DYNAMIC PROVISIONING SYSTEM FOR BANDWIDTH-SCALABLE CORE OPTICAL NETWORK

Kristin Rauschenbach, Regina Hain, Alden Jackson, John Jacob, Will Leland, John Lowry, Walter Milliken, Partha Pal, Ram Ramanathan, Cesar Santivanez
BBN Technologies
10 Moulton St.
Cambridge, MA 02139
krausche@bbn.com
and
Ilia Baldine, Shu Huang
RENCI
Chapel Hill, NC 27517
and
Dan Wood
Verizon Federal Network Systems
Burlington, MA 01803

ABSTRACT

We describe the architecture of PHAROS (Petabit Highly-Agile Robust Optical System), developed under the DARPA CORONET program. PHAROS provides traffic engineering, resource management and signaling solutions for highly-agile, large-capacity core optical networks. PHAROS technology facilitates rapid configuration of network resources to address dynamic traffic needs in future global military and commercial communications, such as localized surges in capacity requirements that result from military operations. PHAROS technology also scales to support bandwidth-intensive, network-centric, collaborative and distributed computing applications, and accommodates the continued growth of video and biometric data services.

I. INTRODUCTION

The core networks supporting the security and defense of our nation must sustain mission-critical traffic despite unprecedented demands for agility and resiliency: DoD no longer has the luxury of relying on network architectures and operations that assume semi-permanent traffic and semi-permanent network resources, where changes in traffic or topology can consume weeks of attention from human experts and network management systems are tied tightly to specific existing network technologies. This situation is already urgent for our national defense, and increasingly pressing for commercial core networks. In response, PHAROS (the Petabit Highly Agile Robust Optical System) has been developed under the DARPA CORONET program to combine agility, efficiency, and resiliency for global-scale terrestrial core networks. PHAROS represents a rethinking of core network architectures to overcome the limitations of existing approaches, fulfill current and future requirements, and leverage emerging technologies.

The PHAROS architecture has been designed with awareness of current commercial core network practice and the practical constraints on core network evolution. It addresses the broad suite of challenges required to meet the critical national need, including the separation of services from transport, fostering new information-enabled applications and service-provider business models. In particular, it meets or exceeds the stringent quantitative performance metrics that CORONET set out, including: network scale >100Tb/s; ratio of reserved protection capacity to provisioned network capacity of less than 0.75; very fast service setup <50 ms + roundtrip time; transport protection from three network element failures; and control and management plane robustness to three simultaneous failures. The design of PHAROS, however,

The research reported in this document was performed in connection with contract number N00173-08-C-2010 with the Defense Advanced Research Projects Agency (DARPA). The views and conclusions contained in this document are those of the authors and should not be interpreted as presenting the official policies or position, either expressed or implied, of DARPA or the U.S. Government unless so designated by other authorized documents. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

goes beyond these specific metrics for symmetric point-to-point intra-domain service to also support asymmetric demands, multicast communications, and cross-domain services. (A *domain* is a network or set of networks under common administrative control.). Thus, PHAROS is the first attempt to provide control and management solutions that support services across global core network dimension with 50-ms-class setup time, and also to respond to multiple network failures in this time frame.

The CORONET program aims to build upon recent research that point to the power of grooming to maximize optical bypass to reduce core network costs. [1-8] The program also exploits the use of optical reconfiguration to provide bandwidth-efficient network equipage that responds gracefully to traffic changes and unexpected network outages. [9-11]

The remainder of this paper is organized as follows. After surveying background work, we begin by highlighting the unique and innovative features that PHAROS incorporates to meet the requirements of the next generation core optical networks. Following that we describe three key components of PHAROS – the cross-layer allocation algorithm, signaling system, and the core node implementation. Finally, we give some preliminary results on performance estimation. We note that PHAROS has many other interesting features not described here due to lack of space.

II. RELATED WORK

We briefly survey prior work on some of the topics discussed in this paper, namely, path computation, protection, and node architectures.

Unlike IP networks, path computation in optical networks involves bi-path computation – working and protection. Approaches can be classified by the nature of the required paths (e.g., node-disjoint, link-disjoint, k-shortest), the order for computing them (e.g., primary-then- protection vs joint-selection), and the cost associated with each path. Some works include [12,13]. Our approach is a hybrid one and uses the concept of joint or shared protection.

The various levels of protection defined for different traffic demands in a core optical network, along with the low-backup-capacity targets, motivate the use of shared-protection schemes for this application. Such techniques fall into broad categories of the various computational and graph-theoretic approaches: constrained shortest paths [14], cycle covers [15], and ILP formulations like the p-cycles [16]. As these techniques can guarantee only single-protection for all the flows, they will have to be augmented to guarantee double or triple protection for the set of flows that require it. In this paper, we have outlined the preliminary formulation of a shared-mesh-protection algorithm based on virtual links and “jointly protected sets” that meets the double- and triple-protection requirements.

The sophistication of optical-network-node architectures has risen as the state of the art for the optical components within these nodes has advanced. Recent advances in optical-switch reliability and functionality, along with the size of the available switch fabrics, have motivated node architectures that allow such multiple functionalities as reconfigurable add/drop, regeneration, and wavelength conversion [17]. The cost, power, size, and reliability calculations for these different implementations are highly technology-dependent and are changing rapidly as new technologies are transitioned into the commercial market. As a result of this rapidly changing trade-space, we have chosen to remain agnostic to the exact switch architecture in our nodes, a feature we discuss further in the next section.

III. INNOVATIVE FEATURES

A basic tenet of the PHAROS architecture is a *technology-agnostic design* that maximizes bypass to achieve lower cost-per-bit core network services and accommodates future generations of switch technology for long-term graceful capacity scaling. Current systems employ some degree of abstraction in managing network resources, using interface adapters that expose a suite of high-level parameters describing the functionality of a node. Such adapters, however, run the twin risks of obscuring key blocking and contention constraints for a specific node implementation, and/or tying their interfaces (and the system’s resource management algorithms) too tightly to a given technology. The PHAROS system avoids both of these problems by using abstract topological representations for all levels of the network. The representations extend down to an abstract network model of the essential contention structure of a node, as illustrated in Figure 1 and extend upward to address successive (virtual) levels of functionality across the entire network.

With a uniform approach, common to all levels of resource representation and allocation, PHAROS accurately exploits the capabilities of all network elements, while remaining independent of the switching technology. At the signaling and control level, the PHAROS architecture also provides a set of common mechanisms for its own internal management functions (such as verification and failover); these mechanisms provide significant architectural immunity to changes in the technologies used in implementing specific PHAROS functional components.

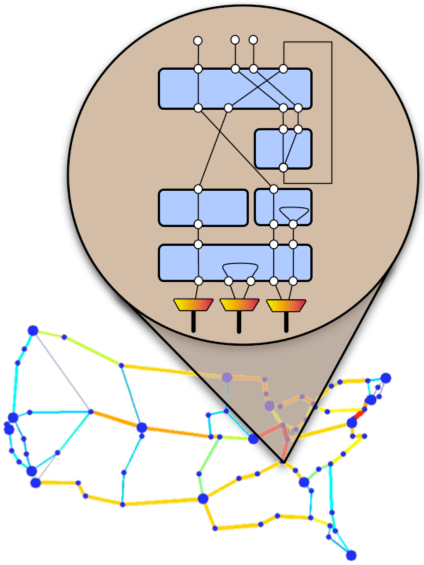

Figure 1. Technology agnosticism through unified multi-level topology abstraction.

The PHAROS architecture uses *multilevel topological abstractions* to achieve efficient integrated resource optimization over the fundamental dimensions of network management: network extent, technology levels, route protection, and timescales. A combination of abstraction and scoping allows a given request to be optimized across the network, simultaneously trading off costs of resources within individual network levels as well as the costs of transit between levels (such as the optical-electrical boundary). Resources of all levels can be considered, including timeslots, wavelengths, grooming ports, and IP capacity. PHAROS optimization unites analysis of the resources needed to deliver the service with any resources required for protection against network element failures. Protection resources (at all levels) are allocated in conjunction with the resources required by other demands and their protection, achieving dramatic reductions in the total resources required for protection (the CORONET B/W metric). Our optimization design allows PHAROS to unify the handling of demand timescales, exploiting current, historical, and predicted future resource availability and consumption. Timescales are also addressed by the overall PHAROS resource management strategy, which selects mechanisms to support available time constraints: for examples, PHAROS employs pre-calculation and tailored signaling strategies for very fast service setup; uses carefully selected topology abstractions to perform more-extensive on-demand optimization where feasible; and evaluates long-term performance out of the critical path to enable rebalancing and improve the efficiency of the on-demand optimizations.

To realize robust, efficient global optimization, the PHAROS architecture adopts a strategy we term *unitary resource management*. This strategy enables PHAROS to autonomously maintain the following three properties across time and network evolution: 1) the integrated cross-layer-resource allocation algorithm is sustained by a resilient hierarchy of cross-layer resource allocator (CRA) instances; 2) for each request for a given combination of service class, source, and destination there is exactly one CRA instance responsible at any time; and 3) for each network resource there is exactly one CRA instance controlling its allocation at any time. An instance in the PHAROS architecture is an individual process running on a physical node and executing the code for a PHAROS function (such as the CRA function). Long-term resource optimization is managed via delegation of resources between CRAs. Unitary management eliminates negotiation, backtracking, and thrashing in responding to service requests, ensuring that PHAROS service setup times are highly deterministic and consistently rapid.

Finally, the PHAROS architecture makes use of a design construct that combines redundancy and cross-checking in a flexible way to mitigate single point of failure and corrupt behavior in a CRA. This design construct, we refer to as *triangulation*, pairs up the consumer of the CRA function (typically a network element controller) with a primary and a verify CRA. The verify CRA checks that the primary CRA is performing correctly, corrupt behavior can be detected by using appropriate protocols amongst the consumer and the primary and verify CRAs.

IV. RESOURCE ALLOCATION

A key architectural decision in any communications network is the organization of the control of resources. Two of the most important aspects are whether global state or just local state is tracked, and how many nodes participate. Based on these and other choices, approaches range from *fully distributed* – where each node participates using local information, and *fully centralized* – where resource control is in the hands of a single node utilizing global information.

The CORONETchallenge presents some unique factors influencing our choice of PHAROS control organization. First, there is adequate signaling bandwidth and processing resources available, which allow for global tracking of resource use if necessary. Second, nodes are neither mobile nor disruption prone, again making it feasible to concentrate control functionality. Third, under high loads, efficient (preferably optimal) allocation is required. Fourth, the stringent service requirements and expectations make the user of the core optical system highly intolerant of stability issues.

We believe that these factors shift the optimum point significantly toward a *centralized* control for PHAROS[2] although not completely. In essence, our approach is to move away from a single point of failure but retain the ability to use global information for resource allocation decisions – resulting in the *unitary* resource management scheme outline in section III. We term our functional module for this purpose the *cross-layer Resource Allocation (CRA).* The CRA uses path-based restoration with shared protection, and playbooks for very fast service setup, and aggregation and grooming services. We describe each of these below.

PHAROS uses the concept of *joint or shared protection,* which significantly reduces the total amount of network resources reserved for protection while providing equal assurance of path restoration after failures. For a given failure or set of failures, only some primary paths are affected, and only some of their protection paths (in the case of multiple failures) are affected. Thus, a protection resource can be reserved for use by an entire set of protection paths if none of the failures under consideration can simultaneously require use of that resource by more than one path in the set. We state our notion more formally below.

Jointly protected set JPS: A set of (primary) paths $\{P_i\}$ is said to be jointly protected if the intersection of all their interiors is not empty, that is, there is at least one link or node that is in the interior of every path in $\{P_i\}$. All paths in a JPS will simultaneously fail if that shared link or node fails.

- $J(P^L)$: The set of all jointly protected sets in P^L.
- A JPS J_i in $J(P^L)$ is *maximal* if there is no other JPS in $J(P^L)$ that properly includes J_i. In other words, there is no path in P^L that can be added to J_i and still have a JPS.
- Mx(J) : The set of all maximal jointly protected sets in J.
- The spare capacity *C* that must be reserved at L due to P^L can be computed as:

$$C(P^L) = \underset{J_i \text{ in } Mx(J(P^L))}{\text{Max}} \quad \sum_{P_i \text{ in } J_i} \text{Demand}(P_i)$$

Intuitively, since every path in a JPS can fail simultaneously due to some single node or link failure, L must have enough capacity to support all paths simultaneously. No single link or interior node can belong to more than one maximal JPS (if it did, they would not be maximal: their union would be non-empty and would properly include both of them). So if L has enough capacity to protect against the sum of the paths in any one maximal JPS, for any maximal JPS drawn from P^L, then L can protect all its protected paths against any single failure. The corresponding result for the capacity needed at L for multicast protection is determined by the same formula, replacing "path" with "tree."

One significant contribution to agility in the PHAROS architecture is a strategy we term *playbooks*. A playbook is a set of pre-calculated alternatives for an action (such as selecting a protection path) that has a tight time budget. The playbook is calculated from the critical path for that action using the CRA function's global knowledge and optimization algorithms. The playbook is stored on each instance that must perform the action; on demand, each instance then makes a fast dynamic selection from among the playbook's alternatives. Playbooks are used to ensure fast, efficient resource use when the time constraints on an action do not allow the computation of paths on demand. In the PHAROS architecture, we use playbooks in two situations.

- *Very fast service setup (VFSS) demands*. For each (source, destination, demand rate), we pre-compute and store several *bi-paths* (a pair of primary and protection path). When an actual demand arrives, the current cost of the stored bi-paths – based on current resource availability – is computed and the lowest-cost chosen.
- *Restoration upon failure*. For each existing demand, there is a playbook entry specifying the path (or paths, for doubly and triply protection demands) to use in case the primary path fails. Each entry specifies the path and regeneration and grooming strategies, and identifies the pool of resources (such as wavelengths) to choose from upon failure. The playbook doesn't specify the resource to use, as such assignment can be made (efficiently under shared protection) only after a failure occurs

Finally, many demands do not fill a full wavelength. If one such demand is uniquely assigned to a full wavelength, without sharing it with other demands, it will result in wasting bandwidth and long-reach transponders. To alleviate this problem, demands can be aggregated into larger flows at the source node. They can also be combined with other nodes' demands at intermediate nodes (a process we refer to as *sub-lambda grooming*, or SLG) so that wavelength utilization at the core is close to 100%. Once demands are sub-lambda-groomed, they can be optically bypassed.

[2] This is in contrast to other scenarios, for example mobile ad hoc networks where the influencing factors are significantly different and favor a more distributed solution.

Within our topology abstraction based architecture, grooming is a generalized operation where each level "packs" its smaller bins into larger bins at the level immediately below. Currently, we have a three-level system where we aggregate and groom sub-lambda demands into full wavelengths, and wavelengths onto fibers. However, aggregation and grooming of smaller bins into larger bins are a fundamental operation that repeats itself at multiple layers.

V. SIGNALING SYSTEM

The PHAROS signaling architecture is designed to support operations in the control as well as management planes. Its function is the delivery of data between the elements of the architecture in a timely, resilient, and secure fashion. The main requirements for the signaling architecture are:

Performance: the architecture must support the stringent timing requirements to connection setup and failure restoration.

Resiliency: the architecture must be resilient to simultaneous failures of several elements and still be able to perform the most critical functions.

Security: the architecture must support flexible security arrangements among architectural elements to allow for proper authentication, non-repudiation and encryption of messages between them.

Extensibility: the architecture must be extensible to be able to accommodate new features and support the evolution of the PHAROS architecture.

The PHAROS signaling and control network (SCN) is the implementation of the PHAROS signaling architecture. It represents a topological overlay over the fiber-span topology, with signaling links segregated from the data plane to minimize the risk of resource exhaustion and interference attacks. For CORONET, providing deterministic and minimal delay in signaling for service setup and fault recovery is paramount. Therefore, our architecture supports the case where the topology of the SCN is mesh-isomorphic to the fiber-span topology, as shown in Figure 2. This approach eliminates additional hops in the signaling plane that cause delays in delivering signaling messages between architectural elements and simplifies the routing of signaling messages. The PHAROS architecture also supports more traditional GMPLS solutions where typically the signaling and the data plane topologies are different. Also, while there is not an architectural requirement that the network element controllers and network elements be co-located, this arrangement minimizes signaling delay.

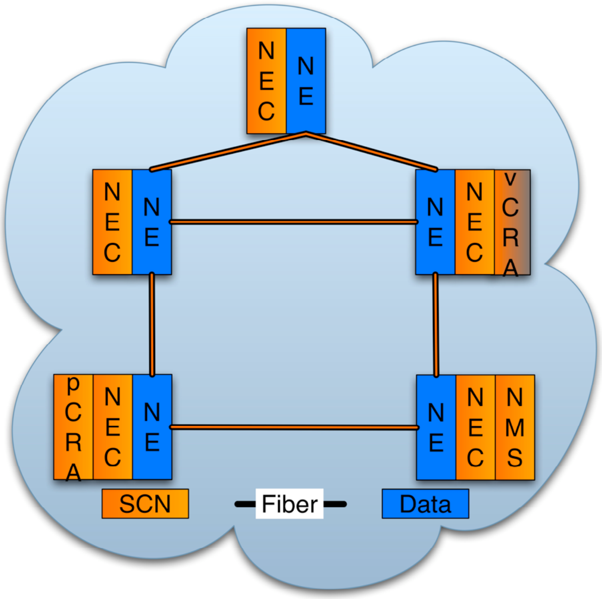

Figure 2. The Signaling and Control Network (SCN) connects network elements (NE) and their associated network element controllers (NEC), cross-layer resource allocator (CRA) and network management system (NMS).

Based on bandwidth sizing estimates that take into account messaging requirements for connection setup, failure signaling and resource assignment, a 1 Gb/s channel is sufficient to maintain stringent timing for set up and restoration under heavy load and/or recovery from multiple fault scenarios. Two performance goals drive the channel size requirements for the PHAROS SCN, very fast service setup and 50-ms-class restoration from simultaneous failure. The sizing estimates assume "worst case" signaling load for a 50-Tb/s-capacity 100-node global fiber network with service granularity ranging from 10 Mb/s to 800 Gb/s. Fibers connecting nodes were presumed to carry one hundred 100-Gb/s wavelengths.

VI. CORE NODE IMPLEMENTATION

The PHAROS core node design focuses on maximizing flexibility and minimizing the complexity of intra-node ports required to provide the complete range of PHAROS services and reducing the capital and operational costs per unit of bits. The primary objectives identified to satisfy this vision include: 1) arrange subscriber traffic onto wavelength and sub-wavelength paths to enable switching at the most economic layer, 2) enable shared protection and 3) enable transponders to be repurposed to service both IP and wavelength services and also service transit OEO (optical-electrical-optical) regeneration functions. When combined with a control plane designed for optimum resource allocation, the PHAROS optical node is highly adaptable to incoming service requests.

The PHAROS node architecture defines the principal hardware systems extending from the fiber connections with the subscriber facility to the fiber connections in the physical plant of the core network, as illustrated in Figure 3.

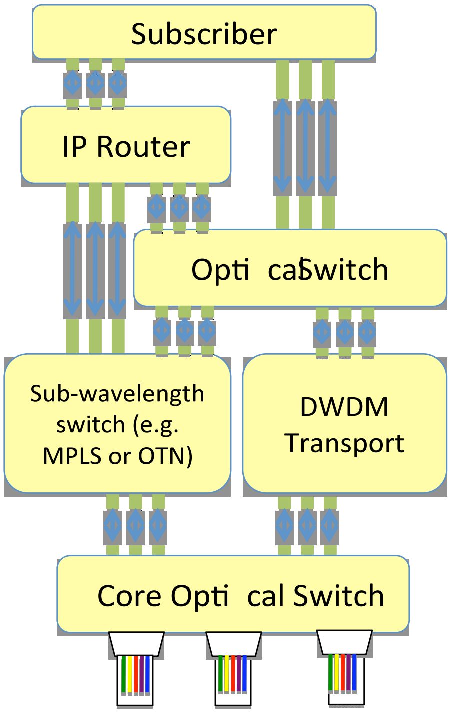

Figure 3. PHAROS core node implementation showing various optical network elements.

The PHAROS node is composed of the following elements:

- Subscriber service layer connections to bring client services into the core node.
- Edge router (packet switch) to support best effort IP services.
- Fast optical switch to allow sharing of sub-wavelength switch and transport ports.
- Sub-lambda grooming switch and DWDM transport platform to support full and sub-wavelength switched (via MPLS, OTN or SONET) and packet services with fast setup, tightly bounded jitter specifications. This equipment also provides OEO regeneration.

Core optical switch to manage optical bypass, optical add/drop, and routing between optical fibers.Note that these elements may or may not be instantiated in the same hardware platform. The PHAROS architecture emphasizes configuration, and can be applied to a variety of different network element configurations.

VII. PRELIMINARY PERFORMANCE ANALYSIS

We have created a high fidelity OPNET simulation of the PHAROS system and are currently evaluating the performance. Figure 4 compares the performance of three protection approaches: 1) *dedicated* protection in which each primary path receives its own protection path; 2) *shared* protection, where a set of protection paths may share a resource as explained in section IV; 3) *opportunistic* shared protection, a sophisticated version of (2) where the protection paths are chosen to maximize shared protection opportunities.

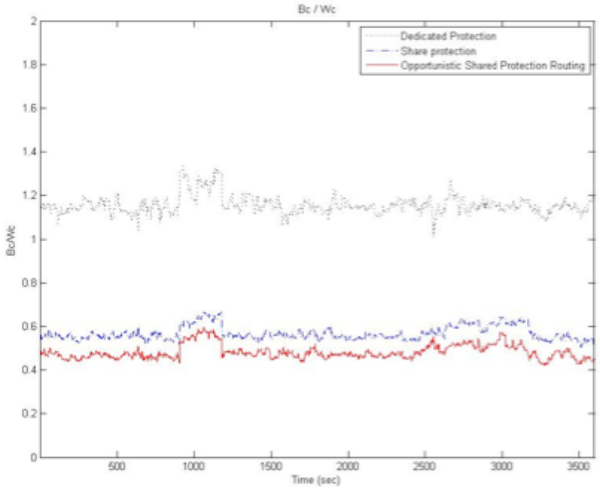

Figure 4. B/W comparison of different protection strategies

Requests for bandwidth are generated over time. For each approach, we plot the *B/W* metric as a function of time. B/W is defined as the Backup (Protection) over Working capacity, which is a rough measure of the relative cost incurred in protection. Thus, lower the B/W, the better.

Results shown here are for a 100 node optical network model with 75 nodes in CONUS, 15 in Europe and 10 in Asia. The line rate is 40 Gbps, and the traffic 20 terabits of which 35% is IP traffic and 65% is wavelength services. 90% of the source-destination pairs are within the U.S. The bit-averaged distance for the intra-CONUS traffic is about 1808 km. The B/W numbers shown in Figure 4 are for CONUS-contained resources only.

We see that the PHAROS shared protection strategies significantly outperform dedicated protection. Specifically, shared protection has about a 50% lower B/W than dedicated, and opportunistic improves this further by about 10%.

VIII. CONCLUSION

In this paper we have described the architecture of a future core network control and management system along with a node implementation that enables future scalable and agile optical networks developed as part of the DARPA/STO CORONET program. This work represents the first attempt to provide control and management solutions that support services across core network dimension with 50-ms-class setup time, and also to respond to multiple network failures in this time frame. It provides a method of cross-layer resource allocation that provides efficient allocation of bandwidth, both working and protect, to services at all layers in the network, including IP and wavelength services. Preliminary evaluations show significant advantages in using PHAROS.

The architecture described in this paper enables core network scale beyond 10X of today's networks by optimizing path selection to maximize optical bypass, and minimize the number of router hops in the network. As a result, a higher capacity of network services can be supported with less network equipage.

IX. REFERENCES

[1] Simmons J., "On determining the optimal optical reach for a long-haul network," *JLT* 23(3), March 2005.

[2] Simmons, J., "Cost vs. capacity tradeoff with shared mesh protection in optical-bypass-enabled backbone networks," *OFC/NFOEC'07, Anaheim, CA,* NThC2 March 2007.

[3] Dutta, R.; Rouskas, G.N., "Traffic grooming in WDM networks: past and future," Network, IEEE, vol.16 no.6 pp 46- 56, Nov/Dec 2002

[4] Iyer, P.; Dutta, R.; Savage, C.D., "On the complexity of path traffic grooming," Broadband Networks, 2005 2nd International Conference, pp 1231-1237 Vol. 2, 3-7 Oct. 2005.

[5] Zhou, L.; Prashant Agrawal; Chava Vijaya Saradhi; Fook, V.F.S., "Effect of routing convergence time on lightpath establishment in GMPLS-controlled WDM optical networks," *Communications, 2005. ICC 2005. 2005 IEEE International Conference on*, vol.3 pp 1692- 1696 Vol. 3, 16-20 May 2005.

[6] Saleh A. and J. Simmons, "Architectural Principles of Optical Regional and Metropolitan Access networks," *JLT*, 17(12), December 1999.

[7] Simmons J. and A. Saleh, "The value of optical bypass in reducing router size in gigabit networks," *Proc. IEEE ICC '99*, Vancouver, 1999.

[8] Saleh A. and Simmons, J., "Evolution toward the next-generation core optical network," *Lightwave Technology, Journal of*, 24(9), September 2006, p. 3303.

[9] Bragg A., I. Baldine, D. Stevenson, "Cost Modeling for Dynamically Provisioned, Optically Switched Networks", *Proc. SCS Spring Sim. Multiconf*, San Diego, April 2005.

[10] Brzezinski, A.; Modiano, E., "Dynamic reconfiguration and routing algorithms for IP-over-WDM networks with stochastic traffic," *Lightwave Technology, Journal of*, vol.23 no.10 pp 3188- 3205, Oct. 2005.

[11] Strand, J..; Chiu, A.., "Realizing the advantages of optical reconfigurability and restoration with integrated optical cross-connects," *Lightwave Technology, Journal of*, 21 (11), November 2003. p.2871.

[12] Chunsheng Xin; Yinghua Ye; Sudhir Dixit; Chunming Qiao, "A joint working and protection path selection approach in WDM optical networks," *Global Telecommunications Conference, 2001. GLOBECOM '01*. IEEE, pp 2165-2168 vol.4, 2001

[13] Kodialam, M.; Lakshman, T.V., "Dynamic routing of bandwidth guaranteed tunnels with restoration," *INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings*. IEEE, pp 902-911 vol.2, 2000

[14] C. Ou; J. Zhang; H. Zang; L.H. Sahasrabuddhe, B. Mukherjee, "New and improved approaches for shared-path protection in WDM mesh networks," *Lightwave Technology*, *Journal of*, vol.22 no.5, pp 1223- 1232, May 2004.

[15] Ellinas, G.; Hailemariam, A.G.; Stern, T.E., "Protection cycles in mesh WDM networks," *Selected Areas in Communications*, *IEEE Journal on*, vol.18 no.10 pp 1924-1937, Oct 2000

[16] Kodian, A.; Sack, A.; Grover, W.D., "p-cycle network design with hop limits and circumference limits," *Broadband Networks, 2004. BroadNets 2004. Proceedings. First International Conference on*, vol. no. pp 244- 253, 25-29 Oct. 2004

[17] J. Gripp, M. Duelk, J.E. Simsarian, A. Bhardwaj, P. Bernasconi, O. Laznicka and M. Zirngibl, "Optical Switch Fabrics for Ultra-High-Capacity IP Routers", *J. Lightwave Technol.*, vol. 21 no. 11, p. 2839, (2003).