience in simulation performance modeling and network design optimization for optical and wireless networks.

John G. Klincewicz received the S.B. degree in mathematics from the Massachusetts Institute of Technology (MIT), Cambridge, and the Ph.D. degree in operations research from Yale University, New Haven, CT.

He is a Principal Member of Technical Staff in the Network Design and Performance Analysis Department of AT&T Labs–Research. Since joining AT&T in 1979, he has worked on a variety of applications, including IP backbone network design, cache server location, ring network planning, capacity expansion and logistics. His research interests include telecommunication network design, facility location models and heuristics for combinatorial problems. He has over 20 refereed publications and is a member of INFORMS.

Guangzhi Li (M'04) received the Ph.D. degree in computer science from the College of William and Mary, Williamsburg, VA.

He is a Principal Member of Technical Staff at AT&T Labs–Research. His research interests include IP-based control plane for optical networks, restoration and protection schemes and algorithms, network simulation and performance evaluation, and network-related applications.

Ronald A. Skoog (M'89) received the B.S. degree from Oregon State University, Corvallis, and the M.S. and Ph.D. degrees in electrical engineering (control and systems theory) from the Massachusetts Institute of Technology (MIT), Cambridge.

He is a Senior Scientist at Telcordia Technologies, Inc. Prior to joining Telcordia, he spent 29 years at Bell Laboratories/AT&T Bell Laboratories/AT&T Laboratories working in the areas of transport network design, signaling network design, protocols, and performance/reliability studies. He has worked in the areas of optical networking architectures, Internet protocol/wavelength-division-multiplexing network architectures and evolution studies, optical network management and control, emerging network technology studies and reliability studies for optical networks and optical network elements.

John Strand (SM'02) received the Ph.D. degree in mathematics from the University of California at Berkeley, and degrees in economics from Harvard and Georgetown.

He is a consultant to AT&T Labs–Research. His research interests include both theoretical and practical aspects of transport network architecture and planning.

Ann Von Lehmen (M'91–SM'06–F'08) received the Ph.D. degree in applied physics from Cornell University, Ithaca, NY.

She is Executive Director of the Broadband Networking Research Department at Telcordia Technologies, Red Bank, NJ. Her current research interests are in the area of optical networking include control and management for dynamic IP/optical networks.

Dahai Xu (M'04) received the B.Eng. degree in applied electronics in 1996, the M.Eng. degree in computer science and engineering from Shanghai Jiao Tong University, China, in 1999, and the Ph.D. degree in computer science and engineering from the State University of New York at Buffalo in 2005.

He is currently a Senior Member of Technical Staff at AT&T Labs–Research. His research interests include Internet design, control and management, algorithm design, and fast implementation, large-scale nonlinear network optimization, and secure communication in wireless ad hoc networks.