Convergence of IP and Optical Networking

Authors: Kristin Rauschenbach and Cesar Santivanez

Non-print items:

Keywords:

IP/Optical convergence

Converged networks

Dynamic optical networks

Network reconfiguration

Network control plane

Optical control plane

Resource assignment

Resource representation

Routing, wavelength assignment, grooming

Recovery, restoration and protection

Abstract:

Rapidly increasing network demand based on unpredictable services has driven research into methods to provide intelligent provisioning, efficient restoration and recovery from failures, and effective management schemes that reduce the amount of "hands-on" activity to plan and run the network. Integrating the service-oriented IP layer together with the efficient transport capabilities of the optical layer is a cornerstone of this research. Converged IP-optical networks are being demonstrated in large multi-carrier and multi-vendor venues. Research is continuing on making this convergence more efficient, flexible, and scalable. In this chapter, we review the current key technologies that contribute to the convergence of IP and optical networks, describing control and management plane technologies, techniques and standards in some detail. We also illustrate current research challenges, and discuss future directions for research.

1 Introduction

Increasing demands on capacity, service delivery speed, and variable quality of service are stressing both IP (Internet Protocol) and optical networks. Methods are being sought to provide intelligent provisioning, efficient restoration and recovery from failures, and effective management schemes that reduce the amount of "hands-on" activity to plan and run the network. Over the past ten years, great strides have been made in integrating the service-friendly IP layer together with the efficient transport capabilities of the optical layer. Converged IP-optical networks are being demonstrated in large multi-carrier and multi-vendor venues. Research is continuing on making this convergence more efficient, flexible, and scalable.

New challenges are emerging in the face of an increasingly heterogeneous future of networking. Certainly any convergence of the IP and optical layers must be done within a context that recognizes the complexity and heterogeneity of all the network services and resources that use and support these layers. In this chapter, we review the current key technologies that contribute to the convergence of IP and optical networks, describing control and management plane technologies, techniques and standards in some detail. We also illustrate current research challenges, and discuss future directions for research. While we focus primarily on the IP and optical layer, we also provide perspectives on research toward an even more heterogeneous future.

Much of the technology and methods described in this chapter apply to networks of both small and large geographic scale, however most activity on converged networks is aimed at large-scale core networks with global reach.

This chapter is organized as follows. First we begin with a description of the services, architectures and technologies that are driving toward converged networks. We then provide the background on the existing and emerging technologies that will contribute to the establishment of widespread converged network implementation and deployment. We discuss relevant standards and industry activities that are facilitating this deployment. We then describe a novel integrated control framework that enables both network convergence, and scalable dynamic network control. Finally, we discuss some research that will both drive and enable future highly flexible and heterogeneous networks.

2 Motivation

2.1 Network services

Cloud-based services, mobility, and video-based content delivery are driving increasingly unpredictable network traffic that is stressing network management and control. Application innovations are driving changes in the make-up of network services. For example, the commercial Amazon™ Elastic Compute Cloud service provides resizable compute capacity to general consumer market users via a simple, quick-response interface. Combining compute and connection resources would appear a simple extension of this service. Likewise, the rapid proliferation of handheld technologies and performance improvement of wired devices such as high definition television and high-density storage systems have caused dramatic changes in how consumers and businesses receive and deliver content. This drives

not only tremendous bandwidth growth, but also presents challenges of supporting mobility and quality on-demand video delivery.

Along with these exciting new services, the traditional challenges of growing network capacity remain. In the past, reducing capital cost of network equipment has been the major solution to address capacity growth. However, today, network scale is demanding that operational expense reduction play a larger role in technology decisions than in the past. This directs more attention to automation and interoperability of network capabilities.

2.2 Network architectures

Today's deployed optical core network architectures rely exclusively on static point-to-point transport infrastructure. Higher-layer services operate according to their place in the traditional OSI (Open Systems Interconnection) network stack (see section 3.1). The stack helps to confine conceptually similar functions into layers, invoking a service model between them to implement the services using multiple layers. However, this practice has lead to stovepiped management, creating multiple parallel networks within a single network operator's infrastructure.

This type of rigidly layered architecture is expensive to build and operate, and will not react quickly to variable traffic and service types. As such, the industry has been calling for “network convergence” to simplify network management and provisioning and ultimately save operational and capital costs.

IP services now dominate network traffic. However, IP, or layer 3, networks utilize stateless per-node forwarding that is costly at high data rates, prone to jitter and packet loss, and ill-suited to global optimization. Layer 2 switching mechanisms

are more deterministic, but they lack fast signaling, which hinders service setup time. GMPLS (generalized multiprotocol label switching) attempts layer 2 and 3 coordination but is not yet mature enough for optical layer 1 and layer 2 shared protection over wide areas. Today's Synchronized-Optical-Network-based (SONET) methods of provisioning protected routes for critical services consumes excessive resources, which drives down utilization, increases cost, and limits the use of the more efficient route protection schemes.

There is a move to integrate multiple layer 1 and layer 2 functions to reduce cost and minimize space and power requirements. These efforts also aim to minimize the costly equipment (router ports, transponders, etc.) in the network by maximizing bypass at the lowest layer possible. These coordination efforts require a control plane that supports dynamic resource provisioning across the layers to support scalable service rates (from 100 Mb/s to 100 Gb/s) and multiple services, e.g., Time Division Multiplexing (TDM), Storage Area Networking (SAN), and IP services. Such a control plane also enables automated service activation and dynamic bandwidth adjustments, reducing both operational and capital costs. Surmounting these challenges requires a re-think of core network architectures to overcome the limitations of existing approaches and better leverage emerging technologies.

2.3 Network technologies

Increased capabilities of the switching and transmission systems, underlying component commoditization, improved and open-sourced software are providing substantial improvements in accessibility, agility and extensibilty of the network

elements, such as switches, routers and transport equipment. A good representation of the latest software and control protocols are presented in reference [1].

Advances in optical technology now allow practical reconfigurable wavelength networks to be constructed. These networks use wavelength-switching components to dynamically route wavelengths across mesh topologies, and provide a level of flexibility and scalability previously not possible.[2-4]

Optical switch architectures that combine Wavelength Selective Switching (WSS) and optical cross-connect switching (OXC) technologies like micro-electro-mechanical switches (MEMs) are emerging to support dynamic routing at the optical layer. [5,6] Currently in deployed networks, WSS's are used to remotely configure optical wavelength bypass in an optical transport node. WSSs can be reconfigured to automatically route wavelengths through mesh networks while minimizing the amount of optical-to-electrical conversion because of this bypass function.

One challenge of the WSS technology is that it is not colorless, because wavelength sensitive elements are used to provide the spatial separation in the switch that routes colors to different output ports. That is, with WSS only certain wavelengths are available at a particular port. Colorless switching is enabled through MEMs devices and other technologies that spatially redirect optical signals independently of the signal wavelength. While colored switching was not an issue with fixed-wavelength transceivers, the advent of tunable transponders has in turn demanded colorless architectures. Cost is an important factor, and recently technology advances have lead to the creation of cost effective colorless switch architectures, just emerging commercially.[7]

In addition to the optical switches, other components are needed to enable more agility in optical networks. These include fast-tunable transmitters and receivers, low-noise optical amplifiers, electronic dispersion compensators, and advanced, programmable modulation techniques. These technologies have existed for several years in research and prototype form, and most are also now beginning to appear in commercial subsystems.[8]

3 Background

Converged networking builds off a long history of network technology development. To understand the challenges and opportunities of network convergence some background in the methods and architectures that provide efficient and reliable network provisioning is useful. In this section, we start with a description of the data and transport network stacks. We then describe the functions and architecture of the management, control and data planes. Traffic management is the key to efficient network operation, and methods for routing, wavelength assignment and grooming are presented in this context. We also discuss restoration- and protection-based recovery methods to achieve high network availability. Finally, we discuss multi-domain network architecture and optimization approaches. In all cases, the focus is on IP, optical, and the emerging converged IP/Optical, networks.

3.1 Network stack

In a multi-layer network, layered models are useful to track how traffic is generated and carried across a heterogeneous infrastructure. Both the data community and the transport community have similar, but not identical, stacks that represent the layered model. Figure 1 shows both the telecommunications standard OSI (Open Systems Interconnection) stack and the data communications TCP/IP (Transport Control Protocol/Internet Protocol) stack. In general, each layer of the stack interacts with the adjacent layers to create an end-to-end connection for the application. While many variations are possible, the boxes that implement the technology in a layer above select, or are assigned, paths from the box technology that implements the lower layer.

As an information packet or circuit passes through a network, it traverses several network nodes, and may at each node pass through interfaces at different layers of the network as shown in Figure 2. The specific equipment (or layer) trace the packet or circuit travels is determined by the routing requirements for that particular information flow.

Various protocols are used to manage the interaction that directs a flow or traffic demand on its path through the network. As an example, an email from the application layer server that distributes email would become encapsulated in a TCP/IP protocol within the server. The TCP/IP packet would then be encapsulated in an Ethernet frame in the local area network to which the server is attached, that Ethernet frame is then encapsulated into a SONET-formatted stream from the enterprise network that connects to the next hop in the IP route as established by

the TCP/IP protocol. At this hop, the stream is demultiplexed, and the IP packet is extracted and sent to the local router for processing and retransmission to the next IP node. This process of encapsulation, multiplexing and de-encapsulation and demultiplexing continues through multiple hops through the IP network, which is typically connected via SONET or OTN (Optical Transport Network) circuits in the network core. Ultimately, the IP packet that contains the mail message is received by the recipient's mail server, and the message can be unpacked at the application layer.

Each layer in the network performs various management functions, and attempts to operate the functions of the layer efficiently and quickly. However, these layer functions are often redundant, and the packing and unpacking process results in unnecessary processing, latency, and loss of efficiency. This has prompted investigation into multi-layer network management and control, which is currently an active area of research. [9-13]

3.2 Management, control and data planes

As shown in Figure 3 (a), originally networks had data planes that carried the traffic, and management planes to conduct the operations of the network. This approach leads to a proliferation of management systems as the network scale and scope increase. Also, because management plane response times are typically slow, especially given the large amount of information they manage, service changes in these architectures are also slow. Increasingly today, there are three distinct planes in a network architecture: the management plane, the control plane and the data

plane as shown in Figure 3 (b). This architecture yields faster and more efficient management and service instantiation.

The management plane typically handles high-level network management and operations including network monitoring and customer billing. The telecommunication FCAPS (Fault, Configuration, Accounting, Performance, Security) model describes the functions of network management. The data plane carries the user data, and also employs interfaces to network transport equipment such as transponders, optical amplifiers and optical and electronic switches and switch ports. The control plane operates between these two layers. The role of the control plane is evolving, but generally, the control plane handles automated network provisioning, some fault recovery, and other administrative control functions like topology management and liveness verification. As shown in Figure 4 illustrations of example control plane architectures, the control-plane data communications can be provided in-band or out-of-band. The control-plane data communications topology can also be isomorphic to the communications data plane, or not, as the use case and technology demand.

The advantage of a control plane is that it can operate much more quickly than a typical management plane, enabling bandwidth to be provisioned on-demand (carrier or customer initiated), novel scheduled services, and more efficient automated restoration and recovery. There is a major push by standards organizations to ensure that the optical control plane is multi-vendor and multi-carrier capable. This will ensure interoperability that will hasten its introduction and use.

3.3 Control plane functions

While the definition of a control plane continues to evolve, in general, the control plane serves to decouple services from service delivery mechanisms and to decouple quality-of-service from its realization mechanisms. It also provides boundaries for policy and information sharing based on trust, business models, and need for scale. It provides end-to-end signaling path and maintains topology and path information across heterogeneous platforms.

A typical control plane handles functions of network element discovery, routing, path computation, and signaling and connection setup. Discovery is the process of learning the links and nodes that are adjacent to a given node. Discovery protocols manage the finding and verification of nodes and links, including real-time updates for liveliness. Discovery is enabled by exchanging address or naming information over either the in-band or out-of-band control channel.

Routing is the process of establishing all link and node connectivity in the network topology, the reachability of a node to other nodes in the network, and resource information such as link bandwidth and node switching capacity and connectivity. The routing information is used to construct a full network topology of the available links and nodes in the network, and is either stored locally in the network elements, maintained in management database, or some combination of these. The management of the topology information database, including where the information is stored, how it is updated, and how it is used for network optimization and service creation drives the control and management plane architecture and implementation. In today's optical networks routing information is typically

gathered infrequently, as network topology does not change frequently. In contrast, in IP networks, routing is handled on a packet-by-packet basis.

To achieve scale, not all resource information is included in the routing protocol. Abstraction, or minimal resource representation, is used to filter information and help scale, albeit at the sacrifice of optimization potential. Path computation provides available source-destination paths to either the end points or the traffic-engineering engine in the network. Path computation takes in routing information, and adds other optimization features, such as cost functions, desired protection schemes, or diverse routing requirements to determine viable network paths.

Signaling serves to implement a desired path set up or tear down, as well as maintaining liveness information of an active path.

3.4 Traffic management

The objective of traffic management is to efficiently assign service demands to available network resources. Grooming, routing and wavelength assignment are methods for assigning traffic to network resources in a multi-layered, multi-node network. Regardless of the underlying transmission method, traffic demands are said to flow from a "source" to a "destination" over a "path" (e.g. see Figure 5). There are different "paths" available in a typical network, including fiber paths (the specific fiber optic strand a signal transits), wavelength paths (the specific wavelength channel a demand is assigned to), SONET paths, etc.

In the recent past, a significant amount of human intervention was involved in the planning of the network capacity and the allocation of wavelengths between

network nodes. As the network migrated from simple SONET-based rings, to singly- and dually-honed multiple-rings, to today's more common mesh architectures, more automated methods have been introduced, because the complexity associated with path selection is significantly higher.

While the terms are often used interchangeably, in general grooming describes the multiplexing of lower rate signals into higher rate signals for subsequent routing and path assignment. The wavelength assignment process determines how to assign desired service paths to specific wavelengths. This is usually constrained to ensuring wavelength continuity (e.g. wavelength remains the same) along the path.

In many studies, routing and wavelength assignment are considered separately from grooming, because wavelengths are considered to be fully packed from source to destination. However, to achieve the best efficiency, the problems should be considered together. Overall, the objective is to either maximize throughput, minimize blocking, or minimize network costs (equipage) for a specific throughput of blocking requirement. Routing and wavelength assignment and grooming are each described in more detail in the following sections.

3.4.1 Routing and wavelength assignment

Wavelength assignment involves determining a path in the network between the two nodes and allocating a free wavelength on all of the links on the path. An all-optical path is commonly referred to as a lightpath and may span multiple fiber links without any intermediate electronic processing, while using one wavelength channel per link. The entire bandwidth on the lightpath is reserved for this

connection until the call is ended and service terminated, at which time the associated wavelengths become available on all the links along the route.

In a typical wavelength division multiplexed network, a fixed number of fixed color wavelengths are available on an optical fiber. Today, wavelength conversion occurs only at transponders. Therefore, routing and wavelength assignment of a demand consists of finding a path with available capacity, and then selecting a single wavelength used along the entire path. This is known as the wavelength continuity constraint. If wavelength conversion is allowed (either via back-to-back optical/electrical conversions or with an all-optical wavelength converter) then different colors can be used on different segments along the path. While wavelength conversion would appear to be a major advantage for the efficiency of traffic assignment, especially since a limited number (<100) of wavelengths are available on a given fiber, in general the gains to be realized by using wavelength conversion are relatively small.[14,15]

Routing and wavelength assignment can be divided into preplanned and on-demand. In preplanned assignment, the traffic is known, and wavelengths are assigned to minimize blocking of the known demand set. This is analogous to the generic bin-packing problem. The solution is conducive to an integer linear program which is NP-complete (NP stands for Non-deterministic Polynomial time), and heuristic approaches are utilized to minimize blocking. [16-18] In on-demand wavelength assignment, paths are selected as the demand appears in the network. Routing paths are typically selected based on shortest path, or least-loaded path

criteria. Methods to approach this problem include layered graphs, and logical-link representation.[19,20]

Routing and wavelength assignment becomes even more complex in an optical mesh network where optical cross-connects are used to re-route wavelengths dynamically onto different fibers. In these cases, the use of a centralized path computation element (e.g. the PCE section 3.3), allows for a near-optimal, deterministic and stable solution. [21,22]

In ASON (Automatically Switched Optical Networks) networks, GMPLS-based (Generalized Multi-Protocol Label Switched) resource reservation schemes are used for wavelength assignment and resource reservation based on either forward or backward reservation protocols. In both cases, signaling establishes available wavelengths on the path from source to destination. At the destination, a single continuous wavelength is selected. In forward reservation protocol, all available wavelengths are reserved on the forward path, and those not used are released. In backward reservation protocol, the backward signal reserves the selected path on the way back to the destination. These distributed algorithms do not guarantee a path will be available until two signaling passes on the path, and instability can arise if multiple reservations compete for a resource during set up.

3.4.2 Grooming

As compared to wavelength routing and assignment, the grooming problem is typically focused on multiple layers in the network. This is because grooming involves the aggregation of multiple lower rate streams onto a single higher rate stream. Often, the highest layers in the network, like layer 3, consist of highly

granular flows with small data rates that are groomed together either at layer 3 or at lower layers to create higher rate streams. Grooming is the optimization of network transmissions that span multiple distinct transmission channels or methods. Grooming can occur within multiple layers of the same technology or between technologies. Grooming is performed at the edge of the network, where tributaries are merged into long-haul connections, and also at intermediate nodes where the electronic transport, switching and routing equipment is capable of converting signals between different wavelengths, channels, or time slots. This intermediate grooming is only advantageous when multiple input ports at the node combine toward common output ports.

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 so that wavelength utilization at the core is close to 100%. However, not all nodes are capable of the required aggregation or disaggregation. Once demands are fully groomed onto a wavelength, the resulting channel can take advantage of optical bypass at intermediate nodes. This can reduce network capital cost, because optically bypassed signals do not require an optical-to-electrical-to-optical conversion (OEO), saving the cost of the electronics required for this function in the node.

Deciding where and when to groom demands is a difficult optimization problem. It must take into account different tradeoffs among capacity available, the

cost (both capital and operational) of the grooming ports and transponders, and the fact that constantly adding or removing demands will unavoidably result in fragmentation inside a wavelength. What may appear to be a good grooming decision in the short term may hurt performance in the future. Grooming decisions, then, must balance medium- to long-term resources tradeoffs and be based in medium-term traffic patterns. Like wavelength assignment, grooming is typically represented as a mixed-integer linear-programming problem, and heuristics and partitioning are used to achieve near-optimal results. Several good papers have been published regarding grooming solutions for multilayer networks, including real-time grooming via on-line algorithms, and static integer linear programming approaches.[23-25]

3.5 Recovery

Another major function of the control and management planes in distributed systems is recovery. Network recovery deals with the ability of a network to recover from the failure of a technology within the network. This includes link failures, transponder failures, node failures, and others. The control and management planes play a role in recovery by identifying faults, notifying appropriate resources about that fault, and establishing the eligible pool of resources used to recover from the fault, either pre-planned or on-demand. Particular implementation methods vary, depending on the speed and determinism demanded of the recovery process. Usually, the bottlenecks for recovery latency are round-trip delays for the signaling messages and the queuing delays for requests at

the failover switching node. There are schemes that address these issues, which are particularly acute in mesh architectures. [26]

Fault recovery can be broadly classified into protection and restoration. Protection relies on dedicated resources that are reserved in advance during connection setup. Restoration takes place immediately after failure detection to discover and reserve capacity and reroute the signal using this capacity. Compared with protection switching, restoration is more efficient in terms of resource utilization, but usually requires longer restoration time.

Typically, the lowest layers of the network (fibers, wavelengths, SONET, OTN) use protection-based schemes to recover from failures, while higher layers in the network (IP, TCP, Application) use restoration-based schemes. Also, because protection schemes are simpler to implement on a link basis, or using ring-based topologies typical of SONET infrastructure, restoration was also implemented in telecommunications digital cross-connect networks.

3.5.1 Recovery approaches

There are three basic approaches to recover an established connection in the face of network node and link failures: link-based, segment-based, and path-based. For purposes of this discussion the *interior* of a path consists of all links on the path and all nodes except the path's endpoints; two paths are said to be *interior-disjoint* when there is no node or link that is in the interior of both paths.

In link-based recovery, for each interior element along the primary path, a backup route is found by omitting the failed element from the network topology and recalculating the end-to-end path. Thus for each working path there is a set of *n*

backup paths where *n* is the number of interior elements on the primary path. These paths need not be (and usually are not) interior-disjoint from the primary path or from one another. For a single failure, link-based recovery may give an efficient alternate route; however, the approach faces combinatorial explosion when protecting against multiple simultaneous failures.

In segment-based recovery, a working path is associated with a set of *n* backup paths, one for each interior link or node. A given backup path is associated with one of these interior elements; it is not based on the end-to-end service requested but simply defines a route around that element. A classic example of segment-based recovery can be found in SONET rings, where any one element can fail and the path is rerouted the other way round the ring. Because segment-based backup paths are independent of any particular working path, they may be defined per failed element instead of per path. However, they can also be highly non-optimal from the perspective of a specific service request, and are ill-suited to protect against multiple simultaneous failures.

Path-based recovery defines one or more backup paths for each working path. A working path with one backup path is said to be *singly protected*; a working path with two backup paths is *doubly protected*; and similarly for higher numbers. Each backup path for a primary is interior-disjoint with the primary path and interior-disjoint with each of the working path's other backup paths (if any). Practical algorithms exist for jointly optimizing a working path and its backup path(s). Relative to link- and segment-based methods, path-based recovery maximizes bandwidth efficiency, provides fast reaction to partial failures, and is

readily extended to recover from multiple simultaneous failures. Its main drawbacks are a high signaling load if the path contains services with many different endpoints. MPLS and dynamic optical mesh networks both tend to favor the use of path-based recovery for these reasons.

3.5.2 Restoration

There are many restoration schemes used today. In general, these schemes use some form of signaling after a failure is detected to determine an alternate path. The different schemes use different mechanisms to flood failure state information, to calculate alternative paths, and to switch the data to the new path. [27-30] In general, it takes on the order of seconds for a restoration event to stabilize, and the network to recover. In large IP/MPLS networks, where IP forwarding tables must be updated, this can take 10's of seconds. In cases where network instability occurs, and/or a large amount of fragmentation of the topology results, the network may not recover without human intervention.

IP networks typically rely on restoration for failure recovery. IP networks are stateless, so in IP restoration, control plane and forwarding plane convergence time contributes to overall traffic restoration times. Control plane protocol updates (e.g. IGP, EGP, PIM) are required to exchange information after a topology change. Control plane convergence is completed when all network elements reflect the updated topology. The forwarding plane allows routers and switches to forward traffic from ingress to egress. The forwarding plane convergence is completed when affected traffic flows are restored.

To help speed recovery in IP-based networks, MPLS (Muli-Protocol Label Switching) is useful. MPLS was designed to provide a connection-oriented framework for the connectionless IP networks running at the layer above. MPLS utilizes labels, attached to flows of multiple IP packets headed to the same destination, to establish virtual paths (called label-switched paths or LSPs) between the MPLS-enabled router nodes.

MPLS Fast Reroute (also called MPLS local restoration or MPLS local protection) is a local restoration network resiliency mechanism. It is a capability represented in the RSVP Traffic Engineering (RSVP-TE) standard. In this case, each LSP passing through a facility is protected by a backup path that originates at the node immediately upstream to that facility.

In theory, labels can be assigned to any of a variety of underlying layer 2 technologies, such as ATM (Asynchronous Transfer Mode) virtual circuits, SONET circuits, or even wavelengths. These paths can be managed like the any other circuit, and so the MPLS layer can be used for traffic engineering, and protection and restoration. Since the MPLS paths are often running over underlying SONET infrastructure, which is already protected, hold-off mechanisms are used to prevent the MPLS layer from reacting to a failure that the lower layer protocol will address.

3.5.3 Protection

In protection schemes, a preplanned alternative, or back-up, path is either pre-provisioned and standing by idle, or reserved, and data is automatically routed to the back-up path in the event a failure is detected. The SONET standard for back-up path switching time is 50-ms. Protection can be implemented so as to be “hitless”,

if the backup path is up and providing redundant data to the receiver (so-called 1+1 protection). Slightly more efficient protection is afforded by schemes where a backup link is used to protect more than one path, as in 1:N SONET protection, or shared mesh protection. As long as only one failure occurs at a time, these shared schemes provide the same level of determinism as the dedicated protection.

Automatic protection switching is the capability of a transmission system to detect a failure on a working link or lightpath and to switch to a standby to recover the traffic. There are two commonly used types of protection for the links in SONET and OTN transport networks. They are differentiated by how they reserve backup resources. One-plus-one protection provides a continuously active backup path for each working path. At the source, the optical signal is split into two signals and sent over both the working and the protection facilities simultaneously, producing a working signal and a protection signal that are identical and always on. At the destination, both signals are monitored independently for failures. The receiving equipment selects either the working or the protection signal. Extending SONET-like protection schemes to WDM mesh architectures involves different algorithms to select both working and disjoint protection paths, but otherwise the switching mechanism is similar. [31-34]

In one-for-N protection (1:N), there is one backup path for several working paths (the range is from 1 to 14). In the 1:N protection architecture, all communication from the source to destination is carried out over the signaling channel. Because this represents a shared protection scheme, all traffic reverts to the working facility as soon as the failure has been corrected.

This approach to sharing protection paths can be extended to meshes. In these shared protection approaches, 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. 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. Several methods to determine shared protection paths have been reported.[35-37]

In general, path-based shared protection requires significantly less reservation of network resources than does dedicated protection. However, in real-world optical networks, a fundamental tradeoff arises between protection sharing and electrical port count: more sharing (shorter protection links) requires more electrical ports (regenerators, OEO converters, electrical switching). These practical considerations imply that not all nodes or regions of the network need the same strategy. Regions of the network that are bandwidth-rich or port-scarce will likely benefit from dedicated protection, while regions that are bandwidth-poor or port-rich will likely benefit from shared protection. Optimization schemes with a broad purview can consider a combination of shared and dedicated protection resource assignment strategies for those areas of the network that benefit from one or the other.

The control plane must be appropriately engineered to implement shared protection in a mesh, especially when fast (sub-100-ms) recovery is needed, because careful synchronization between the network resources and the control system is

required to trigger a protection switch, and signaling to the distributed resources involved in the failover is required.[38]

3.6 Multi-domain

Because of many technological and business reasons, today's global core networks are heterogeneous. Multiple global carriers operate using different network control and management practices. Within a typical carrier, there are again multiple networks operating that serve various purposes including local, metropolitan and wide area or backbone connections, as well as different services such as public IP, private IP and circuit networks. These networks are outfitted with equipment from different box vendors that each utilize proprietary box control (so-called element management), single-box-vendor network control (so-called network management) and interfaces to carrier management systems (so-called north-bound management interfaces). It is useful to delineated separate control regions into "domains", and control and management across different domains is then called multi-domain management. Typically within a domain, there is a common method for resource representation and the control and management functions that enable service creation on the equipment within that domain.

Multi-domain network management and control strives to achieve management and administration to enable functions such as service set-up and restoration end-to-end between a source and a destination that may lie in separate domains. These functions are achieved at domain boundaries through interfaces that contain typically limited information about the internal networks, a so-called abstract representation. The challenge for multi-domain management is that the

abstraction must contain enough information for rapid, efficient and complete satisfaction of the service request, but not so much information to limit scale or sacrifice the privacy concerns of the information within the domain.

The control plane for multidomain networks connecting a user source and destination is described by a user network interface (UNI) and an external network-to-network interface or E-NNI (External-Network-to-Network Interface), as shown in Figure 5. It is important to note that the control plane connections are logical diagrams. The control planes may communicate via the underlying transport networks they control or by a completely separate communication channel, like the control planes for single domain networks described in Section 3.2. Thus, we refer to intra-network control plane "reachability" as the ability of one control plane to communicate with another over an available data communication network. This data communications network used for control signaling may be separate for reasons of practicability, security, survivability availability and other quality of service considerations or scale.

Most of today's applications require end-to-end delivery of information, and so provisioning in optical networks must address the multi-domain issues. Recent research activities have started to consider multiple domain scenarios with a focus on improving routing and signaling protocol to increase network utilization. [39,40] Most multi-domain routing protocols use abstract representations of the local resource representations to advertise domain information to other domains in the network. Since each domain receives only abstract information about the other domains, calculating an optimal end-to-end path is challenging task.

Multi-domain network optimization proceeds much like single-domain optimization, and aims to optimize capacity utilization, reduce blocking, provide low or guaranteed latency for an end-to-end path that crosses multiple domains. The challenge is the lack of complete information across all domains. There are two extreme approaches to resolving resource conflicts, a distributed method that probes the domains on a demand-by-demand basis to establish resource availability, and a more centralized approach that uses control elements in each domain, with full intra-domain knowledge. These control elements then manage resource allocation across domains.

To illustrate a distributed approach, recently, a dynamic optimal end-to-end path computation algorithm for multi-domain optical transport networks was reported [41]. This algorithm determines an a priori optimal path using domain abstraction information. Then, the end-to-end path is re-optimized by dynamically re-assessing the inter-domain paths and domain's egress node after visiting intermediate domains in the path. The algorithm updates the inter-domain paths and domain's egress node after visiting all intermediate domains in the path, so egress can be changed to avoid congestion in a mid-path domain. The update decision considers several parameters such as available bandwidth, shortest hop count, and link failures. The update process occurs as the circuit is set up from source to destination. In simulation of a 100+ node, 4-domain United-States-based network topology, this "two-pass" approach showed improvement in blocking at high loads and no increase in blocking at light loads owing to path computation overhead.

The more centralized approach is covered in the next section.

3.6.1 Path computation element

In a multi-domain scenario, visibility amongst different domains is usually limited. So, to improve accuracy, determinism, efficiency and scale, an architectural element that provides purview within and across different domains is needed. This is the function of the Path Computation Element (PCE) introduced by the IETF (Internet Engineering Task Force) in 2006 and described in RFC 4655. [42] The PCE incorporates the special computation control elements needed to coordinate path selection in multi-domain networks. The IETF PCE model defines elements for computation (centralized or distributed), synchronization, discovery and load balancing, and liveness. It also describes control communications amongst these elements. It also supports coordination amongst multiple, distributed PCE's (both stateful and stateless) with functions including synchronization and monitoring. Finally, it establishes need for policy, confidentiality, and evaluation metrics.

Centralized path computation allows CPU-intensive (control processor unit) calculations to be dedicated to a high-end processor as opposed to the limited computation available in network equipment. It also improves determinism and optimality of resource assignment.

For multi-domain networks with limited visibility, a hybrid of centralized and distributed architecture is envisioned. A collection of PCE's, each with full information of a particular topology, coordinate at domain boundaries to establish end-to-end paths. This semi-distributed approach helps the system scale.

A PCE-based architecture also helps with extensibility, because it eases inclusion of a network node, which may or may not have its own control plane, which lies outside the original domain. The PCE also enables multi-domain protection and restoration schemes, as well as providing an insertion point for cross-domain policy implementation.

The IETF and ITU-T, which addresses PCE functionality in G.7715.2 (ASON routing architecture and requirements for remote route query), have an ongoing collaboration to align routing requirements for large multi-domain networks. The OIF is already working toward implementation and testing in carrier networks.

Work continues on standardization of the following key elements of the PCE architecture:

- Methods for communication between PCEs for policy updates, and between resources and PCE;
- Protocols on support of PCE discovery and signaling of inter-domain paths;
- Metrics to evaluate path quality, scalability, responsiveness, robustness, and policy to support path computation algorithms;
- Management modules related to communication protocols, routing and signaling extensions, metrics, and PCE monitoring information.

4 Standards

Networks have traditionally been built by two distinct communities, the data community and the transport community, each with their own respective standards organizations: the IETF for the data community and the ITU-T for the transport

community. One place these communities converge is in the area of multilayer network control. There is currently much activity focused on converging the IETF control plane model, so-called GMPLS, or generalized multi-protocol label switching, and the ITU control plane model, so-called ASON for automatically switched optical network. These models address large-scaled switched optical networks that would typically contain both data-centric network equipment (IP routers), and transport-centric equipment (OTN-based WDM transport and optical switching).

Merging of the control schemes for the data world and the transport world is a major challenge, because the starting points for the two network approaches are quite different. In traditional data, or IP networking, router nodes function as stateless per-node forwarding engines. There is no separate control plane, all control information is contained within the packet header, and packet forwarding governed by a forwarding table residing in each router. The forwarding table is periodically updated when the underlying topology changes. Generally, there is no management function, and traffic is routed on a "best effort" basis. This kind of approach supports low-cost, on-demand (though not guaranteed) service over a heterogeneous transport infrastructure, and works particularly well when the underlying router connections are low loss and statically connected (so that the router tables do not have to be frequently updated). These attributes have served to make IP the "service layer" of choice for most of today's applications, including computer interconnection, but also voice and video.

In contrast, transport networks were built to support very high-efficiency circuit-oriented connections between telecommunications switching and

aggregation points. There is a nominally-centralized, hierarchical management plane that handles call setup and connection control, which is the term given to the set up, tear down, and management of connections through the transport network. The data transport plane of these networks operates deterministically, using well-defined and well-timed frames to handle multiplexing and demultiplexing functions cost-effectively. These connection-oriented networks operated with a very high level of fidelity (the 5-9's or 99.999% availability standard), and utilized pre-planned protection schemes to achieve this availability in the face of typical failure modes (back-hoe link outages, and equipment failures). The traditional transport network is deterministic, and highly efficient. The equipment is able to cost effectively pack data into transport channels (more than a factor of five times lower cost than the equivalent speed router interface) because the interface cards do not have to examine every packet for the control and routing information. However, these networks are not as agile or flexible to accommodate unplanned growth and new service creation as the IP networks.

Building off the traditional SONET circuit-based standards, the ITU has introduced the OTN standard, which addresses multiple wavelength and multi-service features as compared to the SONET standard. OTN introduces containers, or optical data units (ODU) with different rates (2.5 Gb/s – 100 Gb/s) into which not only traditional SONET-framed data, but others such as Gigabit Ethernet, Fiber Channel, FICON and ESCON can also be conveniently packed. These containers can be configured in a multiplexing hierarchy for grooming and aggregation, and the OTN control plane (G.709) supports discovery, signaling and routing of the

connections establish for containers in the multiplexing hierarchy. There is also recently added and ODUflex channel, that allows for client-specified data rates. The ODUflex type containers necessitate the need for more signal information (data rate and frame size), and the ability to adjust the frame size, which adds significant flexibility to the otherwise conventional transport method.

Packet-based network control is based on multi-protocol label switching protocol (MPLS). MPLS-TP (MPLS-Transport Profile), introduced in 2008, is emerging as a method to converge IP and optical TDM transport control. This standard adds traditional operations and management functions common in ITU-based SONET and OTN standards to the MPLS protocol. The standard is evolving with both IETF and ITU activity.

For IP networks, the IETF has specific RFC's (request for comments) that govern the control plane for heterogeneous, e.g. IP, MPLS, TDM and WDM optical, networks. RFC 3471 describes extensions to Multi-Protocol Label Switching (MPLS) signaling required to support Generalized MPLS. [43] Generalized MPLS extends the MPLS control plane to encompass time-division (e.g. Synchronous Optical Network and Synchronous Digital Hierarchy, SONET/SDH), wavelength (optical lambdas) and spatial switching (e.g., incoming port or fiber to outgoing port or fiber). The GMPLS control plane ensures traffic-grooming capability on edge nodes by operating on a two-layer model; that is, an underlying pure optical wavelength routed network and an electronic grooming layer built over it (MPLS or TDM). In the wavelength routed layer, operating exclusively at lambda granularity, when a transparent light path connects two physically adjacent or distant nodes, these nodes will seem adjacent

for the upper layer. The upper layer can perform multiplexing of different traffic streams into these wavelengths. The GMPLS control plane essentially facilitates routing, resource discovery, and connection management and recovery.

In GMPLS, light paths are established by exchanging control information among nodes, distributing labels, and reserving resources along the path to route appropriately labeled flows. In practice, the signaling protocol is closely integrated with the routing and wavelength assignment protocols. Typical GMPLS signaling protocols include Resource Reservation Protocol (RSVP) and Constraint-Based Label Distribution Protocol (CR-LDP). GMPLS also uses the Link Management Protocol (LMP) to communicate proper cross-connect information between the network elements. LMP runs between adjacent systems for link provisioning and fault isolation. It can be used for any type of network element, particularly in natively photonic switches.

An emerging area of control plane standards is that for managing purely optical layer capability, including optical impairments and, potentially, dynamic optical layer topologies realized through optical switching and reconfigurable add drop function. Activities for this standard include routing and wavelength assignment methods, methods to include impariments, as well as the required signaling extensions to the emerging optical components and their performance monitoring test points. The Wavelength Switched Optical Network (WSON) standard emerging from the IETF provides a framework for applying Generalized Multi-Protocol Label Switching (GMPLS) and the Path Computation Element (PCE) architecture to the control of wavelength switched optical networks.[44]

The relevant ITU (international Telecommunication Union) standards for optical control plane are part of both the architecture for optical transport networks (G.805), the ASON (Automatically switched optical network) standards to govern the architecture for switched optical networks (G.8080), transport network functions (G.807), and various call setup and connection management (G.7713) and discovery (G.771). [45]

The layered-model of the ITU-T standards includes a client-server model, and is recursive such that any particular layer is a server to the layer above, and a client to the layer below. Links consist of a set of ports that connect the edge of a subnetwork to another. Link connections are static, but subnetwork connections are flexible and managed by the management plane. Links and subnetwork connections are delimited by connection points (CPs) in the client layer. The network connection in client layer is delimited by a terminal connection point (TCP). A link connection is represented in the server layer by a pair of adaptation functions and a trail. In the management plane, these reference points are represented by objects called connection termination points and trail termination points (CTP and TTP) for connection points and trails. [46] In G.8080, these concepts are extended to switched network topologies. A subnetwork point (SNP) is an abstraction that represents a connection point or a terminal connection point, and a set of SNPs that are grouped together for routing purposes is called a subnetwork pooi (SNPP). The SNPs may be static, an SNP link connection, or dynamic, and SNP subnetwork connection.

ASON maintains separation of the control plane from the transport plane. The

control plane can be assigned link connections without the link being physically connected. Thus, there are two steps to network discovery. In the first, transport plane discovery, a discovery agent maintains the transport connections for later binding to the associated control plane connections. The second step, control plane discovery, is handled by a link resource manager (LRM) that holds the SNP-SNP link connection information. A termination adapter performer (TAP) maintains the relationship of the control plane and transport plane resource names, which is necessary with the separate control planes.

The OIF (Optical Internetworking Forum), launched in 1998, is an industry group that sponsors internetworking activities and demonstrations, and has forged development of user network interface implementations (UNI), and networking interfaces (E-NNI for intra-carrier and I-NNI and N-NNI for internal networks). The work of the OIF has allowed multi-carrier, multi-domain demonstrations of control plane interoperability, and early implementations of end-to-end circuit set-up and restoration functions in an automated fashion via an optical control plane. This kind of interoperability demonstration has been ongoing since 2004. A diagram that illustrates the purview of the IETF, ITU-T and OIF is shown in Figure 6.

5 Next generation control and management

5.1 Drivers

Much of the industry work to date has focused on optical control plane for integration of the dynamic, flexible IP layer over a static, circuit-oriented

wavelength-division-multiplexed optical transport layer. Research, however, has addressed such challenges as making the optical layer dynamic and responsive to traffic changes [47-51], and also including non-traditional resources into the purview of the network control plane such as compute, storage and sensor resources.[52] This increase in scope puts additional stress on the control plane to handle more and more dynamism and heterogeneity going forward. Fortunately, there is a sound framework, building off the current optical control plane research and development, for managing and implementing services across these large, complex systems. Below we describe one of these frameworks, and then we describe work toward future control and management framework that supports an even greater degree of heterogeneity.

5.2 Novel Framework

A functional architecture that can manage and control the instantiation of services across a large, heterogeneous infrastructure must address several key requirements. It must be technology agnostic, allowing not only for introduction of new generations of traditional optical and routing equipment, but also the ability to add higher layer capabilities from processing and storage and lower layer functions, such as energy and other supporting infrastructure, into the model. The architecture must be able to optimize and manage flexibly across a heterogeneous subset of resources and constraints. Finally, it must be scalable and deterministic or stable.

A functional architecture that meets these objectives was developed in 2008 called PHAROS (Petabit/s Highly-Agile Robust Optical System). [38] While the

PHAROS functional architecture is general, and applies to any large-scale dynamic system, certain specific design choices were made based on the desire to apply the architecture to control of a global-scale (~100-node) dynamic optically-switched wavelength-division multiplexed system as part of DARPA's CORONET project. Here we describe the general framework, and provide some of the specific implementation decisions that apply to a global-scale wavelength division multiplexed optical network.

5.2.1 Governance, decision, action

The PHAROS functional architecture explicitly separates *governance*, *decision-making*, and *action* as three key roles in control and management of multi-layer, multi-domain networks. Their functional relationship is illustrated in Figure 7. They are analogous to the traditional management plane, emerging control plane and existing data plane functions.

The governance function controls the behavior of the full system, establishing which actions and parameters will be performed automatically and which require human intervention. Governance establishes policy and reaction on a human scale. It is not on the critical path for service instantiations. This function contains the primary repository of nonvolatile governance information and is the primary interface between human operators and the network.

The decision function applies the policies established by the governance function to effectively allocate resources to meet service demands. It is highly time-critical. The decision process is on the critical path for realizing each service request on demand: the decision process is applied to each service request and creates

directives for control of network resources. The PCE in the IETF architecture addresses the mechanisms required to carry out the decision role. In PHAROS, the decision process is unitary. That is, one and only one decision maker is assigned to a given resource. Minimizing the negotiations required to make a decision improves both the optimality of the decision and the consistency of the state it was made from, ensures deterministic decision times (without backtracking or thrashing), and enhances speed and resilience by reducing the number of entities on the critical path that need to reach a consensus. This results in globally consistent resource allocation, with consistently fast service setup. However, as is the case with the PCE, the decision function may also be implemented in a distributed fashion.

The challenge of allocating and assigning communications resources across multiple technological layers, rapidly and efficiently, requires careful attention to the functionality of the decision role. The key characteristics of the role are to minimize negotiations while maximizing the horizon of resource-allocation decisions: that is, making each decision with the widest feasible awareness of the total resources in the network and the total demands upon it. Maximizing the horizon of a resource-allocation decision allows consideration of the potential uses of a resource for local as well as for transit and protection functions.

The action function implements decisions made by the decision function quickly and reports any changes in the state of the system. The action function is time-critical. The responsibility of the action role is limited to implementing directives. The network element controllers in a typical router or switch device

would be the primary implementation components responsible for carrying out the action function.

5.2.2 Signaling network

The PHAROS functional architecture includes a signaling network, a closed system connecting all the components required for connecting the elements that perform governance, decision and action. The signaling communication network can be implemented in-band with the transmission network, either with a separate signaling channel like OTN, or within a packet header as in IP, or out-of-band or even on a separate network. Cost, congestion and delay are all factors affecting this design decision. There is also the option to implement signaling on a topology that is isomorphic to the data plane topology. Isomorphism has the advantage that it simplifies routing and speeds up signaling. However, this choice may result in additional delay because it dictates the link distances between control nodes. In the PHAROS functional architecture, the desire to manage resources dynamically, with sub-100-ms-class response times drove a design choice of a data-plane-isomorphic signaling network with dedicated in-fiber, separate wavelength channel, bandwidth.

5.2.3 Resource representation

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 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.

A more generalized and extensible framework for resource representation is based on topology abstraction. Topology abstraction is used to track network resources and route demands. For a multi-layer system, multiple topology abstractions are used, each representing a set of like resources, or addressing a set of like services. Topology abstractions represent these collections of resources by graphs with edges based on their connectivity. Each abstraction then presents a view of resources at different "levels" of the network (these levels may be the traditional network optical, MPLS layers, as shown in Figure 8, but may also be a more complex arrangement of resources) and are tuned for the optimization calculations required for specific tasks.

For example, within an optical wavelength division multiplexed network, there may be a topology abstraction that represents the transparent, non-electrically regenerated, network connections within a particular network configuration. As another example, protection resources may be represented in a "shared protection" topology abstraction that describes the resources available for protection, and may, for example, put lower costs for shorter routes, as opposed to other topology abstractions that may favor lightly loaded links. As a final example, an optical cross connect within a network node may be represented by a topology abstraction that provides all the input output port connections available per wavelength. The topology representation can communicate constraints such as wavebanding through its topology. By using abstract topological representations for all levels of the

network, representations extend down to an abstract network model of the essential contention structure of a node, and extend upward to address successive (virtual) levels of functionality across the entire network, as shown in Figure 9. This method is highly extensible because it uses one approach common to all levels of resource representation and allocation.

Constraints are incorporated by considering the edges available in a particular topology abstraction. Levels are layered as appropriate given the network configuration, and the standard "client-server" model is used for higher layer topology abstraction nodes to select particular "paths" from the layer below. Cost information is passed from lower layers to higher layers, and routing information is passed from higher to lower, as shown in earlier Figure 8.

The critical advantage of using topology abstractions is efficiency and agility. For efficiency, the optimization of resource allocation becomes truly global, with cross-layer and cross-network properties evaluated jointly. For agility, the technology agnosticism within the abstractions ensure that legacy and new technologies are readily incorporated and the requirements of emerging service classes addressed.

Note that topology abstractions keep track of resource usage, and adapt to resource utilization, and each topology abstraction keeps track of constraints at its "level". For example, optical reach and wavelength continuity would be captured in the "available resources" attribute of the edges in the optical transparent topology abstraction described above. Policy and "learning" about network behavior can also be inserted into the topology abstractions to affect routing decision. For example,

as traffic patterns are discovered, expertise can be captured by adding/removing edges from the respective topology abstraction to influence routing decisions from the layer/level above.

5.2.4 Optimization strategies

Whether the objective is efficient grooming or efficient routing and wavelength assignment on static link resources, efficient assignment of dynamic traffic to a dynamic circuit layer, or agile bandwidth assignment to a best-effort routing, layer optimization is achieved via a combination of a resource assignment strategy (centralized, distributed or some hybrid) and underlying optimization algorithms that drive that strategy.

Thus, a key element of a multi-layer resource management system is its choice of setup strategy for allocating resources to a new service request. There are three broad classes of strategy for doing resource allocation when setting up a service: pure centralized (the single master), path threading, and predistributed resources. In addition, hybrid strategies are available, including the one selected for the PHAROS project: *unitary resource allocation* strategy. The engineering tradeoffs of the various strategies are summarized in Table 1, and described further below.

Table 1. Comparison of the setup strategies for resource allocation in a distributed network.

Strategy	***Advantages***	***Disadvantages***
Single Master	• Optimal allocation • Deterministic latency • Fast decision algorithm	• Worst-case: adds round trip to master to setup latency • Can cause focused traffic loads • Potentially limited scalability

		• Vulnerable to single node failure • Vulnerable to network partition
Path Threading	• Fastest latency (most of the time)	• If high call-setup rates, high chance of long setup • Potential thrashing behavior at high setup rates • Additional state distribution (flooding) • Didn't work well in practice • No global optimality
Predistributed	• Fastest latency (most of the time)	• Less optimal • If high setup rates and utilization, maybe long setup • Potential thrashing at high setup rates and utilization • Additional state distribution (flooding)
Unitary	• Optimal allocation • Fast and deterministic latency • Robust and scalable	• Additional state distribution (flooding) • Requires governance (to set parameters controlling assignment of scopes and resources) • More complex implementation than Single Master

5.2.4.1 Single-Master Setup

The *single-master strategy* entails a single control node that receives all setup requests and makes all resource allocation decisions for the network. This approach allows, in principle, global optimization of resource allocation across all network resources. It has the further virtue of allowing highly deterministic setup times: it performs its resource calculation with full knowledge of current assignments and service demands, and has untrammeled authority to directly configure all network resources as it decides. The challenge for this strategy is that a single processing node with sufficient capacity for communications, processing, and memory to encompass the entire network's resources and demands. This node becomes a single point of failure, a risk typically ameliorated by having one or more additional, equally capable standby nodes. Moreover, each service request must interact directly with the master allocator, which not only adds transit time to service requests (which may need to traverse the entire global network) but also can create traffic congestion on the signaling channel, potentially introducing unpredictable delays and so undercutting the consistency of its response time.

5.2.4.2 Path-Threading Setup

The *path-threading strategy* goes to the other extreme: each node controls and allocates its local resources, and a setup request traces a route between source and destination(s). When a request reaches a given node, it reserves resources to meet the request, based on its local knowledge, and determines the next node on the request's path. If a node has insufficient resources to satisfy a request, the request backtracks, undoing the resource reservations, until it fails or reaches a node willing to try sending it along a new candidate path. This strategy can yield very fast service setup, provided enough resources are available and adequately distributed in the network. There is no single point of failure; indeed, any node failure will at most render its local resource unavailable. Similarly, there is no single focus to the control traffic, reducing the potential for congestion in the signaling network.

However, the strategy has significant disadvantages. Setup times can be highly variable and difficult to predict; during times of high request rates, there is an exceptionally high risk of long setup times and potential thrashing, as requests independently reserve, compete for, and release partially completed paths. Because backtracking is more likely precisely during times when there are already many requests being set up, the signaling network is at increased risk of congestive overload due to the nonlinear increase in signaling traffic with increasing request rate. The path-threading strategy is ill-suited to global optimization, as each node makes its resource allocations and next-hop decisions in isolation. This drawback may be ameliorated by global state flooding, though this adds to the risk of congestive overload during times of many service requests. At all times optimization

decisions may suffer as a given node's model of the network is neither internally consistent nor consistent with that of other nodes along a request's path. In practice, the path-threading strategy has not worked well owing to these limitations.

5.2.4.3 Predistributed-Resources Setup

The *predistributed-resources strategy* is an alternative approach to distributed resource allocation. In this strategy, each node "owns" some resources throughout the network. When a node receives a setup request, it allocates resources that it controls and, if they are insufficient, requests other nodes for the resources they own. This strategy has many of the strengths and weaknesses of path-threading. Setup times can be very quick, if sufficient appropriate resources are available, and there is no single point of failure nor a focus for signaling traffic. Under high network utilization or high rates of service requests, setup times are long and highly unpredictable; thrashing is also a risk. Most critically, resource use can be quite suboptimal. Not only is there the issue of local knowledge limiting global optimization, there is also an inherent inefficiency in that a node will pick routes that use resources it owns rather than ones best suited to global efficiency. In effect, every node is reserving resources for its own use that might be better employed by other nodes setting up other requests.

5.2.4.4 Unitary Resource Management

To resolve the tradeoffs among these strategies, the PHAROS functional architecture relies on a resource allocation strategy called *unitary resource management*. This approach involves running the optimization algorithm and

resource management from a resilient hierarchy of control nodes. For each service request there is exactly one control node responsible at any time; and for each network resource there is exactly one control node controlling its allocation at any time. Each control node has an assigned scope that does not overlap with that of any other control node. Scope consists of a service context and a suite of assigned resources. The service context defines the service requests for which the control node will perform setup. For example, a service context may be a set of tuples, each consisting of a service class, a source node, and one or more destination nodes. A scope would typically be based on a meaningful network service region, for example, all service requests whose endpoints fall in the continental U.S. The unitary strategy allows a high degree of optimization and highly consistent setup times, as a control node can execute a global optimization algorithm that takes into account all resources and all service demands within its scope. There is no backtracking or thrashing, and no risk of nonlinear increases in signaling traffic in times of high utilization or of high rates of setup requests. There is some risk of suboptimal resource decisions..

The unitary strategy uses multiple control nodes to avoid many of the problems of the single master strategy. Thus, a standard distributed system failover mechanism is used to manage information updates, and handoff in the event of failure of a control node.

5.2.5 Shared protection

Responding as quickly to a detected fault is an important capability of the network. At the same time, it is important to do so efficiently, using the current

network state where feasible. With PHAROS, speed and efficiency are achieved by using a fault notification scheme that relies on flooding, ensuring that every node is aware of the event as quickly as possible and taking action in parallel across the network to implement service recovery. The data plane nodes responsible for the failed element send a status change to all the adjacent control nodes and this is forwarded along all interfaces that have not acknowledged failover. The switching actions required to respond to a given failure are pre-determined at service setup, and loaded into the switching nodes in the form of playbooks.

Upon a service request, the control node calculates both the primary route (the path taken when there are no failures affecting the service) and the required disjoint protection route(s). The network resources required to instantiate each protection route are identified in terms of pools of reserved resources.

Disjoint routes for shared protection are established by an algorithm running in the control node. If two desired paths have no links or nodes in common, then no single failure can break both of them, and therefore they can be protected using some of the same backup nodes and links. The requirement is that no single failure will cause them to contend for these resources. Resources for the joint protection can be calculated straightforwardly by determining a set of all paths that simultaneously fail when at least one link or node along those paths fail. The capacity of services on this "jointly protected set" then defines the capacity required to protect the set. The protection actions for each failure, along with the required reserve capacity pools are then sent to the local nodes in playbooks that define the actions to take upon failure notification.

Upon receipt of a failure message, each local node will check its playbook to determine if it has protection actions to take due to the status change. If not, then it is done. If a protection action is required, the local node selects from the pool, implements the switch, and signals to its adjacent nodes that the resource is now assigned to the specific service request. Once all connections have been made, subscriber traffic may flow on the connection. The benefit of this solution is that the establishment of the resource pools was made by the element that implements the decision function (the "decision node"), but no decision node needs to be consulted to take action. Each local node will send a message to the decision node and to its neighbors acknowledging the connection; the decision node informs all local nodes of the specific resources to monitor for the flow and the abstracted "failure handler" to report if any failure occurs. Playbook semantics are based on these "failure handlers", which reside in the local nodes. Use of shared protection substantially reduces the amount of spare capacity required in a large core network as compared to dedicated protection, and is a major driver for the economic advantage of mesh-based architectures.

5.2.6 Optimal resource assignment

There are many published approaches to optimal wavelength assignment and grooming, as described earlier in this chapter. To illustrate how a subset of these approaches might come together in a consistent framework, the optimal service assignment approach for the PHAROS project is described further here.

PHAROS integrated routing, wavelength-assignment, grooming and protection strategy is guided by the topology abstractions described above. Each

topology abstraction keeps track of resource availability and constraints (e.g., optical reach and wavelength continuity for level 1 (L1) and wavelength topology abstraction edges) at its level.

Each demand is routed at its associated level, as shown in Figure 8. Note that the corresponding topology abstraction contains all required information for its level as well as all levels below. This information is abstracted into edge attributes (e.g. availability, cost, latency, intersection) that are used for routing and resource allocation decisions. Conversely, once a route is chosen, this route can be mapped into actual resources and cross-connects decisions. For example, a level 2 (L2) edge in a path implies that grooming is performed at the edge's endpoints only. Furthermore, the L2 edge lower level path (e.g. L1 path) carries information of exactly where regeneration and wavelength conversion (at the full wavelength level) is taking place. Similarly, a L1 edge in a path implies OEO at the edge's endpoints and optical bypass along the edge's physical (Level 0) path.

For example, Figure 8 shows an IP demand is routed at level 2 and the lowest-cost path obtained is the (level 2) path ***a – c – g***. Thus, grooming is performed at the source and destination nodes (nodes ***a*** and ***g***) as well as in node ***c***. Furthermore, expanding level 2 edge ***c – g*** revels its associated level 1 path to be ***c – e – g.*** Thus, OEO conversion must also be done at node ***e*** (along with nodes a, c and g). Finally, nodes ***b***, ***d***, and ***f*** are optically bypassed.

PHAROS routing algorithm is based on successive Dijkstra computation [53]: it first computes the working path, then removes those links and computes the first protection path and so forth. Successive Dijkstra was chosen to provide very fast

(sub-100-ms-class) setup times. It should be noted that standard algorithms such as Bhandari are not applicable in our cross-layer setting since they require non-overlapping edges. To minimize the impact of "trap links" and very long protection paths, we used the concept of "forbidden node/forbidden links". Basically, an edge will not be used for routing if its corresponding level 0 path contains a forbidden element. Standards algorithms such as Bhandari can be used to detect trap links and add them to the forbidden set without compromising the cross-layer optimization. [54]

The above formulation allows for trading off different costs at different layers (from bandwidth to transponders to MPLS ports). Also, the costs are not fixed but they can be expressed as a function of resource availability. This allows for not only load balancing at a single layer (e.g. bandwidth congestion) but to change the relative cost of different resources as the network operating point changes (say, from bandwidth rich to bandwidth poor). It also allows for an optimization based on equipment or operational (e.g., energy) costs or total-cost-of-ownership. For example, Figure 10 shows an example of cross-layer decision making. The best path (either the one with 2 OEOs or the one with 1 MPLS port) will depend on the selected cost function. This cost function can be set to optimize cost and/or to minimize the likelihood of future blocking. In any case, the level 2 topology abstraction contains all the required information.

Grooming, i.e. the action of joining several sub-lambda flows into a single wavelength with the purpose of maximizing wavelength utilization, is a relatively expensive operation since it operates at the electrical (and sometimes packet) level.

To be worthwhile, grooming must save more resources (e.g. bandwidth or ports) that it costs, and so are made based on the aggregate traffic traversing a link.

Determination of the cost/benefit of grooming is relatively simple when the traffic is static, but when the traffic is dynamic, as is the PHAROS case, then the decision is more complicated. Grooming determinations need to be made on-the-fly at the moment a demand arrives (and for both working and protection paths) *before* the system knows the total traffic that will flow through a link in the future.

PHAROS addresses this issue by decoupling the timescales. L2 edges are created *a priori* based on expect traffic patterns (e.g. using traffic matrices derived from past history or current expectations). The effect of the existence of a long L2 edge in the algorithm is that it makes available an "express link" that the algorithm can use to pack small flows into. When a demand arrives, the control node uses the level 2 TA to decide the path to use. This decision is made based on network wide, cross-layer criteria (such as resource availability, congestion, etc. as described above). If the path includes a long level 2 edge, the flow will be packed into it. Thus, the TA uses medium-term to long-term information to decide whether to add an edge and the decision to actually use the edge and groom traffic is decided by the routing algorithm on-the-fly.

During path setup, protection resources are pre-staged across the network and assigned to protection pools. The information about when and how to use a protection pool's resources is maintained by local nodes in playbooks provided and updated by the control node. Both playbooks and protection pools are expressed in terms of edges in the TA representation. The playbooks map failure identification(s)

and affected services onto the protection pools to be used to protect those services (i.e. protection pools associated with the edges in the protection path). The playbook does not specify the resource within a protection pool to use. Instead, the assignment is efficiently made only after a failure occurs. At recovery time, each node on the protection path selects the resources for protecting the 'outgoing' direction of a service independently and informs its neighbor nodes. Based on this distributed selection, each intermediate node can complete protection actions by interconnecting the outgoing resources it selected for protection with those incoming resources it was informed about by the neighbor nodes. Note that since each node only chooses resources in one direction ('outgoing'), and since the resources have been pre-assigned and sized for the worst failure event, no conflict occurs. This process is successively repeated across layers. That is, when a L2 resource is needed, first the underlying L1 path is established, and then the L2 resource (e.g. timeslot, label, etc.) is grabbed. Once again, since there is no need to coordinate between endpoints, the outgoing timeslot/label is selected shortly after the outgoing L1 resource is selected (no need to wait for a path-length round trip period).

One unique feature of the PHAROS system is that it allows the sharing of protection resources across layers. E.g., the same wavelength can provide protection to a wavelength service and to a set of level 2 demands, as shown in Figure 11. Each demand, d, whose protection path includes a protection pool, PP e, has an associated "Failure resource allocation matrix" ($FRAM_e^d$). Each entry (i,j,k, …) in $FRAM_e^d$ represents the number resources from PP e that are required if the failure sequence

$(f_i, f_j, f_k, \cdots)$ occurs. A failure f_i may represent a node failure, a span cut, etc.. The size of the FRAM matrices is $|F|^K$, where |F| is the number of failure events and K is the maximum level of protection LoP of any demand. All the FRAMs associated with a PP are combined (simple matrix addition) into the pool's Protection Matrix (M_e). The maximum entry (worst failure sequence) in M_e determines the number of resources needed in the protection pool (i.e., max sum instead of sum max). Note that there is no need to individually track each demand's $FRAM_e^d$, instead their sum (M_e) can be updated each time a demand arrives/departs. The rules determining which entries of M_e to update for a given demand are dependent on the protection policy (i.e. reversion mandatory or not) and the control plane capabilities (i.e. can it distinguish between span and node failures). Therefore, PHAROS is flexible and can support different update rules. For example when the maximum LoP is 2, and under certain conditions, the update rule for a PP e in the protection path of a singly-protected demand d of bit-rate r is equal to $M_e = M_e + FRAM_e^d$, where $FRAM_e^d = r \cdot \delta_{I_e^d}$, $\delta_S(x) = 1$ iff $x \in S$ (zero otherwise), and $I_e^d = P_1 \times P_2^C$, referred to as the "protected index set", is a set containing all failure events (ordered failure sequences) for which the demand d will need protection resources from the PP e. P_1 and P_2 represent the working and protection paths, respectively, P_2^C represents P_2's complement, and x represents the Cartesian product.

Figure 11 provides an illustration of sizing L1 protection resources due to both wavelength services and "L2 protection edges" which serve IP services. For the L1 edge under consideration, each L2 edge traversing it is a "client" (i.e. a demand) the same as a wavelength service demand. The FRAMs associated with each of these L2

edges is derived directly from the L2 edge Protection Matrix M by applying the generalized "ceiling" function, which round up each entry in P into the units of the lower level. Then, the L1 edge's Protection Matrix M is the sum of all the FRAMs of the WS as well as the rounded up Protection Matrices of all L2 edges including the L1 edge in its lower level path. This relatively simple methodology allows for sharing of protection resources between different services classes at different layers (i.e., IP and WS).

5.3 Research extensions: highly heterogeneous networks

Going forward, the trend toward converging "layers" to improve efficiency and improve service delivery times while improving the richness of the service offering is extending beyond the IP and optical layers, as shown in Figure 12. Recently, research has been focusing on dynamic resource management "higher up the stack" as well. This research represents an important step toward large-scale distributed system management that will reduce the amount of human intervention required to establish and use a complex array of resources. The ability to realize complex yet agile "systems-of-systems" autonomously will improve efficiency, and ultimately lead to simpler methods to solve larger and more distributed problems. To illustrate the extensions to the multi-layer and multi-domain network control and management strategies described earlier in this chapter, we describe an emerging approach to heterogeneous resource management being forged in the National Science Foundation's GENI project, the Global Environment for Network Innovation. GENI is a deeply programmable infrastructure suite for performing computer science and networking experimentation at scale. [55]

GENI allows programmable configuration and control of all aspects of a computation network (computation, storage and communication), and provides transparent access to a federation of shareable and sliceable resources in programmable topologies. A federated architecture such as GENI provides a set of mechanisms to allow organizations and users share and collaborate across a set of separately owned and operated resources. Details of the GENI architecture can be found at http://groups.geni.net/geni/wiki/GeniArchitectTeam.

The GENI functional architecture, illustrated in Figure 13, brokers the capabilities of resource owners to the needs of experimenters (users) who access slices of resources that may span several different resource owners. Single-owner resources are managed as aggregates via a GENI aggregate manager. Examples of aggregates are regional network, backbone network, campus networks or computer clusters. As the resource broker, the GENI functional architecture is designed to ensure accountability of the actions taken by experimenters on an aggregate's resources, and manage authentication and authorization services in the GENI clearinghouse. The GENI meta-operations center (GMOC) oversees the health and operations of the aggregates, though not as a replacement for individual aggregated management and operations functions. Experimenters access and create slices via experiment tools. They obtain an identity via a third-party trusted authority that authenticates them to the GENI federation, and credentials that authorize particular actions needed to create slices on GENI aggregates. A picture that illustrates the data flows amongst the principle actors in the GENI functional architecture is shown below.

At the network layer, GENI relies on an emerging technology called OpenFlow (**http://www.openflow.org/**). OpenFlow provides an emerging standard to open the internal flow tables in an Ethernet switch together with an interface to add and remove entries in the table. More generally, OpenFlow is an attempt to unify control across both packets and circuits, which can both be considered "flows". OpenFlow provides a unified method of assigning flows to an underlying switch and transport infrastructure through the OpenFlow interface. To date, the interface is well defined for data center switching, and multiple Ethernet switch vendors are supporting the OpenFlow specification. Research and development are continuing on extending OpenFlow concepts and standards to optical layer. OpenFlow per se does not address the optimization and management function for these open networks, however, the OpenFlow architecture stresses the need for a separate control plane, and enables a nominally centralized resource management strategy, which manifests a trend of the past decade of network research.

Acknowledgements:

The authors acknowledge the important contributions of the PHAROS architecture team for valuable input to this book chapter. In particular, Ilia Baldine, Alden Jackson, Will Leland, Walter Milliken, Ram Ramanathan, and Dan Wood contributed to the PHAROS framework and functional architecture described herein.

6 References

1. Ramamurthy, B., Rouskas, G.N., Sivalingam, K.M., Eds., Next Generation Internet:Architectures and Protocols, Cambridge University Press, 2011

2. Saleh A.; Simmons, J. Evolution toward the next-generation core optical network J. of Lightwave Technology, vol. 24, no. 9, p. 3303, September 2006

3. Gladisch, A.; Braun, R.-P.; Breuer, D.; Ehrhardt, A.; Foisel, H.-M.; Jaeger, M.; Leppla, R.; Schneiders, M.; Vorbeck, S.; Weiershausen, W.; Westphal, F.-J.; , "Evolution of Terrestrial Optical System and Core Network Architecture," Proceedings of the IEEE , vol.94, no.5, pp.869-891, May 2006

4. Strand, J.; Chiu, A.; , "Realizing the advantages of optical reconfigurability and restoration with integrated optical cross-connects," Lightwave Technology, Journal of , vol.21, no.11, pp. 2871- 2882, Nov. 2003

5. Tomlinson, W.J.; , "Requirements, architectures, and technologies for optical cross-connects," Lasers and Electro-Optics Society 2000 Annual Meeting. LEOS 2000. 13th Annual Meeting. IEEE , vol.1, no., pp.163-164 vol.1, 2000

6. Shankar, R.; Florjanczyk, M.; Hall, T.J.; Vukovic, A.; and Hua, H.; “Multidegree ROADMbased on wavelength selective switches: Architectures and scalability,” Optics Communications, vol. 279, no. 1, pp. 94–100, Nov. 2007

7. Jensen, R.A.; , "Optical switch architectures for emerging Colorless/Directionless/Contentionless ROADM networks," Optical Fiber Communication Conference and Exposition (OFC/NFOEC), 2011 and the National Fiber Optic Engineers Conference , vol., no., pp.1-3, 6-10 March 2011

8. Basch, E. B.; Egorov, R.; Gringeri, S.; Elby, S.; “Architectural tradeoffs for reconfigurable dense wavelength-division multiplexing systems,” IEEE Journal of Selected Topics in Quantum Electronics, vol. 12, no. 4, pp. 615–626, Jul. 2006

9. Wilson, B.J.; Stoffel, N.G.; Pastor, J.L.; Post, M.L.; Liu, K.H.; Tsanchi, L.; Walsh, K.A.;

Wei, J.Y.; and Tsai, Y.; "Multiwavelength optical networking management and control," Journal of Lightwave Technology, vol. 18, pp 2038-57, 2000

10. Blight, D.C.; and Czezowski, P.J.; “Management issues for IP over DWDM networks,” Opt. Netw. Mag., vol. 2, no. 1, pp. 81–91, Jan.–Feb. 2001

11. Raptis, L.; Hatzilias, G.; “An integrated network management approach for managing hybrid IP and WDM networks,” IEEE Network, vol. 17, no. 3, pp. 37–43, May–Jun. 2003

12. Schonwaelder, J.; Pras, A.; and Martin-Flatin, J.; “On the future of internet management technologies,” IEEE Communications Magazine., vol. 41, no. 10, pp. 90–97, Oct. 2003

13. Pinart, C.; Giralt, G.J., "On managing optical services in future control-plane-enabled IP/WDM networks," Lightwave Technology, Journal of, vol. 23, no. 10 pp 2868- 2876, Oct. 2005.

14. Barry, R.A.; Humblet, P.A.; , "Models of blocking probability in all-optical networks with and without wavelength changers," Selected Areas in Communications, IEEE Journal on , vol.14, no.5, pp.858-867, Jun 1996

15. Karasan, E.; Ayanoglu, E.; , "Effects of wavelength routing and selection algorithms on wavelength conversion gain in WDM optical networks," Networking, IEEE/ACM Transactions on , vol.6, no.2, pp.186-196, Apr 1998

16. Ramaswamy, R.; Sivarajan, K.N.; “Routing and wavelength assignment in all-optical networks,” IEEE/ACM Transactions on Networking, vol. 3, no.5, pp 489-500, 1995

17. 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

18. Ozdaglar, A.E.; and Bertsekas, D.P.; "Routing and Wavelength Assignment in Optical Networks," IEEE Trans. on Networking, no. 2, pp. 259-272, Apr. 2003

19. Xu, S.; Li, L.; Wang, S.; “Dynamic routing and assignment of wavelength algorithms for multi-fiberwavelength division multiplexing networks,” IEE Journal on Selected Areas in Communication, vol. 18, no.10, pp2130-2137, 2000.

20. Zhou, B.; Bassiouni, M.; Li, G.; “Routing and wavelength assignment in optical networks using logical link representation and efficient bitwise computation,” Photonic Network Communications, vol. 10, no. 3, pp. 333-346, 2005

21. Hu, J.; Lieda, B.; “Traffic grooming, routing, and wavelength assignment in optical WDM mesh networks,” IEEE INFOCOM, 2004

22. Zhau, Y., Zhang, J, Ji, Y., Gu, W., “Routing and wavelength assignment problem in PCE-based wavelength-switched optical networks,” J. Opt. Commun. Netw., vol. 2, no. 4, pp. 196-205, April 2010

23. Keyao Zhu; Mukherjee, B., "Traffic grooming in an optical WDM mesh network," Selected Areas in Communications, IEEE Journal on, vol.20 no.1, pp 122-133, Jan 2002

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

25. 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

26. Assi, C., Hou, W., Shami, A., Ghani, N., “Improving signaling recovery in shared mesh optical networks,” Journal Computer Communications, Volume 29, pp.59-68, December 2005

27. Zhou, L.; Agrawal, P.; Vijaya, C.; Saradhi; Fook, V.F.S.; "Effect of routing convergence time on lightpath establishment in GMPLS-controlled WDM optical networks," 2005 IEEE International Conference on Communications, vol.3 pp 1692-1696, 16-20 May 2005

28. Bouillet, E.; Labourdette, J.-F.; Ramamurthy, R.; Chaudhuri, S., "Enhanced algorithm cost model to control tradeoffs in provisioning shared mesh restored lightpaths," Optical Fiber Communication Conference and Exhibit, 2002. OFC 2002, pp 544- 546, 17-22 Mar 2002

29. Shengli Yuan; Bin Wang; , "Highly Available Path Routing in Mesh Networks Under Multiple Link Failures," Reliability, IEEE Transactions on , vol.60, no.4, pp.823-832, Dec. 2011

30. 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

31. 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

32. Ou, C.; Mukherjee, B.; Zang, H., "Sub-path protection for scalability and fast

recovery in WDM mesh networks," Optical Fiber Communication Conference, OFC 2002, pp 495- 496, 17-22 Mar 2002

33. Lei Guo, Hongfang Yu, and Lemin Li,“Path protection algorithm with trade-off ability for survivable wavelength-division-multiplexing mesh networks”; OPTICS EXPRESS 5834, vol. 12, no. 24, 29, November 2004

34. Pin-Han Ho; Mouftah, H.T., "A framework for service-guaranteed shared protection in WDM mesh networks," Communications Magazine, IEEE, vol.40 no.2, pp 97-103, Feb 2002

35. Canhui Ou; Jing Zhang; Hui Zang; Sahasrabuddhe, L.H.; Mukherjee, B., "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

36. Pin-Han Ho; Mouftah, H.T.; , "Shared protection in mesh WDM networks," Communications Magazine, IEEE , vol.42, no.1, pp. 70- 76, Jan 2004

37. Dikbiyik, F.; Sahasrabuddhe, L.; Tornatore, M.; Mukherjee, B.; , "Exploiting Excess Capacity to Improve Robustness of WDM Mesh Networks," Networking, IEEE/ACM Transactions on , vol.20, no.1, pp.114-124, Feb. 2012

38. I. Baldine, A. Jackson, J. Jacob, W. Leland, J. Lowry, W. Miliken, P.Pal, R. Ramanathan, K. Rauschenbach, C. Santivanez, and D. Wood, “PHAROS: An Architecture for Next-Generation Core Optical Networks,” pp. 154-179, *Next-Generation Internet Architectures and Protocols,* Ed. by Byrav Ramamurthy, George N. Rouskas, and Krishna Moorthy Sivalingam, Cambridge University Press, 2011.

39. Guanglei Liu; Chuanyi Ji; Chan, V.W.S., "On the scalability of network management information for inter-domain light-path assessment," Networking,

IEEE/ACM Transactions on, vol.13 no.1, pp 160- 172, Feb. 2005

40. Berthold, J., Ong, L., “Next-generation optical network architecture and multidomain issues,” Proceeding of the IEEE, vol. 100, no. 5, pp 1130-1139, May 2012

41. Benhaddou, D.; Dandu, S.; Ghani, N.; Subhlok, J.; , "A new dynamic path computation algorithm for multi-domain optical networks," Mediterranean Winter, 2008. ICTON-MW 2008. 2nd ICTON , vol., no., pp.1-6, 11-13 Dec. 2008

42. RFC 4655, “A Path Computation Element (PCE)-Based Architecture,” Farrel, A., Vasseur, J. P., Ash, J., Eds., August 2006

43. RFC 3471, “Generalize Multi-Protocol Label Switching (GMPLS) Signaling Functional Description,” Berger, L., Ed., January 2003

44. RFC6163, “Framework for GMPLS and Path Computation Element (PCE) Control of Wavelength Switched Optical Networks (WSONs),” Lee, Y., Bernstein, G., Imajuku, W., Eds., April 2011

45. ITU-T, “Architecture for the automatically switched optical network (ASON)”, Recommendation G.8080/Y.1304, November 2001 (and Revision, January 2003).

46. Tomsu, P., Schmutzer, C., Next Generation Optical Networks, Prentice Hall PTR, 2002

47. Mahalati, R.; Dutta, R., "Reconfiguration of traffic grooming optical networks," Broadband Networks, 2004. BroadNets 2004. Proceedings. First International Conference, pp 170- 179, 25- 29 Oct. 2004

48. Simeonidou, D.; Nejabati, R.; Zervas, G.; Klonidis, D.; Tzanakaki, A.; O'Mahony, M.J.; , "Dynamic optical-network architectures and technologies for existing and

emerging grid services," Lightwave Technology, Journal of , vol.23, no.10, pp. 3347-3357, Oct. 2005

49. Pandi, Z.; Tacca, M.; Fumagalli, A.; and Wosinska, L.; “Dynamic Provisioning of Availability- Constrained Optical Circuits in the Presence of Optical Node Failures”, J. Lightwave Technol., vol. 24 no. 9, p. 3268, 2006

50. Berthold, J., Saleh, A. Blair, L., Simmons, J., “Optical networking: past, present, and future,” J. Lightwave Technology, vol. 26, no.9, pp.1104-1118, May 2008

51. Sambo, N.; Andriolli, N.; Giorgetti, A.; Valcarenghi, L.; Cerutti, I.; Castoldi, P.; Cugini, F.; , "GMPLS-controlled dynamic translucent optical networks," Network, IEEE , vol.23, no.3, pp.34-40, May-June 2009

52. Nejabati, R.; Escalona, E.; Shuping Peng; Simeonidou, D.; , "Optical network virtualization," Optical Network Design and Modeling (ONDM), 2011 15th International Conference on , vol., no., pp.1-5, 8-10 Feb. 2011

53. Dijkstra, E. W. "A note on two problems in connexion with graphs". Numerische Mathematik vol. 1 pp. 269–271, 1956

54. R. Bhandari, Survivable Networks: Algorithms for Diverse Routing, Kluwer Academic Publishers (1999)