Appeared in *Photonic Network Communications, Vol. 8, Issue 1, June 2004, pp. 7-21.*
Sp. Issue*: IP-Centric Control & Mgt. In WDM Optical Networks.*

On the Issues in Implementing the ***Peer Model*** in Integrated Optical Networks

by

Vishal Sharma[1] (Metanoia, Inc.), Abhimanyu Das[2] (Mahi Networks), Charles Chen[3] (Mahi Networks)

Abstract

This paper contributes to a much-needed understanding of the operation, design, implementation, and evaluation of the *peer model* in integrated optical networks. The *overlay* and *peer* models of operation form the two fundamental architectural alternatives for interworking the control planes of optical TDM/WDM networks with those of packet or cell-based networks. Of these, the overlay model is well understood, having a precedence in IP-over-ATM networks deployed in the mid 1990's. It follows a proven approach to managing multi-area, multi-domain networks. The peer model, on the other hand, has not been implemented yet, and has also not been analyzed adequately in the literature. To enable service providers to implement either model, based on the respective merits of each, it is fundamentally important to develop a working solution for the peer model. The focus of this paper is to provide such a solution, perform a complexity analysis of the solution, and discuss its impact on the design of the protocols and the packet and transport layer devices that must interwork to realize this model.

1 Introduction

As telecommunications carriers look to evolve their existing networks or to build a new integrated network infrastructure, a fundamental paradigm shift is emerging in the way networks are designed. From a layered network model involving the management of network elements individually at each layer (hitherto the guiding principle of network design), the philosophy is shifting to one of an integrated infrastructure, where one can seamlessly manage packets, circuits, and lightpaths. The reasons for this industry trend towards a unified set of mechanisms (the *unified control plane* [1], [2], [7]) that will enable service providers to manage separate network elements in a uniform way can be traced to the historical evolution of transport and packet networks.

The traditional transport network, for example, has not had a control plane at all (it has only had a management infrastructure) because the need for automatic provisioning in SONET/SDH networks was not envisioned in the initial years. With the growth of SONET/SDH networks, however, the manual provisioning process became increasingly slow and often error prone, motivating the need for a better solution. In fact, the ability to manage network elements limited the scalability of such networks, calling for a fundamental shift in network design towards an intelligent way to manage the network.

In parallel, the Internet Protocol (IP) and its associated protocols matured to become the ubiquitous routing and signaling mechanisms for transporting packet data. Further, this packet infrastructure evolved to introduce the idea of decoupling routing from packet forwarding (via technologies such as multi-protocol label switching), thus allowing for virtual circuit setup and traffic engineering [16] in packet networks. Further, the convergence of the traditional transport network and the data network, has produced a dramatic change in the Internet infrastructure, making an integrated control plane technology central to managing a network composed of a wide range of elements, from IP routers, to SONET/SDH cross connects, to optical cross-connects. The design of a control plane is therefore a fundamental architectural issue for the new network infrastructure.

To provide a single, unified control plane for multiple switching layers, generalized multi-protocol label switching (GMPLS)[17][23] has been proposed to extend the MPLS [16] concept of label switched paths and its traffic engineering capabilities to the control of TDM, lambda, and fiber switched networks. Thus,

[1] 888 Villa Street, Suite 200, Mountain View, CA 94041. Email:v.sharma@ieee.org. Ph: 650-641-0082.
[2] 1039 N. McDowell Blvd, Petaluma, CA 94954. Email:adas@mahinetworks.com. Ph: 707-283-1022.
[3] 1039 N. McDowell Blvd, Petaluma, CA 94954. Email:cchen@mahinetworks.com. Ph: (707) 283-1013.

IP-based signaling and routing protocols[9],[10] are being adapted to dynamically control and manage end-to-end data flow in heterogeneous multi-layer networks. The GMPLS architecture [23], however, concerns itself only with protocol extensions to routing and signaling to enable TDM/WDM link advertisement and path-establishment. It does not address issues related to the interaction of the control planes of the nodes in different domains or how this interaction affects forwarding and routing, both of which are the focus of this paper. We emphasize that GMPLS architecture focuses on creating a common control plane for multiple technologies (IP, TDM, ATM, Ethernet, lambdas) in the forwarding plane, and that it levarages IP technology to operate the common control plane. There is, however, no restriction on the technologies in the forwarding plane that could be controlled by such a common control plane.

For the IP routing and signaling messages to flow between such heterogeneous network nodes, either an in-band or out-of-band *IP control channel* (IPCC) is required between every pair of nodes connected by a data bearing link. These heterogeneous transport networks therefore consist of an IP-based control plane and a layered data plane comprised of a mixture of IP, TDM, and WDM data planes, which carry data across the network. There are two architectural choices [5], [6] for the interaction between the packet and the optical TDM or WDM layers, which differ in the manner in which the control plane component of the systems in the network is realized. These are the peer model and the overlay model. (There are, however, other proposals [3] that posit achieving IP-optical integration without either the overlay or peer model of operation.) We briefly discuss these two control-plane operation models in sections 1.2 and 1.3.

1.1 Motivation and Related Work

While a number of works in the literature have dealt with the definition of a common control plane for packet and optical TDM/WDM networks (see for example, [17], [18], [19], [20], [22]), and have touched upon the issues in implementing the overlay and peer models, we are not aware of any work that has provided a solution that specifies how the peer model would operate, and the changes that would be needed in both the protocols and the network elements to realize such a model.

The Optical Internetworking Forum (OIF) has been active in defining control plane extensions for the overlay model of operation for optical networks [24]. Generic protocol extensions to IP signaling and routing protocols to piggyback optical TDM/WDM link properties and topology information have also been extensively researched within the GMPLS framework [20], [23], and these extensions can be used in both the peer and overlay models of operation. However, a detailed analysis of the IP control and forwarding plane interaction between heterogeneous network elements in a peer model has been missing.

In particular, our work is motivated by the following:

i) While there exist known implementations of products and protocols for the overlay model, there is no known implementation, as far as we know, of the peer model yet.

ii) There is a need to examine the impact on the forwarding plane of having common intelligence in the control plane that oversees multiple transport technologies (E.g. IP, ATM, Ethernet, TDM, and lambda). The requirements imposed on the forwarding plane are often a barrier to a practical implementation of the peer model, and need to be examined.

iii) In the peer model, even though routing provides a single topological view of the network, it is necessary that from a forwarding perspective there be a logical separation between the layers. This is because label switched paths of a given type must traverse interfaces of only that type. It should be possible, therefore, to enforce isolation between different forwarding technologies even while having a common view at the control plane. The mechanisms for doing so need to be specified.

The problem of specifying a feasible solution for the working of the peer model is thus an open one. The peer model of operation for a unified packet/optical control plane, while more complex, has certain

advantages over the overlay model (which we enumerate in section 1.3). If service providers are to be able to decide on an evolution path for their networks, and judge which model is most suitable in their particular context, the specification of a solution for peer model operation seems to be a critical task.

In the light of this, the current paper specifies a solution for the operation of the peer model, and evaluates the feasibility of the solution. This, we believe, will allow service providers to make a balanced decision about which particular model (peer or overlay) to deploy in their networks, based on the technical merits of each proposal.

1.2 Overlay Model

Under the overlay model, the heterogeneous network is divided into homogeneous domains, such as IP, TDM, and WDM domains. The routing, topology distribution, and signaling protocols in the IP domain are independent of the routing, topology distribution, and signaling protocols in the transport (TDM and optical) domains. That is, the nodes within a given domain interact (in terms of routing and signaling) only with the nodes in their respective domains, and the IP and optical transport domains are logically separate. The IP routers are unaware of the transport network topology, and form adjacencies with one another over the optical connections provided by the underlying transport network.

This model is conceptually similar to the classical IP-over-ATM or MPOA models. In the overlay model, topology distribution, path computation and signaling protocols would have to be defined for the transport domain. A simple way to achieve this is to run separate instances of an MPLS-based IP control plane for the IP and transport domains.

1.3 Peer Model

In the peer model, the IP router network acts as a peer of the TDM and optical networks. This means that the IP routers, and the TDM and WDM switches are connected to each other in the control plane via routing and signaling adjacencies over the IP control channel. Thus, only a single instance of the IP-based signaling and routing stacks (with optics-based extensions) is required for both the IP and the TDM/optical domains, and the same instance runs in both domains. This is because the objective of the peer model to present a single, unified picture of the IP, TDM, and optical portions of the network to each network node. This allows the routers to compute and signal end-to-end paths across the optical transport network. To do so, a common interior gateway protocol (IGP), such as OSPF-TE, can be used to distribute topology information over the integrated IP-transport network. The implicit assumption in all of the above is that the IP, TDM, and optical networks share a common address space, which can be realized by using IP addresses to number network interfaces (we will see later that only the *control plane* interfaces at a node need be numbered using IP addresses).

Therefore, an objective of a peer model is to have unified control and management planes for both the IP and the TDM/optical transport networks. This helps to increase the interaction between the IP and TDM/optical domains by allowing for a single management interface that can control both, which is not possible in the overlay model. As a result, the peer model has the following advantages over the overlay model:
(a) It provides a common interface to upper management layers, for the management and provisioning of elements that in the overlay model are independently managed. The provider can thus control both packets and circuits (and later, wavelengths).
(b) It therefore simplifies network maintenance and troubleshooting.
(c) It gives the network provider the ability to see the entire (IP and TDM/optical) network topology at any network node.

(d) Knowledge of the entire topology enables better network design through more efficient traffic engineering, due to the visibility into all the data path layers. This allows better monitoring of the layers, and to exploit the correlation between them.
(e) The unified control plane in the peer model facilitates faster circuit provisioning, by employing common signaling protocols across packet and circuit domains, thus streamlining the signaling between network elements, and allowing for a better opportunity to leverage resources in other layers.
(f) The unified control plane also allows for the introduction of protection and restoration schemes that can be coordinated across layers.

The remainder of the paper is organized as follows. We begin in Section 2 with some terminology and assumptions that will be used in the analysis and discussions in subsequent sections. In Section 3, we present a reference architecture for the peer model, and the benefits it offers relative to the overlay model. In Section 4, we present an implementation of the peer model, focusing on specific control plane, data plane, and management plane issues that must be addressed, and we describe how the packet and transport layer elements must work to realize the peer model. In Section 5 we present an analysis of our solution and the changes that it entails to the routing protocols and to the packet forwarding process at the network nodes. Section 6 concludes the paper.

2 Terminology and Assumptions

To set the stage for the discussion in the rest of this paper, we first define some basic terminology and present some assumptions that underlie the remainder of our analysis.

2.1 Terminology

In the following, we define some terms and concepts useful for our subsequent discussions.

Term	Definition/Description
P (or packet) device/node	A term used for an IP router or for the hardware and software within a hybrid system that realize IP router functionality. So a P-device is either a physical or logical system that functions as an IP router.
T (or TDM) device/node	A term used for a TDM cross-connect, or for the hardware and software within a hybrid system that realize TDM cross-connect functionality. So a T-device is either a physical or logical system that functions as a TDM cross-connect.
P (or packet) link	A physical or logical link that connects two P-devices, and carries packet data. (Note that the control and management messages/packets that relate to the forwarding of data packets on a P-link flow over the same P link.)
T (or TDM) link	A physical link that connects two T-devices (or a T device and a P device), and carries TDM data (in the payload portion of the frames running on the T links). Since we do not allow for T devices to be interspersed with P devices, a T link can only be a physical link. Note that the overhead portion of the frames running on a T link may carry control messages.
P (or packet) domain, or Packet network	A collection of P devices and P links that run one instance of the IP routing and signaling protocols.
T (or TDM) domain, or TDM network	A collection of T devices and T links that run one instance of the IP routing and signaling protocols. (In the peer model, of course, the P and T domains both run the *same* instance of the IP routing and signaling stacks.) The T domain merely provides circuit services to connect P

	devices.
Packet forwarding plane	A term used for the entities that are involved in the manipulation and forwarding of data in a P device. This includes hardware and software that is used to classify packets and take actions such as counting, policing, and simple editing, and forwarding to the output.
TDM forwarding plane	A term used for the entities that are involved in the manipulation and forwarding of data in a T device. This includes the hardware and software that is used to receive incoming TDM circuits (say a STS-48 stream), decompose them into their constituent circuits (say STS-1 streams), cross-connect streams from different interfaces, and recombine them into larger circuits on the output link.
Forwarding plane	A term used for the entities and functions that are involved in the manipulation, forwarding, and transmission of data in a P or a T device.
Forwarding or data plane topology	The topology formed by the forwarding planes in the P and T devices and their interconnecting physical, data bearing links.
Control plane	A term used for the entities in a P or T device that run the routing and signaling protocols and are involved in the setting up and tearing down of packet or non-packet (TDM or WDM) LSPs.
Control plane topology	The topology formed by the control planes in the P and T devices and their interconnecting control links/channels.
Management plane	A term used for the entities in a P or T device that are involved with the configuration and management of the P or T device in question, and include functions related to the management of the devices, networking layer, and services.
Management plane topology	The topology formed by the management planes in the T and P devices, and their interconnecting control links/channels. *In our case, the management plane topology is the same as the control plane topology.*
IP control channel (IPCC)	An in-band or out-of-band data communication channel that connects each pair of devices that have a TDM link between them. (This could either be two T devices or a T device and a P device.) The IPCC is used to transmit IP routing and signaling messages.
IPCC topology	The topology formed by all the IPCCs in the network.
IP topology	The topology formed by all the P links and the IPCCs taken together. *In our case the IP topology is the same as the control plane and management plane topologies.*
Interior node	A node in the P (or T) domain that has physical connections only to other nodes in the *same* domain.
Border node	A node in the P (or T) domain that has at least one physical connection to a node in a *different* domain.

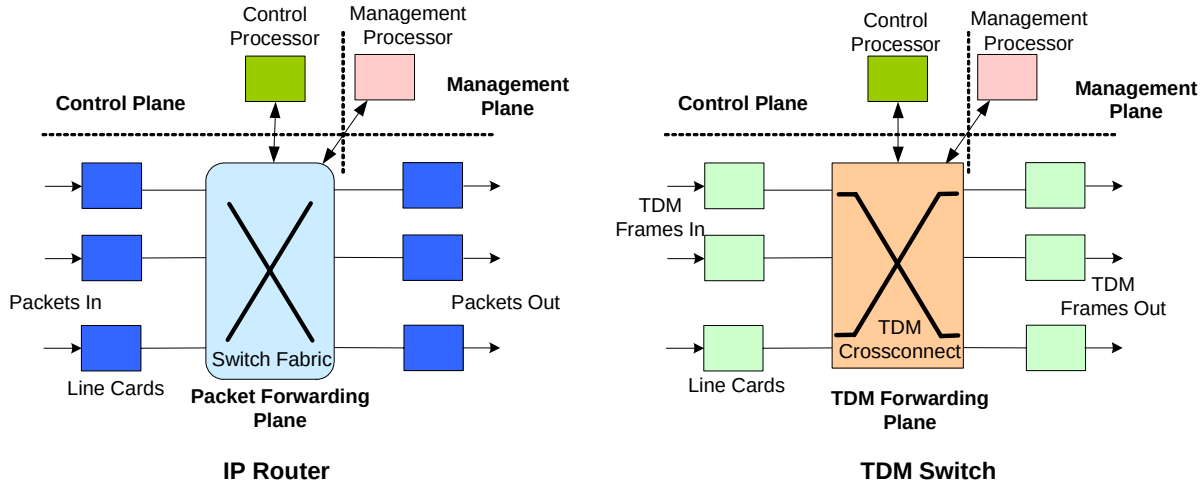

Figure 1. Logical decomposition of a packet (IP) router and a TDM switch to illustrate the control, forwarding, and management planes on each. Note that for simplicity the figures show a separate management processor. It is possible for the management processes and the control processes to run within a single system processor.

2.2 Assumptions

To simplify our exposition of the peer model, we make the following assumptions:

(i) We discuss the peer model in the context of only two domains, the IP and TDM domains. The concepts that we introduce for the two-domain peer model can easily be extended for additional domains, such as a WDM/optical domain.

(ii) We assume that we have a core of T devices (forming the T domain) surrounded by P devices (forming the P domain). Thus, we do not allow for the T devices and the P devices to be interspersed.

(iii) The T domain merely provides circuit services that connect two P devices. In other words, the T devices merely perform Layer 1 or physical switching/forwarding.

(iv) The forwarding plane of a T device cannot communicate directly with the forwarding plane of a P device.

(v) Only the control and management planes of a T device and a P device can communicate with each other.

(vi) The communication between T and P devices may be in-band (over SONET section or line data communication channel (DCC)) or out of band (over an IP generic routing encapsulation (GRE) tunnel).

(vii) Every T link in the network has an associated control channel for carrying control- or management-plane packets.

(viii) The interior gateway protocol (IGP) considered is OSPF with traffic engineering (TE) extensions(OSPF-TE) [8], which is used to advertise both T and P links.

(ix) We assume that the IP addresses used across the network (that is, across the T and P domains) are unique, and cannot be duplicated.

(x) We assume that multicasting belongs only in the P domain. That is, we do not consider multicasting in the T domain.

(xi) We confine our attention to a *single IGP domain*. Inter-domain issues are beyond the scope of this paper.

2.3 Review of Packet Forwarding

It is useful at this point to briefly review the packet forwarding process at a node. This will be important later, because we will see that the peer model requires several modifications to this process and to the various tables that a node must maintain for packet forwarding. The earliest generations of routers did not distinguish between forwarding and control planes, and had monolithic software to handle both tasks, usually in a single processor. Over the last several years, however, the architecture of routers has evolved to have a separation between the control plane and the forwarding plane, thereby allowing the forwarding plane to be distributed on to the line cards, and scale in speed and performance. Thus, the forwarding of most data packets happens in hardware, allowing them to travel via the so-called *fast path* of the router. The control packets (signaling and routing protocol packets for instance) and data packets that cause exceptions (for example, data packets with unrecognized options) are sent to the control processor for processing, and thus travel via the so called *slow path* of the router. It turns out that this architecture fits in well with those of TDM devices, where the forwarding plane (the TDM switch fabric) was already separated from any control/management plane.

The forwarding of packets at a node is based on a *forwarding information base* (or FIB), which is a mapping of the destination addresses to the next hop and outgoing interface[4]. Upon receipt of an IP *data* packet, the node uses the destination address of the packet to extract from the FIB the identity of the outgoing interface to which the packet should be forwarded. Upon the receipt of IP *control* or *management* packets, identified by the processor that processes the incoming packets, the packets are directed to the control processor of the node, which hands the packet to the appropriate software process (for example, an OSPF-TE or RSVP-TE process) responsible for processing that specific type of packet.

Note that in an IP router, the FIB is usually located on the line cards (see Figure 1), so data packets typically only travel through the packet forwarding plane. Control/management packets, however, are directed from the line card to the control/management processor. In a TDM switch, on the other hand, the only IP packets typically arriving are control/management packets, and these are directed from the line card to the control/management processor for further processing.

3 Peer Model Reference Architecture

In traditional IP routing, a single instance of an IGP runs only in one domain, which leads naturally to a single FIB in the forwarding plane of every node, as discussed in Section 2.3. In the peer model, however, since a single IGP instance runs in multiple domains (the P and T domains in our case), this is no longer true. For the peer model to be meaningful, the forwarding planes of the P devices must be fully-connected, making the packet network topology an overlay on the physical transport (TDM and optical) network. Thus, the *peering* between the IP devices and the TDM/optical devices is only in the control and management planes.

Indeed, a node in the peer model has to consider *three topologies* when forwarding IP packets – the *IP topology*, the *P domain topology*, and the *IPCC topology*. This is because, the specific topology that a node needs to consider when forwarding a packet, depends on the type of traffic received. To be able to forward the different types of IP packets towards their destinations, the node has to also maintain different FIBs. These FIBs are created by running a constrained shortest path first (CSPF) algorithm on particular subsets of the IGP link state database at the node. In particular, in the peer model, nodes may require three types of FIBs:

[4] The FIB is derived from a *routing information base* (RIB), which, in turn, is obtained from the link state database (LSDB) built and maintained via the routing protocols.

G-FIB (or *Global FIB*): This is the FIB constructed by running CSPF on the *entire* LSDB at a node, and therefore takes into account the IP topology of the network. This FIB allows packets to be forwarded over both P links and IPCCs.
P-FIB (or *Packet FIB*): This is the FIB constructed by running CSPF on just those LSAs in the LSDB which describe P links. Thus, this FIB allows packet to be forwarded only over P links (and not IPCCs).
T-FIB (or *TDM FIB*): This is FIB constructed by running CSPF on just those LSAs in the LSDB which describe IPCC links. Thus, this FIB allows a packet to be forwarded only over IPCCs.

With this background, we now present a reference architecture for the peer model, which is shown in Figure 2. The figure illustrates a TDM cloud surrounded by an IP cloud, with the interior and border nodes marked. We also illustrate the types of FIBs that each of the four types of network nodes (the interior and border nodes in the T and P domains, respectively) requires.

Figure 3 illustrates the IP topology, the P domain topology, and the IPCC topology for the network in Figure 2.

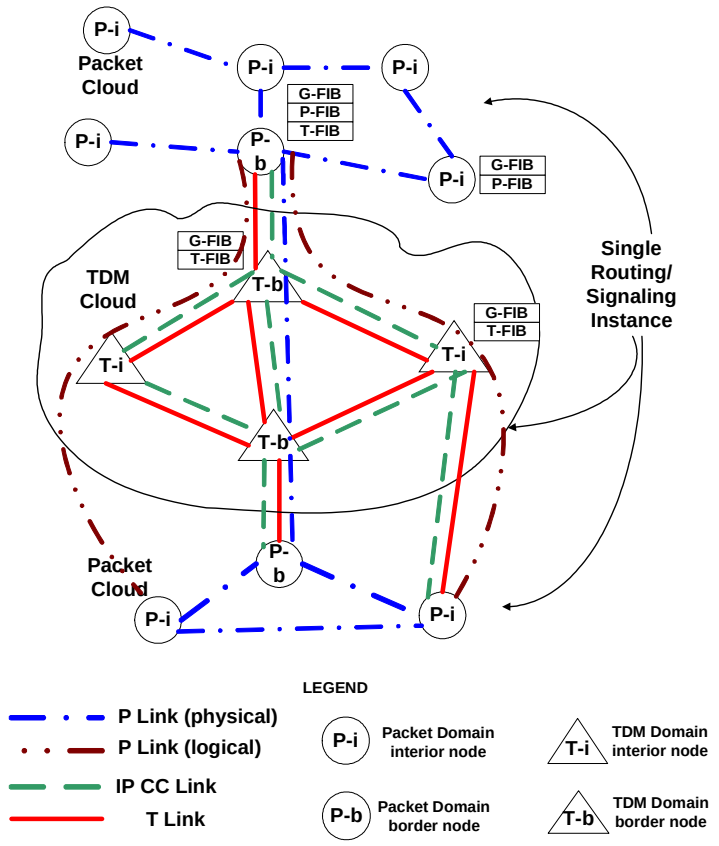

Figure 2. A reference architecture for the Peer Model.

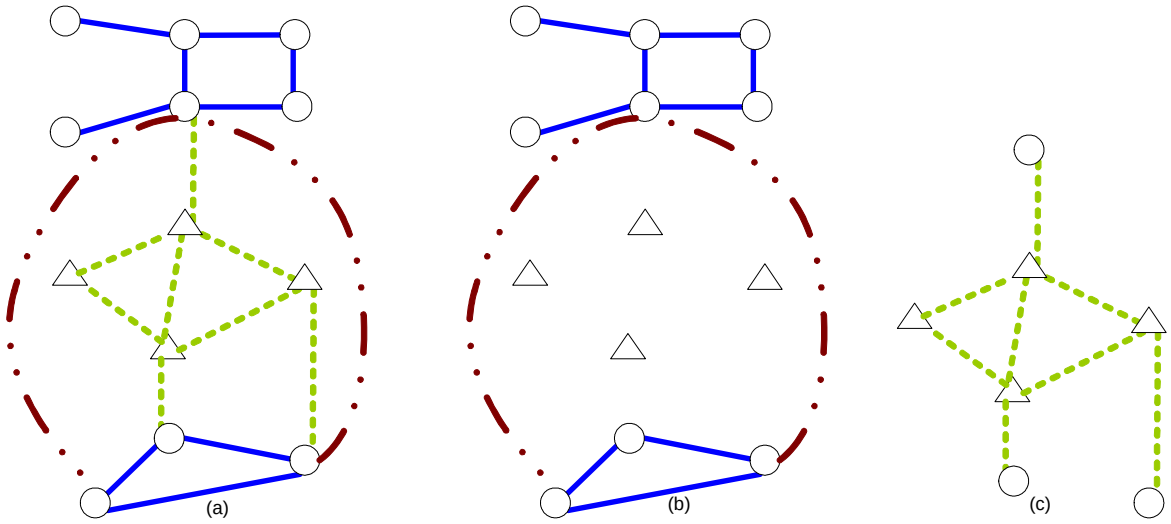

Figure 3. (a) *IP topology* for the network in Figure 2. (This is the same as the control and management plane topology.) This is used to construct the G-FIB. (b) *P domain* topology for the network in Figure 2. This is used to construct the P-FIB. (c) *IPCC topology* for the network in Figure 2. This is used to construct the T-FIB.

4 Operation of the Peer Model

The fundamental consideration in the proper operation of the peer model is that network nodes must be able to correctly forward both data and control packets. As mentioned in Section 3, in the peer model, the specific topology (set of links) over which a packet has to be forwarded depends on the type of traffic that a node has to forward. This means that the node must be capable of classifying incoming packets appropriately to decide which FIB to use to forward each packet. This requires a modification of the forwarding plane architecture as well as the control plane architecture and routing protocols.

4.1 Forwarding Plane Architecture

Since the peer model uses a single instance of routing and signaling, all IP addresses inside the network must be unique. An IP address can therefore belong to either the TDM domain or the IP domain, but not both. This uniqueness gives us an easy mechanism to use the destination address of an IP packet to deduce the domain (T or P) within which the packet's destination lies. This is accomplished by having each node maintain an *IP address-to-domain mapping*. (One way to build this mapping automatically, without having to configure it at every node, is to carry domain information in the IGP LSAs advertised for every link.) This is important, as we shall see later, in classifying the type of an incoming packet, and consequently in deciding which FIB to use at a node for a particular packet-forwarding operation.

To explain the working of the peer model, we view an incoming packet as falling into one of four *packet types*: IP data packets, control packets (routing and signaling packets), management packets (telnet, ftp, or TL1 commands), and ping packets. (Note that since the "ping" operation is used to test both control and data channel reachability, as we explain shortly, we will need to treat "ping" separately from other management traffic). Further, we view an incoming packet as belonging to one of eight *traffic classes*: IP data traffic, routing traffic, P- or T-domain destined signaling traffic, P- or T-domain destined management traffic, and P- or T-domain destined ping packets. The *traffic class* of a packet is a function of its *packet type* and *destination domain* (that is, the domain, P or T, in which its destination is located).

Figure 4 and Figure 5 illustrate how the traffic class of a packet is determined at a P and T node, respectively. The protocol and (optionally) the TCP/UDP port tuple of the incoming packet is used to obtain the *packet type*, which tells whether the packet is a routing, signaling, management or ping packet.

At the same time, the destination address to domain mapping function is used to obtain its destination domain. That is, whether the packet is destined to an address in the P or T domain. Having obtained the traffic class of a packet, the node uses the rules shown in the figures to look up the appropriate FIB.

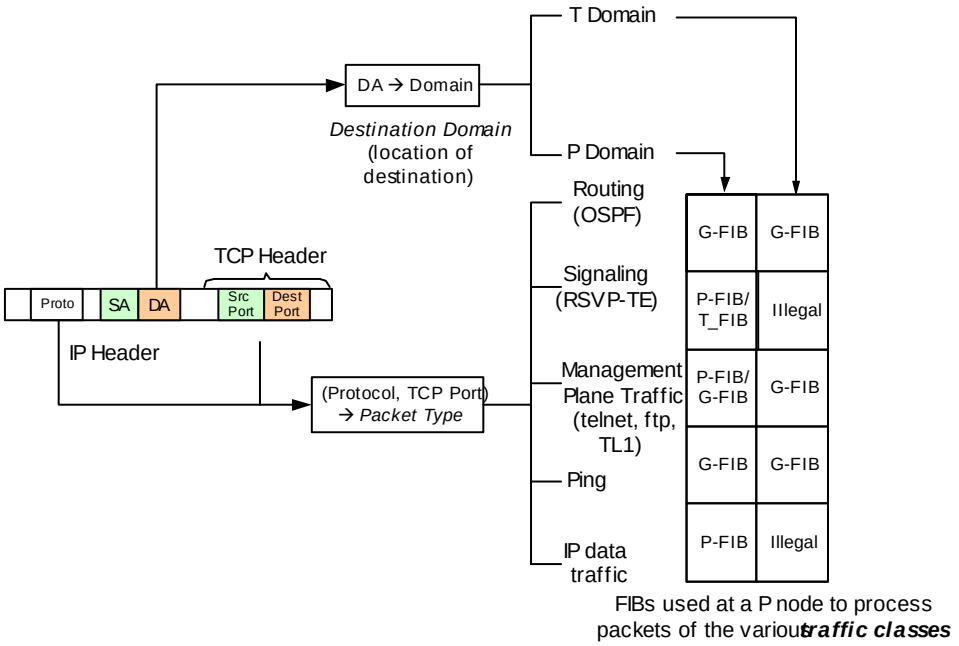

Figure 4. Packet classification and processing at a *P node*. The table on the right shows the FIBs that are consulted at the P node. Note how this agrees with the FIBs maintained at the border and interior P nodes shown in the reference architecture of Figure 2. Note that an "Illegal" entry indicates that a packet of that class should not have arrived at a P node. If it did, this is a misrouted packet and must be dropped.

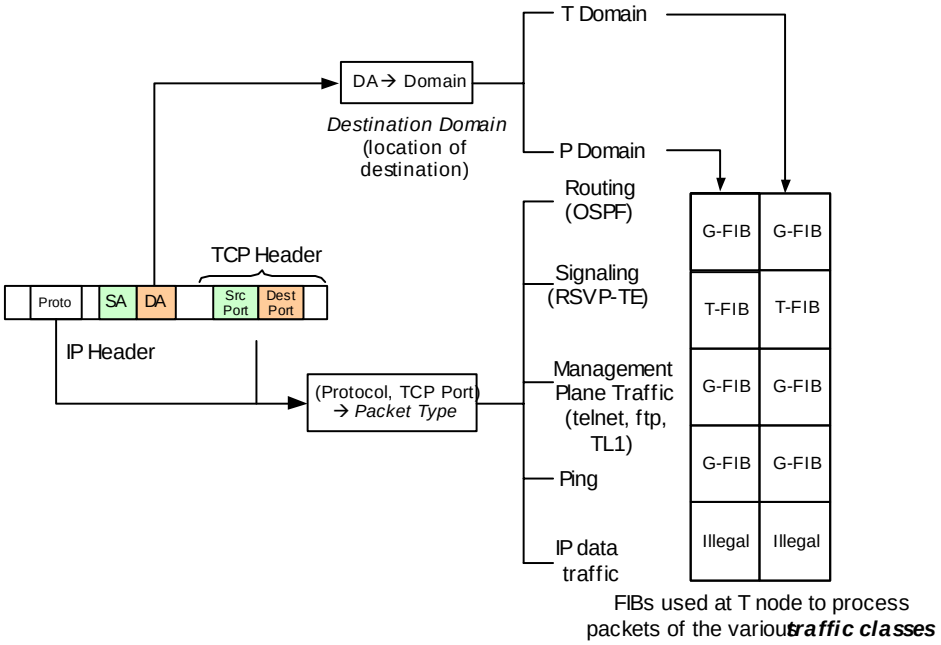

Figure 5. Packet classification and processing at a *T node*. The table on the right shows the FIBs that are consulted at the T node. Note how this agrees with the FIBs that are maintained at the border and interior T nodes

shown in the reference architecture of Figure 2. Note that an "Illegal" entry indicates that a packet of that class should not have arrived at a T node. If it did, this is a misrouted packet and must be dropped.

Next, we discuss the operation of the P and T domain nodes.

4.1.1 Operation of P-domain Nodes

The forwarding plane of a P node has to maintain multiple logical FIBs. An interior P node needs to maintain a P-FIB and a G-FIB, while a border P node must maintain a T-FIB in addition to the other two. This is because an interior P node requires the P-FIB to forward IP data packets, and a G-FIB to forward management plane and routing traffic (see Figure 4; we discuss the case of "ping" packets shortly). The border P node has to maintain the T-FIB so that it may route signaling packets (for setting up a forwarding adjacency) over the TDM infrastructure. We now consider how a P node forwards the various packet types.

All routing traffic is forwarded using the G-FIB, because a single instance of routing is used across the P and T domains. Likewise, all signaling (e.g., RSVP-TE [10]) traffic that is directed at another P node should be limited to traversing only P nodes (with one exception to be discussed next), since MPLS tunnels cannot traverse non IP-capable links. Thus, signaling traffic must use the P-FIB for forwarding. The only exception is signaling traffic that is exchanged between two border P nodes for the purpose of setting up a *forwarding adjacency* between them. Such traffic has to be explicitly routed over T domain nodes, and so should be forwarded using the T-FIB at the first P node. (Note that subsequent nodes, being T nodes, would automatically use the T-FIB for forwarding such traffic, until the packet emerges at the border P node at the far end.)

All management traffic destined to a T node, such as telnet, ftp, or TL1 commands, should be routed using the G-FIB, because there is no issue with using a P link to forward management packets (since, unlike IPCCs, the P links are not bandwidth limited), if that link happens to fall on the shortest path. An implicit assumption in the above of course is that there are no malicious users in the network who open a telnet or ftp session to an IPCC address with the intent of congesting IPCC bandwidth. That is, all telnet, ftp, or other management traffic destined to a T domain address is assumed to be valid, low-bandwidth traffic. For management traffic destined to a P node, it is preferable that no IPCCs be used (because of the limited bandwidth of the IPCCs and the resultant potential for non-T domain related telnet and ftp applications to clog the IPCCs). Thus, for such traffic a P node must consult the P-FIB.

The forwarding of IP data traffic (which can only have a destination address in the P domain) should not use any IPCCs due to their limited bandwidth. (IPCCs are typically low bandwidth tunnels or IP-over-DCC channels, which should be restricted to carrying control and management traffic specific to the T domain.) To forward such traffic, the node must therefore use the P-FIB, which is comprised only of P-links.

Ping packets are treated differently from other management applications, for the following reason. In a traditional IP network, the control plane and data plane topologies are identical, since each IP link carries both control and data traffic. Ping can thus be used to troubleshoot both IP control-plane and IP data-plane connectivity in a network. In a peer network, however, the presence of IPCCs makes the topology of the control plane a superset of that of the IP data plane[5]. Ideally, therefore, we would require *two separate* versions of ping, one to test IP control channel reachability and the other to test IP data channel reachability. One version of ping would thus use the IP topology to route ping packets, while another

[5] Note that this is true in our case where we have assumed the existence of only the IP and TDM domains, and within the TDM domain have assumed that the IPCCs lie along the TDM links. In a more general setting, the IPCC topology may be disjoint from the IP data plane, since the IPCCs may be realized over a data communication network (DCN).

version of ping would only use P links (and not IPCCs). Adding a new ping-like protocol that would provide data plane connectivity for just the peer model is probably impractical. So, we propose using ping to test only control plane reachability. This means that all ping packets must be routed using the G-FIB. This implies that even if there is no P link between two routers connected by a TDM cloud with IPCCs, ping packets will be able to use the IPCCs to successfully travel between the nodes. This is in line with the philosophy of the peer model, one of whose goals is to unify network management.

4.1.2 Operation of T-domain Nodes

The forwarding plane of a T node has to maintain only two FIBs, the G-FIB and the T-FIB. Since a T node never forwards IP data packets to P nodes, it does not require a P-FIB (see Figure 5). Assuming that the P-domain nodes have successfully implemented their FIBs and their forwarding rules, IP data traffic should never reach a T node, so we need not deal with how it to forward it.

As in the case of P nodes, all routing traffic is forwarded using the G-FIB. Signaling traffic destined to P nodes (with the exception of GMPLS signaling traffic for setting up a forwarding adjacency between two P nodes), should never reach a T node. Signaling traffic destined to a T node should of course be forwarded using the T-FIB, since it must travel over IPCCs. (Observe that T domain destined signaling traffic should not be forwarded over P links since the circuit being established by such signaling will only traverse T links. Thus, the corresponding signaling traffic (RSVP-TE messages, for example) also should be restricted to following the same sequence of T nodes through which the data forwarding path passes. Otherwise, we could have TDM circuit related signaling state at P nodes, which is an error.) The only signaling traffic destined to P nodes that would be seen at an intermediate T node is one which is used to establish a forwarding adjacency between two P nodes across the T domain. Since such traffic is normal signaling traffic as far as the T domain is concerned, it too should be forwarded using the T-FIB.

All management traffic, such as telnet, ftp, or TL1 commands, whether destined to a P or T node should be routed using the G-FIB, as before.

Ping traffic in the T domain also follows the same forwarding rules as other management traffic. In other words, the G-FIB is used for routing ping packets. Note that any ping traffic originating from the T domain, of necessity, only tests control plane connectivity (there is no way for T nodes to test IP forwarding plane connectivity) so, unlike the P domain, there is no issue with using the G-FIB to route ping packets.

4.2 Control Plane Architecture and Protocol Enhancements

In this paper, we do not focus on the standard GMPLS enhancements to OSPF-TE and RSVP-TE that support topology discovery and path establishment across TDM nodes, since these changes, while necessary for IP-controlled integrated optical networks, are not specific to the peer model, have been discussed extensively elsewhere (e.g. [9], [10]), and are not central to our discussion here.

An important consideration in the peer model is to distinguish between P-links and IPCC links when routing packets. This is essential not only because some traffic must necessarily traverse only P domain links (such as signaling traffic for MPLS LSP setup) but also because, in a peer network, it is essential to protect IPCCs from being congested by high bandwidth data traffic. We realize this in our implementation by maintaining, at each node, separate FIBs for the IPCC links and P-links, as described in the previous section. We implicitly assumed, however, that each of the network nodes was able to process routing advertisements in a way that allowed it to distinguish between the two types of links and construct the various FIBs it requires. This is not automatic, however, and requires enhancements to the OSPF-TE LSAs as well as to the rules for processing them at a node.

Segregation of normal IP data traffic from the IPCCs requires that all of the network nodes adhere to the following two rules.

Rule1: When constructing the routing topology, keep the OSPF-TE LSAs describing P-links and IPCCs separate from one another.

This can be achieved by having, at each node, a way to maintain logically distinct LSDBs describing: the IP topology, the P domain topology, and the T-domain topology, respectively. The node has to run different instances of the SPF/CSPF algorithm on the appropriate LSDB to derive the G-FIB, P-FIB, and T-FIB, respectively. Note that as explained in Section 4, a particular node may have to maintain only some subset of these three FIBs.

This rule ensures that a node never executes an SPF/CSPF to construct a P-FIB on the mistaken assumption that a remote IPCC is a P-link that can carry IP data traffic. Note that a particular node A that violates this rule (while adhering to Rule 2 to be outlined next) cannot cause an error in the FIB computations of a remote node B. It can only create incorrect FIBs itself, and thereby send data packets destined for an IP address that it (erroneously) believes is reachable, but that is in reality reachable only via an IPCC (and thus not a valid destination for IP data packets). This can cause node A's data packets to go into a black hole, when they are dropped by a remote node B that does not have an entry in its P-FIB corresponding to the IPCC that node A assumed was a valid P link.

Rule 2: Enable every receiving node to differentiate between OSPF-TE-LSAs describing IPCCs from those describing P links.

This rule in fact allows Rule 1 to be realized at each network node. Note that a transmitting node A that violates this rule can potentially create an error in the FIB computations and resultant SPF calculations of a remote node B. This is because a remote node B that could not distinguish the different types of LSAs describing links at node A may perform SPF computations assuming that some IPCCs were bona fide P-links (that is, packet-capable data channels). Correspondingly, assuming that node A follows Rule 1, its violation of Rule 2 will not result in an error in its own FIB computations (assuming, of course, that node A itself can distinguish between LSAs describing its IPCCs from its P links).

In the peer model, Rule 2 can be realized by introducing a flag in the OSPF-TE LSA that distinguishes LSAs describing IPCCs from those describing P links. (This may be achieved by extending the code points of the Link Type field [15] in the Router LSA used in OSPF-TE.) Once a receiving node can properly classify a received LSA, Rule 1 can be realized by maintaining logically separate LSDBs for the IPCCs and P links, respectively, and by running a separate instance of SPF/CSPF on the appropriate LSDB to derive the T-FIB and P-FIB, respectively. The G-FIB can be obtained by running an SPF/CSPF on an LSDB that is the union of the two logical LSDBs.

Recall that the packet classification process described in Section 4 relied on the destination address-to-domain mapping. This mapping needs to be built at each node, either through static configuration or automatically. An easy way to propagate such a mapping automatically through OSPF-TE is to introduce another bit in the Link Type field of the Router LSA to distinguish the domain, T or P, from which an LSA originated. In this way, a receiving node can immediately identify the domain to which the originating end of the LSA belongs.

5 Analysis of our Proposed Solution: Complexities and Feasibility

In this section, we evaluate the complexity and feasibility of our proposed solution for the peer model. We specifically concentrate on the changes required in the IP control and forwarding plane in a P-domain

or T-domain node. The large installed base of routers requires that any solution for the peer model needs to be evaluated in the context of the complexity involved in upgrading the control and data plane of these routers. We also look at the minor protocol additions we proposed to OSPF-TE and their feasibility.

5.1 Forwarding Plane Changes

A major change in the forwarding plane to support the peer model is the use of multiple FIBs at a node (a G-FIB and P-FIB at packet nodes, and a G-FIB and T-FIB at TDM nodes). In addition, the data path processing of an IP packet (at both a P and a T node) now requires a multi-field classification operation prior to the FIB lookup. This classification must be based on the protocol type, incoming link, and destination address, as illustrated in Figures 4 and 5, and requires a policy-driven classifier in the data path, which is used to ascertain the appropriate FIB to use for forwarding the packet.

For a TDM node, which does not expect to forward packets at a fast rate (typically TDM control channels are low bandwidth links that only carry control and management packets), all packet classification and forwarding can be performed in software. So the additional packet classification and multiple FIBs that we propose in this paper, require no changes to the data plane, and are easily implementable, with relatively low overhead, in a software-forwarding engine.

For a P-Node (an IP router), the changes we propose do impact packet processing in the fast forwarding path. Any additional processing related to using multiple FIBs and packet classification to choose the correct FIB can adversely influence a router's data forwarding rate. Advances in hardware, however, have enabled many current routers to maintain and dynamically select between multiple FIBs in their data planes. Virtual Routers and VRFs (Virtual Route Forwarding tables) for example are implemented in several service-provider edge routers and used to offer customers VPN services[11], [12], [13].

The other operation we are concerned with is the classification operation, performed on an incoming packet, in the fast path, to select a particular FIB. Most current routers allow for configuration of "access control lists" in the fast forwarding path to selectively allow or deny an incoming packet to be forwarded based on a number of layer-3 or layer-4 parameters in the packet's headers [14]. Our additional multi-field classification for incoming packets therefore would not require substantial changes in a router's forwarding plane to support it.

We therefore observe that for TDM nodes, our forwarding plane changes can be entirely supported in software and are easily implementable. For IP nodes, the changes that we require in a router's forwarding plane can be built on top of existing features in current routers, which are already in place for other applications (such as VPNs and access lists).

5.2 Control Plane and Protocol Changes

In addition to the GMPLS enhancements already proposed to routing and signaling, the major change involved in making the peer model work in heterogeneous networks is the use of two bits in the Link Type field in OSPF-TE LSAs. One of these bits is needed to distinguish between IPCC links and P links, while the other bit is needed to identify whether the source of the LSA is a P node or a T node.

Accommodating these two additional bits in an OSPF-TE LSA is not difficult. One way to store the two bits is simply by extending the code points of the Link Type [15] field in the Router LSA used in OSPF-TE. For example, two unused bits in the Link Type octet could be used for our purposes. The node originating these LSAs would simply have to fill in these two bits, based on which domain the LSAs belong to and whether the link they are advertising is an IPCC link or a P link.

At a receiving node, the additional work involved in the control plane would be to: classify all LSAs based on the IPCC/P-link bit, build two separate logical Link State Databases on which to run separate CSPF instances, and obtain the G-FIB and the P-FIB/T-FIB (based on whether the node is a P-domain node or a T-domain node). The control plane can use the second additional bit in the router LSA to dynamically build its destination-address to domain mapping table.

Conceptually, our proposal, which requires maintaining separate logical Link State Databases and running different SPF/CSPF instances on them to obtain the multiple FIBs, is not very different from the virtual router model used for VPNs [12]. In this model, one creates multiple virtual routing instances to obtain separate routing tables for each customer VPN. Many current routers are capable of managing multiple routing instances for VPN services, so in this regard also our proposal appears feasible.

Our analysis of the control-plane, the forwarding-plane and the protocol changes proposed indicate that from a technical or engineering perspective, it is indeed feasible to interwork IP signaling, routing, and packet forwarding between P-domain and T-domain nodes in a peer-model configuration. However, in this paper we only provide an architectural solution to allow service providers to create and operate a peer-model based, unified control plane for heterogeneous TDM/packet nodes. Actual deployment of the peer model in heterogeneous networks is a decision that requires additional analysis by the service providers regarding traffic patterns, network management, security and cost issues, which we have not addressed in this paper.

Other practical factors pertinent to deployment of the peer model would also need to be considered, despite some important advantages of the peer model, as specified in section 1.3. For instance, TDM equipment in a carrier network would need to be enhanced with a standardized IP-based control plane (in an overlay model, each domain could potentially have its own proprietary control or management plane and only the border nodes would need to run standardized IP based signaling and routing protocols). Using the same routing and signaling instance for packet and TDM domains could possibly accentuate network security and robustness concerns, from a service provider's point of view. Furthermore, a network service provider might want to keep TDM and packet network interaction to a minimum, for administrative or billing purposes.

The aim of this paper, however is to allow for a clear understanding of the architectural issues in building an intelligent, unified control plane, which would facilitate service providers in their decision process to deploy the peer or overlay model.

6 Conclusions

The overlay model of network operation has been implemented, deployed, and proven to working the classical IP over ATM context, and there has also been significant work in the optical networking on the overlay model for integrated optical networks. The peer model for heterogeneous optical networks however, has not been discussed or analyzed in detail in the literature. We discussed some of the issues in operating the peer model, identified the pros and cons of this model, and analyzed the additional changes needed to the protocols and the optical transport and IP devices to create a functional network based on the peer model. As we observed, the peer model involves a different dimension of complexity relative to the overlay model, with the complexity shifting from the control plane to the forwarding plane. However, as we demonstrated, several of these changes appear to be implementable with current router technology. We believe that the final decision about the model to adopt lies with the service providers. Our work in this paper contributes to a clearer understanding of the peer model, and is designed to help service providers make this determination.

Several extensions of our work are possible. The extension to more than two domains is one which we believe is fairly simple. Examining the peer model in the context of other IGPs is another. Finally, accounting for *inter-domain* operation with multiple IGP domains is an issue that will become increasingly important for inter-carrier operations. This is a bigger challenge, however.

7 References

[1] "The Cisco IP+Optical Unified Control Plane: Accelerating Service Velocity with IP-enabled Provisioning," Cisco Whitepaper, http://www.cisco.com/warp/public/779/servpro/solutions/optical/docs/ucp_wp.pdf

[2] "The Unified Control Plane for IP and Optical Networks," Cisco Whitepaper, http://www.cisco.com/warp/public/cc/pd/olpl/metro/on15327/prodlit/ipto_wp.pdf

[3] "A Management Architecture for IP-over-WDM integration," http://www.terena.nl/conferences/tnc2001/proceedings/PaperKarayannis.pdf

[4] "The Evolution of Telecommunications Transport Architecture: from megabits to terabits," http://www.iee.org/Publish/Support/Auth/ece0101-33.pdf

[5] "Optical Fiber Communication: From transmission to networking" http://www.comsoc.org/livepubs/ci1/public/anniv/rama.html

[6] Interworking with the Intelligent Optical Layer http://www.lightreading.com/opticalintellect/document.asp?doc_id=67&mode=print

[7] V. Sharma, A. Das, C. Chen, "Leveraging IP signaling and routing to manage UPSR-based SONET networks," submitted *ICC 2003*, August 2002.

[8] D. Katz, D. Yeung, K. Kompella, "Traffic engineering extensions to OSPF version 2," draft-katz-yeung-ospf-traffic-07.txt, http://www.ietf.org/internet-drafts/draft-katz-yeung-ospf-traffic-07.txt August, 2002.

[9] Kompella, K., et al, "OSPF extensions in support of Generalized MPLS," draft-ietf-ccamp-ospf-gmpls-extensions-08.txt," http://www.ietf.org/internet-drafts/draft-ietf-ccamp-ospf-gmpls-extensions-08.txt, August 2002.

[10] L. Berger (Editor), "Generalized MPLS: Signaling – RSVP-TE extensions," draft-ietf-mpls-generalized-rsvp-te-08.txt, http://www.ietf.org/internet-drafts/draft-ietf-mpls-generalized-rsvp-te-08.txt, August 2002.

[11] B. Gleeson et al, "A Framework for IP Based Virtual Private Networks" RFC2764, http://www.ietf.org/rfc/rfc2764.txt

[12] Hamid-Ould Brahim et al, "Network based IP VPN Architecture using Virtual Routers", draft-ietf-ppvpn-vpn-vr-01.txt, http://www.ietf.org/proceedings/01dec/I-D/draft-ietf-ppvpn-vpn-vr-01.txt

[13] E. Rosen and Y. Rekhter, "BGP/MPLS VPNs", RFC2547, http://www.ietf.org/rfc/rfc2547.txt

[14] P. Morrissey, "Demystifying access control lists", http://www.networkcomputing.com/907/907ws1.html

[15] J. Moy, "OSPF Version 2" RFC 2328, October 1998.

[16] G. Swallow, "MPLS advantages for traffic engineering," *IEEE Commun. Mag*, Vol. 37, Issue 12, December 1999, pp. 54-57.

[17] D. Awduche, Y. Rekhter, "Multiprotocol Lambda Switching: combining MPLS traffic engineering control with optical crossconnects," *IEEE Commun. Mag*, Vol. 39, Issue 3, March 2001, pp. 111-116.

[18] G. Bernstein, J. Yates, D. Saha, "IP-Centric Control and Management of Optical Networks," *IEEE Commun. Mag.*, Vol. 38, Issue 10, Oct. 2000, pp. 161-167.

[19] G. Bernstein, E. Mannie, V. Sharma, "Framework for MPLS-based Control of SDH/SONET Optical Networks," *IEEE Network*, July/August 2001, pp. 20-26.

[20] E. Mannie, D. Papadimitrou (Editors), "Generalized Multi-Protocol Label Switching Extensions for SDH/SONET Control," Work in progress, Internet Draft, draft-ietf-ccamp-gmpls-sonet-sdh-08.txt, February 2003.

[21] A. Banerjee, et al, "Generalized Multi-protocol Label Switching: An Overview of Signaling Enhancements and Recovery Techniques," *IEEE Commun. Mag.*, Vol. 39, Issue 7, July 2001, pp. 144-152.

[22] A. Banerjee, et al, "Generalized Multi-protocol Label Switching: An Overview of Routing and Management Enhancements," *IEEE Commun. Mag.*, Vol. 39, Issue 1, Jan. 2001, pp. 144-150.

[23] E. Mannie, "Generalized Multi-Protocol Label Switching Architecture", draft-ietf-ccamp-gmpls-architecture-03.txt, http://www.ietf.org/internet-drafts/draft-ietf-ccamp-gmpls-architecture-03.txt

[24] Optical Internetworking Forum, "OIF-UNI 1.0 Signaling Specification" http://www.oiforum.com/public/documents/OIF-UNI-01.0.pdf