Network Working Group
Request for Comments: 5212
Category: Informational

K. Shiomoto
NTT
D. Papadimitriou
Alcatel-Lucent
JL. Le Roux
France Telecom
M. Vigoureux
Alcatel-Lucent
D. Brungard
AT&T

July 2008

Requirements for GMPLS-Based Multi-Region and Multi-Layer Networks (MRN/MLN)

Status of This Memo

Abstract

Most of the initial efforts to utilize Generalized MPLS (GMPLS) have been related to environments hosting devices with a single switching capability. The complexity raised by the control of such data planes is similar to that seen in classical IP/MPLS networks. By extending MPLS to support multiple switching technologies, GMPLS provides a comprehensive framework for the control of a multi-layered network of either a single switching technology or multiple switching technologies.

In GMPLS, a switching technology domain defines a region, and a network of multiple switching types is referred to in this document as a multi-region network (MRN). When referring in general to a layered network, which may consist of either single or multiple regions, this document uses the term multi-layer network (MLN). This document defines a framework for GMPLS based multi-region / multi-layer networks and lists a set of functional requirements.

Table of Contents

1. Introduction

Generalized MPLS (GMPLS) extends MPLS to handle multiple switching technologies: packet switching, Layer-2 switching, TDM (Time-Division Multiplexing) switching, wavelength switching, and fiber switching (see [RFC3945]). The Interface Switching Capability (ISC) concept is introduced for these switching technologies and is designated as follows: PSC (packet switch capable), L2SC (Layer-2 switch capable), TDM capable, LSC (lambda switch capable), and FSC (fiber switch capable).

The representation, in a GMPLS control plane, of a switching technology domain is referred to as a region [RFC4206]. A switching type describes the ability of a node to forward data of a particular data plane technology, and uniquely identifies a network region. A layer describes a data plane switching granularity level (e.g., VC4, VC-12). A data plane layer is associated with a region in the control plane (e.g., VC4 is associated with TDM, MPLS is associated with PSC). However, more than one data plane layer can be associated with the same region (e.g., both VC4 and VC12 are associated with TDM). Thus, a control plane region, identified by its switching type value (e.g., TDM), can be sub-divided into smaller-granularity component networks based on "data plane switching layers". The Interface Switching Capability Descriptor (ISCD) [RFC4202], identifying the interface switching capability (ISC), the encoding type, and the switching bandwidth granularity, enables the characterization of the associated layers.

In this document, we define a multi-layer network (MLN) to be a Traffic Engineering (TE) domain comprising multiple data plane switching layers either of the same ISC (e.g., TDM) or different ISC (e.g., TDM and PSC) and controlled by a single GMPLS control plane instance. We further define a particular case of MLNs. A multi-region network (MRN) is defined as a TE domain supporting at least two different switching types (e.g., PSC and TDM), either hosted on the same device or on different ones, and under the control of a single GMPLS control plane instance.

MLNs can be further categorized according to the distribution of the ISCs among the Label Switching Routers (LSRs):

- Each LSR may support just one ISC.
  Such LSRs are known as single-switching-type-capable LSRs. The MLN may comprise a set of single-switching-type-capable LSRs some of which support different ISCs.

- Each LSR may support more than one ISC at the same time. Such LSRs are known as multi-switching-type-capable LSRs, and can be further classified as either "simplex" or "hybrid" nodes as defined in Section 4.2.

- The MLN may be constructed from any combination of single-switching-type-capable LSRs and multi-switching-type-capable LSRs.

Since GMPLS provides a comprehensive framework for the control of different switching capabilities, a single GMPLS instance may be used to control the MLN/MRN. This enables rapid service provisioning and efficient traffic engineering across all switching capabilities. In such networks, TE links are consolidated into a single Traffic Engineering Database (TED). Since this TED contains the information relative to all the different regions and layers existing in the network, a path across multiple regions or layers can be computed using this TED. Thus, optimization of network resources can be achieved across the whole MLN/MRN.

Consider, for example, a MRN consisting of packet-switch-capable routers and TDM cross-connects. Assume that a packet Label Switched Path (LSP) is routed between source and destination packet-switch-capable routers, and that the LSP can be routed across the PSC region (i.e., utilizing only resources of the packet region topology). If the performance objective for the packet LSP is not satisfied, new TE links may be created between the packet-switch-capable routers across the TDM-region (for example, VC-12 links) and the LSP can be routed over those TE links. Furthermore, even if the LSP can be successfully established across the PSC-region, TDM hierarchical LSPs (across the TDM region between the packet-switch capable routers) may be established and used if doing so is necessary to meet the operator’s objectives for network resource availability (e.g., link bandwidth). The same considerations hold when VC4 LSPs are provisioned to provide extra flexibility for the VC12 and/or VC11 layers in an MLN.

Sections 3 and 4 of this document provide further background information of the concepts and motivation behind multi-region and multi-layer networks. Section 5 presents detailed requirements for protocols used to implement such networks.

1.1. Scope

Early sections of this document describe the motivations and reasoning that require the development and deployment of MRN/MLN. Later sections of this document set out the required features that the GMPLS control plane must offer to support MRN/MLN. There is no intention to specify solution-specific and/or protocol elements in

this document. The applicability of existing GMPLS protocols and any protocol extensions to the MRN/MLN is addressed in separate documents [MRN-EVAL].

This document covers the elements of a single GMPLS control plane instance controlling multiple layers within a given TE domain. A control plane instance can serve one, two, or more layers. Other possible approaches such as having multiple control plane instances serving disjoint sets of layers are outside the scope of this document. It is most probable that such a MLN or MRN would be operated by a single service provider, but this document does not exclude the possibility of two layers (or regions) being under different administrative control (for example, by different Service Providers that share a single control plane instance) where the administrative domains are prepared to share a limited amount of information.

For such a TE domain to interoperate with edge nodes/domains supporting non-GMPLS interfaces (such as those defined by other standards development organizations (SDOs)), an interworking function may be needed. Location and specification of this function are outside the scope of this document (because interworking aspects are strictly under the responsibility of the interworking function).

This document assumes that the interconnection of adjacent MRN/MLN TE domains makes use of [RFC4726] when their edges also support inter-domain GMPLS RSVP-TE extensions.

2. Conventions Used in This Document

Although this is not a protocol specification, the key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" are used in this document to highlight requirements, and are to be interpreted as described in RFC 2119 [RFC2119].

In the context of this document, an end-to-end LSP is defined as an LSP that starts in some client layer, ends in the same layer, and may cross one or more lower layers. In terms of switching capabilities, this means that if the outgoing interface on the head-end LSR has interface switching capability X, then the incoming interface on the tail-end LSR also has switching capability X. Further, for any interface traversed by the LSP at any intermediate LSR, the switching capability of that interface, Y, is such that Y >= X.

2.1. List of Acronyms

ERO: Explicit Route Object
FA: Forwarding Adjacency
FA-LSP: Forwarding Adjacency Label Switched Path
FSC: Fiber Switching Capable
ISC: Interface Switching Capability
ISCD: Interface Switching Capability Descriptor
L2SC: Layer-2 Switching Capable
LSC: Lambda Switching Capable
LSP: Label Switched Path
LSR: Label Switching Router
MLN: Multi-Layer Network
MRN: Multi-Region Network
PSC: Packet Switching Capable
SRLG: Shared Risk Link Group
TDM: Time-Division Multiplexing
TE: Traffic Engineering
TED: Traffic Engineering Database
VNT: Virtual Network Topology

3. Positioning

A multi-region network (MRN) is always a multi-layer network (MLN) since the network devices on region boundaries bring together different ISCs. A MLN, however, is not necessarily a MRN since multiple layers could be fully contained within a single region. For example, VC12, VC4, and VC4-4c are different layers of the TDM region.

3.1. Data Plane Layers and Control Plane Regions

A data plane layer is a collection of network resources capable of terminating and/or switching data traffic of a particular format [RFC4397]. These resources can be used for establishing LSPs for traffic delivery. For example, VC-11 and VC4-64c represent two different layers.

From the control plane viewpoint, an LSP region is defined as a set of one or more data plane layers that share the same type of switching technology, that is, the same switching type. For example, VC-11, VC-4, and VC-4-7v layers are part of the same TDM region. The regions that are currently defined are: PSC, L2SC, TDM, LSC, and FSC. Hence, an LSP region is a technology domain (identified by the ISC type) for which data plane resources (i.e., data links) are represented into the control plane as an aggregate of TE information

associated with a set of links (i.e., TE links). For example, VC-11 and VC4-64c capable TE links are part of the same TDM region. Multiple layers can thus exist in a single region network.

Note also that the region may produce a distinction within the control plane. Layers of the same region share the same switching technology and, therefore, use the same set of technology-specific signaling objects and technology-specific value setting of TE link attributes within the control plane, but layers from different regions may use different technology-specific objects and TE attribute values. This means that it may not be possible to simply forward the signaling message between LSRs that host different switching technologies. This is due to changes in some of the signaling objects (for example, the traffic parameters) when crossing a region boundary even if a single control plane instance is used to manage the whole MRN. We may solve this issue by using triggered signaling (see Section 4.3.1).

3.2. Service Layer Networks

A service provider's network may be divided into different service layers. The customer's network is considered from the provider's perspective as the highest service layer. It interfaces to the highest service layer of the service provider's network. Connectivity across the highest service layer of the service provider's network may be provided with support from successively lower service layers. Service layers are realized via a hierarchy of network layers located generally in several regions and commonly arranged according to the switching capabilities of network devices.

For instance, some customers purchase Layer-1 (i.e., transport) services from the service provider, some Layer 2 (e.g., ATM), while others purchase Layer-3 (IP/MPLS) services. The service provider realizes the services by a stack of network layers located within one or more network regions. The network layers are commonly arranged according to the switching capabilities of the devices in the networks. Thus, a customer network may be provided on top of the GMPLS-based multi-region/multi-layer network. For example, a Layer-1 service (realized via the network layers of TDM, and/or LSC, and/or FSC regions) may support a Layer-2 network (realized via ATM Virtual Path / Virtual Circuit (VP/VC)), which may itself support a Layer-3 network (IP/MPLS region). The supported data plane relationship is a data plane client-server relationship where the lower layer provides a service for the higher layer using the data links realized in the lower layer.

Services provided by a GMPLS-based multi-region/multi-layer network are referred to as "multi-region/multi-layer network services". For example, legacy IP and IP/MPLS networks can be supported on top of multi-region/multi-layer networks. It has to be emphasized that delivery of such diverse services is a strong motivator for the deployment of multi-region/multi-layer networks.

A customer network may be provided on top of a server GMPLS-based MRN/MLN which is operated by a service provider. For example, a pure IP and/or an IP/MPLS network can be provided on top of GMPLS-based packet-over-optical networks [RFC5146]. The relationship between the networks is a client/server relationship and, such services are referred to as "MRN/MLN services". In this case, the customer network may form part of the MRN/MLN or may be partially separated, for example, to maintain separate routing information but retain common signaling.

3.3. Vertical and Horizontal Interaction and Integration

Vertical interaction is defined as the collaborative mechanisms within a network element that is capable of supporting more than one layer or region and of realizing the client/server relationships between the layers or regions. Protocol exchanges between two network controllers managing different regions or layers are also a vertical interaction. Integration of these interactions as part of the control plane is referred to as vertical integration. Thus, this refers to the collaborative mechanisms within a single control plane instance driving multiple network layers that are part of the same region or not. Such a concept is useful in order to construct a framework that facilitates efficient network resource usage and rapid service provisioning in carrier networks that are based on multiple layers, switching technologies, or ISCs.

Horizontal interaction is defined as the protocol exchange between network controllers that manage transport nodes within a given layer or region. For instance, the control plane interaction between two TDM network elements switching at OC-48 is an example of horizontal interaction. GMPLS protocol operations handle horizontal interactions within the same routing area. The case where the interaction takes place across a domain boundary, such as between two routing areas within the same network layer, is evaluated as part of the inter-domain work [RFC4726], and is referred to as horizontal integration. Thus, horizontal integration refers to the collaborative mechanisms between network partitions and/or administrative divisions such as routing areas or autonomous systems.

This distinction needs further clarification when administrative domains match layer/region boundaries. Horizontal interaction is extended to cover such cases. For example, the collaborative mechanisms in place between two LSC areas relate to horizontal integration. On the other hand, the collaborative mechanisms in place between a PSC (e.g., IP/MPLS) domain and a separate TDM capable (e.g., VC4 Synchronous Digital Hierarchy (SDH)) domain over which it operates are part of the horizontal integration, while it can also be seen as a first step towards vertical integration.

3.4. Motivation

The applicability of GMPLS to multiple switching technologies provides a unified control and management approach for both LSP provisioning and recovery. Indeed, one of the main motivations for unifying the capabilities and operations of the GMPLS control plane is the desire to support multi-LSP-region [RFC4206] routing and TE capabilities. For instance, this enables effective network resource utilization of both the Packet/Layer2 LSP regions and the TDM or Lambda LSP regions in high-capacity networks.

The rationales for GMPLS-controlled multi-layer/multi-region networks are summarized below:

- The maintenance of multiple instances of the control plane on devices hosting more than one switching capability not only increases the complexity of the interactions between control plane instances, but also increases the total amount of processing each individual control plane instance must handle.

- The unification of the addressing spaces helps in avoiding multiple identifiers for the same object (a link, for instance, or more generally, any network resource). On the other hand such aggregation does not impact the separation between the control plane and the data plane.

- By maintaining a single routing protocol instance and a single TE database per LSR, a unified control plane model removes the requirement to maintain a dedicated routing topology per layer and therefore does not mandate a full mesh of routing adjacencies as is the case with overlaid control planes.

- The collaboration between technology layers where the control channel is associated with the data channel (e.g., packet/framed data planes) and technology layers where the control channel is not directly associated with the data channel (SONET/SDH, G.709, etc.)

is facilitated by the capability within GMPLS to associate in-band control plane signaling to the IP terminating interfaces of the control plane.

- Resource management and policies to be applied at the edges of such an MRN/MLN are made more simple (fewer control-to-management interactions) and more scalable (through the use of aggregated information).

- Multi-region/multi-layer traffic engineering is facilitated as TE links from distinct regions/layers are stored within the same TE Database.

4. Key Concepts of GMPLS-Based MLNs and MRNs

A network comprising transport nodes with multiple data plane layers of either the same ISC or different ISCs, controlled by a single GMPLS control plane instance, is called a multi-layer network (MLN). A subset of MLNs consists of networks supporting LSPs of different switching technologies (ISCs). A network supporting more than one switching technology is called a multi-region network (MRN).

4.1. Interface Switching Capability

The Interface Switching Capability (ISC) is introduced in GMPLS to support various kinds of switching technology in a unified way [RFC4202]. An ISC is identified via a switching type.

A switching type (also referred to as the switching capability type) describes the ability of a node to forward data of a particular data plane technology, and uniquely identifies a network region. The following ISC types (and, hence, regions) are defined: PSC, L2SC, TDM capable, LSC, and FSC. Each end of a data link (more precisely, each interface connecting a data link to a node) in a GMPLS network is associated with an ISC.

The ISC value is advertised as a part of the Interface Switching Capability Descriptor (ISCD) attribute (sub-TLV) of a TE link end associated with a particular link interface [RFC4202]. Apart from the ISC, the ISCD contains information including the encoding type, the bandwidth granularity, and the unreserved bandwidth on each of eight priorities at which LSPs can be established. The ISCD does not "identify" network layers, it uniquely characterizes information associated to one or more network layers.

TE link end advertisements may contain multiple ISCDs. This can be interpreted as advertising a multi-layer (or multi-switching-capable) TE link end. That is, the TE link end (and therefore the TE link) is present in multiple layers.

4.2. Multiple Interface Switching Capabilities

In an MLN, network elements may be single-switching-type-capable or multi-switching-type-capable nodes. Single-switching-type-capable nodes advertise the same ISC value as part of their ISCD sub-TLV(s) to describe the termination capabilities of each of their TE link(s). This case is described in [RFC4202].

Multi-switching-type-capable LSRs are classified as "simplex" or "hybrid" nodes. Simplex and hybrid nodes are categorized according to the way they advertise these multiple ISCs:

- A simplex node can terminate data links with different switching capabilities where each data link is connected to the node by a separate link interface. So, it advertises several TE links each with a single ISC value carried in its ISCD sub-TLV (following the rules defined in [RFC4206]). An example is an LSR with PSC and TDM links each of which is connected to the LSR via a separate interface.

- A hybrid node can terminate data links with different switching capabilities where the data links are connected to the node by the same interface. So, it advertises a single TE link containing more than one ISCD each with a different ISC value. For example, a node may terminate PSC and TDM data links and interconnect those external data links via internal links. The external interfaces connected to the node have both PSC and TDM capabilities.

Additionally, TE link advertisements issued by a simplex or a hybrid node may need to provide information about the node’s internal adjustment capabilities between the switching technologies supported. The term "adjustment" refers to the property of a hybrid node to interconnect the different switching capabilities that it provides through its external interfaces. The information about the adjustment capabilities of the nodes in the network allows the path computation process to select an end-to-end multi-layer or multi-region path that includes links with different switching capabilities joined by LSRs that can adapt (i.e., adjust) the signal between the links.

4.2.1. Networks with Multi-Switching-Type-Capable Hybrid Nodes

This type of network contains at least one hybrid node, zero or more simplex nodes, and a set of single-switching-type-capable nodes.

Figure 1 shows an example hybrid node. The hybrid node has two switching elements (matrices), which support, for instance, TDM and PSC switching, respectively. The node terminates a PSC and a TDM link (Link1 and Link2, respectively). It also has an internal link connecting the two switching elements.

The two switching elements are internally interconnected in such a way that it is possible to terminate some of the resources of, say, Link2 and provide adjustment for PSC traffic received/sent over the PSC interface (#b). This situation is modeled in GMPLS by connecting the local end of Link2 to the TDM switching element via an additional interface realizing the termination/adjustment function. There are two possible ways to set up PSC LSPs through the hybrid node. Available resource advertisement (i.e., Unreserved and Min/Max LSP Bandwidth) should cover both of these methods.

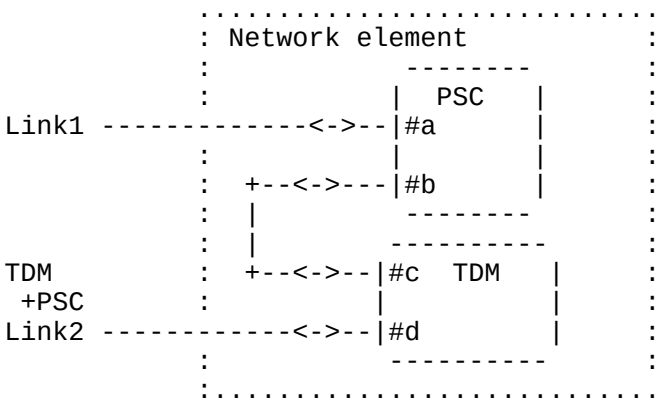

Figure 1. Hybrid node.

4.3. Integrated Traffic Engineering (TE) and Resource Control

In GMPLS-based multi-region/multi-layer networks, TE links may be consolidated into a single Traffic Engineering Database (TED) for use by the single control plane instance. Since this TED contains the information relative to all the layers of all regions in the network, a path across multiple layers (possibly crossing multiple regions) can be computed using the information in this TED. Thus, optimization of network resources across the multiple layers of the same region and across multiple regions can be achieved.

These concepts allow for the operation of one network layer over the topology (that is, TE links) provided by other network layers (for example, the use of a lower-layer LSC LSP carrying PSC LSPs). In turn, a greater degree of control and interworking can be achieved, including (but not limited to):

- Dynamic establishment of Forwarding Adjacency (FA) LSPs [RFC4206] (see Sections 4.3.2 and 4.3.3).

- Provisioning of end-to-end LSPs with dynamic triggering of FA LSPs.

Note that in a multi-layer/multi-region network that includes multi-switching-type-capable nodes, an explicit route used to establish an end-to-end LSP can specify nodes that belong to different layers or regions. In this case, a mechanism to control the dynamic creation of FA-LSPs may be required (see Sections 4.3.2 and 4.3.3).

There is a full spectrum of options to control how FA-LSPs are dynamically established. The process can be subject to the control of a policy, which may be set by a management component and which may require that the management plane is consulted at the time that the FA-LSP is established. Alternatively, the FA-LSP can be established at the request of the control plane without any management control.

4.3.1. Triggered Signaling

When an LSP crosses the boundary from an upper to a lower layer, it may be nested into a lower-layer FA-LSP that crosses the lower layer. From a signaling perspective, there are two alternatives to establish the lower-layer FA-LSP: static (pre-provisioned) and dynamic (triggered). A pre-provisioned FA-LSP may be initiated either by the operator or automatically using features like TE auto-mesh [RFC4972]. If such a lower-layer LSP does not already exist, the LSP may be established dynamically. Such a mechanism is referred to as "triggered signaling".

4.3.2. FA-LSPs

Once an LSP is created across a layer from one layer border node to another, it can be used as a data link in an upper layer.

Furthermore, it can be advertised as a TE link, allowing other nodes to consider the LSP as a TE link for their path computation [RFC4206]. An LSP created either statically or dynamically by one instance of the control plane and advertised as a TE link into the same instance of the control plane is called a Forwarding Adjacency LSP (FA-LSP). The FA-LSP is advertised as a TE link, and that TE link is called a Forwarding Adjacency (FA). An FA has the special

characteristic of not requiring a routing adjacency (peering) between its end points yet still guaranteeing control plane connectivity between the FA-LSP end points based on a signaling adjacency. An FA is a useful and powerful tool for improving the scalability of GMPLS-TE capable networks since multiple higher-layer LSPs may be nested (aggregated) over a single FA-LSP.

The aggregation of LSPs enables the creation of a vertical (nested) LSP hierarchy. A set of FA-LSPs across or within a lower layer can be used during path selection by a higher-layer LSP. Likewise, the higher-layer LSPs may be carried over dynamic data links realized via LSPs (just as they are carried over any "regular" static data links). This process requires the nesting of LSPs through a hierarchical process [RFC4206]. The TED contains a set of LSP advertisements from different layers that are identified by the ISCD contained within the TE link advertisement associated with the LSP [RFC4202].

If a lower-layer LSP is not advertised as an FA, it can still be used to carry higher-layer LSPs across the lower layer. For example, if the LSP is set up using triggered signaling, it will be used to carry the higher-layer LSP that caused the trigger. Further, the lower layer remains available for use by other higher-layer LSPs arriving at the boundary.

Under some circumstances, it may be useful to control the advertisement of LSPs as FAs during the signaling establishment of the LSPs [DYN-HIER].

4.3.3. Virtual Network Topology (VNT)

A set of one or more lower-layer LSPs provides information for efficient path handling in upper layer(s) of the MLN, or, in other words, provides a virtual network topology (VNT) to the upper layers. For instance, a set of LSPs, each of which is supported by an LSC LSP, provides a VNT to the layers of a PSC region, assuming that the PSC region is connected to the LSC region. Note that a single lower-layer LSP is a special case of the VNT. The VNT is configured by setting up or tearing down the lower-layer LSPs. By using GMPLS signaling and routing protocols, the VNT can be adapted to traffic demands.

A lower-layer LSP appears as a TE link in the VNT. Whether the diversely-routed lower-layer LSPs are used or not, the routes of lower-layer LSPs are hidden from the upper layer in the VNT. Thus, the VNT simplifies the upper-layer routing and traffic engineering decisions by hiding the routes taken by the lower-layer LSPs. However, hiding the routes of the lower-layer LSPs may lose important information that is needed to make the higher-layer LSPs reliable.

For instance, the routing and traffic engineering in the IP/MPLS layer does not usually consider how the IP/MPLS TE links are formed from optical paths that are routed in the fiber layer. Two optical paths may share the same fiber link in the lower-layer and therefore they may both fail if the fiber link is cut. Thus the shared risk properties of the TE links in the VNT must be made available to the higher layer during path computation. Further, the topology of the VNT should be designed so that any single fiber cut does not bisect the VNT. These issues are addressed later in this document.

Reconfiguration of the VNT may be triggered by traffic demand changes, topology configuration changes, signaling requests from the upper layer, and network failures. For instance, by reconfiguring the VNT according to the traffic demand between source and destination node pairs, network performance factors, such as maximum link utilization and residual capacity of the network, can be optimized. Reconfiguration is performed by computing the new VNT from the traffic demand matrix and optionally from the current VNT. Exact details are outside the scope of this document. However, this method may be tailored according to the service provider's policy regarding network performance and quality of service (delay, loss/disruption, utilization, residual capacity, reliability).

5. Requirements

5.1. Handling Single-Switching and Multi-Switching-Type-Capable Nodes

The MRN/MLN can consist of single-switching-type-capable and multi-switching-type-capable nodes. The path computation mechanism in the MLN should be able to compute paths consisting of any combination of such nodes.

Both single-switching-type-capable and multi-switching-type-capable (simplex or hybrid) nodes could play the role of layer boundary. MRN/MLN path computation should handle TE topologies built of any combination of nodes.

5.2. Advertisement of the Available Adjustment Resources

A hybrid node should maintain resources on its internal links (the links required for vertical integration between layers). Likewise, path computation elements should be prepared to use information about the availability of termination and adjustment resources as a constraint in MRN/MLN path computations. This would reduce the probability that the setup of the higher-layer LSP will be blocked by the lack of necessary termination/adjustment resources in the lower layers.

The advertisement of a node's MRN adjustment capabilities (the ability to terminate LSPs of lower regions and forward the traffic in upper regions) is REQUIRED, as it provides critical information when performing multi-region path computation.

The path computation mechanism should cover the case where the upper-layer links that are directly connected to upper-layer switching elements and the ones that are connected through internal links between upper-layer element and lower-layer element coexist (see Section 4.2.1).

5.3. Scalability

The MRN/MLN relies on unified routing and traffic engineering models.

- Unified routing model: By maintaining a single routing protocol instance and a single TE database per LSR, a unified control plane model removes the requirement to maintain a dedicated routing topology per layer, and therefore does not mandate a full mesh of routing adjacencies per layer.

- Unified TE model: The TED in each LSR is populated with TE links from all layers of all regions (TE link interfaces on multiple-switching-type-capable LSRs can be advertised with multiple ISCDs). This may lead to an increase in the amount of information that has to be flooded and stored within the network.

Furthermore, path computation times, which may be of great importance during restoration, will depend on the size of the TED.

Thus, MRN/MLN routing mechanisms MUST be designed to scale well with an increase of any of the following:

- Number of nodes
- Number of TE links (including FA-LSPs)
- Number of LSPs
- Number of regions and layers
- Number of ISCDs per TE link.

Further, design of the routing protocols MUST NOT prevent TE information filtering based on ISCDs. The path computation mechanism and the signaling protocol SHOULD be able to operate on partial TE information.

Since TE links can advertise multiple Interface Switching Capabilities (ISCs), the number of links can be limited (by combination) by using specific topological maps referred to as VNTs

(Virtual Network Topologies). The introduction of virtual topological maps leads us to consider the concept of emulation of data plane overlays.

5.4. Stability

Path computation is dependent on the network topology and associated link state. The path computation stability of an upper layer may be impaired if the VNT changes frequently and/or if the status and TE parameters (the TE metric, for instance) of links in the VNT changes frequently. In this context, robustness of the VNT is defined as the capability to smooth changes that may occur and avoid their propagation into higher layers. Changes to the VNT may be caused by the creation, deletion, or modification of LSPs.

Protocol mechanisms MUST be provided to enable creation, deletion, and modification of LSPs triggered through operational actions. Protocol mechanisms SHOULD be provided to enable similar functions triggered by adjacent layers. Protocol mechanisms MAY be provided to enable similar functions to adapt to the environment changes such as traffic demand changes, topology changes, and network failures. Routing robustness should be traded with adaptability of those changes.

5.5. Disruption Minimization

When reconfiguring the VNT according to a change in traffic demand, the upper-layer LSP might be disrupted. Such disruption to the upper layers must be minimized.

When residual resource decreases to a certain level, some lower-layer LSPs may be released according to local or network policies. There is a trade-off between minimizing the amount of resource reserved in the lower layer and disrupting higher-layer traffic (i.e., moving the traffic to other TE-LSPs so that some LSPs can be released). Such traffic disruption may be allowed, but MUST be under the control of policy that can be configured by the operator. Any repositioning of traffic MUST be as non-disruptive as possible (for example, using make-before-break).

5.6. LSP Attribute Inheritance

TE link parameters should be inherited from the parameters of the LSP that provides the TE link, and so from the TE links in the lower layer that are traversed by the LSP.

These include:

- Interface Switching Capability
- TE metric
- Maximum LSP bandwidth per priority level
- Unreserved bandwidth for all priority levels
- Maximum reservable bandwidth
- Protection attribute
- Minimum LSP bandwidth (depending on the switching capability)
- SRLG

Inheritance rules must be applied based on specific policies. Particular attention should be given to the inheritance of the TE metric (which may be other than a strict sum of the metrics of the component TE links at the lower layer), protection attributes, and SRLG.

As described earlier, hiding the routes of the lower-layer LSPs may lose important information necessary to make LSPs in the higher-layer network reliable. SRLGs may be used to identify which lower-layer LSPs share the same failure risk so that the potential risk of the VNT becoming disjoint can be minimized, and so that resource-disjoint protection paths can be set up in the higher layer. How to inherit the SRLG information from the lower layer to the upper layer needs more discussion and is out of scope of this document.

5.7. Computing Paths with and without Nested Signaling

Path computation can take into account LSP region and layer boundaries when computing a path for an LSP. Path computation may restrict the path taken by an LSP to only the links whose interface switching capability is PSC. For example, suppose that a TDM-LSP is routed over the topology composed of TE links of the same TDM layer. In calculating the path for the LSP, the TED may be filtered to include only links where both end include requested LSP switching type. In this way hierarchical routing is done by using a TED filtered with respect to switching capability (that is, with respect to particular layer).

If triggered signaling is allowed, the path computation mechanism may produce a route containing multiple layers/regions. The path is computed over the multiple layers/regions even if the path is not "connected" in the same layer as where the endpoints of the path exist. Note that here we assume that triggered signaling will be invoked to make the path "connected", when the upper-layer signaling request arrives at the boundary node.

The upper-layer signaling request MAY contain an ERO (Explicit Route Object) that includes only hops in the upper layer; in which case, the boundary node is responsible for triggered creation of the lower-layer FA-LSP using a path of its choice, or for the selection of any available lower-layer LSP as a data link for the higher layer. This mechanism is appropriate for environments where the TED is filtered in the higher layer, where separate routing instances are used per layer, or where administrative policies prevent the higher layer from specifying paths through the lower layer.

Obviously, if the lower-layer LSP has been advertised as a TE link (virtual or real) into the higher layer, then the higher-layer signaling request MAY contain the TE link identifier and so indicate the lower-layer resources to be used. But in this case, the path of the lower-layer LSP can be dynamically changed by the lower layer at any time.

Alternatively, the upper-layer signaling request MAY contain an ERO specifying the lower-layer FA-LSP route. In this case, the boundary node MAY decide whether it should use the path contained in the strict ERO or re-compute the path within the lower layer.

Even in the case that the lower-layer FA-LSPs are already established, a signaling request may also be encoded as a loose ERO. In this situation, it is up to the boundary node to decide whether it should create a new lower-layer FA-LSP or it should use an existing lower-layer FA-LSP.

The lower-layer FA-LSP can be advertised just as an FA-LSP in the upper layer or an IGP adjacency can be brought up on the lower-layer FA-LSP.

5.8. LSP Resource Utilization

Resource usage in all layers should be optimized as a whole (i.e., across all layers), in a coordinated manner (i.e., taking all layers into account). The number of lower-layer LSPs carrying upper-layer LSPs should be minimized (note that multiple LSPs may be used for load balancing). Lower-layer LSPs that could have their traffic re-routed onto other LSPs are unnecessary and should be avoided.

5.8.1. FA-LSP Release and Setup

If there is low traffic demand, some FA-LSPs that do not carry any higher-layer LSP may be released so that lower-layer resources are released and can be assigned to other uses. Note that if a small fraction of the available bandwidth of an FA-LSP is still in use, the nested LSPs can also be re-routed to other FA-LSPs (optionally using

the make-before-break technique) to completely free up the FA-LSP. Alternatively, unused FA-LSPs may be retained for future use. Release or retention of underutilized FA-LSPs is a policy decision.

As part of the re-optimization process, the solution MUST allow rerouting of an FA-LSP while keeping interface identifiers of corresponding TE links unchanged. Further, this process MUST be possible while the FA-LSP is carrying traffic (higher-layer LSPs) with minimal disruption to the traffic.

Additional FA-LSPs may also be created based on policy, which might consider residual resources and the change of traffic demand across the region. By creating the new FA-LSPs, the network performance such as maximum residual capacity may increase.

As the number of FA-LSPs grows, the residual resources may decrease. In this case, re-optimization of FA-LSPs may be invoked according to policy.

Any solution MUST include measures to protect against network destabilization caused by the rapid setup and teardown of LSPs as traffic demand varies near a threshold.

Signaling of lower-layer LSPs SHOULD include a mechanism to rapidly advertise the LSP as a TE link and to coordinate into which routing instances the TE link should be advertised.

5.8.2. Virtual TE Links

It may be considered disadvantageous to fully instantiate (i.e., pre-provision) the set of lower-layer LSPs that provide the VNT since this might reserve bandwidth that could be used for other LSPs in the absence of upper-layer traffic.

However, in order to allow path computation of upper-layer LSPs across the lower layer, the lower-layer LSPs may be advertised into the upper layer as though they had been fully established, but without actually establishing them. Such TE links that represent the possibility of an underlying LSP are termed "virtual TE links". It is an implementation choice at a layer boundary node whether to create real or virtual TE links, and the choice (if available in an implementation) MUST be under the control of operator policy. Note that there is no requirement to support the creation of virtual TE links, since real TE links (with established LSPs) may be used. Even if there are no TE links (virtual or real) advertised to the higher layer, it is possible to route a higher-layer LSP into a lower layer on the assumption that proper hierarchical LSPs in the lower layer will be dynamically created (triggered) as needed.

If an upper-layer LSP that makes use of a virtual TE link is set up, the underlying LSP MUST be immediately signaled in the lower layer.

If virtual TE links are used in place of pre-established LSPs, the TE links across the upper layer can remain stable using pre-computed paths while wastage of bandwidth within the lower layer and unnecessary reservation of adaptation resources at the border nodes can be avoided.

The solution SHOULD provide operations to facilitate the build-up of such virtual TE links, taking into account the (forecast) traffic demand and available resources in the lower layer.

Virtual TE links can be added, removed, or modified dynamically (by changing their capacity) according to the change of the (forecast) traffic demand and the available resources in the lower layer. It MUST be possible to add, remove, and modify virtual TE links in a dynamic way.

Any solution MUST include measures to protect against network destabilization caused by the rapid changes in the VNT as traffic demand varies near a threshold.

The concept of the VNT can be extended to allow the virtual TE links to form part of the VNT. The combination of the fully provisioned TE links and the virtual TE links defines the VNT provided by the lower layer. The VNT can be changed by setting up and/or tearing down virtual TE links as well as by modifying real links (i.e., the fully provisioned LSPs). How to design the VNT and how to manage it are out of scope of this document.

In some situations, selective advertisement of the preferred connectivity among a set of border nodes between layers may be appropriate. Further decreasing the number of advertisements of the virtual connectivity can be achieved by abstracting the topology (between border nodes) using models similar to those detailed in [RFC4847].

5.9. Verification of the LSPs

When a lower-layer LSP is established for use as a data link by a higher layer, the LSP may be verified for correct connectivity and data integrity before it is made available for use. Such mechanisms are data-technology-specific and are beyond the scope of this document, but the GMPLS protocols SHOULD provide mechanisms for the coordination of data link verification.

5.10. Management

An MRN/MLN requires management capabilities. Operators need to have the same level of control and management for switches and links in the network that they would have in a single-layer or single-region network.

We can consider two different operational models: (1) per-layer management entities and (2) cross-layer management entities.

Regarding per-layer management entities, it is possible for the MLN to be managed entirely as separate layers, although that somewhat defeats the objective of defining a single multi-layer network. In this case, separate management systems would be operated for each layer, and those systems would be unaware of the fact that the layers were closely coupled in the control plane. In such a deployment, as LSPs were automatically set up as the result of control plane requests from other layers (for example, triggered signaling), the management applications would need to register the creation of the new LSPs and the depletion of network resources. Emphasis would be placed on the layer boundary nodes to report the activity to the management applications.

A more likely scenario is to apply a closer coupling of layer management systems with cross-layer management entities. This might be achieved through a unified management system capable of operating multiple layers, or by a meta-management system that coordinates the operation of separate management systems each responsible for individual layers. The former case might only be possible with the development of new management systems, while the latter is feasible through the coordination of existing network management tools.

Note that when a layer boundary also forms an administrative boundary, it is highly unlikely that there will be unified multi-layer management. In this case, the layers will be separately managed by the separate administrative entities, but there may be some "leakage" of information between the administrations in order to facilitate the operation of the MLN. For example, the management system in the lower-layer network might automatically issue reports on resource availability (coincident with TE routing information) and alarm events.

This discussion comes close to an examination of how a VNT might be managed and operated. As noted in Section 5.8, issues of how to design and manage a VNT are out of scope for this document, but it should be understood that the VNT is a client-layer construct built from server-layer resources. This means that the operation of a VNT

is a collaborative activity between layers. This activity is possible even if the layers are from separate administrations, but in this case the activity may also have commercial implications.

MIB modules exist for the modeling and management of GMPLS networks [RFC4802] [RFC4803]. Some deployments of GMPLS networks may choose to use MIB modules to operate individual network layers. In these cases, operators may desire to coordinate layers through a further MIB module that could be developed. Multi-layer protocol solutions (that is, solutions where a single control plane instance operates in more than one layer) SHOULD be manageable through MIB modules. A further MIB module to coordinate multiple network layers with this control plane MIB module may be produced.

Operations and Management (OAM) tools are important to the successful deployment of all networks.

OAM requirements for GMPLS networks are described in [GMPLS-OAM]. That document points out that protocol solutions for individual network layers should include mechanisms for OAM or make use of OAM features inherent in the physical media of the layers. Further discussion of individual-layer OAM is out of scope of this document.

When operating OAM in a MLN, consideration must be given to how to provide OAM for end-to-end LSPs that cross layer boundaries (that may also be administrative boundaries) and how to coordinate errors and alarms detected in a server layer that need to be reported to the client layer. These operational choices MUST be left open to the service provider and so MLN protocol solutions MUST include the following features:

- Within the context and technology capabilities of the highest technology layer of an LSP (i.e., the technology layer of the first hop), it MUST be possible to enable end-to-end OAM on a MLN LSP. This function appears to the ingress LSP as normal LSP-based OAM [GMPLS-OAM], but at layer boundaries, depending on the technique used to span the lower layers, client-layer OAM operations may need to mapped to server-layer OAM operations. Most such requirements are highly dependent on the OAM facilities of the data plane technologies of client and server layers. However, control plane mechanisms used in the client layer per [GMPLS-OAM] MUST map and enable OAM in the server layer.

- OAM operation enabled per [GMPLS-OAM] in a client layer for an LSP MUST operate for that LSP along its entire length. This means that if an LSP crosses a domain of a lower-layer technology, the client-layer OAM operation must operate seamlessly within the client layer at both ends of the client-layer LSP.

- OAM functions operating within a server layer MUST be controllable from the client layer such that the server-layer LSP(s) that support a client-layer LSP have OAM enabled at the request of the client layer. Such control SHOULD be subject to policy at the layer boundary, just as automatic provisioning and LSP requests to the server layer are subject to policy.

- The status including errors and alarms applicable to a server-layer LSP MUST be available to the client layer. This information SHOULD be configurable to be automatically notified to the client layer at the layer boundary and SHOULD be subject to policy so that the server layer may filter or hide information supplied to the client layer. Furthermore, the client layer SHOULD be able to select to not receive any or all such information.

Note that the interface between layers lies within network nodes and is, therefore, not necessarily the subject of a protocol specification. Implementations MAY use standardized techniques (such as MIB modules) to convey status information (such as errors and alarms) between layers, but that is out of scope for this document.

6. Security Considerations

The MLN/MRN architecture does not introduce any new security requirements over the general GMPLS architecture described in [RFC3945]. Additional security considerations form MPLS and GMPLS networks are described in [MPLS-SEC].

However, where the separate layers of an MLN/MRN network are operated as different administrative domains, additional security considerations may be given to the mechanisms for allowing LSP setup crossing one or more layer boundaries, for triggering lower-layer LSPs, or for VNT management. Similarly, consideration may be given to the amount of information shared between administrative domains, and the trade-off between multi-layer TE and confidentiality of information belonging to each administrative domain.

It is expected that solution documents will include a full analysis of the security issues that any protocol extensions introduce.

7. Acknowledgements

The authors would like to thank Adrian Farrel and the participants of ITU-T Study Group 15, Question 14 for their careful review. The authors would like to thank the IESG review team for rigorous review: special thanks to Tim Polk, Miguel Garcia, Jari Arkko, Dan Romascanu, and Dave Ward.

8. References

8.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC3945] Mannie, E., Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Architecture", RFC 3945, October 2004.

[RFC4202] Kompella, K., Ed., and Y. Rekhter, Ed., "Routing Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4202, October 2005.

[RFC4206] Kompella, K. and Y. Rekhter, "Label Switched Paths (LSP) Hierarchy with Generalized Multi-Protocol Label Switching (GMPLS) Traffic Engineering (TE)", RFC 4206, October 2005.

[RFC4397] Bryskin, I. and A. Farrel, "A Lexicography for the Interpretation of Generalized Multiprotocol Label Switching (GMPLS) Terminology within the Context of the ITU-T's Automatically Switched Optical Network (ASON) Architecture", RFC 4397, February 2006.

[RFC4726] Farrel, A., Vasseur, J.-P., and A. Ayyangar, "A Framework for Inter-Domain Multiprotocol Label Switching Traffic Engineering", RFC 4726, November 2006.

8.2. Informative References

[DYN-HIER] Shiomoto, K., Rabbat, R., Ayyangar, A., Farrel, A. and Z. Ali, "Procedures for Dynamically Signaled Hierarchical Label Switched Paths", Work in Progress, February 2008.

[MRN-EVAL] Le Roux, J.L., Ed., and D. Papadimitriou, Ed., "Evaluation of existing GMPLS Protocols against Multi Layer and Multi Region Networks (MLN/MRN)", Work in Progress, December 2007.

[RFC5146] Kumaki, K., Ed., "Interworking Requirements to Support Operation of MPLS-TE over GMPLS Networks", RFC 5146, March 2008.

[MPLS-SEC] Fang, L., Ed., "Security Framework for MPLS and GMPLS Networks", Work in Progress, February 2008.

[RFC4802] Nadeau, T., Ed., and A. Farrel, Ed., "Generalized Multiprotocol Label Switching (GMPLS) Traffic Engineering Management Information Base", RFC 4802, February 2007.

[RFC4803] Nadeau, T., Ed., and A. Farrel, Ed., "Generalized Multiprotocol Label Switching (GMPLS) Label Switching Router (LSR) Management Information Base", RFC 4803, February 2007.

[RFC4847] Takeda, T., Ed., "Framework and Requirements for Layer 1 Virtual Private Networks", RFC 4847, April 2007.

[RFC4972] Vasseur, JP., Ed., Leroux, JL., Ed., Yasukawa, S., Previdi, S., Psenak, P., and P. Mabbey, "Routing Extensions for Discovery of Multiprotocol (MPLS) Label Switch Router (LSR) Traffic Engineering (TE) Mesh Membership", RFC 4972, July 2007.

[GMPLS-OAM] Nadeau, T., Otani, T. Brungard, D., and A. Farrel, "OAM Requirements for Generalized Multi-Protocol Label Switching (GMPLS) Networks", Work in Progress, October 2007.

9. Contributors' Addresses

Eiji Oki
NTT Network Service Systems Laboratories
3-9-11 Midori-cho, Musashino-shi
Tokyo 180-8585
Japan
Phone: +81 422 59 3441
EMail: oki.eiji@lab.ntt.co.jp

Ichiro Inoue
NTT Network Service Systems Laboratories
3-9-11 Midori-cho, Musashino-shi
Tokyo 180-8585
Japan
Phone: +81 422 59 3441
EMail: ichiro.inoue@lab.ntt.co.jp

Emmanuel Dotaro
Alcatel-Lucent
Route de Villejust
91620 Nozay
France
Phone: +33 1 3077 2670
EMail: emmanuel.dotaro@alcatel-lucent.fr

Authors' Addresses

Kohei Shiomoto
NTT Network Service Systems Laboratories
3-9-11 Midori-cho, Musashino-shi
Tokyo 180-8585
Japan
EMail: shiomoto.kohei@lab.ntt.co.jp

Dimitri Papadimitriou
Alcatel-Lucent
Copernicuslaan 50
B-2018 Antwerpen
Belgium
Phone : +32 3 240 8491
EMail: dimitri.papadimitriou@alcatel-lucent.be

Jean-Louis Le Roux
France Telecom R&D
Av Pierre Marzin
22300 Lannion
France
EMail: jeanlouis.leroux@orange-ftgroup.com

Martin Vigoureux
Alcatel-Lucent
Route de Villejust
91620 Nozay
France
Phone: +33 1 3077 2669
EMail: martin.vigoureux@alcatel-lucent.fr

Deborah Brungard
AT&T
Rm. D1-3C22 - 200
S. Laurel Ave.
Middletown, NJ 07748
USA
Phone: +1 732 420 1573
EMail: dbrungard@att.com

Full Copyright Statement

Intellectual Property