3rd OPTICAL SIGNALING, ROUTING AND MANAGEMENT Test Event

July 18 – 22, 2005

White Paper

OSRM Test Event
InterOperability Laboratory
Research Computing Center
University of New Hampshire

121 Technology Drive Suite 2
Durham, NH 03824
+1-603-862-0090
http://www.iol.unh.edu

Executive Summary

Service providers have long looked to Generalized Multi-Protocol Label Switching (GMPLS) for its promise to reduce the time involved in provisioning new services from months to minutes, thus saving significant network operation and management costs. These benefits are expected to result from using GMPLS as a common control plane for multi-layer set up and teardown of networks. Services such as cross-layer traffic engineering and multi-layer operation and restoration, in theory, should facilitate rapid and dynamic service provisioning and the roll-out of new revenue generating services more or less on demand.

However, testing next-generation network solutions such as GMPLS presents many new challenges. The networks in question are large and complex, and service provider needs, including new and existing quality of service (QoS) requirements, are paramount.

To meet these requirements, both for service providers deploying GMPLS and vendors bringing GMPLS-capable network devices to market, the University of New Hampshire InterOperability Laboratory (UNH-IOL) has worked with Japanese service provider Nippon Telegraph and Telephone Corp. (NTT) and the test participants to design and deploy multi-vendor test scenarios. These scenarios go beyond protocol verification to delve into the areas of the technology relevant to deploying GMPLS technologies in real carrier networks. The goals of these test scenarios, arguably the most extensive multi-vendor deployment-style GMPLS tests to date, are to showcase and explore new service and control capability concepts from a carrier perspective, to demonstrate GMPLS functionalities and level of maturity, and to promote GMPLS implementation among fully interoperable products.

The third test event in this series included NTT, and testing, optical equipment, and IP routing companies Agilent Technologies, Juniper Networks, Sycamore Networks, and Spirent Communications along with Japanese distributor Toyo Corporation. This year's testing, held July 18th-22nd, was designed with a new focus on the failure recovery mechanisms of a GMPLS network. The event provided a vendor-neutral setting that gave participants an opportunity to assess interoperability and valuable feedback to assist them in refining their implementations.

The neutral testing at the UNH-IOL included interconnected products demonstrating the functionality of various aspects of explicit route and label control to set up GMPLS traffic-engineered most suitable path, control channel failure recovery, data plane failure recovery by multi-layer traffic engineering, end-to-end protection in signaling, and the ability of GMPLS to manage diverse networks with increased scalability.

Introduction

Born from Multi-Protocol Label Switching (MPLS) technology, which was designed as a next-generation traffic engineering technology, GMPLS consists of a suite of optical signaling, routing, and management (referred to in this white paper as OSRM) protocols that enable dynamic end-to-end provisioning, maintenance, and teardown of connections across the electrical and optical transport domains. In effect, GMPLS

merges IP-based routing, signaling, and management with the optical realm. Based on the standards efforts of the Common Control and Measurement Plane Working Group (CCAMP-WG) of the Internet Engineering Task Force (IETF), GMPLS also provides a foundation for multi-vendor interoperability.

The test items implemented at the latest UNH-IOL GMPLS test event provided an exceptionally realistic and demanding test suite in line with service providers' operational demands, especially the need for failure recovery mechanisms.

Participants included NTT, as well as telecom equipment vendors Agilent Technologies, Juniper Networks, Sycamore Networks, and Spirent Communications, along with its ClearSight products and Japanese distributor Toyo Corporation.

Test Methodology

I. TE Link Configuration

The GMPLS architecture offers OSRM functionality over a variety of data-plane resources, called TE links. TE links can be configured to support many attributes, including numbered links, unnumbered links, bundled links, Forwarding Adjacency (FA)-LSPs, link protection types, etc. Support for all of these attributes allows service providers the maximum flexibility in establishing a GMPLS-LSP (G-LSP) end-to-end. However, not all features are suitable for all types of devices. For example, implementing numbered links is a likely requirement for a Packet Switch Capable (PSC) device, but unnumbered links are more applicable to Time Division Multiplexed (TDM), Lambda Switch Capable (LSC), and Fiber Switch Capable (FSC) devices. As TE links become more complex, proper encoding and decoding of the sub-TLVs in Open Shortest Path First - Traffic Engineering (OSPF-TE) LSAs is vital to interoperability among multi-layer devices.

Test Case #1. TE Links Advertisement
Properly interconnected devices were configured to exchange TE links via OSPF-TE LSAs**.** Numbered and unnumbered TE links were tested. The various configurable parameters characterizing the numbered and unnumbered TE links were properly exchanged. FA-LSPs were also verified as part of Test Case #8.

II. Bidirectional LSPs with Graceful Deletion

In a realistic GMPLS network, it is highly preferred that existing LSPs may be gracefully torn down without generating undesired alarms. In an optical environment, alarms (e.g., LOS) may be generated even when a G-LSP is administratively removed in the absence of any link failure or degradation. While automated generation of alarms is valuble information to network operators in other scenarios, it is unnecessary when intentionally removing an LSP. To prevent these alarms from being generated during a G-LSP removal process, RFC 3473 defines a set of procedures that GMPLS-capable nodes should employ when administratively removing a G-LSP. When implemented properly, a G-LSP may be torn down without causing the optical equipment to generate these alarms, unless configured otherwise. Graceful Deletion of Bidirectional LSPs was tested in two separate scenarios. Both scenarios are important as they represent different options in which the operator may control the state of the LSPs.

Test Case #2. Bidirectional LSP Setup and Teardown with Graceful Deletion – Requested by Initiator

The initiator indicated the removal of the LSP by including an Admin_Status object in its Path message with the Reflect (R) and Delete (D) bits set. Upon receiving a Resv message from the terminator with an Admin_Status object with the (D) bit set, the initiator proceded to delete the LSP by sending out a PathTear message. The LSP was torn down successfully.

Test Case #3. Bidirectional LSP Setup and Teardown with Graceful Deletion – Requested by Terminator

The terminator signaled its desire to have the LSP removed by inserting an Admin_Status object in its Resv message upstream with the (R) and (D) bits set. The ingress then performed the LSP removal by transmitting a PathTear message downstream. The LSP was torn down successfully.

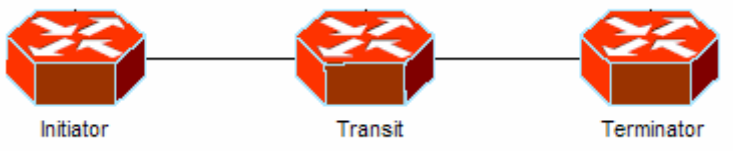

Figure 1: Bidirectional LSP Topology

III. Bidirectional LSPs with Explicit Route Control

Explicit Route Control is a preferred technique employed by service providers to fine-tune the datapath used for a predetermined set of traffic flows. In RSVP-TE, Explicit Route Control is achieved by implementing the Explicit Route Object (ERO), in which a list of network hops are specified when signaling an LSP. Explicit Route Control enables a network operator to dictate some or all hops in the path of the LSP. This allows operators to easily segment traffic flows and efficiently utilize network resources in providing transport services. Two types of explicit hops are defined: Strict is used when it is necessary to specify a directly connected node as the next hop; Loose is used when

it is sufficient to specify a node in the path of the LSP that may be connected via one or more nodes.

Test Case #4. Bidirectional LSP Setup with Explicit Route Control – Strict ERO
Two separate datapaths were set up between the initiator and the terminator, where one path had a lower OSPF cost than the other. The route of the LSP was explicitly (manually) configured for each hop, and the path chosen was a less optimal path. The LSP was observed to establish the connection over the less optimal path, as specified by the strict ERO.

Test Case #5. Bidirectional LSP Setup with Explicit Route Control – Loose ERO
Two separate datapaths were set up between the initiator and the terminator, where one path had a lower OSPF cost than the other. The route was explicitly configured for certain hops, and the devices calculated the most optimal path to the hop(s) specified by the loose ERO configuration. The LSP was observed to take the most optimal path available to the hop(s) specified by configuration.

IV. Bidirectional LSPs with Explicit Label Control

In GMPLS, it is sometimes necessary to fine-tune a datapath beyond what Explicit Route Control can provide. Explicit Label Control is a technique introduced with GMPLS that allows a network operator to control a specific wavelength or timeslot that an LSP must use at a certain hop. For example, Path Computation Element (PCE) can integrate the route and label information as criteria in determining an LSP path. Explicit Label Control augmented by PCE can create the optimal end-to-end path across a large GMPLS network without complicated manual configurations by an operator.

Test Case #6. Bidirectional LSP Setup with Explicit Label Control
The initiator signaled a bidirectional LSP with an ERO in its Path message. The ERO included sub-objects specifying the interface address of the hops, followed by a label to be used at each of the hops. Record Route Object (RRO) recording was also verified during the process. The LSP was allowed to establish using the specified labels.

Test Case #7. Bidirectional LSP Setup with Egress Label Control
The initiator signaled a bidirectional LSP with an ERO in its Path message. The ERO included sub-objects specifying the interface address of the hops, followed by a label to be used at each of the hops. RRO recording was also verified during the process. In addition, the initiator specified a label to be used by the outgoing interface of the egress node. The LSP was allowed to establish using the specified labels.

V. Hierarchical LSPs

In the GMPLS architecture, a higher layer LSP may be nested within a lower layer LSP. Fundamentally, hierarchical LSPs are FA-LSPs – the higher layer LSP traverses nodes that appear to be another TE link, but are, in fact, parts of a

lower layer LSP. One obvious advantage to the deployment of hierarchical LSPs is that multiple LSPs can be grouped into a very high bandwidth pipe, namely, the lower layer LSP. From a network level perspective, hierarchical LSPs offer enhanced scalability. Another benefit of hierarchical LSPs for service providers is the ability to flexibly establish an LSP across a diverse network with various switching types.

Test Case #8. Hierarchical LSP Setup and Teardown (Graceful)
A lower layer LSP was first established within the core network and advertised as an FA via OSPF. After the routing tables synchronized, a comparatively higher layer LSP was then signaled, with an ERO specifying the FA as a hop between the edge end points. The hierarchical LSP was verified in both the control plane and data plane.

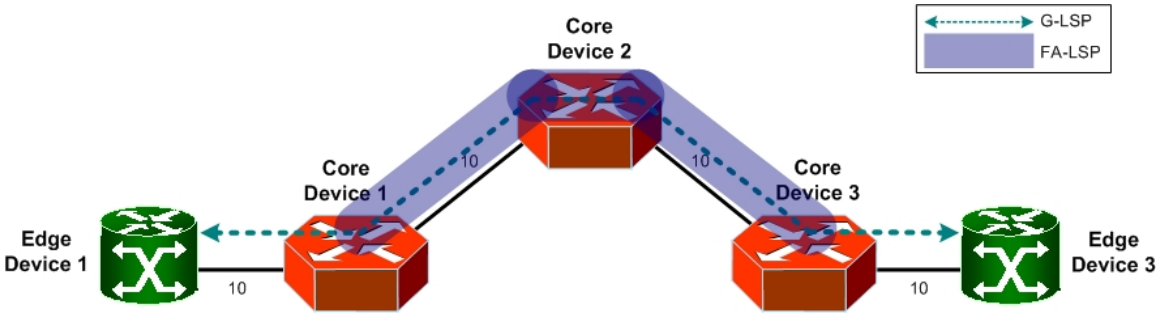

Figure 2: Hierarchical LSP Topology

VI. Failure Recovery – Control Plane

GMPLS partitions its control plane and data plane into two separate channels, thus allowing a service provider to dedicate appropriate network resources for the control plane and data plane as necessary. For example, the control plane and the data plane are likely to have very different protection needs at the link layer. However, one requirement in a realistic GMPLS network is that temporary failures in one of the planes must not cause problems in the other. In the control plane, Graceful Restart allows a pair of GMPLS nodes to preserve their MPLS forwarding state throughout a temporary control plane failure (e.g., control link pull and control plane reset), thus avoiding the need to reinitialize the signaling process of the LSP.

RSVP Graceful Restart deals with two types of failures – nodal faults and control channel faults. In the event of a nodal fault, a restart of the control plane has occurred and all RSVP states are lost, but the data forwarding state is maintained. In this case a recovery label (label used by the restarting router prior to the restart) is sent by the neighbor to the restarting node to recover its RSVP states. Control channel faults are restricted to control channel failures, in which the data forwarding state is also maintained. In addition to the nodal fault case, the RSVP states are also maintained. In both cases, RSVP routes or MPLS labels should be maintained to continue forwarding traffic in the data plane without disruption. In addition to preserving forwarding across restarts, RSVP Graceful Restart provides a helper capability, wherein a node can help a

restarting neighbor with its restart and recovery procedures. RSVP will advertise restart and recovery time in the Hello messages.

Test Case #9. Graceful Restart
The neighbor of a restarting node that supports state recovery as described in RFC 3473 Section 9.3 is referred to as a helper node in this document. During a control channel failure, the helper node sent Hello messages with Dst_Instance equal to zero and the Src_Instance unchanged, and the restarting node sent Hello messages with a non-zero Recovery Time. Before the Recovery Time expired, the helper node and the restarting node were able to preserve the MPLS forwarding state.

VII. Multi-Layer Failure Recovery – Data Plane

Failure recovery in the data plane is extremely important as subscribers are demanding services with virtually no interruptions. In GMPLS, this requirement is furthur complicated with a multi-layered infrastructure. Multi-layer failure recovery enables an operator to manage both the optical and packet paths dynamically, and to flexibly recover from failures, by coordinating the optical and packet layers. Single layer recovery methods may not be as flexible due to limited end-to-end paths that span across a carrier network over just one layer. While protection and restoration (P&R) schemes providing zero or minimum (e.g., sub 5-ms recovery) packet loss are important, it is economically attractive to service providers to offer failure recovery methods that do not use dedicated backup paths, which introduces tolerable service disruptions. By not dedicating resources primarily to restoration purposes reduces redundancy in the network, which, in turn, lowers infrastructure costs, especially in a large scale network. This event offered participants the opportunity to demonstrate an interoperable, multi-vendor, multi-layer failure recovery mechanism using high bandwidth video streaming in the data plane.

Test Case #10. Multi-Layer Traffic Engineering
An FSC-LSP was established as an FA. A PSC-LSP was then established end-to-end to interface with the packet-based content server and client, using the FA-LSP as a TE-link. High bandwidth video was streamed from the content server to the client via the PSC-LSP. To verify the recovery mechanisms, the end-to-end datapath was disrupted by physically removing a fiber. During the recovery process, CSPF calculated an alternate datapath for the FSC-LSP and automatically established a new FSC-LSP between available nodes. Immediately after the FSC-LSP established successfully, a new PSC-LSP was established to provide the appropriate end-to-end transport for the content server and client. During the failure recovery process, video streaming was disrupted for a few seconds, and the TE-link disconnect was “transparent” to the PSC-LSP.

VIII. Scalability – LSP Scalability

In order for a new transport technology to be adopted by large carriers, it must be capable of scaling to meet growing needs. LSP is the core concept of the MPLS/GMPLS transport method, therefore the LSPs must be scalable. In addition,

MPLS/GMPLS must be able to establish and tear down large numbers of LSPs without negative impact on the existing network. Other forms of scalability are also important, for example, the number of nodes in a GMPLS domain, the number of TE links the OSPF tables can manage, or the convergence time of the GMPLS network itself. In this event, with the equipment available, an initial LSP scalability test was executed to gauge the impact of multiple, simultaneous LSP creations and deletions on various vendors' GMPLS implementations. Also, a total of five 10x10 grid simulated nodes were verified, and all node IDs and TE links were properly exchanged in the OSPF tables.

Discoveries for Further Investigation

List of Issues Encountered

Problem Area	Problem in General	Problem in Test Event	Proposed Solution
Graceful Restart	The total duration of the Recovery Period is advertised by the recovering node in the Recovery Time field of the Restart_Cap object. In addition, the Recovery Time is used by a restarting node to notify its neighbors whether the forwarding state was preserved. A Recovery Time value of zero (0) is defined by RFC 3473 Section 9.1 to indicate that the MPLS forwarding state was not preserved across a particular reboot. However, it does not define its meaning except for the above scenario. The Recovery Period is the period during which a node is performing the recovery process, from the instant Hello synchronization is re-established to the instant when pre-failure conditions are restored – namely the MPLS forwarding state.	A Recovery Time value of zero (0) was advertised by the restarting node prior to the Recovery Period. Upon receiving a Recovery Time of zero (0), the receiver reported an error because a non-zero Recovery Time was expected.	The meaning and/or interpretation of a Recovery Time of zero (0) prior to and after the Recovery Period must be defined in the relevant standards. If the Recovery Time is intended to be meaningless prior to and after the Recovery Period, a node receiving a Restart_Cap object with a Recoery Time of zero (0), or any value, SHOULD silently ignore the encoded value.
RSVP Hello	RFC 3209 Section 5.3 states "The sender also fills in the Dst_Instance field with the Src_Instance value most recently received from the neighbor." It does not explicitly specify that a node MUST fill in the Dst_Instance field with the Neighbor_Src_Instance value most recently	A node waited for several Hello messages before it attempted to form a Hello session with its neighbor, as a method to ensure the neighbor is reasonably stable. During this period, the node continues to send out Hellos with a Dst_Instance value of zero (0), since a Hello session has not established	If a node supports GMPLS, it MUST always fill in the Dst_Instance field with the Neighbor_Src_Instance value most recently received from its neighbor. In such a case, it must not use a Dst_Instance value of zero (0) unless it has determined communication with its neighbor has been lost.

	received from the neighbor. The current text in RFC 3209 regarding the implementation of the Dst_Instance field lacks enforcement. As a result, some implementations do not always update the Dst_Instance field immediately in a Hello ACK.	yet. This behavior is not explicitly prohibited by RFC 3209, as polling is allowed. On the other hand, its neighbor generated an error message because the Dst_Instance field was not updated to the most recent Src_Instance value it sent out, as other sections of RFC 3209 specify. In addition, continuing to send out Hello messages with a Dst_Instance value of zero (0) also causes a receiver to interpret that the sender is performing the Graceful Restart procedure (i.e.,as if it has restarted its control plane), as RFC 3473 Section 9.3 states that during a recovery process, "all Hello messages MUST be sent with a Dst_Instance values set to zero (0)."	
RSVP ERO/RRO	The order in which Route and Label sub-objects are recorded into the RECORD_ROUTE object and the EXPLICIT_ROUTE object is not clearly understood.	RFC 3473 Section 5.1 describes the Label sub-object following a sub-object containing the IP address, or the interface identifier. RFC 3209 Section 4.4 describes the Label Record sub-object being pushed onto the RECORD_ROUTE object prior to the node's IP address. During the test event, it was observed that some vendors' implementations recorded the Label sub-object following a sub-object containing the IP address.	When the node supports label recording, the Label sub-object for RRO follows a sub-object containing the IP address, or the interface identifier.
RSVP ERO/RRO	The Label Recording flag is only set in the Path message, not the Resv message. Therefore, in the Resv direction, the internal nodes do not record the Label sub-object and the sender node does not receive Label sub-objects, even if the sender node desires label recording.	RFC 3209 states that if the node also desires label recording, it sets the Label_recording flag in the SESSION_ATTRIBUTE object, and when the Label Recording flag is set in the SESSION_ATTRIBUTE object, nodes doing route recording SHOULD include a Label Record subobject. But the SESSION_ATTRIBUTE object exists only in Path messages, not in Resv messages. Therefore, in the Resv direction, the internal	When the Label Recording flag is set in the SESSION_ATTRIBUTE object of the Path message, internal nodes performing route recording SHOULD include a Label Record sub-object in the Resv direction as well as in the downstream direction.

		nodes do not record the Label sub-object, even if the sender node desires label recording.	

Conclusion

The third OSRM test event at the UNH-IOL confirmed a number of GMPLS capabilities. Next generation networks with GMPLS technologies enable carriers to manage both optical and packet paths dynamically, and recover them flexibly, by coordinating optical and packet layers in a way that single layer recovery can not.

At the conclusion of the test event, the interoperability of many fundamental functions of GMPLS were verified, such as Bidirectional LSP Setup and Teardown with Graceful Deletion, Hierarchical LSPs, and LSP Setup with Explicit Route and Label Specification. Proper behaviors of the selected advanced features were also confirmed, such as consistent route and label recording, precise egress label control, and multi-layer protection and restoration features; in addition to an initial study of multi-vendor LSP scalability and GMPLS network scalability.

Carriers can benefit from these functions, and very likely from other features of GMPLS, when employing the common optical control plane for multi-layer and multi-service networks. Uniting disparate networks under a single control plane allows for better resource utilization, more flexible provisioning, and highly intelligent failure recovery. GMPLS allows service providers to more economically transport large-scale, IP-based packet traffic while achieving high reliability and multi-QoS (quality of service) support.

Recommendations for Further Investigation

As a result of the OSRM test methodologies and findings described above, several facets of OSRM technology emerged as compelling candidates for further testing:

- LSP formation with all switching capabilities end-to-end: Testing with multiple PSC/TDM/LSC/FSC nodes in one LSP
- Hierarchical LSPs with all switching capabilities
- Interoperable protection and restoration scenarios – single layer and multi-layer
- Increased scalability: Testing with a greater number of nodes, TE links, LSPs and setup/teardown patterns
- Link Management Protocol (LMP)
- GMPLS OAM
- GMPLS and UNI and NNI interworking

The UNH-IOL looks forward to working with service providers and all participants in the OSRM test events to further investigate these and other aspects of GMPLS that are equally important to validation in complex operational networks.

Glossary

Abbreviation	Definition
FSC	**F**iber **S**witch **C**apable device. An optical cross connect that switches the contents of a whole fiber to another fiber.
G-LSP	**G**MPLS **LSP**. NOT an IETF defined terminology. Used in this document to explicitly distinguish a GMPLS LSP from a conventional RFC 3031 LSP.
GMPLS	**G**eneralized **M**ulti-**P**rotocol **L**abel **S**witching. An architecture that extends the MPLS architecture defined by RFC 3031 to support transport links.
LSC	**L**ambda **S**witch **C**apable device. An optical cross connect that switches incoming data based on wavelengths.
LSP	**L**abel **S**witched **P**ath. A virtual tunnel across an MPLS domain that switches data by labels in the Shim header.
MPLS	**M**ulti-**P**rotocol **L**abel **S**witching. An architecture that introduces the concept of forwarding packets across a network by labels instead of routing.
NNI	**N**etwork-to-**N**etwork **I**nterface. The interface between two network nodes.
OSPF	**O**pen **S**hortest **P**ath **F**irst. A routing protocol that calculates the least-cost path to all network points within a routing area.
OSPF-TE	**T**raffic **E**ngineering Extensions to **O**pen **S**hortest **P**ath **F**irst.
PSC	**P**acket **S**witch **C**apable device. A device that switches incoming data based on the packet header.
RFC	Request For Comments. A document that specifies the standard behaviors of a protocol.
RSVP-TE	Resource **R**e**S**er**V**ation **P**rotocol – **T**raffic **E**ngineering Extensions
TDM	**T**ime **D**ivision **M**ultiplexed device. A device that switches incoming data based on a specific slot in time.
TLV	**T**ype/**L**ength/**V**alue. A message format in which OSPF elements are specified.
UNI	**U**ser-to-**N**etwork **I**nterface. The interface between an edge node and a network node.

References

Request for Comments 2205 – Resource ReSerVation Protocol (RSVP) -- Version 1 Functional Specification

Request for Comments 2328 – Open Shortest Path First (OSPF) Version 2

Request for Comments 3031 – Multi-Protocol Label Switching Architecture

Request for Comments 3032 – MPLS Label Stack Encoding

Request for Comments 3209 – RSVP-TE: Extensions to RSVP for LSP Tunnels

Request for Comments 3471 – Generalized Multi-Protocol Label Switching (GMPLS) Signaling Functional Description

Request for Comments 3473 – Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions

Request for Comments 3477 – Signalling Unnumbered Links in Resource ReSerVation Protocol – Traffic Engineering (RSVP-TE)

Request for Comments 3945 – Generalized Multi-Protocol Label Switching (GMPLS) Architecture

Request for Comments 4003 - GMPLS Signaling Procedure for Egress Control

Internet Draft draft-ietf-ccamp-gmpls-recovery-e2e-signaling-03.txt – RSVP-TE Extensions in support of End-to-End GMPLS-based Recovery

Internet Draft draft-ietf-ccamp-gmpls-routing-09.txt – Routing Extensions in Support of Generalized MPLS

Internet Draft draft-ietf-ccamp-gmpls-recovery-terminology-03.txt – Recovery (Protection and Restoration) Terminology for Generalized Multi-Protocol Label Switching (GMPLS)

Internet Draft draft-ietf-ccamp-ospf-gmpls-extensions-12.txt – OSPF Extensions in Support of Generalized Multi-Protocol Label Switching

Internet Draft draft-ietf-mpls-lsp-hierarchy-08.txt – LSP Hierarchy with Generalized MPLS TE

Contributors

The content of this whitepaper is an accumulation of agreements, input, and comments from test event participants. Takumi Ohba and Kaori Shimizu from NTT Network Service Systems Laboratories and Henry He from UNH-IOL MPLS Services Consortium wrote the initial content. Special thanks to Chris Volpe from UNH-IOL for his work, especially on the Executive Summary and Conclusion sections. Much appreciation for the contributions of John Allen from Juniper Networks and Ankur Chadda from Spirent Communications to the Graceful Restart sections. Many thanks to Scott Larson from Sycamore Networks for his contributions on many sections of this paper.