Network Working Group Kohei Shiomoto (NTT)
Internet-Draft Dimitri Papadimitriou (Alcatel-Lucent)
Intended Status: Informational Jean-Louis Le Roux (France Telecom)
Martin Vigoureux (Alcatel-Lucent)
Deborah Brungard (AT&T)

Expires: February 2008 August 2007

# Requirements for GMPLS-based multi-region and multi-layer networks (MRN/MLN)

draft-ietf-ccamp-gmpls-mln-reqs-04.txt

## 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 a 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 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 (Time Division Multiplex 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  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 technologies (e.g. PSC + TDM) hosted on the same device (referred to as multi-switching-type-capable LSRs, see below) 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 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
  that 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 controlling the MLN/MRN 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 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. Further, 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 resources availability (e.g., link bandwidth, or adaptation ports between regions) across the regions. The same considerations hold when VC4 LSPs are provisioned to provide extra flexibility for the VC12 and/or VC11 layers in an MLN.

## 1.1 Scope

This document describes the requirements to support multi-region/multi-layer networks. 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.

For such TE domain to interoperate with edge nodes/domains supporting

interfaces
by other SDOs e.g. ITU-T and OIF, 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].

### 2.1.List of acronyms

MLN: Multi-Layer Network
MRN: Multi-Region Network

ISC: Interface Switching Capability
ISCD: Interface Switching Capability Descriptor
PSC: Packet Switching Capable
L2SC: Layer-2 Switching Capable
TDM: Time-Division Switch Capable
LSC: Lambda Switching Capable
FSC: Fiber Switching Capable
SRLG: Shared Risk Ling Group
VNT: Virtual Network Topology
FA: Forwarding Adjacency
FA-LSP: Forwarding Adjacency Label Switched Path
TE: Traffic Engineering
TED: Traffic Engineering Database
LSP: Label Switched Path
LSR: Label Switching Router

## 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 LSR hosting different switching technologies
because change 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 the issue by using triggered signaling (See 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 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 [MPLS-GMPLS]. 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 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 lambda switching capable areas relate to horizontal integration. On the other hand, the collaborative mechanisms in place between a packet switching capable (e.g. IP/MPLS) domain over a different time division switching capable (eg VC4 SDH) domain is part of the horizontal integration while it can be seen as a first step towards vertical integration.

## 3.4.Motivation

The applicability of GMPLS to multiple switching technologies provides the unified control management approach for both LSP provisioning and recovery.

Indeed, one of the main motivations for unifying the capabilities and operations GMPLS control plane is the desire to support multi LSP-region [RFC4206] routing and Traffic Engineering (TE) capability. For instance, this enables effective network resource utilization of both the Packet/Layer2 LSP regions and the Time Division Multiplexing (TDM) or Lambda LSP regions in high capacity networks.

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

- The maintenance of multiple instances of the control plane on devices hosting more than one switching capability not only increases the complexity of their interactions but also increases the total amount of processing individual instances would handle.
- The unification of the addressing spaces helps in avoiding multiple identification 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 and the data plane.
- By maintaining a single routing protocol instance and a single TE database per LSR, a unified control plane model prevents from maintaining a dedicated routing topology per layer and therefore does not mandate a full mesh of routing adjacencies as it is the case with overlaid control planes.

- The collaboration between associated control planes (packet/framed data planes) and non-associated control planes (SONET/SDH, G.709, etc.) is facilitated due to the capability of hooking the associated in-band signaling to the IP terminating interfaces of the control plane.
- Resource management and policies to be applied at the edges of such environment is facilitated (less 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 sub-set 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, 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. For example, 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 adaptation capabilities between the switching technologies supported. That is, the node's capability to perform layer border node functions.

### 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 adaptation 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/adaptation 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.

```
                            Network element
                       .............................
                       :            --------       :
                       :           |  PSC   |      :
          Link1 -------------<->--|#a      |      :
                       :  +--<->---|#b      |      :
                       :  |         --------       :
          TDM          :  |        ----------      :
           +PSC        :  +--<->--|#c  TDM   |     :
```

```
         Link2 ------------<->--|#d         |     :
                     :          ----------     :
                     :.............................
```

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 inter-working can be achieved, including (but not limited too):

- 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 [AUTO-MESH]. 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 Traffic Engineering (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 of 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 virtual network topology 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 virtual network topology is configured by setting up or tearing down the lower layer LSPs. By using GMPLS signaling and routing protocols, the virtual network topology 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 virtual network topology may be triggered by traffic demand changes, topology configuration changes, signaling requests from the upper layer, and network failures. For instance, by reconfiguring the virtual network topology 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 Adaptation Resource

A hybrid node SHOULD maintain resources on its internal links (the links required for vertical (layer) integration) and SHOULD advertise the resource information for those links. Likewise, path computation elements SHOULD be prepared to use the availability of termination/adaptation resources as a constraint in MRN/MLN path computations to reduce the higher layer LSP setup blocking probability caused by the lack of necessary termination/adaptation resources in the lower layer(s).

The advertisement of the adaptation capability to terminate LSPs of lower-region and forward traffic in the upper-region is REQUIRED, as it provides critical information when performing multi-region path computation.

The mechanism SHOULD cover the case where the upper-layer links which are directly connected to upper-layer switching element and the ones which 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 a unified traffic engineering and routing model.

- Unified routing model: by maintaining a single routing protocol instance and a single TE database per LSR, a unified control plane model prevents from maintaining 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 links interfaces on multiple-switching capability LSR can be advertised with multiple ISCD). This may lead to a large 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 (ISC), the number of links can be limited (by combination) by using specific topological maps referred to as VNT (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.

Creation, deletion, and modification of LSPs MAY be triggered by adjacent layers or through operational actions to meet traffic demand changes, topology changes, signaling requests from the upper layer, and network failures. Routing robustness SHOULD be traded with adaptability with respect to the change of incoming traffic requests.

## 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 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 MAY take into account LSP region and layer boundaries when computing a path for an LSP. For example, path computation MAY restrict the path taken by an LSP to only the links whose interface switching capability is PSC.

Interface switching capability is used as a constraint in path computation. For example, 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 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 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 is responsible for decision as to which it should use the path contained in the strict ERO or it should re-compute the path within in the lower-layer.

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

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

It MUST be possible to utilize network resources efficiently. Particularly, 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

Statistical multiplexing can only be employed in PSC and L2SC regions. A PSC or L2SC LSP may or may not consume the maximum reservable bandwidth of the TE link (FA LSP) that carries it. On the other hand, a TDM, or LSC LSP always consumes a fixed amount of bandwidth as long as it exists (and is fully instantiated) because statistical multiplexing is not available.

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 resource 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, and 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 assumptions 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 ports 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 resource in the lower-layer.

Virtual TE-links MAY be added, removed or modified dynamically (by changing their capacity) according to the change of the (forecast) traffic demand and the available resource in the lower-layer. The maximum number of virtual TE links that can be defined SHOULD be configurable.

Any solution MUST include measures to protect against network destabilization caused by the rapid changes in the virtual network topology 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.

### 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. Such mechanisms are data technology-specific and are beyond the scope of this document, but may be coordinated through the GMPLS control plane.

## 6. Security Considerations

The current version of this document does not introduce any new security considerations as it only lists a set of requirements.

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

## 7. IANA Considerations

This informational document makes no requests to IANA for action.

## 8. References

### 8.1. Normative Reference

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

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

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

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

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

[RFC4397] I.Bryskin 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.

## 8.2. Informative References

[MRN-EVAL] Le Roux, J.L., Brungard, D., Oki, E., Papadimitriou, D., Shiomoto, K., Vigoureux, M.,"Evaluation of existing GMPLS Protocols against Multi Layer and Multi Region Networks (MLN/MRN)", draft-ietf-ccamp-gmpls-mrn-eval, work in progress.

[MPLS-GMPLS] K. Kumaki (Editor), "Interworking Requirements to Support operation of MPLS-TE over GMPLS networks", draft-ietf-ccamp-mpls-gmpls-interwork-reqts, work in progress.

[DYN-HIER] Shiomoto, K., Rabbat, R., Ayyangar, A., Farrel, A. and Ali, Z., "Procedures for Dynamically Signaled Hierarchical Label Switched Paths", draft-ietf-ccamp-lsp-hierarchy-bis, work in progress.

[AUTO-MESH] Vasseur, JP., Le Roux, JL., et al., "Routing extensions for discovery of Multiprotocol (MPLS) Label Switch Router (LSR) Traffic Engineering (TE) mesh membership", draft-ietf-ccamp-automesh, work in progress.

## 9. 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-ft.com

Martin Vigoureux
Alcatel-Lucent
Route de Nozay, 91461 Marcoussis cedex, France
Phone: +33 (0)1 69 63 18 52
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

10.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 Nozay,
91461 Marcoussis cedex,
France
Phone : +33 1 6963 4723
Email: emmanuel.dotaro@alcatel-lucent.fr

## 11. Intellectual Property Considerations

general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at http://www.ietf.org/ipr.

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement this standard. Please address the information to the IETF at ietf-ipr@ietf.org.

## 12. Full Copyright Statement