Generalized Multiprotocol Label Switching: An Overview of Signaling Enhancements and Recovery Techniques

Ayan Banerjee, John Drake, Jonathan Lang, and Brad Turner, Calient Networks
Daniel Awduche and Lou Berger, Movaz Networks
Kireeti Kompella and Yakov Rekhter, Juniper Networks

ABSTRACT

Generalized multiprotocol label switching, also referred to as multiprotocol lambda switching, is a multipurpose control plane paradigm that supports not only devices that perform packet switching, but also devices that perform switching in the time, wavelength, and space domains. The development of GMPLS necessitates enhancements to existing IP signaling and routing protocols. An overview of the MPλS concept was presented in [1], and a summary of the extensions to IP routing protocols (OSPF [2] and IS-IS [3]) and the new Link Management Protocol to support GMPLS were presented in a companion paper in the January issue of this magazine [4]. In this article we present enhancements to two commonly used IP signaling protocols, RSVP and LDP, to support GMPLS. We illustrate the concept of hierarchical label switched path setup with an example, discuss mechanisms for bidirectional LSP setup, and describe the applications of suggested labels. We also discuss the important problem of protection and restoration in the GMPLS context. Finally, we describe how various recovery mechanisms can be implemented within the GMPLS framework.

INTRODUCTION

In a typical connectionless network (i.e., the Internet) packet forwarding is performed independently at each router in the network and is based on the destination address carried in the packet. This packet forwarding supports only one connectivity abstraction: multipoint-to-point path.[1] Recently, substantial effort has been expended to evolve conventional IP routing architecture and protocols by augmenting them with additional functionality under the umbrella of multiprotocol label switching (MPLS). One of the key aspects of MPLS is the addition of a new connectivity abstraction: explicitly routed point-to-point path. This is accomplished by the concept of explicitly routed label switched paths (LSPs). As we discuss below, this connectivity abstraction enables support for constraint-based routing, which in turn is the foundation for GMPLS.

MPLS is based on the following key concepts:

- Separation of forwarding information (label) from the content of the IP header
- Use of a single forwarding paradigm (label swapping) at the data plane to support multiple routing paradigms at the control plane
- Use of different technologies and link layer mechanisms to realize the label swapping forwarding paradigm; examples are the "shim header" in Ethernet and packet over synchronous optical network (SONET) networks, the data link connection identifier in frame relay networks, and virtual circuit/path identifier (VCI/VPI) in asynchronous transfer mode (ATM) networks
- Flexibility in the formation of forwarding equivalence classes (FECs)
- The concept of a forwarding hierarchy via label stacking

One application of MPLS is constraint-based routing, which is used to compute paths that satisfy various requirements subject to a set of constraints. Constraint-based routing is employed for two main purposes in contemporary networks: traffic engineering and fast reroute; future applications include diversity routing, which is the process by which disjoint paths are computed for protection purposes. One of the merits of MPLS is that it allows the elimination of redundant network layers by migrating some of the functions

[1] *This is in the context of unicast forwarding.*

0163-6804/01/$10.00 © 2001 IEEE

provided by ATM and SONET/SDH layers to the IP/MPLS control plane.

With MPLS constraint-based routing the extensions to Open Shortest Path First (OSPF) and Intermediate System to Intermediate System (IS-IS) allow nodes to exchange information not just about network topology, but also about resource availability and administrative constraints. This information is used as input to a constraint-based path computation algorithm that computes paths subject to topology, resource, and administrative constraints. After computation of an appropriate path, a signaling protocol such as Resource Reservation Protocol with Traffic Engineering (RSVP-TE) or Constraint-Based Routing Label Distribution Protocol (CR-LDP) is then used to instantiate a label forwarding state along the path.

Recent work has been done to extend and adapt the MPLS control plane, and specifically MPLS constraint-based routing, so that it can be used not just with routers and ATM switches, but also with optical crossconnects (OXCs) [4]. This is a fundamental step in the evolution and integration of data and optical network architectures. Using MPLS as the foundation for connection establishment and a common control plane addresses several issues related to this network evolution. First, a common control plane simplifies network operations and management, which ultimately results in reduced operational costs. Second, a common control plane provides a wide range of deployment scenarios, ranging from overlay to peer, where the overlay model is realized using only a subset of the functionality needed to implement the peer model. This allows the choice of peer or overlay (or a combination of both) deployment models to be driven by business and engineering considerations, rather than constrained by the current approach of stratification of subnetworks into technology domains. At the same time, developing a common control plane by reusing and extending existing routing and signaling protocols avoids the need to "reinvent the wheel," thus minimizing risks associated with protocol development and reducing the time to market for advanced optical switching equipment. Some enhancements are clearly required to the existing MPLS routing and signaling protocols to address the peculiar characteristics of optical transport networks. These protocol extensions are being standardized by the Internet Engineering Task Force (IETF) under the framework of generalized MPLS (GMPLS). These extensions can be summarized as follows:

- Enhancements to the RSVP-TE and CR-LDP signaling protocols to allow the signaling and instantiation of optical channel trails in optical transport networks and other connection-oriented networking environments [5–7]
- Enhancements to OSPF and IS-IS interior gateway routing protocols (IGPs) to advertise availability of optical resources in the network (e.g., bandwidth on wavelengths, interface types) and other network attributes and constraints [8, 9]
- A new link management protocol, LMP, designed to address the issues related to link management in optical networks [7]

Additional functionality has been added to GMPLS to address some limitations of the MPLS control plane, such as the inability to establish bidirectional connections in a single request, and the absence of mechanisms to account for protection bandwidth so that it can be used for lower-priority traffic. Using the MPLS framework and its signaling protocols, a link or node failure (e.g., power outage) along the routes of established service connections could only be handled locally, or along the nodes of the path.[2] However, in the GMPLS framework, additional capability has been added such that failures which impact service connections can also be reported to a predefined alarm center (e.g., a centralized management system). Thus, the devices in the network can detect a failure, report the failure, quickly determine whether spare capacity is available on other routes, and then use signaling to restore the service connection onto a fault-free route, circumventing the point of failure. In this article we provide an overview of the protection and restoration techniques used by GMPLS devices, highlighting the following aspects:

- Fault isolation
- Fault localization
- Fault notification
- Fault mitigation (via protection and restoration schemes)

The reader is referred to [4] for the enhancements to the routing and link management protocols and a brief overview of MPLS. This article discusses the enhancements made to the signaling protocols in support of GMPLS. It also illustrates the GMPLS mechanisms that can be used for protection and restoration.

MPLS is designed so that the control plane is logically separate from the data plane. GMPLS extends this concept, allowing the control plane to be physically diverse from the associated data plane.

Enhancements to Signaling

GMPLS requires that an LSP start and end on similar types of devices. In certain technologies, the data plane that carries the traffic may be transparent (i.e., the device is unable to terminate the traffic). However, in order to set up LSPs between transparent devices, signaling requests need to be terminated; this necessitates a separate control plane transport network to convey signaling messages. MPLS is designed so that the control plane is logically separate from the data plane. GMPLS extends this concept, allowing the control plane to be physically diverse from the associated data plane. In this section we discuss the enhancements that have been made to the label distribution protocols (RSVP-TE and CR-LDP) to support GMPLS. We further discuss the methodology by which the concepts of hierarchical connection setup, suggested label, and upstream label interact to support a *scalable*, *generalized*, and *manageable architecture*.

Hierarchical LSP Setup

GMPLS supports the concept of hierarchical LSPs [11], which occurs when a new LSP is tunneled inside an existing higher-order LSP so that the preexisting LSP serves as a link along the path of the new LSP. In this section we illustrate how lower-order LSPs trigger the formation of higher-order LSPs. The ordering of LSPs is

[2] *Note however, that a link or node failure is flooded in link state updates throughout the IS-IS/OSPF area the link (or node) is in, and therefore such failure is visible to all nodes within the area.*

Generalized MPLS signaling allows a label to be suggested by an upstream node. This suggestion is an optimization that may be overridden by a downstream node, but in most cases at the cost of higher LSP setup time, and perhaps sub-optimal allocation of network resources.

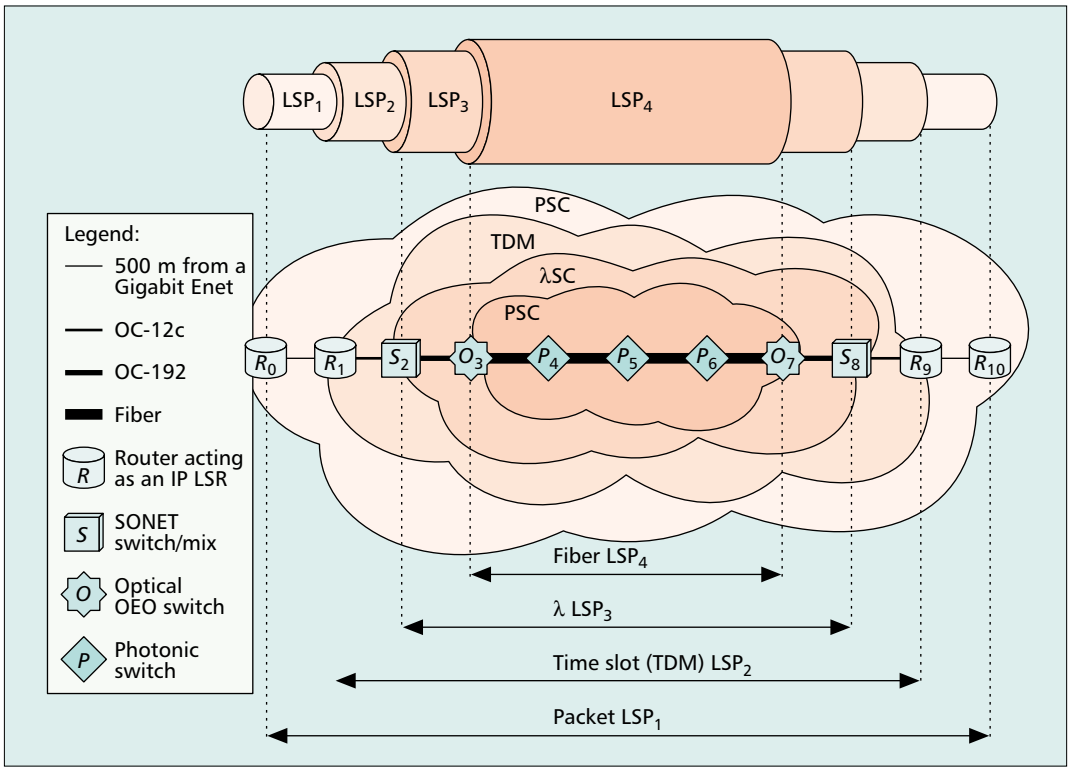

Figure 1. *LSP_1 is set up from R_0 to R_{10} for 500 Mb/s of bandwidth. LSP_1 is nested in the subordinate LSPs 2, 3, and 4, respectively, as depicted at the top of the figure. If an existing LSP has capacity, the subordinate LSP is added to it; otherwise, the setup of a new LSP of the appropriate type is triggered. The endpoints of LSP_1 classify packet data and source it into LSP_1. R_1 and R_9 are packet LSRs; S_2 and S_8 are SONET path levels switches. Devices O_3 and O_7 are optical or lambda switched, providing SONET/SDH section level signals (e.g., OC-192 including all overheads); they also have WDM capabilities between photonic switches P_4–P_6. The link between R_0 and R_1 is Gigabit Ethernet, between R_1 and S_2 an OC-48, between S_2 and O_3 an OC-192, between O_3 and P_4 a WDM multiplex of 16 OC-192 signals which remains intact through to O_7 (P_4–P_6 are pure photonic switches). The link between O_7 and S_8 is an OC-192, between S_8 and R_9 an OC-48, and between R_9 and R_{10} Gigabit Ethernet.*

based on the link multiplexing capabilities [4] of the nodes. Nodes at the border of two regions, with respect to multiplexing capabilities, are responsible for forming higher-order LSPs and aggregating lower-order LSPs.

As an example, the bandwidth signaled for LSP1 in Fig. 1 is 500 Mb/s; all links traversed by the LSP must be large enough to support the requested bandwidth. R_0 classifies and maps packets into LSP_1 and shapes the data to 500 Mb/s if necessary. The 500 Mb/s for LSP_1 will be allocated from LSP_2, which is running at the next larger SONET increment, an OC-12c. The S_2 switch grooms the OC-12c into LSP_3, which is an OC-192 between S_2 and O_3. The optical switch, O_3, takes this OC-192 and switches it through LSP_4, a WDM channel toward P_4. The OC-192 corresponding to LSP_3 is switched through P_4, P_5, and P_6 to O_7. Continuing in this fashion, O_7 selects the correct lambda and passes the signal to the port adjacent to S_8. S_8 will select the appropriate OC-12c from the OC-192 and pass this to R_9. Finally, R_9 will take the packets from the OC-12c and forward them on R_{10}.

We now illustrate the process of creating an LSP (Fig. 2) using the RSVP-TE signaling extensions[3] defined in GMPLS, assuming that the requested bandwidth is available on each of the links. In addition to the formation of LSP_1, this process also triggers the establishment of the following additional LSPs: LSP_2, an STS-12c connection between R_1 and R_9; LSP_3, an OC-192 connection between S_2 and S_8; and LSP_4, WDM channels between O_3 and O_7. Note that the residual bandwidth available in the LSP hierarchy is advertised by the IGP as follows:

- Node R_1 announces a packet-switch-capable (PSC) "link" from R_1 to R_9 with bandwidth equal to the difference between the STS-12c (622 Mb/s) capacity of the link and the 500 Mb/s that has been allocated to LSP_1.
- Node S_2 announces the equivalent 180 STS-1s worth of bandwidth for a time-division multiplex (TDM) link.
- Node O_3 announces 15 lambdas, each of which has OC-192 bandwidth for a lambda-switch-capable (LSC) link.

The Suggested Label

GMPLS signaling allows a label to be suggested by an upstream node. This suggestion is an optimization that may be overridden by a downstream node, but in most cases at the cost of higher LSP setup time and perhaps suboptimal allocation of network resources. The suggested label is particularly valuable when it is desired to set up a bidirectional LSP using paired transmit (Tx) and receive (Rx) interfaces to the same

[3] *In this document, PATH and RESV messages of the RSVP-TE signaling protocol may be replaced with REQUEST and MAPPING messages of the CR-LDP signaling protocol.*

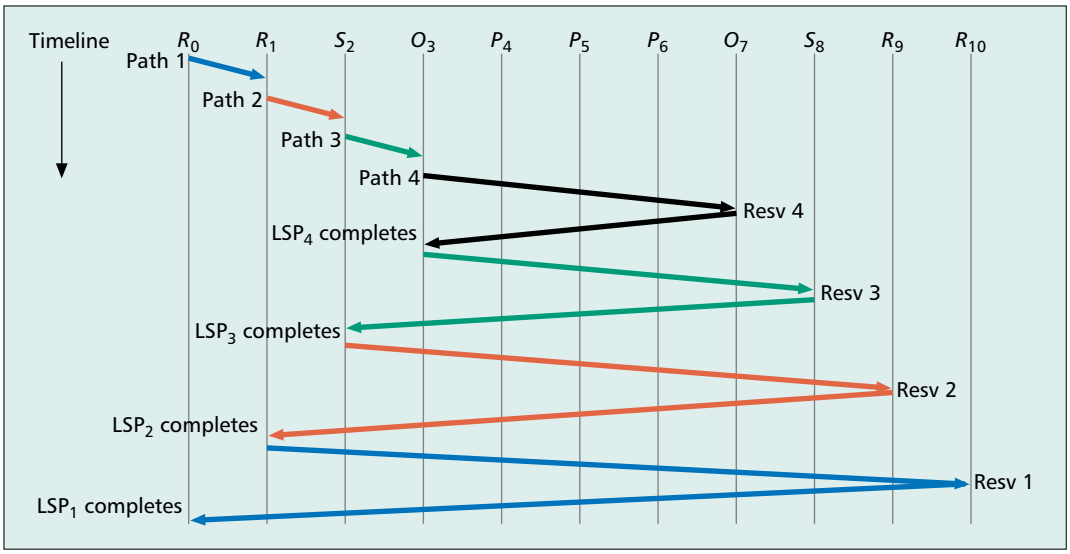

Figure 2. *A Path request (Path 1) is generated at* R0 *that is sent to* R_1. *At node* R_1 *(a boundary node) this arrival triggers a requirement for a new LSP (*LSP_2*) from* R_1 *to* R_9. *These dynamic LSP creation requests are triggered until Path 4 is generated at* O_3. *Following successful establishment of* LSP_4, *the Path 3 message is tunneled through* LSP_4. *This process of LSP formation, and a lower-level LSP creation request being tunneled through the higher-level LSP so formed, continues until the initial LSP (*LSP_1*) is successfully created, thus forming a hierarchy.*

The Notify message has been added to RSVP-TE for GMPLS to provide a mechanism for informing non-adjacent nodes of LSP related failures; a similar mechanism has not been defined for CR-LDP.

physical port (e.g., WDM transponders Tx/Rx pair) or to set up an LSP transiting certain kinds of optical switching equipment where there is some latency associated with configuring the switching fabric. The suggested label is also useful in optical subnetworks with limited wavelength conversion capability where wavelength assignment can be performed by the originating node of an optical LSP to minimize blocking probability. The suggested label concept permits an upstream node along a service path to start configuring its hardware with the suggested label before the downstream node communicates a label to it. For example, micro mirrors in a MEMS switch have to be positioned, and this physical motion and subsequent damping may consume some time. Early configuration offered by a suggested label can reduce setup latency, and may be important for restoration purposes as well, where alternate LSPs may need to be rapidly established. However, if a downstream node rejects the suggested label and passes a different label upstream, the upstream node must accept the label specified by the downstream node, thereby maintaining the downstream control of label allocation. In that situation, the switching fabric is configured in the reverse direction (the norm), and the label binding operation and propagation of the Resv/Mapping message upstream may need to be delayed at each hop in order to establish a usable forwarding path.

Bidirectional LSP Setup

Bidirectional optical LSPs (or lightpaths) are a requirement for many optical networking service providers. This section discusses how GMPLS supports bidirectional LSPs. It is assumed that both directions of such LSPs have the same traffic engineering requirements, including fate sharing, protection and restoration, and resource requirements (e.g., latency and jitter). The term *initiator* is used to refer to a node that starts the establishment of an LSP, and *terminator* is used to refer to the LSP destination node. Note that for a bidirectional LSP, there is only one initiator and one terminator.

In the basic MPLS architecture, LSPs are unidirectional, so in order to establish a bidirectional LSP using either [6] or [7], two unidirectional LSPs in opposite directions must be established independently. This approach has the following disadvantages:

- The latency to establish the bidirectional LSP is equal to one round-trip signaling time plus one initiator-terminator signaling transit delay. This increases the setup latency for LSP establishment, and extends the worst-case latency for discovering an unsuccessful LSP. These delays are particularly significant during network faults when LSPs need to be restored quickly.
- The control overhead is twice that of a unidirectional LSP. This is because separate control messages (e.g., Path/Request and Resv/Mapping) must be generated for both segments of the bidirectional LSP.
- Since the resources are established in separate segments, route selection can be quite complicated (e.g., Tx and Rx might be assigned to separate transponders on a WDM). There may be race conditions for resource allocation, which decreases the overall probability of successful establishment of the bidirectional connection.
- It is more difficult to provide a clean interface for SONET equipment which may rely on bidirectional hop-by-hop paths for protection switching. Existing SONET equipment transmits control information in-band using overhead bytes. This implies that connections should remain paired; hence, bidirectional establishment is highly desirable in this case.

A key requirement for the development of a common control plane for both optical and electronic networks is the need for features in the signaling, routing, and link management protocols to enable intelligent fault management.

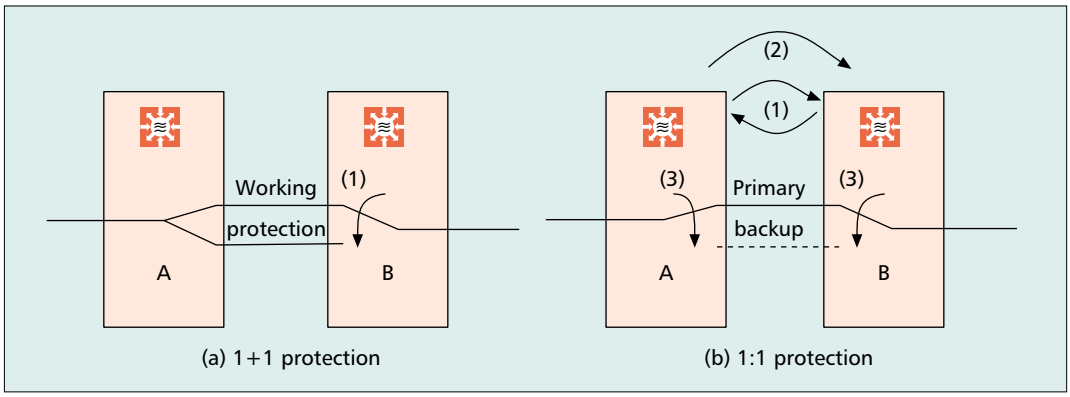

Figure 3. *a) In 1+1 span protection a connection is transmitted simultaneously over two disjoint channels (one working, one protection) and a selector is used at the receiving node to choose the best signal. If a failure occurs, (1) receiving node B switches from the working to protect channel. b) In 1:1 span protection (special case of* M:N *span protection) one dedicated backup channel is preallocated for the primary channel. If a failure occurs, (1) LMP is used to localize the failure. Once the failure has been localized, (2) an RSVP refresh message can be used to indicate path switchover, and (3) both nodes must switch to the backup channel.*

Additional methods have been defined to allow bidirectional LSPs' downstream and upstream data paths to be established using a single set of Path/Request and Resv/Mapping messages. This reduces the setup latency to essentially one initiator-terminator round-trip time plus processing time, and limits the control overhead to the same number of messages as a unidirectional LSP.

Notify Messages

One key requirement for providing network reliability is that reaction to network failures must be quick and decisions must be made intelligently. As part of failure notification, a node passing transit connections (i.e., connections that neither originate nor terminate at that node) should be able to notify the node(s) responsible for restoring the connections when failures occur, without intermediate nodes processing the messages or modifying the state of the affected connections. This is important because unnecessary message processing at the intermediate nodes may delay notification and even alter the state of the connection at the intermediate nodes. The Notify message has been added to RSVP-TE for GMPLS to provide a mechanism for informing nonadjacent nodes of LSP-related failures; a similar mechanism has not been defined for CR-LDP. The Notify message does not replace existing RSVP [12] error messages; however, it differs from them in that it can be "targeted" to any node other than the immediate upstream or downstream neighbor.

An important application for the nonadjacent Notify message is to notify when the control plane of a link has failed but the data plane (LSP) is still functional; a link in this condition is referred to as a *degraded link*. This is important because with GMPLS the control and data planes can be physically diverse and fail independently. In many cases, it is unacceptable to tear down an LSP just because the control plane has failed. Control plane failures, however, may limit the management features (e.g., failure localization, local restoration) provided for an LSP. As part of the notification process, the affected LSP and failed resource are identified in the Notify message, and new error codes have been added to identify degraded links.

GMPLS Protection and Restoration Techniques

A key requirement for the development of a common control plane for both optical and electronic networks is the need for features in the signaling, routing, and link management protocols to enable intelligent fault management. At the connection level, fault management consists of four primary steps:

- Detection
- Localization
- Notification
- Mitigation

Fault detection should be handled at the layer closest to the failure; for optical networks this is the physical (optical) layer. One measure of fault detection at the physical layer is detecting loss of light (LOL); other techniques based on optical signal-to-noise ratio (OSNR), optically measured bit error rate (BER), dispersion, crosstalk, and attenuation are still being developed.

Fault localization requires communication between nodes to determine where the failure has occurred (e.g., the SONET alarm indication signal, AIS, is used to localize failures between spans). One interesting consequence of using LOL to detect failures is that LOL propagates downstream along the connection's path, and therefore all downstream nodes may detect the failure. The Link Management Protocol (LMP) includes a fault localization procedure designed to localize failures in both transparent (all-optical) and opaque (optoelectrical) networks. This is done by sending LMP ChannelFail messages between adjacent nodes over a control channel maintained separately from the data-bearing channels. This separation of the control and data plane allows a single message set to be used for fault localization, irrespective of the encoding scheme of the data plane.

Once a failure has been detected and localized, protection and restoration are used to mitigate the failure. The distinction between protection and restoration is centered on the different time scales in which they operate. Protection requires preallocated resources (typically requiring 100 percent resource redundancy) and is designed to react to failures rapidly (less than a couple of hundred milliseconds). For example, SONET automatic protection switching (APS) is designed to switch traffic from a primary path to a secondary path in less than 50 ms. This requires simultaneous transmission along both a primary and a secondary path (called *1+1 protection*, see below) with a selector at the receiving node deciding which path to use. This approach requires twice as many network resources as a non-APS protected path. Restoration, on the other hand, relies on dynamic resource establishment, and it may take up to an order of magnitude longer to restore the connection than protection switching. Restoration may also involve dynamic route calculation, which can be computationally expensive if the backup paths are not precalculated or the precalculated resources are no longer available.

Protection and restoration have traditionally been addressed using two techniques: path switching and line switching. In path switching, the failure is addressed at the path endpoints (i.e., the path initiating and terminating nodes), whereas in line switching the failure is addressed at the transit node where the failure is detected. Path switching can be further subdivided into path protection, where secondary protection paths are preallocated, and path restoration, where connections are rerouted, either dynamically or using precalculated (but not preallocated) paths. Line switching can be subdivided into span protection, where traffic is switched to an alternate parallel channel or link connecting the same two nodes, and line restoration, where traffic is switched to an alternate route between the two nodes (this involves passing through additional intermediate nodes).

To effectively use protection, there must be mechanisms to:

- Distribute relevant link properties, such as protection bandwidth and protection capabilities
- Establish secondary paths through the network
- Signal a switch from the primary path to secondary paths and back

In the remainder of this section we focus on the second and third mechanisms as they relate to protection switching in GMPLS signaling. Specifically, we highlight the features of RSVP-TE signaling that will be used to provide protection and restoration. We discuss span and path protection, and focus on the signaling mechanisms used for fault notification and fast path switchover. We examine span and path restoration, where new resources are dynamically established when failures occur.

Protection Mechanisms

The nomenclature for protection mechanisms is as follows:

- 1+1 protection: Payload data is transmitted simultaneously over two disjoint paths (one working, one protection), and a selector is used at the receiving node to choose the best signal.
- *M*:*N* protection: *M* preallocated backup paths are shared between *N* primary paths; however, data is not replicated onto a backup path, but is assigned and transmitted over the backup path only on the failure of the primary path.
- 1:*N* protection: 1 preallocated backup path is shared among *N* primary paths.
- 1:1 protection: 1 dedicated backup path is preallocated for 1 primary path.

Note that 1:*N* and 1:1 protection are special cases of *M*:*N* protection.

Span Protection — Span protection is carried out between two adjacent nodes and involves switching to a backup channel or link when a failure occurs. As part of the GMPLS routing extensions ([8, 9]) the link protection type (LPT) is advertised so that span protection can be used in route calculation. Once a route is selected, the connection is signaled using RSVP-TE or CR-LDP, which includes a protection bit vector indicating which LPTs are acceptable for the connection.

Each node providing dedicated 1+1 span protection must replicate the data onto two separate channels (Fig. 3a). This requires using twice the connection bandwidth between the node pair and the ability to replicate the data onto both channels. When a failure is detected on the receiving node, it must switch from the working channel to the protect channel. Although the switchover itself does not require internode signaling, it is often done to inform the transmitting node that a switchover has occurred. This mechanism is used for connections requiring the highest availability over a span, and the working/protect channels must not be fate sharing so that a single failure across the link will not remove both working and protect channels.

For shared *M*:*N* span protection, *M* backup channels are shared between *N* primary channels (Fig. 3b shows the special case of 1:1 span protection). Since data is not replicated on both the primary and backup channels, failures must first be localized before the switchover can occur. Once a failure has been localized, the upstream node can initiate local span protection by sending an RSVP Path refresh message. Path refresh messages are a distinct feature of RSVP that allows intermediate nodes to refresh the state of an LSP. This feature allows for a switchover from the primary to the backup channel. Note that the benefit of exchanging the shared protection configuration in advance using LMP is that it minimizes the potential backup channel (label) conflict when protection switching. When the downstream node receives the Path message with the new objects, it verifies the parameters, updates the signaling state, and either responds with a Resv message with a new label or generates an error message.

Path Protection — Path protection is addressed at the end nodes (i.e., initiator and terminator nodes) and requires switching to an alternate path when a failure occurs.

As part of the GMPLS routing extensions the link protection type is advertised so that span protection can be used in route calculation. Once a route is selected, the connection is signaled using RSVP-TE or CR-LDP.

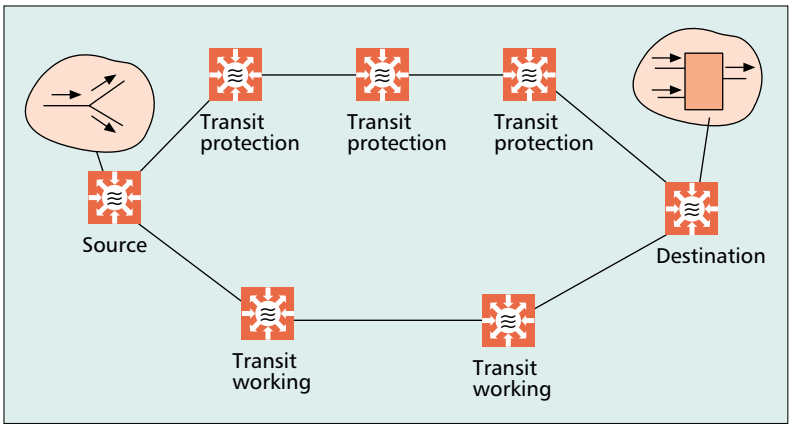

Figure 4. *1+1 path protection: the destination node switches to receive along the protection path when a failure is detected on the primary working path. Note that the signal is carried over both paths to the destination.*

After the two paths are computed, the source originates two explicitly routed connections with the *dedicated 1+1* and *unprotected* bits, respectively, set in the protection bit vector of the corresponding signaling setup message. The setup indicates that these two paths desire shared reservations. For 1+1 path protection, the connection is transmitted simultaneously over two disjoint paths, and a selector is used at the terminator node to choose the best signal (Fig. 4). At each node where the two paths branch out, the node must replicate the data into both branches. At each node where the two paths merge, the node must select the data from one path based on the integrity of the signal.

For *M*:*N* path protection, *N* different connections are transmitted along *N* disjoint paths, and *M* disjoint paths are preestablished for shared protection switching for the *N* primary paths. An interesting feature of GMPLS is that it allows preconfiguring backup paths to protect primary paths. These backup paths, called *secondary paths*, are used for fast switchover when the primary path fails. Although the resources for these backup paths are preallocated, lower-priority traffic may use the resources with the caveat that the lower-priority traffic will be preempted if there is a failure on the primary path. Note that the precomputed backup path cuts down on the restoration time in the event of a failure.

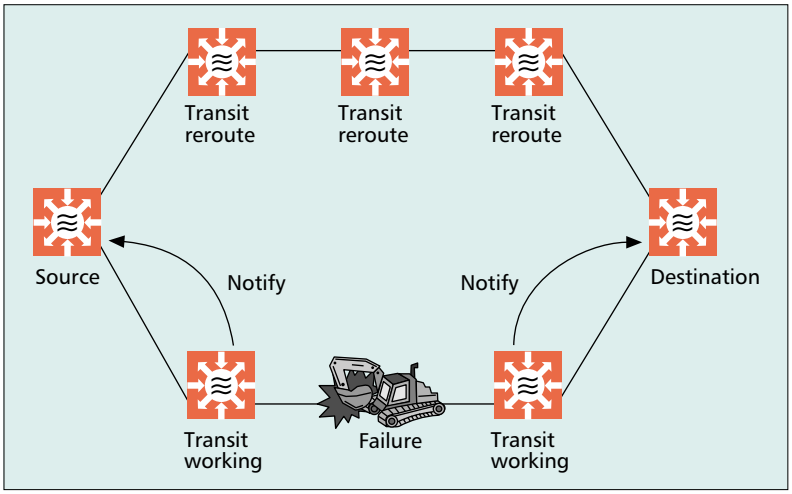

Figure 5. *Path restoration. On receipt of a failure notification, the source node computes the path to be used dynamically and signals for a new connection to be set up. Not shown is the possibility of reuse of some segments of the original path. Note that the original signal is not carried over both paths to the destination.*

Restoration Mechanisms

Restoration is designed to react to failures quickly and use bandwidth efficiently, but typically involves dynamic resource establishment and route calculation, and therefore takes more time to switch to an alternate path than protection techniques. Similar to protection techniques described previously, restoration can be implemented at the source or an intermediate node once the responsible node has been notified. Failure notification is done using the notify procedures discussed earlier, or using standard error messages.

To support line restoration, where traffic is switched via an alternate route around a failure, a new path is selected at an intermediate node. This involves passing traffic through additional transit nodes. Line restoration may be beneficial for connections that span multiple hops and/or large distances because the latency incurred for failure notification may be significantly reduced. In this case only segments of the connection are rerouted instead of the entire path. Line restoration, however, may break TE requirements if a strict-hop explicit route is defined for the connection. Furthermore, the constraints used for routing the connection must be forwarded so that an intermediate node doing line restoration is able to calculate an appropriate alternate route. This is similar to the problem of establishing/maintaining TE requirements that span multiple areas.

Path restoration, on the other hand, switches traffic to an alternate route around a failure, where the new path is selected at the source node. Optimizations may be invoked to speed the restoration process; for example, alternate routes may be precomputed by the head-end of the connection and cached for future use. A restored path may reuse nodes in the original path and/or include additional intermediate nodes (Fig. 5). It should be noted that resources at the downstream nodes are reused (shared) whenever possible, and resources at intermediate nodes that are no longer needed are freed. This sharing of resources increases the chances of the connection to get the requisite resources when rerouting is in progress. Note that if the paths are precomputed and resources preallocated, this enables faster reroute since those resources are guaranteed unless they fail or are claimed by higher-priority connections.

Conclusions

GMPLS will constitute an integral part of next-generation data and optical networks. It provides the necessary linkage between the IP and photonic layers, allowing interoperable, scalable, parallel, and cohesive evolution of networks in the IP and photonic dimensions. The functionality delivered by GMPLS allows network operators to scale their networks well beyond current limitations implicitly created by the segregation of the transport network from the data above. The signaling capabilities of GMPLS will allow service providers to quickly build out high-capacity agile

infrastructures that support fast provisioning of connection services. Now that demand for data transport has eclipsed that for voice, it is appropriate that the control plane mechanism expand to embrace data transport more closely while still serving the incumbent needs of voice transport. Service providers can incrementally deploy GMPLS-based products in existing networks to decrease costs without impacting service quality. Furthermore, the flexible *M*:*N* protection and restoration capabilities of GMPLS allow efficient addressing of network survivability, while opening the door to new types of services.

ACKNOWLEDGMENTS

The authors would like to thank George Swallow, Peter Ashwood-Smith, Eric Mannie, Yahne Fan, Greg Bernstien, Bala Rajagopalan, Debanjan Saha, Bo Tang, and Vishal Sharma, co-authors of the IETF drafts on GMPLS signaling.

REFERENCES

[1] D. Awduche and Y. Rekhter, "Multiprotocol Lambda Switching: Combining MPLS Traffic Engineering Control with Optical Cross-connects," *IEEE Commun. Mag.*, Mar. 2001.
[2] J. Moy, "OSPF Version 2," IETF RFC 2328.
[3] D. Oran, "OSI IS-IS Intra-Domain Routing Protocol," IETF RFC 1142.
[4] A. Banerjee *et al.*, "Generalized Multiprotocol Label Switching: An Overview of Routing and Management Enhancements," *IEEE Commun. Mag.*, vol. 39, no. 1, Jan. 2001.
[5] P. Ashwood-Smith et al., "Generalized MPLS — Signaling Functional Description," Internet draft, draft-ietf-mpls-generalized-signaling-03.txt, Apr. 2001, work in progress.
[6] P. Ashwood-Smith *et al.*, "Generalized MPLS Signaling — RSVP-TE Extensions," Internet draft, draft-ietf-mpls-generalized-rsvp-te-02.txt, Apr. 2001, work in progress.
[7] P. Ashwood-Smith et al., "Generalized MPLS Signaling —CR-LDP Extensions," Internet draft, draft-ietf-mpls-generalized-cr-ldp-02.txt, Apr. 2001, work in progress.
[8] K. Kompella *et al.*, "IS-IS Extensions in Support of Generalized MPLS," Internet draft, draft-ietf-gmpls-isis-extensions-02.txt, Mar. 2001, work in progress.
[9] K. Kompella *et al.*, "OSPF Extensions in Support of Generalized MPLS," Internet draft, draft-kompella-ospf-gmpls-extensions-01.txt, Mar. 2001, work in progress.
[10] J. Lang et al., "Link Management Protocol," Internet eraft, draft-lang-mpls-lmp-02.txt, Mar. 2001, work in progress.
[11] K. Kompella *et al.*, "LSP Hierarchy with MPLS TE," Internet draft, draft-ietf-mpls-lsp-hierarchy-02.txt, Mar. 2001, work in progress.
[12] R. Braden, Ed., "Resource Reservation Protocol (RSVP)," IETF RFC 2205.

BIOGRAPHIES

AYAN BANERJEE (abanerjee@calient.net) is a systems engineer at Calient Networks, where he assists in network and switch architecture design and analysis. Prior to joining Calient, he was a network scientist at BBN Technologies where he was involved in the design, analysis, and implementation of networking architectures and protocols. He received a B.Tech. degree from the Indian Institute of Technology, Kharagpur, and his M.S. and Ph.D. from the University of California at Santa Barbara. His research interests include switch architectures, constraint-based routing, topology control, dynamic communication algorithms, and software radios.

JOHN DRAKE [SM] (jdrake@calient.net) is chief architect at Calient Networks, where he is responsible for the development of the company's routing and networking architectures, and technical management of vendor partnerships. Before joining Calient, he was distinguished engineer in the Office of CTO at FORE Systems, a member of senior technical staff at IBM, and lead architect for IBM's Broadband Network Services. Some of his most significant accomplishments include the development of FORE's MPLS strategy, and leading the group that developed FORE's PNNI, MPOA, and LANE v2.0 implementations. He acquired 25 broadband networking patents during his tenure with FORE and IBM (e.g., distributed management communications network; quality of service management for source routing multimedia packet networks). He holds a B.S. in physics and mathematics from Carnegie-Mellon University.

JONATHAN P. LANG (jplang@calient.net) is systems engineer of network architecture at Calient Networks, where he focuses on the design of Calient's network and routing architecture. He is actively involved in the MPLS working group of the IETF, and has submitted numerous drafts extending MPLS to support intelligent optical networks. He has made several contributions to the OIF, and participates in the organization's architecture and signaling groups. Before joining Calient he worked for Whitetree, Inc., developing rate allocation algorithms for ATM. As a researcher and member of the Multidisplinary Optical Switching Technology center at the University of California, Santa Barbara, he was responsible for developing the switch architecture and networking protocols for a new optical packet switch. He received an M.S. and a Ph.D. in electrical engineering from the University of California, Santa Barbara, and a B.S. in electrical engineering from the University of California, San Diego.

BRAD TURNER (bturner@calient.net) is a senior product manager for Calient Networks, where he works on routing and signaling software for Calient's photonic switch platform. Prior to joining Calient, he held product management positions at Cisco Systems and 3Com Corporation, where he specialized in layer 3 routing and layer 2 switching systems. He has a B.S. in computer science from Birmingham-Southern College.

DANIEL AWDUCHE (awduche@movaz.com) is vice president of network architecture at Movaz Networks, an optical networking equipment company. Prior to Movaz, he served as distinguished technical member and acting director of Global Network Architecture at UUNET, a WorldCom company, a global provider of Internet communications services. At UUNET, he led a team of engineers engaged in architecture, design, and development activities aimed at UUNET's next-generation network. He is active in the IETF where he chairs the IP-over-Optical (IPO) working group. He contributed to the development of MPLS and helped pioneer the concept of multiprotocol lambda switching (MPLambdaS). A respected authority on traffic engineering, he has been very prominent in articulating some of the modern concepts of Internet traffic engineering.

LOU BERGER (lb@movaz.com) is associated with Movaz Networks.

KIREETI KOMPELLA (kireeti@juniper.net) is a senior protocols engineer at Juniper Networks. His current interests are all aspects of MPLS, including traffic engineering, GMPLS, and MPLS applications such as VPNs. He is active at the IETF where he is a co-chair of the CCAMP Working Group. Previously, he worked in the area of filesystems at Network Appliance and SGI. He received his B.S. in EE and M.S. in C.S. at the Indian Institute of Technology, Kanpur; and his Ph.D. in C.S. at the University of Southern California.

YAKOV REKHTER (yakov@juniper.net) is a distinguished engineer at Juniper Networks. Prior to Juniper he was a Cisco Fellow at Cisco Systems. He was one of the leading architects and a major software developer of the NSFNET Backbone Phase II. He co-designed the Border Gateway Protocol (BGP) which is the standard exterior gateway routing protocol for the Internet. He was also one of the lead designers of Tag Switching, BGP/MPLS-based VPNs, and MPLS traffic engineering. His other contributions to contemporary Internet technology include Classless Inter-Domain Routing (CIDR) and IP address allocation for private Internets. Among his most recent activities is the work on GMPLS (aka MPLambdaS). He is author or co-author of over 40 IETF RFCs, as well as numerous papers and articles on TCP/IP and the Internet.

Now that demand for data transport has eclipsed that for voice, it is appropriate that the control plane mechanism expand to embrace data transport more closely while still serving the incumbent needs of voice transport.