AN OVERVIEW OF OPTICAL NETWORK BANDWIDTH AND FAULT MANAGEMENT

JUNAID AHMED ZUBAIRI

Department of Computer Science, State University of New York at Fredonia, Fredonia NY 14063, USA

zubairi@fredonia.edu

ABSTRACT: This paper discusses the optical network management issues and identifies potential areas for focused research. A general outline of the main components in optical network management is given and specific problems in GMPLS based model are explained. Later, protection and restoration issues are discussed in the broader context of fault management and the tools developed for fault detection are listed. Optical networks need efficient and reliable protection schemes that restore the communications quickly on the occurrence of faults without causing failure of real-time applications using the network. A holistic approach is required that provides mechanisms for fault detection, rapid restoration and reversion in case of fault resolution. Since the role of SDH/SONET is diminishing, the modern optical networks are poised towards the IP-centric model where high performance IP-MPLS routers manage a core intelligent network of IP over WDM. Fault management schemes are developed for both the IP layer and the WDM layer. Faults can be detected and repaired locally and also through centralized network controller. A hybrid approach works best in detecting the faults where the domain controller verifies the established LSPs in addition to the link tests at the node level. On detecting a fault, rapid restoration can perform localized routing of traffic away from the affected port and link. The traffic may be directed to pre-assigned backup paths that are established as shared or dedicated resources. We examine the protection issues in detail including the choice of layer for protection, implementing protection or restoration, backup path routing, backup resource efficiency, subpath protection, QoS traffic survival and multilayer protection triggers and alarm propagation. The complete protection cycle is described and mechanisms incorporated into RSVP-TE and other protocols for detecting and recording path errors are outlined. In addition, MPLS testbed configuration procedure is outlined with suggested topologies. Open issues in this area are identified and current work is highlighted. It is expected that this paper will serve as a catalyst to accelerate the research and development activities in high speed networking.

1. TRAFFIC GROOMING IN OPTICAL NETWORKS

Newer web based applications are pushing the network bandwidth demands higher and higher. Optical networks can transfer information at a very high rate with little or no errors. The dependability and reliability of optical networks has resulted in their increased deployment across the world. Optical networks transfer information through lightpaths. A lightpath is a wavelength routed optical channel that runs at 10Gbps and a fiber may contain hundreds of lightpaths. This explains the reason behind the current interest in optical networks.

Traffic Engineering issues in optical networks are unique. For example, we cannot take low speed microflows and assign one wavelength to each microflow. Without traffic grooming, most of this huge bandwidth may be wasted. Traffic grooming in WDM mesh networks is an active area of work. One way to groom the traffic is to accumulate the electronic bits in a buffer and transmit the buffer contents in a burst on optical side once it reaches a threshold size. It is known as OBS (Optical burst switching). This scheme counts on the presence of electronic buffers. In OPS (Optical packet switching), a packet is transmitted as soon as it is received from the IP layer and no buffers are deployed. This approach will work as long as there is no contention. In case of contention, the wavelength for one of the packets may be changed. This solution requires wavelength converter and thus it may prove to be expensive. Another solution is to delay one packet through the delay line or use path deflection to let the network change the path.

The end-to-end all optical network is quite enticing because of low overhead, huge bandwidth and little or no errors. However, the signal to noise ratio increases with the distance. Bit errors accumulate along the lightpath, making it unusable. This problem is not present in O-E-O network because of the fact that in an O-E-O network, optical signal is regenerated at intermediate nodes. In an O-O network, the quality factor must be considered and impairment aware routing should be performed for wavelengths.

2. TRAFFIC ENGINEERING: THE MPLS FACTOR

MPLS (Multiprotocol Label Switching) is developed [1,2] for automated control and management of integrated networks. MPLS can combine various traffic streams inside a single LSP (Label switched path). Since an LSP receives similar routing and priority treatment across a domain, specific level of performance can be assured with an upper bound on delay, jitter and loss through the network.

MPLS combines the L3 routing and L2 switching into "L2.5 forwarding". It attaches its shim header between the headers of layer 2 and layer 3 protocols. It works on the principle of providing a "virtual path" from an ingress node to an egress node in an MPLS domain. This idea of providing a virtual path is not new and it has been used in the Internet to provide VPN tunnels across the public network and IPv6 tunnels across IPv4 networks. In tunneling, the packets originating at a router and destined for a specific node are labeled in such a way that the intermediate routers forward them towards a common destination through the same path. Thus, tunneling implicitly introduces the notion of a connection because all packets in the tunnel follow the same path and experience the same routing through the network.

MPLS introduces connection oriented LSP's in a connectionless network. Traditional Internet routers are connectionless devices. When a packet arrives at an intermediate router, the router selects the best route for the packet and then forwards it on the link that falls on the selected route. In MPLS, the route for a flow is decided in advance of transmitting it and all the packets that belong to the flow are tagged with a label consisting of LSP ID and relevant information. Due to advance path computation, the intermediate routers can forward the traffic on the fly without any processing delays. Thus the role of intermediate routers is simplified to switching. In MPLS, the paths can be mapped on ATM network because ATM also selects paths before start of transmission and the labels can be matched to VPI/VCI numbers. The network administrators can manually lay down paths through the MPLS domain based on the organization policy. The LSP's are allocated to requesting flows that may enter the domain via an ingress node and exit via an egress node. Once an LSP is installed, it is monitored and terminated when the flow transmission has been completed. All packets within this LSP are treated in a similar manner. One of the most important capabilities of MPLS is to map traffic on paths established with traffic engineering principles, resulting in load balancing and fault tolerance of the underlying network.

MPLS labels are stackable and thus it does not suffer from scalability issues. Let us analyze this benefit by assuming that a router is required to track all the flows that use its attached links. As the number of flows increases, the amount of work in tracking the flows increases rapidly thus overloading the router. In ATM, the amount of work is controlled by using two levels of hierarchy. Several virtual circuits can be placed inside a virtual path. The scalability issue does not arise in MPLS due to the fact that the label switched paths (LSP's) can be organized in a hierarchy of arbitrary levels. Figure 1 section-A shows an example of placing 3 virtual circuits inside a virtual path in ATM. Section-B of Fig. 1 clearly demonstrates the scalability of MPLS by stacking three LSP's inside LSP 20. While creating a higher level, a new label can be pushed on the packet. The packet will be handled throughout the current MPLS domain using the top label. At the penultimate router, the top label will be popped and the packet will be treated according to the next top label. This definition essentially creates a stack of labels on a packet. When all the labels have been popped, the packet will be forwarded based on the IP routing principles. Thus MPLS manages to seamlessly integrate newer MPLS domains with older IP based subnets.

MPLS can be combined with Diffserv to realize Diffserv-aware MPLS network. In such a network, the FEC (forwarding equivalence class) is the common factor between the Diffserv and MPLS. In MPLS, the packets that are within the same FEC are assigned the same label and thus experience the same routing through the network. In a Diffserv-aware MPLS domain, packets with the same FEC also fall within the same Diffserv class, as inferred from the TOS field in the IP header.

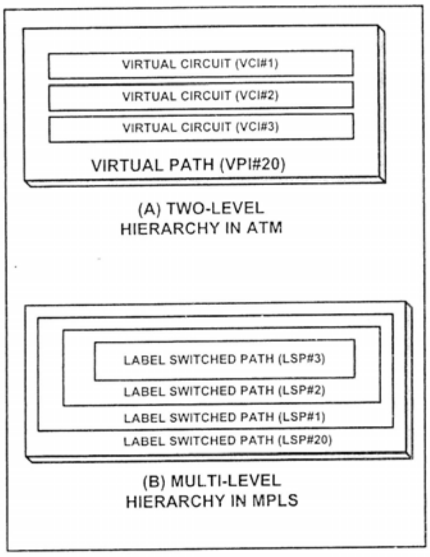
VIRTUAL CIRCUIT (VCI#1)
VIRTUAL CIRCUIT (VCI#2)
VIRTUAL CIRCUIT (VCI#3)
VIRTUAL PATH (VPI#20)
(A) TWO-LEVEL HIERARCHY IN ATM
LABEL SWITCHED PATH (LSP#3)
LABEL SWITCHED PATH (LSP#2)
LABEL SWITCHED PATH (LSP#1)
LABEL SWITCHED PATH (LSP#20)
(B) MULTI-LEVEL HIERARCHY IN MPLS

Fig. 1: ATM and MPLS Hierarchies

3. FROM MPLS TO GMPLS

Since the optical modules forward the data based on wavelength, MPLS does not fare well in optical communication. The earlier efforts to incorporate wavelength based switching resulted in a proposed name change from MPLS to MPλS. It was soon realized that a general protocol is needed to address the needs of devices that do not deal with individual packets. Later, the IETF released a draft [3] that suggested extending MPLS to GMPLS or Generalized MPLS. The target of GMPLS is to provide a generalized control plane that can be used to perform switching based on packet labels or labels that identify wavelengths, time slots or ports. Thus GMPLS extends MPLS in order to provide support for modules that cannot identify cell and packet boundaries. The following types of modules are identified as part of a node in a GMPLS environment.

- PSC (Packet-Switch Capable) For example, switching based on MPLS shim header or ATM's VPI/VCi
- L2SC (Layer-2 Switch Capable) For example, Ethernet bridges
- TDMSC (Time-Division-Multiplex-Switch-Capable) For example, SDH/SONET cross-connect or G.709 interface
- LSC (Lambda -Switch-Capable) For example, interfaces that switch based on a wavelength or waveband
- FSC (Fiber-Switch-Capable) For example, photonic cross-connects that switch one fiber to another

An LSP can only be defined between interfaces of the same type. Since GMPLS allows hierarchical LSP's, it is easy to define a hierarchy that covers all types of interfaces listed above. At the top of the hierarchy are FSC interfaces, followed by LSC interfaces, followed by TDM interfaces, followed by L2SC, and followed by PSC [3]. GMPLS consists of several building blocks most of which are the extended forms of Internet routing and signaling protocols. A new specialized protocol LMP (Link Management Protocol) [4] has been developed for GMPLS. LMP automates the link provisioning and fault isolation tasks. It also provides bundling of links because maintaining link adjacencies is no more scalable when technologies like DWDM (Dense Wavelength Division Multiplexing) are used resulting in thousands of links between adjacent nodes. Wavelength division multiplexing works by transmitting several wavelengths on a single fiber simultaneously with each wavelength allocated a part of the optical spectrum. Scalability solutions in GMPLS include the introduction of FA-LSP (Forward Adjacency LSP) that can be joined by additional lower level LSP's on intermediate nodes. Traffic grooming techniques are applied in order to fully utilize existing FA-LSP's before setting up new tunnels [5].

It is anticipated that the MPLS/GMPLS would replace ATM and WDM/DWDM would replace SDH because of the fact that MPLS provides virtual paths (LSP's) and WDM provides OAM (operation, administration and Maintenance) functions. Therefore it is expected that instead of IP/ MPLS over ATM over SDH/SONET, the future networks will operate with IP/MPLS over WDM. This approach is illustrated in Fig. 2.

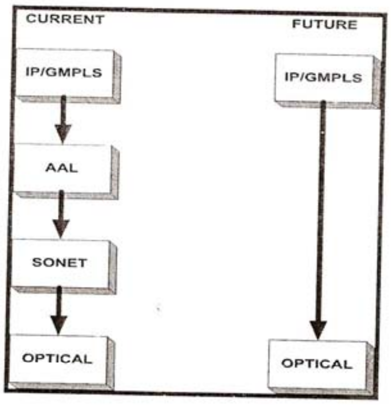

Fig. 2: IP Over Glass Scenario

GMPLS extends the MPLS labels by adding time-slots, wavelengths and port identification. Thus it is possible to identify a single fiber within a bundle, a single wavelength or a time slot from the information contained in the label. GMPLS adapts MPLS routing and signaling messages to the optical network and defines an architecture framework for their functional inter-relationship to achieve effective traffic engineering across both the IP and the optical networks. However, it doesn't define the relationship between the control and the management planes and leaves it to the vendors to implement. This ambiguity leads to sub-optimal utilization of network resources and results in interoperability issues when interconnecting optical .networks of different service providers. It is expected that the research conducted in these topics will accelerate the development of network management solutions for next generation IP optical networks

[6]. Work can be done to define standardized roles and inter-relationships of the control and management planes in GMPLS.

4. THE THREE MANAGEMENT MODELS FOR IP-OPTICAL NETWORKS

The management of integrated IP and optical networks is a technical challenge and its successful implementation is essential for the success of next generation networks [6]. A traditional management oriented approach is the development of the TMF (Tele Management Forum) MTNM (Multi Technology Network Management) interface. Other, more ambitious, approaches push for a control plane that moves many management functions to the network such as the ASTN and ASON frameworks along with the OIF UNI (and the associated GMPLS and LMP protocols).

Three models have been proposed for management and control of the IP-optical networks. In an *overlay network*, each layer controls itself independently from other layers. The IP network behaves like a client and requests transport services from the optical network. The overlay is evolved into *augmented model* where some information is shared between layers. The peer model runs unified protocols across all the network layers and uses an integrated control plane [7].

It is believed that the peer model is the most complex to operate as compared to overlay or augmented models. The peer model divides the network into a single unified control plane and a data plane. Its single GMPLS control plane runs both the IP and Optical layers and treats both as a single integrated network. All IP and optical devices and links have IP addresses and share the same addressing space. In other words, optical devices and IP routers are peers. The same instance of the routing and signaling protocols run across both control planes. A routing protocol such as OSPF or IS-IS with appropriate extensions can be used to distribute topology information. Hence, it is possible for an IP router to signal an OXC (Optical Cross Connect) to set up a lightpath that would carry a group of logical LSPs as defined in MPLS. IP routers do not send explicit requests over a user network interface (UNI) for services (e.g., like setting up an optical path as in the overlay model) to the optical network, rather CR-LDP or RSVP-TE signaling is used to set-up end-to-end LSPs. All of the MPLS traffic engineering principles are applicable with the OXC being equivalent to an LSR and lightpaths considered similar to LSP's. However, modifications to CR-LDP or RSVP-TE are needed to handle specific optical layer devices and links. The peer model provides the IP routers with enough flexibility to specify attributes for LSPs. For example, an LSP or a group of LSPs can be set up with specific high security requirements, a traffic-engineering rule, or other pre-defined QoS policy. The flexibility of this model comes with a price: added complexity. However, it is expected that the networks will eventually migrate to the peer model [8] despite the long standardization effort that it will take.

The main goal of the network management function is to improve the efficiency of the network via automating network control functions, and optimizing the network resources. We discuss how this is achieved using a GMPLS control plane. At a fundamental level there are three basic requirements for the control plane:

1) capability to set-up optical channels in real time (dynamic provisioning),
2) capability to provide protection and restoration upon fault detection,
3) capability to provide traffic engineering with extensions to optical network elements such as fibers, wavelengths, time division switching elements, etc.

The basic architecture of the GMPLS control plane consists of four main modules: resource discovery, state information dissemination, path computation and path management. Resource discovery is responsible for maintaining link adjacencies. State information dissemination propagates information about topology and resource availability. Explicit routing is invoked by the path computation module to select paths. Path management module provides protection, restoration and deletion functions for paths established in the network.

5. GMPLS OPEN ISSUES

One of the main issues is to study the different methods of integrating the control and management planes in GMPLS and state which ones are considered best to provide the expected new services in an efficient manner [6]. In general, GMPLS control plane does not assume the presence of a centralized network manager, though does not preclude its use either. It is based upon distributed control plane where functions are implemented within the nodes without exchanging messages with a global network manager. In that regard, distributed control is not as optimum as a centralized approach that has a global view of the network, but on the other hand it is fast, and "accurate enough" to provide approximate solutions that utilize the network local resources efficiently. Also, centralized management is difficult to implement because of the fact that in a domain, equipment from different vendors may be used, raising interoperability issues. Each LSR, or OXC has a control plane that implements network management functions, and runs the same routing, discovery, and state dissemination protocols and out-of-band dedicated wavelengths or channels can be used to carry and distribute control traffic between the control planes.

There are several issues and questions about the proper implementation of the specific functions and services. Below, a summary of these questions and issues is listed:

- Access Security
- Fault detection, analysis and notification
- Fast P & R (Protection and Restoration)
- Automatic topology discovery
- Rapid provisioning
- Combining various data planes seamlessly
- Managing thousands of paths for operational status and TE-compliance

Security in GMPLS is an issue because the IP header carries globally valid source and destination information as opposed to GMPLS/MPLS where locally valid labels are swapped on each intermediate node. Label switched routers thus cannot enforce authentication based on labels and it must be done on the edge of the network. On the other hand, authentication on the edge adds considerable processing overhead to the edge router.

Most of the network traffic is based on TCP/IP, the most popular protocol suite for the Internet. Since IP is based on packet switching, IP packets must be broken down into ATM cells before being repacked into SONET frames. This situation is unacceptable in the long run as a significant portion of the transmitted bits do not contain useful information. There is a large administrative overhead in dealing with three different technologies compounded by redundant packetizing and de-packetizing. With the development of MPLS, network operation can be simplified to a great extent. IP over MPLS can directly run over DWDM. Before removing SONET/SDH from the protocol stack, we have to keep in mind that SONET has well tested OAM (Operation, Administration and Management) functions that are still under development for packet-based network. The current OAM functions in ATM and IP are inefficient. For example, the time to restore a virtual circuit in ATM runs into several seconds and for an IP route, it may take several minutes before an alternate route is installed between two nodes that are in different autonomous systems. Additionally, IP based restoration may be unable to provide bandwidth recovery for bandwidth sensitive applications [9].

6. OPTICAL NETWORK FAULT MANAGEMENT

Fault detection and notification in the IP over optical networks with SONET uses forward error correction. Fault handling without SONET layer becomes a bit more complicated. Efforts are underway to implement "IP over glass". As shown in Fig. 2, this term refers to removing ATM and SONET from the GMPLS layers and carrying IP directly over WDM links. However, SONET is very popular among the vendors due to its traffic grooming and protection features. The key benefits of SONET are its survivability and its fast response to link and node failures, around 60ms in path restoration. Therefore, it is argued that SONET will continue to be used for some time. The main issue in GMPLS is to provide a SONET-like or better protection and restoration mechanism that can deal with all the layers within the transport plane. Recent extensions in LMP (LMP-DWDM) try to address control and management issues related to link failures and faults. These extensions can to be studied and tested for fault detection. Rapid restoration issues deal with the degree of redundancy provided with the path. In any case, the protection of sub-wavelength flows should not be neglected. There is a strong demand for devising grooming, protection and restoration schemes for the components of a wavelength channel.

6.1 Fault Identification, Localization, Restoration and Reversion

Strong fault management functions must be developed that can handle defects that may arise in MPLS packet based networks. Some of the defects associated with LSP's include [10]:

1) Broken connection to next hop router
2) Mismerging of two LSP's into one
3) Wrong labels causing destinations to be swapped
4) Infinite looping

Broken connection fault is illustrated in Fig. 3. Physically broken links are easy to identify and isolate. All LSP's using the broken link will experience sudden interruption in traffic. An OAM packet traversing any of the affected LSP's can detect the failure of the LSP. The MPLS OAM requirements specify the localization of the fault in addition to detection of failure [11]. Other requirements include measurement of latency, loss and jitter as per SLA, path characterization, alarm suppression; LSP based accounting and preventing denial of service attacks.

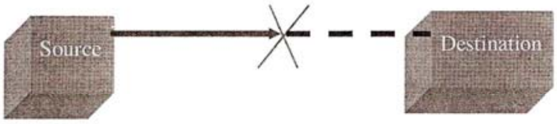

Fig. 3: Broken Connection Effect

Broken connections due to misconfiguration may lead to a more serious problem. The separation of control and data planes in MPLS means the control information is exchanged separately from the data information. Thus, specially formatted OAM packets are sent on the data path. This may lead to a scenario in which an LSP carrying only data packets may experience faults due to misconfiguration but the OAM packets related to this LSP may be traversing a separate healthy LSP due to ECMP (Equal Cost MultiPath) sharing between two nodes [12]. Since control information keeps flowing, LSR's continue to send data on the data LSP's, assuming everything to be correct. This may cause the problem of creating a traffic black hole in the middle of a network domain as shown in Fig. 4. A temporary solution would be to disallow ECMP by resetting the first nibble of MPLS payload to 0x0 or 0x1 instead of 0x4 or 0x6. This is the location where IP version number is expected [13]. Thus the LSP will not be handled with ECMP and the traffic black holing problem will not show up. Additionally, the service latency and jitter will be much more predictable for time-sensitive applications.

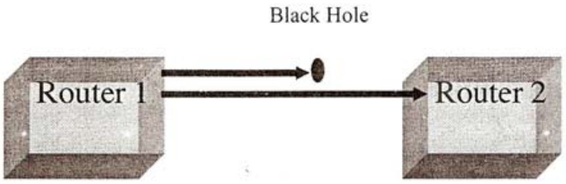

Fig. 4: Black Hole Effect

Misconnection faults may allow one LSP to be connected to a wrong node that cannot swap its label. Figures 5 and 6 depict the misconfigured and swapped LSP scenarios. In either case, the node may simply remove the label and use the IP routing to forward the traffic on this LSR correctly. Thus, this fault will be hard to detect. In a similar way, swapping or mismerging would lead to inability on behalf of the recipient LSR to swap the labels of the packets. The recipient LSP may use IP routing to forward the mismerged or misconfigured LSP's and it may at the same time, inform the ingress of the problem through a special OAM packet. For infinite looping problem (Fig. 7), the well tested TTL (Time to Live) counter can be used which allows the packet to be discarded on the value of zero. Additionally, there should be a mechanism to notify the ingress router about the packet that is killed due to the zero value of TTL counter.

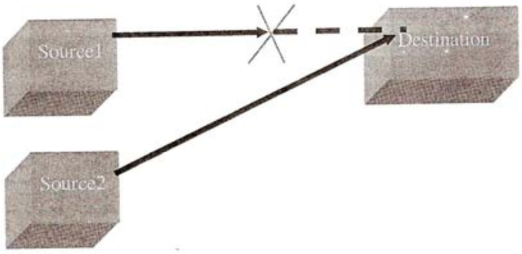

Fig. 5: MisconFig.d LSP

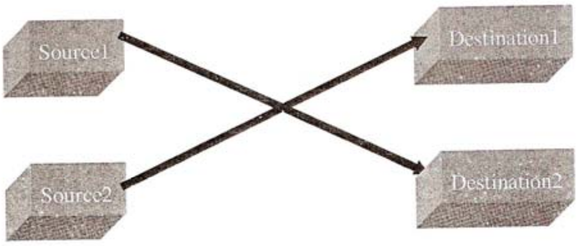

Fig. 6: Swapped LSP Fault

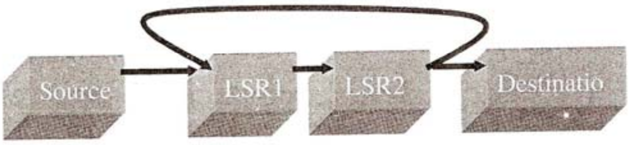

Fig. 7: Looping Fault

Fault localization means identifying the exact faulty node or link. Hop-by-Hop manual localization is tedious and impractical. Therefore tools and techniques are needed to automate the fault detection and localization process. Special OAM packets may be generated by the ingress router and forwarded on all the LSP's that originate at this node. The egress router must be configured to handle the received OAM packets and acknowledge the correctness of the established LSP's. It may be desirable to record the complete path traversed by the OAM packet so that the mismerging or misconfiguration faults can be identified. Each intermediate LSR should be capable of modifying the OAM packet and forwarding it to the next hop.

6.2 Protection and Restoration

High speed optical networks are expected to be HA (Highly Available) systems that can meet the critical resilience requirements of today's real-time and mission critical applications such as VoIP and telemedicine. In order to guard against failures, a number of protection and restoration techniques have been presented. Protection and restoration are separate techniques that yield different results. Restoration reacts to a failure by computing the path around the failed node or link after the failure has occurred. On the other hand, protection is a proactive technique that precomputes and installs backup paths for the high priority LSPs. In most cases, the protection paths may be utilized by best effort traffic that can be preempted if the original high priority path fails. In case of extremely sensitive high priority traffic, protection paths can be used to carry duplicate streams so that the communication is assured even if one of the original LSPs fails.

Following degrees of protection are defined in GMPLS, with the notation of W:B, where W is the working LSP and B is the backup LSP:

1:0 No backup LSP's are provisioned. Signaling will be done on working LSP failure to try to get a backup LSP. It is known as best effort scheme.

1:1 One working LSP is protected by one backup LSP

1+1 Traffic is sent on both LSP's and the best one is chosen at the destination

N:1 N working LSP's are protected by one backup LSP that has been provisioned in advance

N:M N working LSP's are protected by M backup LSP's

1:M One working LSP is protected by M backup LSP's

It may be desirable in an MPLS-Diffserv domain to use the class of service as the criteria for choosing the protection scheme provided that the LSP's are established with same class of traffic. For example, an LSP with EF class of service may be provided with 1:M, 1:1 or 1+1 [10] scheme, the highest levels of protection. In 1:M scheme, several backup paths may be provisioned to protect one original path. In 1:1 scheme, one backup path may be fully reserved for one original path and in 1+1 scheme, traffic may be transmitted on both paths so that the received data with least delay, jitter and loss can be selected or the recipient can switch from the primary path to the backup path on failure notification. The N:1 protection reserves one backup path for N working LSP's and N:M scheme reserves M backup paths for N working LSP's that may belong to various levels of AF class. Finally, the LSP's with best effort traffic may not be protected and they may have to rely on dynamic restoration through higher layers such as IP dynamic routing.

Determination of SRG (shared risk group) and degrees of protection are the key issues in P & R. If two LSP's share a node or a link, it may become a single point of failure for both of them. Therefore, it is important to identify node-disjoint and link-disjoint paths in a network for the set of protection paths. The node-disjoint and link-disjoint paths should belong to different shared risk groups. For example, all the links that share a common medium can be considered as belonging to the same SRG. The determination of SRG can be followed by setting up the protection paths.

Another important issue is to select the layer on which the protection is provided. In the overlay model, IP/MPLS network is controlled separately from the optical network. Thus

SONET/SDH controls the optical network and MPLS controls the upper layer, an overlay over the optical network [15]. The physical links between the ADM(Add Drop Multiplexer) are lightpaths that are configured and protected by SONET/SDH. The logical links between LSR (Label Switched Router) are LSPs (Label switched paths) that are configured for carrying the IP traffic and protected by MPLS and associated protocols. Figure 8 illustrates the overlay model where the four LSPs between LSR1 and LSR2 are carried over a single lightpath between ADM1 and ADM2. Since LSPs can be aggregated due to the stackable labels, there may be hundreds of LSPs using one lightpath. If that lightpath is broken or goes down, MPLS based protection would trigger hundreds of fault handling events. On the other hand, if the optical layer activates a backup lightpath to replace a primary lightpath, only one event would be triggered. Clearly, the optical layer based protection would be preferable to keep the protection overhead to a minimum. Also, only one path needs to be recovered instead of recovering hundreds of LSPs each using a potentially different backup LSP. This results in a faster switchover to the backup path in less than 50 ms. However, on the other hand, the optical layer protection mechanism would reserve one backup lightpath for each primary lightpath. This would result in wasting the network resources as the backup path is configured at the optical layer level and reserved only for the failure scenario. It would not carry traffic when the primary path is active.

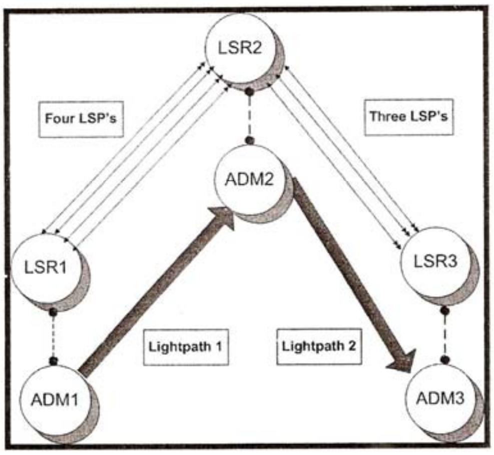

Fig. 8: IP/MPLS Over WDM Network

In MPLS based protection, backup LSPs are set up with additional overhead of identifying PSL, PML and RNT. When backup LSPs are installed, the start and end of the backup LSP is marked by the PSL (Protection Switch LSR) and PML (Protection Merge LSR). PSL needs to select the primary or the backup path for forwarding the traffic. The fault information is propagated to the PSL using RNT (reverse notification tree) [16] so that it can divert the traffic from the primary path to the backup path.

Some recent work has focused on using integrated and multilayered protection approaches [15]. The integrated approach can only be applied to the networks which are configured with the peer model instead of the overlay model. In the peer model, the LSRs and OXCs are connected together as peers and run the same instance of control protocols. GMPLS, along with enhanced OSPF enables uniform access to the load and bandwidth information on both layers. The integrated approach uses a mix of existing and new lightpaths in order to provide protection paths. On the other hand, the multilayered approach calls for protecting the high priority traffic with optical layer protection and the low priority traffic with IP/MPLS layer protection. Lighpaths are divided into three categories. The primary and backup lightpaths are reserved for high priority traffic and its protection. When the primary lighpaths fail, the optical layer switches the traffic to the backup paths without informing the upper layer of any problem. When an ordinary lightpath fails, the optical layer notifies MPLS of its failure and does not take any corrective action.

In the following sections, we present an overview of the protection methods at the layers below the IP layer and routing for global and local protection. Next the protection triggers and alarm management and sub-wavelength protection is discussed and the RSVP-TE protection signaling is highlighted as an example. In the end, we identify some of the issues that are currently under investigation.

6.2.1. Protection at the Optical Layer

In the wavelength convertible WDM networks, end to end protection can be provided at the lightpath level (PAL) or at the connection level (PAC) [19]. In PAL, backup lightpath does not carry traffic during the normal operation. Thus under PAL, end to end protection is provided through a sequence of backup lightpaths that are fiber disjoint with the primary lightpaths.

In PAC, the end to end protection path is a sequence of unprotected lightpaths that are simply fiber disjoint with the primary lightpaths. Thus the protection path and the primary path carry traffic during the normal operation. The protection path may set aside some of its bandwidth for protecting against failure of the primary lightpaths. When a failure is detected, PAL would simply switch to the protection lightpath and all the connections using the affected lightpath continue to function normally. On the other hand, in PAC, several connections that were using the failed optical channel would switch to the backup channel. Given that most of the connections are low bandwidth and many connections share a lightpath, PAL outperforms PAC for wavelength convertible shared mesh protection WDM networks. Some additional schemes are also presented in [22, 23] for shared and dedicated protection for lightpaths.

6.2.2. Protection Below Layer 3

Due to the slow convergence of L3 updates in case of failures, industry has been looking for ways to provide protection below layer 3. We discuss the RPR (802.17), BIP, FRR and RB protection [17, 18] below.

IEEE 802.17 RPR WRAPPING AND STEERING PROTECTION

In resilient packet ring network (RPR), there are clockwise and counterclockwise paths from a node to any other node. Since both directional paths are established, nodes can take corrective action in case of single link failure. In wrapping protection, both the nodes that share the failed link can wrap the traffic around to the other direction. In steering protection, the nodes injecting the traffic into the failed link start sending the traffic in the opposite direction of the fault.

BUNDLED INTERFACE PROTECTION

Some simple techniques that can avoid L3 delays include protecting aggregated links between two routers. This can be accomplished by programming two FIB (forwarding information base) entries in the router [17]. One entry is for the working path and the other entry is for the protection path. In case of failure of the working path, the traffic destined for the failed link can be redirected to the backup link within several milliseconds without L3 updates. Similar strategy is used when an IP router attaches to an ADM using two SONET APS links one of which is a backup link. This would allow for recovery before L3 layer takes any notice of the fault.

MPLS FRR

Since MPLS labels are stackable, traffic can be switched from the primary LSP to the backup LSP or backup segment LSP by a simple push of a label. Since each LSP is destined towards a specific node, any link failures that occur before the egress can be resolved by FRR (fast reroute) technique that creates a bypass tunnel around the fault. The bypass tunnel would resolve the local failures without involving L3. If the bypass tunnel rejoins the primary tunnel at the next hop, it only protects against link failures. If the bypass tunnel does not join the primary tunnel until after several hops, the PML is configured to identify and remove the backup label so that the original tunnel is identified at the primary egress node. Issues pertaining to the bypass tunnel include the scope of the label space and modification of FIB entries. In order to avoid signaling for label exchange, global label space should be used for the domain. In case of failure, FIB entries in the router are modified to push the bypass tunnel first hop label on each packet carrying the primary tunnel label and to point to the new interface through which the traffic would pass.

MPLS RB

In MPLS, a global backup path may be established for a working path however the fault indication may be implemented with simple path continuity tests. Thus, it may result in larger recovery time. Using RB (reverse backup) can speed up recovery in the following way [18]. When an outgoing link goes down, the current node can utilize a reverse backup LSP to direct the traffic back to the ingress of the working path. This node can also propagate the FIS (fault indication signal) to the ingress in addition to the traffic that can now be directed through the global backup path. Thus no packets are lost and the FIS is propagated to the ingress within the shortest possible time. For example, in Figure 9, consider the primary LSP R1-R3-R6 that is protected by the global backup LSP R1-R2-R4-R6. If the link between R3 and R6 goes down, R3 can no longer send the traffic on the primary LSP forward to R6. Utilizing a reverse backup LSP segment to R1, R3 redirects the traffic back to the PSL of the global backup LSP. In addition, R3 also sends the fault indication to R1 that can direct the reversed traffic as well as any new traffic to the pre-established global backup LSP. This technique has a larger overhead in terms of additional backup LSP reverse segments added to the protection topology but it ensures that no packets are lost.

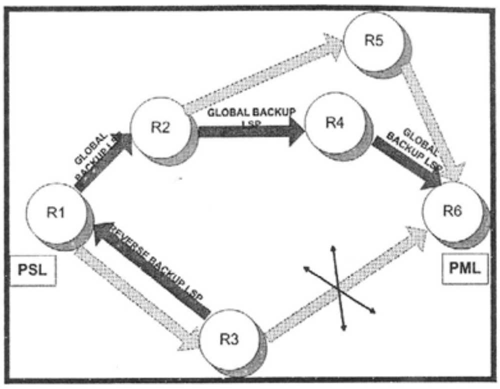

Fig. 9: Reverse Backup LSP Scheme

MPLS MULTIPLE FAILURE PROTECTION

In case of multiple failures, a combination of FRR local repair and global backup paths can be utilized to overcome the problem [18]. If a link in the working path fails, traffic is diverted to the global backup path. If the global backup path also develops a link failure, local recovery path can be established around the faulty link and traffic can continue to flow. For example, if the primary path in Fig. 9 develops a fault, the global protection path R1-R2-R4-R6 is utilized. If the link R2-R4 in the protection path also fails, MPLS FRR can locally repair the failed link by routing around it as shown in Fig. 10. Note the new PSL for the locally repaired problem that would select the rerouted path.

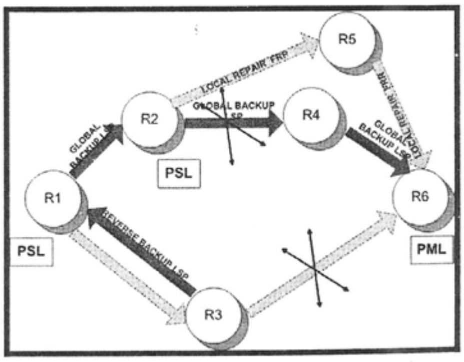

Fig. 10: Handling Multiple Failures

6.2.3. Protection Path Routing

As discussed in the previous section, protection paths can be global, locally rerouted or reverse. Based on the QoS needs of the streams, the FRR protection can be set up for premium time sensitive traffic and RB can be provided to loss intolerant traffic [18]. The protection paths can be routed in the form of link disjoint LSP for each original LSP or protected group of LSPs. Three schemes are defined [14] as below.

In SLB (Single Link Basis) scheme, an original path consisting of 'L' links is protected by 'L' protection paths. Each protection path excludes one of the links of the original path. Protection paths may share links or nodes together. As seen in the diagram, the original LSP R1-R2-R4 is protected by two backup LSPs because it has two links. The first protection path through R1-R3-R2-R4 protects link from R1 to R2 whereas the second one R1-R3-R4 protects the link from R2 to R4. SLB requires extensive computation for protection paths. Protection path computations can be carried out in the network model by simply removing the protected link from the modeled network. SLB requires faulty link identification in case of link failure in order to migrate the traffic to the protection path.

In DP (Disjoint Path) scheme, one link disjoint path is computed for one original path thus removing the requirement of faulty link identification in case of path failure. The protection path can be computed by removing the links that are members of the active path from the modeled network and applying routing algorithm. For example, in Fig. 11, links R1-R2 and R2-R4 can be pruned to yield a link-disjoint backup path R1-R3-R4 under the DP scheme.

In LR (Local Repair) scheme, instead of protecting paths, links are protected at the local level. For each link that is a member of an active path, resources and paths are reserved that would allow the local node to reroute traffic destined for the faulty link. Thus, if a link is a member of 'M' active paths, it would have 'M' LR paths computed under this scheme. The computation is similar to SLB scheme however in LR scheme, path segments are computed instead of full paths and each link may have several such segments depending on the number of active paths through that link.

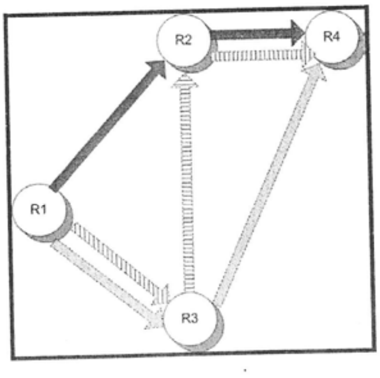

Fig. 11: SLB and DP Protection Paths

6.2.4. Triggers, alarms and timers

Since protection may be provided at multiple layers in a network, care should be taken to guard against triggering of protective methods on more than one layer. In general, lower layer protection should be activated first [17]. For example, routers connected through a SONET protected network should delay their triggers to allow the lower layer to correct the problem. On the other hand, routers connected directly may trigger protection as soon as a defect is detected. Thus, the protection trigger is supposed to be at the location closest to the defect. The defect is converted to alarm after being observed for 2-3 seconds [20]. At this point, it can be reported to the user where it is maintained for 9.5 to 10.5 seconds even if it disappears sooner.

An important issue is to determine the appropriate time to revert back to the original path once it has been repaired. We discuss two approaches to reversion that safeguard against oscillations [17]. When a working path has a WTR (Wait to Restore) timer on it and it recovers from a failure, the timer must start at recovery and run out before this path can be used again. In "Fast down slow up" approach, a failed path is held down for a time that would increase exponentially with each failure on the same path. In case of failure of protection path while WTR timer is running, WTR can be cleared immediately and reversion can take place thus making WTR preferable over slow up approach.

6.2.5. Sub-wavelength protection

Since most of the connections require only a fraction of the wavelength capacity, it is preferable to provide traffic grooming in WDM mesh networks. In traffic grooming, multiple connections are carried on a single wavelength. Grooming can be done at the electronic level or the optical level. With all optical grooming, the conversions between electronic and optical domains can be avoided. For efficient protection at the sub-wavelength level, some of the sub-wavelength protection schemes offer partial protection in which the backup bandwidth is less than or equal to the primary bandwidth [21]. This allows reduction in the connection blocking probability and improvement in utilization of the capacity of the network. However, this comes at the cost of replacing the 100% recovery guarantee for fully protected connections with the QoP (Quality of Protection) parameter. Each connection is admitted with its desired QoP (guaranteed or best effort). Providing this kind of protection in sub-wavelength connections is still under investigation.

6.2.6. RSVP fault notification messaging

In IP/MPLS networks, there are multiple mechanisms to detect and localize defects. ICMP extensions allow adding MPLS stack information to the control packets generated by LSRs. GTTP (Generic Tunnel Trace Protocol) [24] can be used to trace MPLS tunnels. ITU-T has specified OAM packets for MPLS in Y.1711 [25]. GMPLS specifies LMP (Link Management Protocol) that verifies link connectivity. Once a fault is detected, it can be notified in RSVP to the upstream nodes by "PathErr" message and to the downstream nodes by "ResvErr" message. A "Notify" message is added to the RSVP-TE for GMPLS in order to notify the non-adjacent nodes of LSP related failures [26]. The notify message can be targeted to any node that is not an upstream or downstream neighbor.

RSVP-TE implements tunnel rerouting, an important requirement of Traffic Engineering [27] , in order to migrate traffic to a better route or to resolve a failure along the established route. The "make-before-break" strategy is used in RSVP-TE with shared resources on links common to the old and new tunnel.

6.3 Fault Detection and Localization Tools

For IP networks, connection-testing tools such as ping and traceroute are used frequently to check the reachability and time delay of destinations. Ping generates ICMP packets that are forwarded by each intermediate router until they hit the destination or a router that does not have an entry for or towards the destination in its routing table. Traceroute reports on each intermediate router encountered in the journey. On similar lines, MPLS ping [29] proposal has been developed. In MPLS ping, UDP echo request/reply packets are exchanged for testing a prefix reachability via the LSP under test. The MPLS ping packets are set with an IP TTL of 1 so that they can be discarded at the egress router of the MPLS domain. Prefix could be an IPv4 or IPv6 address besides other options. Each intermediate LSR runs several checks on receiving the MPLS ping packet and reports the results back to the originating LSR. MPLS LSP traceroute [29] operates with similar UDP packets. Originating LSR sets the label TTL to 1 for the first packet so that the first packet cannot travel beyond the next hop. On receiving reply packet from the downstream neighbor, the originating LSR would direct the next echo request packet to the downstream TLV (Type Length Value) as received in the echo reply packet from the neighboring LSR and at the same time increment the label TTL value by 1. It can be seen that the MPLS traceroute operates in a way similar to the IP traceroute. LSP ping and traceroute tools can be used to localize failed LSP's. LSP ping overcomes many challenges including the fact that labels have only local significance and LSP's may merge in the middle of the domain.

IETF has developed the bi-directional forwarding detection tool [30] to test connectivity. It uses MPLS ping to establish a BFD session. Probe packets can be sent in asynchronous mode or in echo mode with specific thresholds for loss of connectivity. In asynchronous mode, BFD packets are sent continuously to the remote node. The remote node would declare the connection lost when a threshold of lost packets is reached. In echo mode, the remote node transmits the packets back to the source. A self-test for label switched router is also developed [31] that can be invoked by a LSR to test its label bindings. This test requires passing a test packet to the upstream router that returns the packet back to originating LSR. The packet then travels to the downstream router which decrements the TTL to zero and issues a special response to the original LSR.

ITU (International Telecommunication Union) has proposed its own protocol for fault detection [32]. ITU specifies CV (connectivity verification) packets that can be sent by the ingress router to the egress router at regular intervals. It also specifies FFD (fast failure detection) function that operates at a higher frequency.

6.4 Current Issues In Fault Management

Current challenges in fault management include protection path routing for node or link disjoint backup paths, multi-layer survivability schemes, inter-AS fault management, foiling denial of service attacks and self-healing with AI (Artificial Intelligence). Protection path routing can be dealt with the help of graph theory. All the nodes in the domain are mapped to vertices and links are mapped to edges of a graph. Protection path can be computed when the original path has been identified so that it does not share the edges and vertices with the original path. Heuristic algorithms may be developed that provide protection paths for 1:1, 1:M or N:M schemes. Multi-layer survivability schemes should address the issue of coordinating fault handling in different layers. In [22], fault hold off time is clearly marked before propagating the fault signal to the PSL (Path switch LSR). This hold-off timé is added due to the fact that the lower layer fault handling function may have already kicked off. As an example, a fiber-cut fault may be handled at the WDM layer in an optical network. On the other hand, data error faults may be dealt with at higher layers. Multi-layer survivability schemes should also implement alarm suppression because of the fact that a single failure may generate fault signals at many layers. These different layer signals must be correlated so that a single fault event is generated. Fault signaling between different autonomous systems should be implemented smoothly so that end-to-end services can be maintained. Other considerations include dealing with denial of service attacks and hiding network topology from intruders while making it visible to network operators who wish to investigate into the available resources.

7. CONFIGURING MPLS NETWORK TESTBED

Developers who wish to run bandwidth allocation and management experiments need testbeds for testing their algorithms. MPLS testbeds can be configured on Linux PC's without incurring a large expenditure. In order to install MPLS on Linux, one should obtain and install Fedora Linux core 3 [33]. Since Fedora is based on Red Hat Linux, one should use the RPM package manager to install and configure MPLS. MPLS for Linux is now in its revision 1.946. It can be obtained from [34] and the developer guide is available at [35]. This implementation uses RFC 3036 LDP (Label Distribution Protocol) but it is still under development. Until the LDP enters beta stage, it is recommended to set up the labels manually. MPLS for Linux does not "support" RSVP-TE RFC 3029 however some implementations of RSVP have shown up. Testbed projects for configuring MPLS domains should follow some popular topologies such as the Fish (see Fig. 12), irregular ISP domain and others. For example, in the fish network, we will need at least six Linux PC's running MPLS and connected together as MPLS routers.

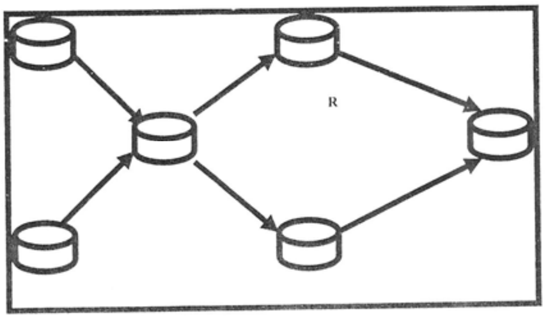

Fig. 12: The Fish Network with a bottleneck router and two alternate paths.

Traffic may be generated and admitted across the domain using software running on the same hosts that serve as MPLS routers. CBR traffic can be generated by writing a simple socket based application that opens connection to the remote host and sends out packets at regular intervals. It is not required to limit the domain to synthetic traffic as there are several network applications on Linux that can be configured and started. The configuration of MPLS domain and labels is straightforward. In the MPLS users' guide, replace mplsadm by mpls as mplsadm is deprecated. For example, one or more interfaces can be assigned to a label space with the following commands:

```
mpls labelspace add dev eth0 labelspace 0
```

Consider an example from [34], as shown in Fig. 13. It is desired to set up an LSP between the two LSR's shown. The network address is 11.0.1.0/24. The node 'A' on the left is 11.0.1.1 and node 'B' on the right is 11.0.1.2. The following commands are used on node 'A'. As a result, a generic label of value 10000 is generated in the labelspace 0 and it will be used on all packets leaving via interface eth1 for the next hop which is 11.0.1.2.

```
mpls nhlfe add key 0

Key: 0x00000002

mpls nhlfe change key 0x2 instructions push gen 10000 nexthop eth1 ipv4 i1.0.1.2

ip route add 11.0.1.2/32 via 11.0.1.2 spec_nh 0x8847 0x2
```

On the node 'B', following instructions are executed to configure an entry in the ILM (Incoming Label Mapping) table for label 10000 in labelspace 0 on interface eth1.

```
mpls labelspace add dev eth1 labelspace 0

mpls ilm add label gen 10000 labelspace 0
```

At this point, any network application can be configured and started to send data from 'A' to 'B'. A program such as tcpdump [36] or ethereal [37] can be used to monitor the traffic and extract the packet fields. Another LSP would be configured in the reverse direction in order for the node 'B' to send something to the node 'A'. It is also possible to set up tunnels and hierarchies. For further information, please refer to [35].

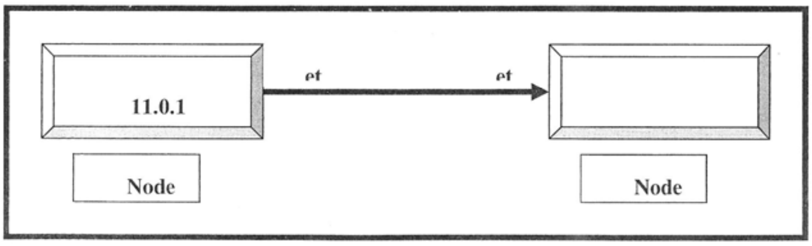

Fig. 13: Setting up an LSP in MPLS for Linux.

8. CONCLUSION

This paper has outlined the main issues in deployment of IP-Optical network in which the core of the network is all optical and voice, video and data are integrated within the same network. It can be seen from the above that there are operation and administration issues that must be resolved before the dream of "IP over Glass" can be fully realized. OAM functions are still evolving for packet based networks and a lot of work needs to be done. Protection mechanisms and strategies in modern high speed optical networks have been outlined. Important considerations such as the choice of layer, protection path routing, bandwidth guarantees and reversion delays are discussed. Pros and cons of provisioning protection paths at the lower layers and higher layers are described. Important areas for further research include the coordination among multiple layers for fault handling and selection of appropriate protection strategy for various classes of Diffserv in Diffserv enabled networks. Innovative sub-wavelength protection schemes are needed that are scalable and efficient for ensuring the backup resources for low speed connections that have been groomed through the grooming ports of optical cross connects. Partial protection schemes have been developed recently that offer partial bandwidth protection to sub-wavelength connections. These schemes must be evaluated and enhanced in order to reduce the blocking probabilities for new connections and efficient usage of available resources.

Those researchers who wish to get involved in development of operation and administration methods for next generation networks should identify an area of their interest in relevant workgroups. Major bodies involved in the development of optical networks include Optical Internetworking Forum (http://www.oiforum.com), Internet Engineering Task Force (http://www.ietf.org) and International Telecommunications Union (http://www.itu.int/home/index.html). Several mailing lists are managed by IETF in which current problems are discussed and solutions presented. IETF documents including RFC's (finalized versions) and Internet Drafts (current temporary versions) are available to everyone from www.ietf.org. On the other hand, ITU recommendations are more restricted and only member organizations can view their documents. For developers who wish to test new traffic and bandwidth management algorithms in MPLS domains, procedures of configuring an MPLS testbed are given with examples.

REFERENCES

[1] D. Awduche *et al* "Requirements for Traffic Engineering Over MPLS", RFC 2702, IETF, Sep 1999

[2] E. Rosen *et al* , "Multiprotocol Label Switching Architecture", RFC 3031, IETF, Jan 2001

[3] E. Mannie *et al*, "Generalized Multi-Protocol Label Switching (GMPLS) Architecture" IETF, Internet Draft, Work in progress, May 2003 <draft-ietf-ccamp-gmpls-architecture-07.txt>

[4] A. Banerjee *et al*, "Generalized Multiprotocol Label Switching: An Overview of Routing and Management Enhancements" IEEE Communications Magazine Vol.39 Issue 1, January 2001, pp 144-150.

[5] J. Comellas *et al* , "Integrated IP/WDM Routing in GMPLS Based Optical Networks" IEEE Network March/April 2003, Pages 22-27.

[6] W. Smari and J. A. Zubairi, "Developing Network Management Solutions For Next Generation IP Optical Networks" NSF STTR Research Proposal.

[7] J. Comellas *et al* , "Integrated IP/WDM Routing in GMPLS Based Optical Networks" IEEE Network March/April 2003, Pages 22-27

[8] S. Dixit & Y. Ye, "Streamlining the Internet-Fiber Connection" IEEE Spectrum Online Weekly Feature April 2001.

[9] V. Sharma *et al*, "Framework for Multi-Protocol Label Switching (MPLS) Based Recovery", IETF RFC 3469, Feb 2003.

[10] D. Cavendish *et al*, "Operation, Administration and Maintenance in MPLS Networks" IEEE Communications, Oct'04, PP 91-99.

[11] G. Swallow *et al*, "Label Switching Router Self Test", IETF Internet Draft, Work in Progress, Oct'03

[12] M. Morrow *et al*, "OAM in MPLS-Based Networks", IEEE Communications, Oct'04 PP 88-90 guest editorial.

[13] G. Swallow *et al*," Avoiding Equal Cost Multipath Treatment in MPLS Networks", IETF Internet Draft, Work in Progress, Sep'04.

[14] G. Conte *et al*, "Strategy for Protection and Restoration of Optical Paths in WDM Backbone Networks for Next-Generation Internet Infrastructures" Journal Of Lightwave Technology, Vol. 20, No. 8, August 2002, PP 1264-1276.

[15] Q. Zheng *et al*, "Protection Approaches for Dynamic Traffic in IP/MPLS Over WDM Networks" in IEEE Optical Communications May '03 PP S24-S29.

[16] C. Huang *et al*., "Building Reliable MPLS Networks Using A Path Protection Mechanism." IEEE Communications Magazine. Mar. 2002, pp. 156-62.

[17] G. Suwala *et al*, "SONET/SDH Like Resiliency for IP Networks: A Survey of Traffic Protection Mechanisms" IEEE Network 3-4, 2004, PP. 20-25.

[18] J. Marzo *et al*, "QoS Online Routing and MPLS Multilevel Protection: A Survey" IEEE Communications Magazine Oct 2003, PP. 126-132.

[19] C. Ou *et al*, "Survivable Traffic Grooming in WDM Mesh Networks" in Proc. OFC 2003 PP 624-625.

[20] Telcordia GR-253, "SONET Generic Criteria," issue 3, Sept. 2000.

[21] J. Fang *et al*, "On Partial Protection in Groomed Optical WDM Mesh Networks" in Proc. IEEE DSN'05.

[22] C. Ou, K. Zhu, and et al. "Traffic grooming for survivable WDM networks - shared protection" IEEE Journal on Selected Areas in Communications, pages 1367–1383, November 2003.

[23] C. Ou, K. Zhu, and et al. "Traffic grooming for survivable WDM networks - dedicated protection" Journal of Optical Networking, pages 50–74, January 2004.

[24] Bonica *et al*., "Generic Tunnel Tracing Protocol (GTTP) Specification", Internet Draft, draf-bonica-hmnpmto-Ol.txt, work in progress, July 2001.

[25] ITU-T Draft Recommendation Y.1711, "OAM mechanism for MPLS networks", work in progress, 2002.

[26] M. Brunner and C. Hullo, “GMPLS Fault Management and Its Impact on Service Resilience Differentiation” in Proc. IEEE INM’03.

[27] D. Awduche et al , “RSVP-TE: Extensions to RSVP for LSP Tunnels” IEF RFC 3209 Dec 2001.

[28] Y. Lee *et al*, ”Traffic Engineering in Next Generation Optical Networks”, IEEE Communications Surveys and Tutorials, QII, 2004, PP 16-33.

[29] K. Kompella *et al*, ”Detecting MPLS Data Plane Failures” IETF Internet Draft, Work in Progress, Oct’03.

[30] D. Katz and D. Ward, “Bidirectional Forwarding Detection” IETF Internet Draft, Work in Progress, Aug’03.

[31] G. Swallow *et al*, “Label Switching Router Self Test”, IETF Internet Draft, Work in Progress, Oct’03.

[32] ITU-T Rec. Y.1711, “Operation and Maintenance Mechanism for MPLS Networks”, Feb 2004.

[33] Fedora project, sponsored by RedHat http://fedora.redhat.com/

[34] MPLS for Linux Homepage http://sourceforge.net/projects/mpls-linux/

[35] MPLS for Linux Developer’s Guide http://perso.enst.fr/~casellas/mpls-linux/index.html

[36] TCPDUMP Homepage

[37] Ethereal Homepage