

CCAMP Working Group
Internet-Draft
Intended status: Proposed Standard
Expires: September 15, 2011

Ashok Kunjidhapatham
Rajan Rao
Biao Lu
Snigdho Bardalai
Khuzema Pithewan
Infinera Corp
John E Drake
Juniper Networks
Steve Balls
Metswitch Networks
March 14, 2011

OSPF TE Extensions for Generalized MPLS (GMPLS) Control of
G.709 Optical Transport Networks
draft-ashok-ccamp-gmpls-ospf-g709-03.txt

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at
<http://www.ietf.org/lid-abstracts.html>

The list of Internet-Draft Shadow Directories can be accessed at
<http://www.ietf.org/shadow.html>

Copyright and License Notice

Copyright (c) 2011 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect

to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Abstract

As OTN network capabilities continue to evolve, there is an increased need to support GMPLS control for the same. [RFC4328] introduced GMPLS signaling extensions for supporting the early version of G.709 [G.709-v1]. The basic routing considerations from signaling perspective is also specified in [RFC4328].

The recent revision of ITU-T Recommendation G.709 [G.709-v3] and [GSUP.43] have introduced new ODU containers (both fixed and flexible) and additional ODU multiplexing capabilities, enabling support for optimal service aggregation.

This document describes OSPF protocol extensions to support Generalized MPLS (GMPLS) control for routing services over the standardized OTU/ODU containers in support of ODU based TDM switching. Routing support for Optical Channel Layer switching (Lambda switching) is not covered in this document.

Table of Contents

1. Introduction	4
2. Conventions used in this document	5
3. OTU/ODU Link Representation	5
3.1. OTUk TE-Link	5
3.2. ODUk TE-Link	6
3.3. ODUj TE-Link	6
3.4. Bundled TE-Link	7
3.5. OTU/ODU Link Property Agreement	7
4. OTU/ODU Link Bandwidth Model	8
5. OSPF TE-LSA Extension	9
5.1. Maximum Bandwidth	9
5.2. Maximum Reservable Bandwidth	9
5.3. Unreserved Bandwidth	9
5.4. Interface Switch Capability Descriptor	9
5.4.1 ODU Switching	11
5.4.2. ODUk Switch Capability Specific Information	11
5.4.2.1 Bandwidth sub TLV for fixed ODUj	12
5.4.2.2 Bandwidth sub-TLV for ODUflex	13
5.5. Interface Multiplexing Capability Descriptor	13
5.5.1 Multiplex Layers and Hierarchical LSP	14

5.5.2 IMCD format	15
5.5.2.1 G-PID	16
5.5.2.2 Available Bandwidth	17
5.5.3 Controlling IMCD advertisement	17
5.5.4 How to use IMCDs for FA creation	18
5.5.5 IMCD and non OTN services	18
6. Examples	19
6.1. Network with no IMCD advertisement (no FA support)	19
6.2. Network with IMCD advertisement for FA support	20
6.3. Link bundle with similar muxing capabilities	22
6.4. Link bundle with dissimilar muxing capabilities: Layer relation	23
9. IANA Considerations	24
10. References	25
10.1. Normative References	25
10.2. Informative References	25
11. Acknowledgements	26
Author's Addresses	26
Appendix A: Abbreviations & Terminology	28
A.1 Abbreviations:	28
A.2 Terminology	28
Appendix B : Optimization Techniques	30
B.1 Multiple ISCDs Vs. Single ISCD	30
B.1 Multiple IMCDs Vs. Single IMCD	30
B.1 Eight priorities Vs. restricted number of priorities	30
Appendix C: Relation with MLN & MRN	30
Appendix D : AMP, BMP & GMP Mapping	30

1. Introduction

Generalized Multi-Protocol Label Switching (GMPLS) [RFC3945] extends MPLS from supporting Packet Switching Capable (PSC) interfaces and switching to include support of four new classes of interfaces and switching: Layer-2 Switching (L2SC), Time-Division Multiplex (TDM), Lambda Switch (LSC), and Fiber-Switch (FSC) Capable. A functional description of the extensions to MPLS signaling that are needed to support these new classes of interfaces and switching is provided in [RFC3471]. OSPF extensions for supporting GMPLS are defined in [RFC4203].

ITU-T Recommendations G.709 and G.872 provide specifications for OTN interface and network architecture respectively. As OTN network capabilities continue to evolve; there is an increased need to support GMPLS control for the same.

GMPLS signaling extensions to support G.709 OTN interfaces are specified in [RFC4328]. The basic routing considerations from signaling perspective is specified. G.709 specifications evolved rapidly over the last few years. Following are the features added in OTN since the first version [G.709-v1].

- (a) OTU Containers:
 - Pre-existing Containers: OTU1, OTU2 and OTU3
 - New Containers introduced in [G.709-v3]: OTU2e and OTU4
 - New Containers introduced in [GSUP.43]: OTU1e, OTU3e1 and OTU3e2
- (b) Fixed ODU Containers:
 - Pre-existing Containers: ODU1, ODU2 and ODU3
 - New Containers introduced in [G.709-v3]: ODU0, ODU2e and ODU4
 - New Containers introduced in [GSUP.43]: ODU1e, ODU3e1 and ODU3e2
- (c) Flexible ODU Containers:
 - ODUflex for CBR and GFP-F mapped services. ODUflex uses 'n' number of OPU Tributary Slots where 'n' is different from the number of OPU Tributary Slots used by the Fixed ODU Containers.
- (d) Tributary Slot Granularity:
 - OPU2 and OPU3 support two Tributary Slot Granularities:
 - (i) 1.25Gbps and (ii) 2.5Gbps.
- (e) Multi-stage ODU Multiplexing:
 - Multi-stage multiplexing of LO-ODUs into HO-ODU is supported. Also, multiplexing could be heterogeneous (meaning LO-ODUs of different rates can be multiplexed into a HO-ODU).

OTN networks support switching at two layers: (i) ODU Layer - TDM

Switching and (ii) OCH Layer - Lambda (LSC) Switching. The nodes on the network may support one or both the switching types. When multiple switching types are supported MRN/MLN based routing [RFC5212] and [RFC6001] is assumed.

This document covers OSPF extensions to support routing over the standardized OTU/ODU containers in support of ODU Layer based TDM switching as outlined in the framework document [G.709-FRAME]. The Interface Switch Capability Descriptor extensions for ODU Layer switching and bandwidth representation for ODU containers are defined in this document.

Routing support for Optical Channel Layer switching (LSC) is beyond the scope of this document. Refer to [WSON-FRAME] for further details.

2. Conventions used in this document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document is to be interpreted as described in RFC-2119 [RFC2119].

In addition, the reader is assumed to be familiar with the terminology used in ITU-T [G.709-v3], [G.872] and [GSUP.43], as well as [RFC4201] and [RFC4203]. Abbreviations used in this document is detailed in Appendix A.

3. OTU/ODU Link Representation

G.709 OTU/ODU Links are represented as TE-Links in GMPLS Traffic Engineering Topology for supporting ODU layer switching. These TE-Links can be modeled in multiple ways. Some of the prominent representations are captured below.

3.1. OTUk TE-Link

OTUk Link can be modeled as a TE-Link. Switching at ODUk layer and ODUj layer (including multi-stage multiplexing) can be managed on OTUk TE-Link. Figure-1 below provides an illustration of this link type.

When a LO-ODU layer being switched on an OTUk interface involves multi-stage multiplexing, all the HO-ODU layer(s) should necessarily terminate between the same pair of nodes as the OTUk layer in this case. For example, if ODU1 layer switching is configured on a OTU3 link via multiplexing hierarchy

ODU3<-ODU2<-ODU1, HO-ODUs (namely ODU3 & ODU2) should terminate between the same pair of nodes as OTU3 layer.

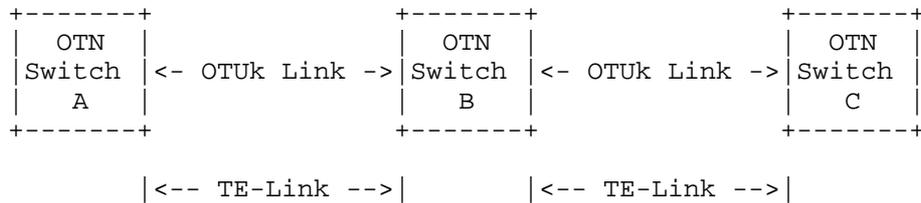


Figure-1: OTUk TE-Link

3.2. ODUk TE-Link

When ODUk layer does not terminate on the same pair of nodes as OTUk layer, ODUk link should be modeled as a TE-Link. As bandwidth is directly managed on the ODUk link, associated OTUk links are not significant in this case. Switching at ODUj layer (including multi-stage multiplexing) can be managed on ODUk TE-Link. Figure-2 below provides an illustration of this link type.

When a LO-ODU layer being switched on the ODUk interface involves multi-stage multiplexing, all the HO-ODU layer(s) should necessarily terminate between the same pair of nodes as ODUk in this case. For example, if ODU1 layer switching is configured on an ODU3 link via multiplexing hierarchy ODU3<-ODU2<-ODU1, HO-ODU (namely ODU2) should terminate between the same pair of nodes as ODU3.

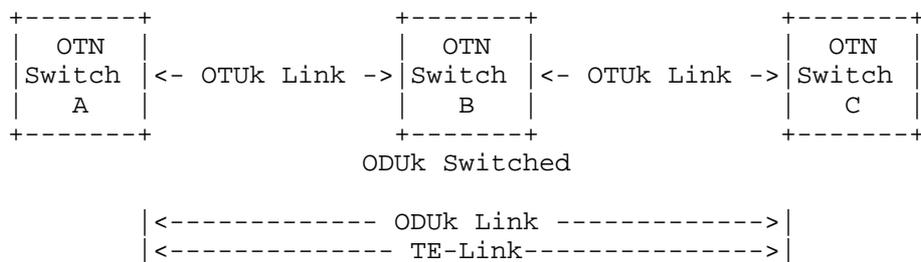


Figure-2: ODUk TE-Link

3.3. ODUj TE-Link

When a LO-ODUj within a HO-ODUk does not terminate on the same pair of nodes as HO-ODUk layer, Separate TE-Links needs to be modeled for ODUk link and ODUj link. Also, ODUk link shall no longer manage the bandwidth associated with the ODUj link. Switching at sub-ODUj layer (including multi-stage multiplexing)

can be supported on this ODUj TE-Link. Figure-3 below provides an illustration of this link type.

When a LO-ODU layer being switched on an ODUj interface involves multi-stage multiplexing, all the HO-ODU layer(s) should necessarily terminate between the same pair of nodes as ODUj in this case. For example, if ODU0 layer switching is configured on an ODU2 link via multiplexing hierarchy ODU2<-ODU1<-ODU0, HO-ODU (namely ODU1) should terminate between the same pair of nodes as ODU2.

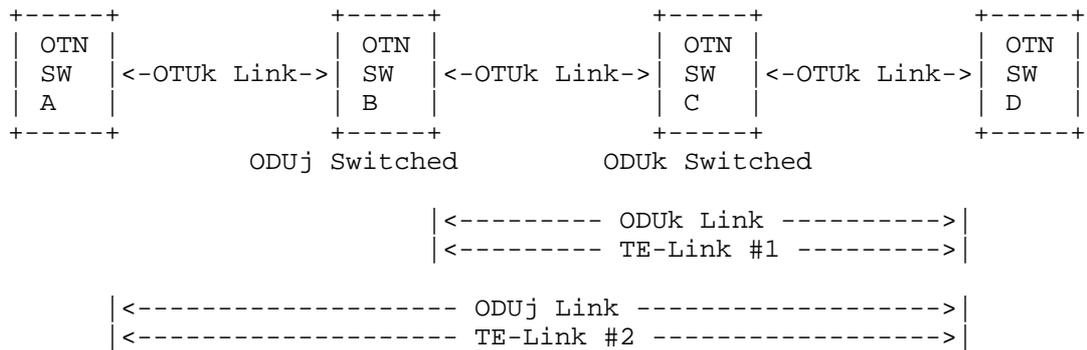


Figure-3: ODUj TE-Link

3.4. Bundled TE-Link

Any mix of OTU and ODU links of dissimilar rates that terminates on same pair of nodes and meets the entire bundling criterion specified in TE-Link Bundling specification [RFC4201] can be pulled together to form a Bundle TE-Link. This is required for better scalability.

3.5. OTU/ODU Link Property Agreement

The OTN interfaces (associated with peer nodes) participating in a TE-Link may not be fully compatible in terms of OTN interface properties. The lowest common denominator between the two links endpoints need to be used for forming the TE link. Some of OTN specific link properties that need to be agreed upon between the two link endpoints (on peer nodes) are:

- (a) OPU Tributary Slot Granularity for OPU2 and OPU3.
- (b) Multiplexing hierarchies supported - both number of stages and specific LO-ODUs supported in each stage. This includes both Fixed and Flexible ODU containers.

These link properties either can be configured or discovered through Link discovery mechanism. The details of such mechanism is beyond the scope of this document.

4. OTU/ODU Link Bandwidth Model

Bandwidth allocation/management on OTU/ODU links is done in terms of discrete units called OPU Tributary Slots. OPU Tributary Slots occurs in two granularities (1.25Gbps and 2.5Gbps) and the actual bit-rate of the OPU Tributary Slot slightly varies for different ODU container types (i.e., ODU1, ODU2, ODU3 and ODU4). As a result of this disparity, number of Tributary Slots required to map a LO-ODU on different HO-ODU container types could vary. For example, ODU2e requires 9 OPU TSs on ODU3 and 8 OPU TSs on ODU4.

The basic objectives of OTN interface bandwidth model are as follows:

- (a) Support ODU multi-stage multiplexing hierarchy and yet not require advertisement of complete hierarchy tree.
- (b) Account for bandwidth fragmentation that can result due to the restricted multiplexing hierarchy supported on an OTN interface. For example, assume that an ODU3 interface supports direct multiplexing of ODU2 only. Here, mapping of ODU1 and ODU0 is possible only through second stage multiplexing underneath ODU2. If two ODU1 are created under two different ODU2, only two ODU2 can be created further on the interface although 28 Tributary Slots (1.25Gbps) are available on the interface (ODU hierarchy).
- (c) Hide the complexities in Tributary Slot Granularities (1.25Gbps and 2.5Gbps) from bandwidth model and thereby simplify the end-to-end path computation. As explained in the previous section, this needs to be negotiated as a part of link discovery or pre-configured locally on the either ends.
- (d) Hide the complexities in Tributary Slot Size disparities (among ODU containers) and number of Tributary Slots required to map a LO-ODU. This can be achieved by advertising the number of LO-ODU containers that can be mapped on an OTN interface rather than number of Tributary Slots or absolute bandwidth in bytes/sec.
- (e) For ODU-Flex service, Absolute bandwidth required (for CBR or GFP mapped service) needs to be mapped to 'n' Tributary Slots of certain bit rate. This needs Tributary Slot bit-rate and number of Tributary slots to be advertised.

5. OSPF TE-LSA Extension

This section describes the OSPF TE-LSA Extensions to support bandwidth encoding for OTU/ODU TE-Links.

5.1. Maximum Bandwidth

The format and interpretation of this attribute must be consistent with OSPF-TE Extension [RFC3630] and TE-Link Bundling Support [RFC4201] specifications. The OPUk payload nominal rate (in bytes per sec) as specified in [G.709-v3] shall be encoded in this attribute.

5.2. Maximum Reservable Bandwidth

The format and interpretation of this attribute must be consistent with OSPF-TE Extension [RFC3630] and TE-Link Bundling Support [RFC4201] specifications.

5.3. Unreserved Bandwidth

The format and interpretation of this attribute must be consistent with OSPF-TE Extension [RFC3630] and TE-Link Bundling Support [RFC4201] specifications.

Unreserved Bandwidth in bytes per second is not of much value for OTU/ODU interfaces. Unreserved Bandwidth per ODU rate is more appropriate and useful in this case. Implementations may choose to ignore this attribute and consider per ODU-rate Unreserved Bandwidth defined in Interface Switch Capability Descriptor for "G.709 ODUk" encoding type. See section 5.4.1 for details.

5.4. Interface Switch Capability Descriptor

The Interface Switching Capability Descriptor describes switching capability of an interface [RFC 4202]. This document defines a new Switching Capability value for OTN [G.709-v3] as follows:

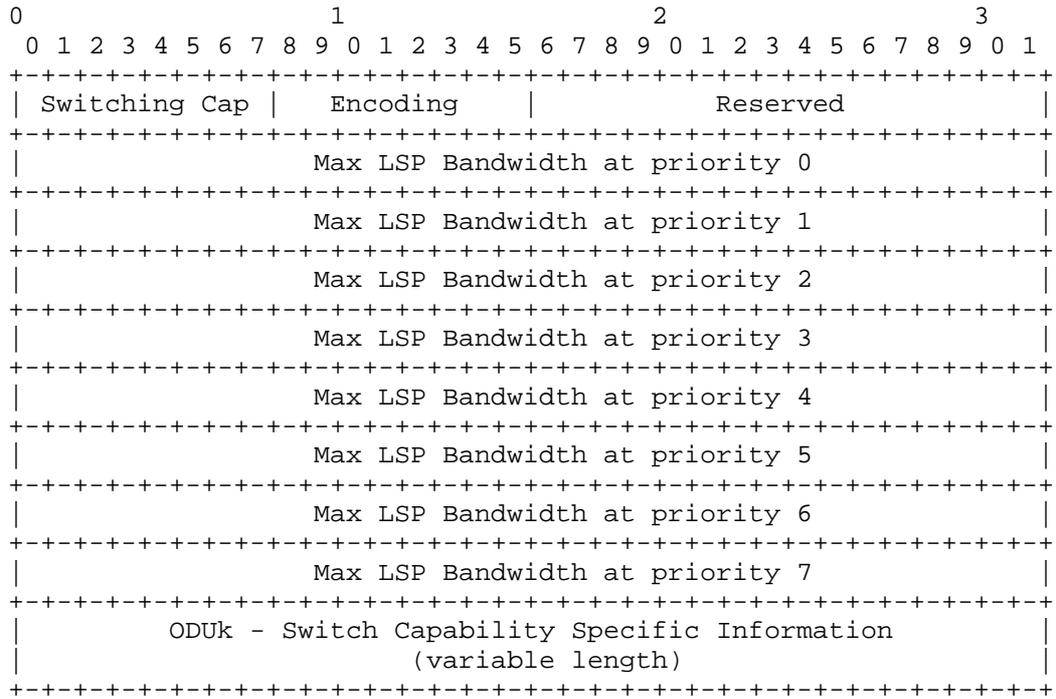
Value	Type
250	OTN-TDM capable (OTN-TDM)

Nodes advertising ODUk switching BW for its links must use Switching Type and Encoding values as follows:

Switching Type = OTN-TDM
Encoding Type = G.709 ODUk (Digital Path) [as defined in RFC4328]

Both fixed ODUk (where k=0,1,2,3,4,1e,2e) and flexible ODUs (ODUflex) use same switching type and encoding values.

When Switching Type and Encoding fields are set to values as stated above, the Interface Switching Capability Descriptor should be interpreted as follows:



Maximum LSP Bandwidth

This field should be encoded with Nominal Rate of the ODUj (j<= k) for which Bandwidth is advertised. The bandwidth unit is in bytes per second & the encoding is in IEEE floating point format [RFC 3471]. The discrete values

for varous ODUj(s) is shown in the table below.

For an unbundled link, the Maximum LSP Bandwidth at priority 'p' is set to Nominal rate of the ODUj for which bandwidth is advertised in Switch Capability Specific Information (SCSI).

For bundled link too, the Maximum LSP Bandwidth at priority 'p' is set to Nominal rate of the ODUj for which bandwidth is advertised in Switch Capability Specific Information (SCSI).

ODU type	Nomial Rate(bytes/s)	Value in Byes/Sec (IEEE format)
ODU0	15552000	
ODU1	312346890.75	
ODU2	1254659240.50	
ODU2e	1299940664.50	
ODU1e		
ODU3	5039902372.875	
ODU4	13099305726.875	
ODUflex	Any	

The Maximum LSP bandwidth field is used to identify the ODUj type.

5.4.1 ODU Switching

When Switching Capability is set to OTN-TDM, it means the node is capable of

- terminating OTUk layer
- Switching of HO-ODU (ODUk)
- switching of LO-ODU (ODUj) if HO-ODU supports mux/demux (termination of HO-ODU is required for mux/demux operation)

Multiple ISCDs would be advertised if an interface supports more than one type of ODUk switching. There would be one ISCD advertisement per ODUj independent of the OTN multiplexing branch it belongs to.

For e.g. If an OTU3 interface supports ODU0, ODU1 and ODU2 switching, there would be three ISCDs one for each ODU type.

Refer to examples in section 7.0.

5.4.2. ODUk Switch Capability Specific Information

This SCSE field contains bandwidth information for fixed ODUj(j=0,1,2,3,4,2e,1e) or ODUflex.

The type of ODUj/ODUflex is identified by Maximum LSP bandwidth field and BW sub TLV Type field as follows.

If bandwidth advertisement is for fixed size ODUj, then

- set BW sub TLV Type = 1
- Encode nominal rate of the ODUj in Max LSP BW field
- Encode available number of ODUj(s) as shown below

If bandwidth advertisement is for ODUflex, then

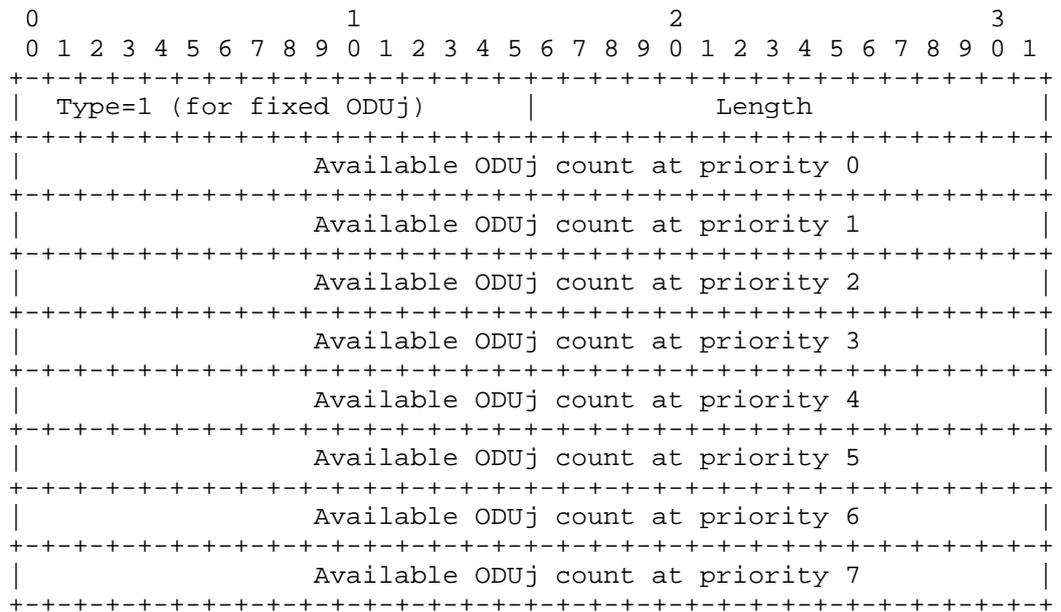
- set BW sub TLV Type = 2
- Encode available BW in Max LSP BW field

- Encode available Bandwidth as shown below

The SCSI field must be included when Switching Capability is "OTN-TDM".

5.4.2.1 Bandwidth sub TLV for fixed ODUj

The format of Bandwidth sub TLV for fixed size ODUj is shown below; (j=0,1,2,3,4,2e,1e). The TLV Type must be set to 1 for fixed ODUs.



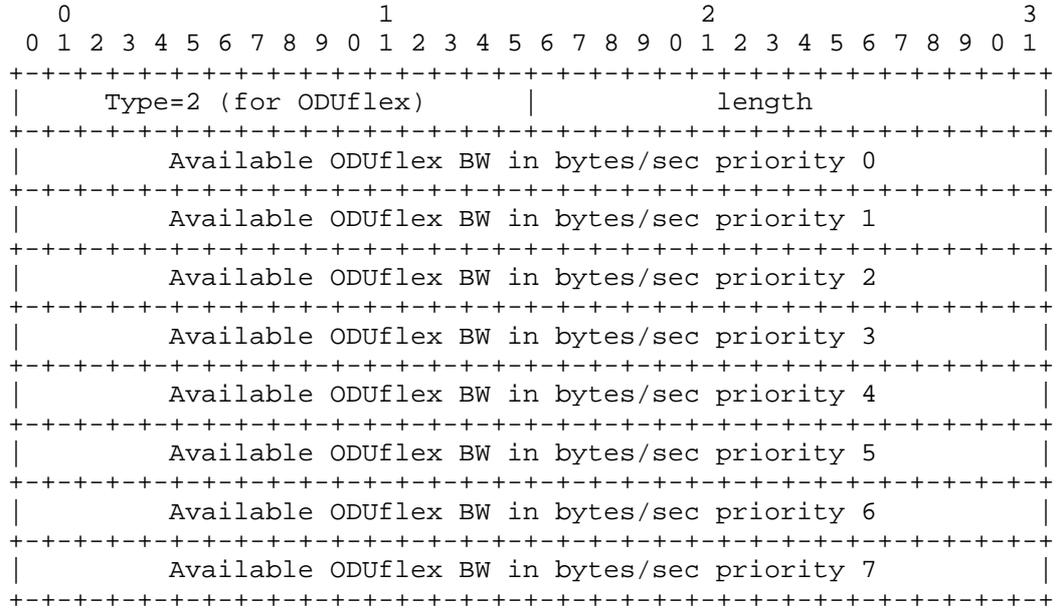
Available ODUj(s):

This field (32 bits) indicates the maximum number of Containers of a given ODUj-Type at priority 'p' available on this TE-Link.

The "Available ODUj(s)" of a bundled link at priority p is defined to be the sum of "Available ODUj(s)" at priority p of all of its component links.

5.4.2.2 Bandwidth sub-TLV for ODUflex

The format of Bandwidth sub TLV for ODUflex is shown below.
 The TLV Type is set to 2 for flexible ODUs.



Available BW (in bytes/sec)

Available BW (in bytes/sec) is represented in IEEE float-point format similar to Max-Lsp-Bandwidth in ISCD.

The "Available BW" of a bundled link at priority p is defined to be the sum of "Available BW" at priority p of all of its component links.

This information may be used to route LSPs over links which have most bandwidth available.

5.5. Interface Multiplexing Capability Descriptor

The OTN multiplexing hierarchy involves one or more ODU layers. The server ODU layer is called the higher order ODU(HO-ODU) and the layer multiplexed into a server ODU layer is called lower order ODU (LO-ODU).

A HO-ODU can carry (mux/demux) one or more LO-ODUs as specified in G.709. Termination of HO-ODU is necessary to mux/demux LO-ODUs. For e.g.

a) on a OTU2 interface with OTU2-ODU2-ODU0 muxing stack, it is necessary to terminate ODU2(H) in order to mux/demux contained ODU0s.

b) on a OTU2 interface with OTU2-ODU2-ODU1-ODU0 muxing stack, it is necessary to terminate ODU2 and ODU1 layers to mux/demux contained ODU0s.

An OTN interface supporting multi-stage multiplexing requires termination of more than one HO-ODU to access one or more LO-ODUs for switching purposes. For e.g. on an interface with OTU3-ODU3-ODU2-ODU0 multiplexing stack/hierarchy,

ODU3 and ODU2 layers should be terminated to access ODU0s for switching purposes.

5.5.1 Multiplex Layers and Hierarchical LSP

It is possible to construct H-LSP(s) using different HO-ODU muxing layer(s). While creation of H-LSP is optional, it becomes necessary in network scenarios where switching restrictions exist for LO-ODUs.

Example #1:

- Nodes A, B, D & E are ODU0 and ODU2 switching capable;
- Node C is ODU2 switching capable only.

An ODU2-FA between nodes B & D is necessary to support E2E ODU0-LSP(s)

```
A-----B-----C-----D-----E
          <-----ODU2-FA----->
<-----ODU0-LSP ----->
```

Example #2: ODU0-LSP over G.709-v1 capable node (legacy node)

- Nodes A, B, D & E are ODU0 & ODU1 switch capable nodes;
- Node C is ODU1 switching capable

An ODU1-FA between nodes B & D is necessary to support E2E ODU0-LSPs

```
A-----B-----C-----D-----E
          <-----ODU1-FA----->
<-----ODU0-LSP ----->
```

In order to support identification of potential FA boundary points, it is

necessary to flood mux/demux information. This includes information about:

- the HO-ODU layer which can be terminated
- the LO-ODUs available upon HO-ODU termination (muxing hierarchy)

The multiplexing hierarchy provides information about specific branch(es) of the OTN muxing hierarchy. This includes

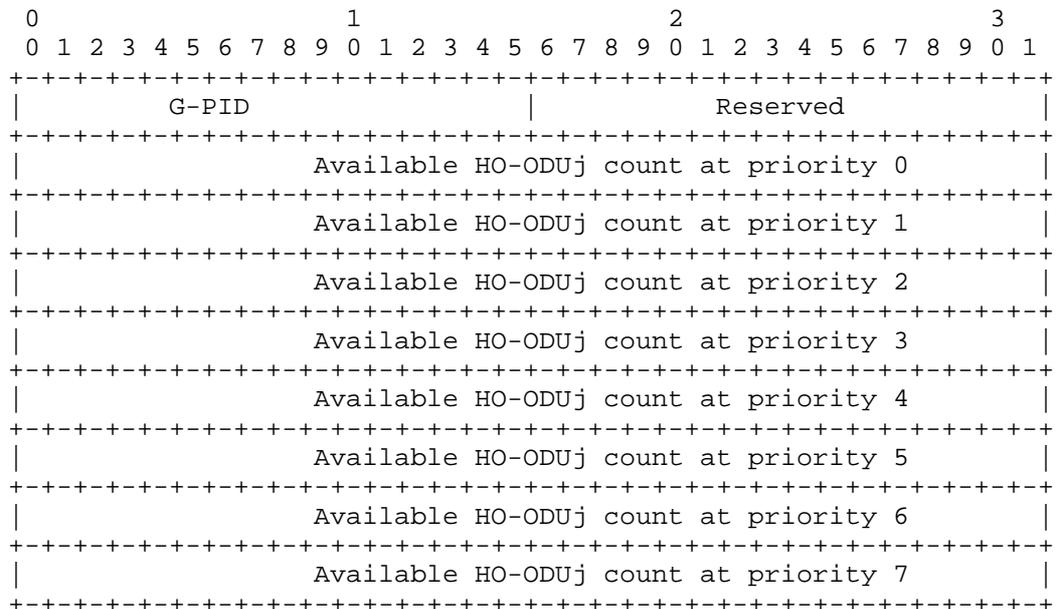
- one or more HO-ODU(s) which needs to be terminated and
- a LO-ODU layer which can be accessed after termination

The HO-ODUs which are terminate-able are potential FA end points. FA becomes necessary when switching bandwidth is not available at all nodes along the path for an LSP (specifically for LSPs at LO-ODU layers).

This draft proposes the use of IMCD (Interface Multiplexing Capability Descriptor) to distribute OTN mux/demux information of Te-end points.

5.5.2 IMCD format

The Interface Multiplexing Capability Descriptor (IMCD) describes "Multiplexing" capability of an interface. It is a sub-TLV of the Link TLV (Type TBD). The format of value field is as shown below:



5.5.2.1 G-PID

The G-PID field is a 16 bit field as defined in [RFC3471].
 New G-PID values are defined in addition to those defined in [RFC3471].
 Within OTN context, the new G-PID values identify multiplexing stack supported by the Te-end point.

The table below shows newly defined values for G-PID:

Value	G-PID	Meaning
60	ODU1-ODU0	ODU1 termination required
61	ODU2-ODU0	ODU2 termination required
62	ODU2-ODU1	ODU2 termination required
63	ODU2-ODU1-ODU0	ODU2 & ODU1 termination required
64	ODU2-ODUflex	ODU2 termination required
65	ODU3-ODU0	ODU3 termination required
66	ODU3-ODU1	ODU3 termination required
67	ODU3-ODU1-ODU0	ODU3 & ODU1 termination required
68	ODU3-ODU2	ODU3 termination required
69	ODU3-ODU2-ODU0	ODU3 & ODU2 termination required
70	ODU3-ODU2-ODU1	ODU3 & ODU2 termination required
71	ODU3-ODU2-ODU1-ODU0	ODU3 & ODU2 & ODU1 termination required
72	ODU3-ODU2-ODUflex	ODU3 & ODU2 termination required
73	ODU3-ODUflex	ODU3 termination required
74	ODU3-ODU2e	ODU3 termination required
75	ODU4-ODU0	ODU4 termination required
76	ODU4-ODU1	ODU4 termination required
77	ODU4-ODU1-ODU0	ODU4 & ODU1 termination required
78	ODU4-ODU2	ODU4 termination required
79	ODU4-ODU2-ODU0	ODU4 & ODU2 termination required
80	ODU4-ODU2-ODU1	ODU4 & ODU2 termination required
81	ODU4-ODU2-ODU1-ODU0	ODU4 & ODU2 & ODU1 termination required
82	ODU4-ODU2-ODUflex	ODU4 & ODU2 termination required
83	ODU4-ODU3	ODU4 termination required
84	ODU4-ODU3-ODU0	ODU4 & ODU3 termination required
85	ODU4-ODU3-ODU1	ODU4 & ODU3 termination required
86	ODU4-ODU3-ODU1-ODU0	ODU4 & ODU3 & ODU1 termination required
87	ODU4-ODU3-ODU2	ODU4 & ODU3 termination required
88	ODU4-ODU3-ODU2-ODU0	ODU4 & ODU3 & ODU2 termination required
89	ODU4-ODU3-ODU2-ODU1	ODU4 & ODU3 & ODU2 termination required
90	ODU4-ODU3-ODU2-ODU1-ODU0	ODU4 & ODU3 & ODU2 & ODU1 termination required
91	ODU4-ODU3-ODU2-ODUflex	ODU4 & ODU3 & ODU2 termination required
92	ODU4-ODU3-ODUflex	ODU4 & ODU3 termination required
93	ODU4-ODU3-ODU2e	ODU4 & ODU3 termination required
94	ODU4-ODUflex	ODU4 termination required
95	ODU4-ODU2e	ODU4 termination required

96	ODU1	ODU1 termination required
97	ODU2	ODU2 termination required
98	ODU3	ODU3 termination required
99	ODU4	ODU4 termination required
100	ODU2-GFP-10GBE	ODU2 termination for Ethernet
101	ODU2e-10GBE	ODU2e termination for Ethernet
102	ODU2-OC192	ODU2 termination for SONET

5.5.2.2 Available Bandwidth

The available bandwidth advertised in "Available HO-ODUj" field indicates the number of "Terminations" possible at HO-ODUj layer. The HO-ODUj layer (Parent ODU) is identified by G-PID field.

This field (32 bits) indicates maximum number of Containers of a given HO-ODUj at priority 'p' available on the TE-Link; where {j=1,2,3,4}.

The "Available HO-ODUj(s)" of a bundled link at priority 'p' is defined to be the sum of "Available HO-ODUj(s)" at priority 'p' of all of its component links for that specific G-PID.

Example#1: Unbundled link with ODU2-ODU0 muxing hierarchy support

A ----- B

IMCD advertised would be as follows:

- o G-PID= ODU2-ODU0
- o Available HO-ODUj count = 1 (refers to ODU2 layer)

The ODU2 termination implies ability to mux/demux 8xODU0s.

Example#2: Bundled Te-link with ODU2-ODU0 muxing hierarchy support (3 links)

A ===== B

IMCD advertised would be as follows:

- o G-PID= ODU2-ODU0
- o Available HO-ODUj count = 3 (refers to ODU2 layer)

The ODU2 termination implies ability to mux/demux 24xODU0s in total.

5.5.3 Controlling IMCD advertisement

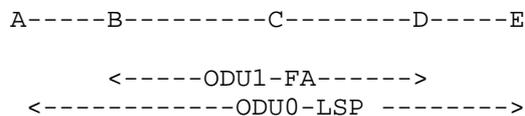
The IMCD advertisement is not mandatory and it is required only when FA support is needed.

The network operators can selectively enable IMCD advertisement for specific HO-ODU mux layer(s). This can be done on a link by link basis, node basis or network basis. The mechanism to achieve this is outside the scope of this document.

5.5.4 How to use IMCDs for FA creation

When computing path for an FA (induced or otherwise), the path computing node should look for matching G-PIDs at the FA boundary nodes. For example, to create ODU1-FA for ODU0 service, the path computation should look for matching G-PID = ODU1-ODU0 at nodes B & D

The need for FA is due to Node-C's ability to switch ODU1 only.



5.5.5 IMCD and non OTN services

In certain deployments it may be beneficial to advertise ODU termination bandwidth without the LO-ODU information. The intent is to allow signaling to decide non-OTN signal to adapt at the time of path establishment.

The G-PID values 96, 97, 98, 99 defined in section 5.5.2.1 are meant for this purpose.

The path computation can also be preformed for specific clients over an ODUj using G-PID values 100, 101 & 102 (e.g. 10GBE mapping to ODU2 using GFP).

6. Examples

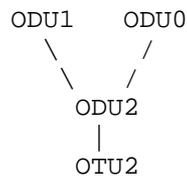
This sections presents some use-cases for bandwidth encoding and usage.

6.1. Network with no IMCD advertisement (no FA support)

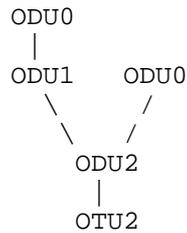
A-----B-----C-----D-----E

Suppose Muxing Hierarchy supported at the end points as shown:

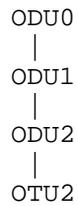
Link A-B: Mux hierarchy at A & B ends is as follows:



Link B-C: Mux hierarchy at B & C ends is as follows:



Link C-D: mux hierarchy at C & D ends is as follows:



a) The ISCD advertisement by nodes A, B, C & D would be as follows

ISCD1:
Max LSP BW = ODU2 nominal rate in bytes/sec
Available ODU2 count at priority 'p' = 1

ISCD2:
Max LSP BW = ODU1 nominal rate in bytes/sec
Available ODU1 count at priority 'p' = 4

ISCD3:
Max LSP BW = ODU0 nominal rate in bytes/sec
Available ODU0 count at priority 'p' = 8

b) BW advertisement after an ODU0-LSP creation from A to D.
The bandwidth is no longer available at ODU2 rate.

ISCD1:
Max LSP BW = ODU1 nominal rate in bytes/sec
Available ODU1 count at priority 'p' = 3

ISCD2:
Max LSP BW = ODU0 nominal rate in bytes/sec
Available ODU0 count at priority 'p' = 7

6.2. Network with IMCD advertisement for FA support

```
A-----B-----C-----D-----E
      <---ODU1-FA--->
<----- ODU0-LSP ----->
```

The above network can support FA at ODU2 and ODU1 layers.
To support FA origination/termination, the IMCDs would be advertised
as follows. This is in addition to ISCD advertisement.

The ISCD1, ISCD2 & ISCD3 advertisement by A, B, C & D is same as in section 7.1

The IMCD advertisement by A & B for link A-B:

IMCD1:
G-PID = ODU2-ODU1
Available HO-ODUj count at Pi = 1 (ODU2)

IMCD2:
G-PID = ODU2-ODU0
Available HO-ODUj count at Pi = 1 (ODU2)

The IMCD advertisement by B & C for link B-C:

IMCD1:

G-PID = ODU2-ODU1
Available HO-ODUj count at Pi = 1 (ODU2)

IMCD2:
G-PID = ODU2-ODU0
Available HO-ODUj count at Pi = 1 (ODU2)

IMCD3:
G-PID = ODU1-ODU0
Available HO-ODUj count at Pi = 4 (ODU1)

The IMCDs advertised by C & D for link C-D would be as follows:

IMCD1:
G-PID = ODU2-ODU1
Available HO-ODUj count at Pi = 1 (ODU2)

IMCD2:
G-PID = ODU2-ODU1-ODU0
Available HO-ODUj count at Pi = 1 (ODU2)

IMCD3:
G-PID = ODU1-ODU0
Available HO-ODUj count at Pi = 4 (ODU1)

The IMCD advertisement by B & C for link B-C after ODU1-FA creation:

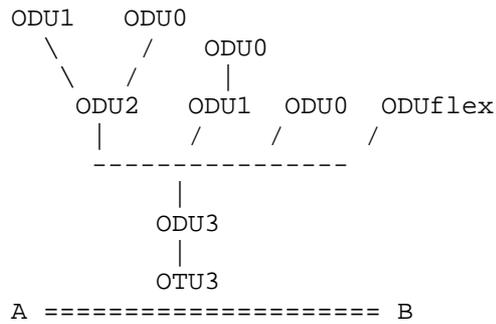
IMCD1:
G-PID = ODU1-ODU0
Available HO-ODUj count at Pi = 3 (ODU1)

The IMCD advertisement by C & D for link C-D after ODU1-FA creation:

IMCD1:
G-PID = ODU1-ODU0
Available HO-ODUj count at Pi = 3 (ODU1)

6.3. Link bundle with similar muxing capabilities

Consider a Bundled Te-link with 2xOTU3 links between Nodes A & B with multiplexing hierarchy as shown:



The ISCDs and IMCDs advertised by A & B would be as follows:

- ISCD1:
Max LSP BW = ODU3 nominal rate in bytes/sec
Available ODU3 count at priority 'p' = 2
- ISCD2:
Max LSP BW = ODU2 nominal rate in bytes/sec
Available ODU2 count at priority 'p' = 8
- ISCD3:
Max LSP BW = ODU1 nominal rate in bytes/sec
Available ODU1 count at priority 'p' = 32
- ISCD4:
Max LSP BW = ODU0 nominal rate in bytes/sec
Available ODU0 count at priority 'p' = 64
- ISCD5:
Max LSP BW = ODU3 nominal rate in bytes/sec
Available ODUflex BW = 2xODU3 BW in byte/sec

To support FAs at ODU3, ODU2 & ODU1 rates, the following IMCDs are advertised

- IMCD1:
G-PID = ODU3-ODU2
Available HO-ODUj count at Pi = 2 (ODU3)

IMCD2:
G-PID = ODU3-ODU2-ODU1
Available HO-ODUj count at Pi = 2 (ODU3)

IMCD3:
G-PID = ODU3-ODU2-ODU0
Available HO-ODUj count at Pi = 2 (ODU3)

IMCD4:
G-PID = ODU3-ODU1
Available HO-ODUj count at Pi = 2 (ODU3)

IMCD5:
G-PID = ODU3-ODU1-ODU0
Available HO-ODUj count at Pi = 2 (ODU3)

IMCD6:
G-PID = ODU3-ODU0
Available HO-ODUj count at Pi = 2 (ODU3)

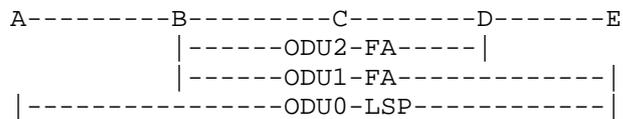
IMCD7:
G-PID = ODU2-ODU1
Available HO-ODUj count at Pi = 8 (ODU2)

IMCD8:
G-PID = ODU2-ODU0
Available HO-ODUj count at Pi = 8 (ODU2)

IMCD9:
G-PID = ODU1-ODU0
Available HO-ODUj count at Pi = 32 (ODU1)

IMCD9:
G-PID = ODU3-ODUflex
Available HO-ODUj count at Pi = 3 (ODUflex)

6.4. Link bundle with dissimilar muxing capabilities: Layer relation



Link A-B: Hierarchy at both ends is OTU2-ODU2-ODU0
Link B-C: Is a bundled Te-link with 3 component links with multiplexing hierarchy at both ends as shown:

Component link#1: OTU2 link with mux hierarchy: OTU2-ODU2-ODU1-ODU0
Component link#2: OTU2 link with mux hierarchy: OTU2-ODU2-ODU1
Component link#3: OTU1 link with mux hierarchy: OTU1-ODU1-ODU0

Link C-D:

- Hierarchy at C end is OTU2-ODU2
- Hierarchy at D end is OTU2-ODU2-ODU1

Link D-E:

- Hierarchy at D end is OTU1-ODU1
- Hierarchy at E end is OTU1-ODU1-ODU0

The IMCDs advertised for B-C would include the following:

IMCD1:

G-PID = ODU2-ODU1
Available HO-ODUj count at Pi = 2 (ODU2)

IMCD2:

G-PID = ODU1-ODU0
Available HO-ODUj count at Pi = 5 (ODU1)

IMCD3:

G-PID = ODU2-ODU1-ODU0
Available HO-ODUj count at Pi = 1 (ODU2)

In this example, we need two FAs to originate from the same point (at node-B). It is necessary to advertise IMCD3 as we can not conclude full mux relation from IMCD1 & IMCD2.

7. Backward Compatibility

If backwards compatibility is required with G.709-v1, then [RFC4328] based ISCDs should be advertised in addition to ISCDs/IMCDs specified in this document.

8. Security Considerations

There are no additional security implications to OSPF routing protocol due to the extensions captured in this document.

9. IANA Considerations

The memo introduces two new sub-TLVs of the Interface Switch Capability Descriptor Sub-TLV of TE-LSA. [RFC3630] says that the sub-TLVs of the TE Link TLV in the range 10-32767 must be assigned by Expert Review, and must be registered with IANA.

10. References

10.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels".
- [RFC3630] Katz, D., Kompella, K., and D. Yeung, "Traffic Engineering (TE) Extensions to OSPF Version 2", RFC 3630
- [RFC3471] Berger, L., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Functional Description", RFC 3471, January 2003.
- [RFC4201] Kompella, K., Rekhter, Y., and L. Berger, "Link Bundling in MPLS Traffic Engineering (TE)"
- [RFC4203] Kompella, K. and Y. Rekhter, "OSPF Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)"
- [RFC4204] Lang, J., Ed., "Link Management Protocol (LMP)", RFC 4204, October 2005.
- [RFC4328] Papadimitriou, D., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Extensions for G.709 Optical Transport Networks Control", RFC 4328, January 2006.
- [RFC5212] Shiimoto, K., Papadimitriou, D., Le Roux, JL., Vigoureux, M., and D. Brungard, "Requirements for GMPLS-Based Multi-Region and Multi-Layer Networks (MRN/MLN)", RFC 5212, July 2008.
- [RFC5339] Le Roux, JL. and D. Papadimitriou, "Evaluation of Existing GMPLS Protocols against Multi-Layer and Multi-Region Networks (MLN/MRN)", RFC 5339, September 2008.
- [RFC6001] D. Papadimitriou, et al, Generalized MPLS (GMPLS) Protocol Extensions for Multi-Layer and Multi-Region Networks (MLN/MRN)
- [G.709-v3] ITU-T, "Interfaces for the Optical Transport Network (OTN)", G.709 Recommendation, December 2009.
- [GSUP.43] ITU-T, "Proposed revision of G.sup43 (for agreement)", December 2008.

10.2. Informative References

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

[G.709-v1] ITU-T, "Interface for the Optical Transport Network (OTN)," G.709 Recommendation (and Amendment 1), February 2001 (October 2001).

[G.872] ITU-T, "Architecture of optical transport networks", November 2001 (11 2001).

[G.709-FRAME] F. Zhang, D. Li, H. Li, S. Belotti, "Framework for GMPLS and PCE Control of G.709 Optical Transport Networks", draft-zhang-ccamp-gmpls-g709-framework-02, work in progress.

[WSON-FRAME] Y. Lee, G. Bernstein, W. Imajuku, "Framework for GMPLS and PCE Control of Wavelength Switched Optical Networks (WSON)", draft-ietf-ccamp-rwa-wson-framework, work in progress.

11. Acknowledgements

Special thanks to Daniele Ceccarelli, Lyndon Ong, Sergio Belotti, Pietro Grandi, Jonathan Sadler, Remi Theillaud, Fatai Zhang and Diego Caviglia for discussions on various modeling options.

Authors would like to thank Lou Berger, Ping Pan, Radhakrishna Valiveti and Mohit Misra for review comments and suggestions.

Author's Addresses

Ashok Kunjidhpatham
Infinera Corporation
169, Java Drive
Sunnyvale, CA-94089
USA
Email: akunjidhpatham@infinera.com

Rajan Rao
Infinera Corporation
169, Java Drive
Sunnyvale, CA-94089
USA
Email: rrao@infinera.com

Snigdho Bardalai
Infinera Corporation

169, Java Drive
Sunnyvale, CA-94089
USA
Email: sbardalai@infinera.com

Khuzema Pithewan
Infinera Corporation
169, Java Drive
Sunnyvale, CA-94089
USA
Email: kpithewan@infinera.com

Biao Lu
Infinera Corporation
169, Java Drive
Sunnyvale, CA-94089
USA
Email: blu@infinera.com

John Drake
Juniper Networks
USA
Email: jdrake@juniper.net

Steve Balls
Metaswitch Networks
100 Church Street
Enfield
EN2 6BQ U.K.
Email: steve.balls@metaswitch.com

Xihua Fu,
ZTE
China
fu.xihua@zte.com.cn

Appendix A: Abbreviations & Terminology

A.1 Abbreviations:

CBR	Constant Bit Rate
GFP	Generic Framing Procedure
HO-ODU	Higher Order ODU
LSC	Lambda Switch Capable
LSP	Label Switched Path
LO-ODU	Lower Order ODU
ISCD	Interface Switch Capability Descriptor
OCC	Optical Channel Carrier
OCG	Optical Carrier Group
OCh	Optical Channel (with full functionality)
OChr	Optical Channel (with reduced functionality)
ODTUG	Optical Data Tributary Unit Group
ODU	Optical Channel Data Unit
OMS	Optical Multiplex Section
OMU	Optical Multiplex Unit
OPS	Optical Physical Section
OPU	Optical Channel Payload Unit
OSC	Optical Supervisory Channel
OTH	Optical Transport Hierarchy
OTM	Optical Transport Module
OTN	Optical Transport Network
OTS	Optical Transmission Section
OTU	Optical Channel Transport Unit
OTUkV	Functionally Standardized OTUk
SCSI	Switch Capability Specific Information
TDM	Time Division Multiplex

A.2 Terminology

1. ODU_k and ODU_j

ODU_k refers to the ODU container that is directly mapped to an OTU container. ODU_j refers to the lower order ODU container that is mapped to an higher order ODU container via multiplexing.

2. LO-ODU and HO-ODU

LO-ODU refers to the ODU client layer of lower rate that is mapped to an ODU server layer of higher rate via multiplexing. HO-ODU refers to the ODU server layer of higher rate that supports mapping of one or more ODU client layers of lower rate.

In multi-stage multiplexing case, a given ODU layer can be a client for one stage (interpreted as LO-ODU) and at the same

time server for another stage (interpreted as HO-ODU). In this case, the notion of LO-ODU and HO-ODU needs to be interpreted in a recursive manner.

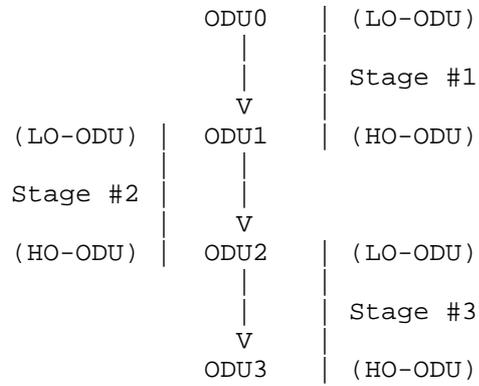


Figure-4 : LO-ODU and HO-ODU

Appendix B : Optimization Techniques

Optimization techniques can be used to reduce TE-LSA size. The following sub sections describe available options.

B.1 Multiple ISCDs Vs. Single ISCD

It is possible to encode ISCDs corresponding to different ODU layers into SCSI field of a single ISCD. This options was shown in previous version of this draft (draft-ashok-ccamp-gmpls-ospf-g709-02).

Doing so will reduce the LSA size by a factor of:
10 words x (#ODUjs - 1)

It is possible to reduce LSA size further by reducing the size of BW field to half word. Doing so will reduce LSA size by a factor of:
4 words x (#ODUjs)

B.1 Multiple IMCDs Vs. Single IMCD

This optimization doesn't save much. The shrinking of BW field to 1/2 word helps reduce LSA size to some extent. The size reduction depends on the number of ODUs supported.

4 words x (#ODUjs)

B.1 Eight priorities Vs. restricted number of priorities

It is possible to further optimize by advertising BW only for supported priorities. This can be easily achieved by having a bit vector as described in previous version of this draft.

Appendix C: Relation with MLN & MRN

The ISCD and IMCDs defined in this draft doesn't replace IACDs. All three (ISCD, IMCD & IACD) can co-exist in a network and serve different purposes.

Appendix D : AMP, BMP & GMP Mapping

The G.709 defines various mapping schemes for LO-ODUs into HO-ODUs. From G.709 descriptions, the AMP & GMP mapping appears to be fixed for a given LO-ODU to HO-ODU based on the time slot granularity. Since the mapping is fixed we do not see value in advertising this information in TE-LSAs.

CCAMP Working Group
Internet-Draft
Intended status: Informational
Expires: September 12, 2011

S. Belotti
P. Grandi
Alcatel-Lucent
D. Ceccarelli
D. Caviglia
Ericsson
F. Zhang
D. Li
Huawei Technologies
March 11, 2011

Information model for G.709 Optical Transport Networks (OTN)
draft-bccg-ccamp-otn-g709-info-model-04

Abstract

The recent revision of ITU-T recommendation G.709 [G.709-v3] has introduced new fixed and flexible ODU containers in Optical Transport Networks (OTNs), enabling optimized support for an increasingly abundant service mix.

This document provides a model of information needed by the routing and signaling process in OTNs to support Generalized Multiprotocol Label Switching (GMPLS) control of all currently defined ODU containers.

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on September 12, 2011.

Copyright Notice

Copyright (c) 2011 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1.	Introduction	3
1.1.	Terminology	4
2.	OSPF-TE requirements overview	4
3.	RSVP-TE requirements overview	5
4.	G.709 Digital Layer Info Model for Routing and Signaling	5
4.1.	Tributary Slot type	8
4.2.	Tributary Port Number	9
4.3.	Signal type	9
4.4.	Bit rate and tolerance	11
4.5.	Unreserved Resources	11
4.6.	Maximum LSP Bandwidth	11
4.7.	Distinction between terminating and switching capability	12
4.8.	Priority Support	14
4.9.	Multi-stage multiplexing	14
4.10.	Generalized Label	14
5.	Security Considerations	15
6.	IANA Considerations	15
7.	Contributors	15
8.	Acknowledgements	15
9.	References	15
9.1.	Normative References	15
9.2.	Informative References	16
	Authors' Addresses	17

1. Introduction

GMPLS[RFC3945] extends MPLS to include Layer-2 Switching (L2SC), Time-Division Multiplexing (e.g., SONET/SDH, PDH, and OTN), Wavelength (OCh, Lambdas) Switching and Spatial Switching (e.g., incoming port or fiber to outgoing port or fiber).

The establishment of LSPs that span only interfaces recognizing packet/cell boundaries is defined in [RFC3036, RFC3212, RFC3209]. [RFC3471] presents a functional description of the extensions to Multi-Protocol Label Switching (MPLS) signaling required to support GMPLS. Resource Reservation Protocol-Traffic Engineering (RSVP-TE) -specific formats, mechanisms and technology specific details are defined in [RFC3473].

From a routing perspective, Open Shortest Path First-Traffic Engineering (OSPF-TE) generates Link State Advertisements (LSAs) carrying application-specific information and floods them to other nodes as defined in [RFC5250]. Three types of opaque LSA are defined, i.e. type 9 - link-local flooding scope, type 10 - area-local flooding scope, type 11 - AS flooding scope.

Type 10 LSAs are composed of a standard LSA header and a payload including one top-level TLV and possible several nested sub-TLVs. [RFC3630] defines two top-level TLVs: Router Address TLV and Link TLV; and nine possible sub-TLVs for the Link TLV, used to carry link related TE information. The Link type sub-TLVs are enhanced by [RFC4203] in order to support GMPLS networks and related specific link information. In GMPLS networks each node generates TE LSAs to advertise its TE information and capabilities (link-specific or node-specific) through the network. The TE information carried in the LSAs are collected by the other nodes of the network and stored into their local Traffic Engineering Databases (TED).

In a GMPLS enabled G.709 Optical Transport Networks (OTN), routing and signaling are fundamental in order to allow automatic calculation and establishment of routes for ODUk LSPs. The recent revision of ITU-T Recommendation G.709 [G709-V3] has introduced new fixed and flexible ODU containers that augment those specified in foundation OTN. As a result, it is necessary to provide OSPF-TE and RSVP-TE extensions to allow GMPLS control of all currently defined ODU containers.

This document provides the information model needed by the routing and signaling processes in OTNs to allow GMPLS control of all currently defined ODU containers.

OSPF-TE and RSVP-TE requirements are defined in [OTN-FWK], while

protocol extensions are defined in [OTN-OSPF] and [OTN-RSVP].

1.1. Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

2. OSPF-TE requirements overview

[OTN-FWK] provides a set of functional routing requirements summarized below :

- Support for link multiplexing capability advertisement: The routing protocol has to be able to carry information regarding the capability of an OTU link to support different type of ODUs
- Support of any ODUk and ODUFlex: The routing protocol must be capable of carrying the required link bandwidth information for performing accurate route computation for any of the fixed rate ODUs as well as ODUFlex.
- Support for differentiation between switching and terminating capacity
- Support for the client server mappings as required by [G.7715.1]. The list of different mappings methods is reported in [G.709-v3]. Since different methods exist for how the same client layer is mapped into a server layer, this needs to be captured in order to avoid the set-up of connections that fail due to incompatible mappings.
- Support different priorities for resource reservation. How many priorities levels should be supported depends on operator policies. Therefore, the routing protocol should be capable of supporting either no priorities or up to 8 priority levels as defined in [RFC4202].
- Support link bundling either at the same line rate or different line rates (e.g. 40G and 10G). Bundling links at different rates makes the control plane more scalable and permits better networking flexibility.

3. RSVP-TE requirements overview

[OTN-FWK] also provides a set of functional signaling requirements summarized below :

- Support for LSP setup of new ODUk/ODUflex containers with related mapping and multiplexing capabilities
- Support for LSP setup using different Tributary Slot granularity
- Support for Tributary Port Number allocation and negoziation
- Support for constraint signaling

4. G.709 Digital Layer Info Model for Routing and Signaling

The digital OTN layered structure is comprised of digital path layer networks (ODU) and digital section layer networks (OTU). An OTU section layer supports one ODU path layer as client and provides monitoring capability for the OCh. An ODU path layer may transport a heterogeneous assembly of ODU clients. Some types of ODUs (i.e., ODU1, ODU2, ODU3, ODU4) may assume either a client or server role within the context of a particular networking domain. ITU-T G.872 recommendation provides two tables defining mapping and multiplexing capabilities of OTNs, which are reproduced below.

ODU client	OTU server
ODU 0	-
ODU 1	OTU 1
ODU 2	OTU 2
ODU 2e	-
ODU 3	OTU 3
ODU 4	OTU 4
ODU flex	-

Figure 1: OTN mapping capability

ODU client	ODU server
1,25 Gbps client	ODU 0
-	
2,5 Gbps client	ODU 1
ODU 0	
10 Gbps client	ODU 2
ODU0,ODU1,ODUflex	
10,3125 Gbps client	ODU 2e
-	
40 Gbps client	ODU 3
ODU0,ODU1,ODU2,ODU2e,ODUflex	
100 Gbps client	ODU 4
ODU0,ODU1,ODU2,ODU2e,ODU3,ODUflex	
CBR clients from greater than 2.5 Gbit/s to 100 Gbit/s: or GFP-F mapped packet clients from 1.25 Gbit/s to 100 Gbit/s.	ODUflex
-	

Figure 2: OTN multiplexing capability

How an ODUk connection service is transported within an operator network is governed by operator policy. For example, the ODUk connection service might be transported over an ODUk path over an OTUk section, with the path and section being at the same rate as that of the connection service (see Table 1). In this case, an entire lambda of capacity is consumed in transporting the ODUk connection service. On the other hand, the operator might exploit different multiplexing capabilities in the network to improve infrastructure efficiencies within any given networking domain. In

this case, ODUk multiplexing may be performed prior to transport over various rate ODU servers (as per Table 2) over associated OTU sections.

From the perspective of multiplexing relationships, a given ODUk may play different roles as it traverses various networking domains.

As detailed in [OTN-FWK], client ODUk connection services can be transported over:

- o Case A) one or more wavelength sub-networks connected by optical links or
- o Case B) one or more ODU links (having sub-lambda and/or lambda bandwidth granularity)
- o Case C) a mix of ODU links and wavelength sub-networks.

This document considers the TE information needed for ODU path computation and parameters needed to be signaled for LSP setup.

The following sections list and analyze each type of data that needs to be advertised and signaled in order to support path computation and LSP setup.

4.1. Tributary Slot type

ITU-T recommendations define two types of TS but each link can only support a single type at a given time. The rules to be followed when selecting the TS to be used are:

- If both ends of a link can support both 2.5Gbps TS and 1.25Gbps TS, then the link will work with 1.25Gbps TS.
- If one end can support the 1.25Gbps TS, and another end the 2.5Gbps TS, the link will work with 2.5Gbps TS.

In case the bandwidth accounting is provided in number of TSs, the type of TS is needed to perform correct routing operations. Currently such information is not provided by the routing protocol and not taken into account during LSP signaling.

The tributary slot type information is one of the parameters needed to correctly configure physical interfaces, therefore it has to be signaled via RSVP-TE. This allows the end points of the FA know which TS should be used.

[editor note]: SG15 ITU-T G.798 describes the so called PT=21-to-

PT=20 interworking process that explains how two equipments with different PayloadType, and hence different TS granularity (1.25Gbps vs. 2.5Gbps), can be coordinated so to permit the equipment with 1.25 TS granularity to adapt his TS allocation accordingly to the different TS granularity (2.5Gbps) of a neighbour. Therefore, in order to let the NE change TS granularity accordingly to the neighbour requirements, the AUTOpayloadtype needs to be configured and the HO ODU source can be either not provisioned (i.e. TS not allocated) or configured following a specific mapping depending of the type of LO ODU carried. In this case the process of auto-negotiation makes the system self consistent and the only reason for signaling the TS granularity is to provide the correct label (i.e. label for PT=21 has twice the TS number of PT=20). On the other side, if the AUTOpayloadtype is not configured, the RSVP-TE consequent actions in case of TS mismatch need to be defined.

4.2. Tributary Port Number

[RFC4328] supports only the deprecated auto-MSI mode which assumes that the Tributary Port Number is automatically assigned in the transmit direction and not checked in the receive direction.

As described in [G709-V3] and [G798-V3], the OPUk overhead in an OTUk frame contains n (n = the total number of TSs of the ODUk) MSI (Multiplex Structure Identifier) bytes (in the form of multi-frame), each of which is used to indicate the association between tributary port number and tributary slot of the ODUk.

The association between TPN and TS has to be configured by the control plane and checked by the data plane on each side of the link. (Please refer to [OTN-FWK] for further details). As a consequence, the RSVP-TE signaling needs to be extended to support the TPN assignment function.

4.3. Signal type

From a routing perspective, [RFC 4203] allows advertising foundation G.709 (single TS type) without the capability of providing precise information about bandwidth specific allocation. For example, in case of link bundling, dividing the unreserved bandwidth by the MAX LSP bandwidth it is not possible to know the exact number of LSPs at MAX LSP bandwidth size that can be set up. (see example fig. 3)

The lack of spatial allocation heavily impacts the restoration process, because the lack of information of free resources highly increases the number of crank-backs affecting network convergence time.

Moreover actual tools provided by OSPF-TE only allow advertising signal types with fixed bandwidth and implicit hierarchy (e.g. SDH/SONET networks) or variable bandwidth with no hierarchy (e.g. packet switching networks) but do not provide the means for advertising networks with mixed approach (e.g. ODUflex CBR and ODUflex packet).

For example, advertising ODU0 as MIN LSP bandwidth and ODU4 as MAX LSP bandwidth it is not possible to state whether the advertised link supports ODU4 and ODUflex or ODU4, ODU3, ODU2, ODU1, ODU0 and ODUflex. Such ambiguity is not present in SDH networks where the hierarchy is implicit and flexible containers like ODUflex do not exist. The issue could be resolved by declaring 1 ISCD for each signal type actually supported by the link.

Supposing for example to have an equivalent ODU2 unreserved bandwidth in a TE-link (with bundling capability) distributed on 4 ODU1, it would be advertised via the ISCD in this way:

MAX LSP Bw: ODU1

MIN LSP Bw: ODU1

- Maximum Reservable Bandwidth (of the bundle) set to ODU2
- Unreserved Bandwidth (of the bundle) set to ODU2

Moreover with the current IETF solutions, ([RFC4202], [RFC4203]) as soon as no bandwidth is available for a certain signal type it is not advertised into the related ISCD, losing also the related capability until bandwidth is freed.

In conclusion, the OSPF-TE extensions defined in [RFC4203] require a different ISCD per signal type in order to advertise each supported container. This motivates attempting to look for a more optimized solution, without proliferations of the number of ISCD advertised. The OSPF LSA is required to stay within a single IP PDU; fragmentation is not allowed. In a conforming Ethernet environment, this limits the LSA to 1432 bytes (Packet_MTU (1500 Bytes) - IP_Header (20 bytes) - OSPF_Header (28 bytes) - LSA_Header (20 bytes)).

With respect to link bundling, the utilization of the ISCD as it is, would not allow precise advertising of spatial bandwidth allocation information unless using only one component link per TE link.

On the other hand, from a signaling point of view, [RFC4328] describes GMPLS signaling extensions to support the control for G.709 OTNs [G709-V1]. However, [RFC4328] needs to be updated because it

does not provide the means to signal all the new signal types and related mapping and multiplexing functionalities.

4.4. Bit rate and tolerance

In the current traffic parameters signaling, bit rate and tolerance are implicitly defined by the signal type. ODUflex CBR and Packet can have variable bit rates and tolerances (please refer to [OTN-FWK] table 2); it is thus needed to upgrade the signaling traffic parameters so to specify requested bit rates and tolerance values during LSP setup.

4.5. Unreserved Resources

Unreserved resources need to be advertised per priority and per signal type in order to allow the correct functioning of the restoration process. [RFC4203] only allows advertising unreserved resources per priority, this leads not to know how many LSPs of a specific signal type can be restored. As example it is possible to consider the scenario depicted in the following figure.

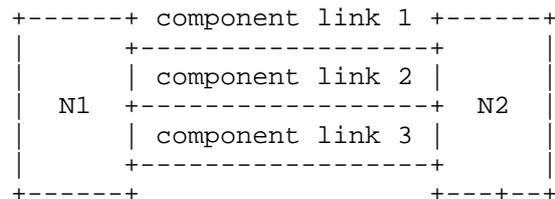


Figure 3: Concurrent path computation

Suppose to have a TE link comprising 3 ODU3 component links with 32TSS available on the first one, 24TSS on the second, 24TSS on the third and supporting ODU2 and ODU3 signal types. The node would advertise a TE link unreserved bandwidth equal to 80 TSS and a MAX LSP bandwidth equal to 32 TSS. In case of restoration the network could try to restore 2 ODU3 (64TSS) in such TE-link while only a single ODU3 can be set up and a crank-back would be originated. In more complex network scenarios the number of crank-backs can be much higher.

4.6. Maximum LSP Bandwidth

Maximum LSP bandwidth is currently advertised in the common part of the ISCD and advertised per priority, while in OTN networks it is only required for ODUflex advertising. This leads to a significant

waste of bits inside each LSA.

4.7. Distinction between terminating and switching capability

The capability advertised by an interface needs further distinction in order to separate termination and switching capabilities. Due to internal constraints and/or limitations, the type of signal being advertised by an interface could be just switched (i.e. forwarded to switching matrix without multiplexing/demultiplexing actions), just terminated (demuxed) or both of them. The following figures help explaining the switching and terminating capabilities.

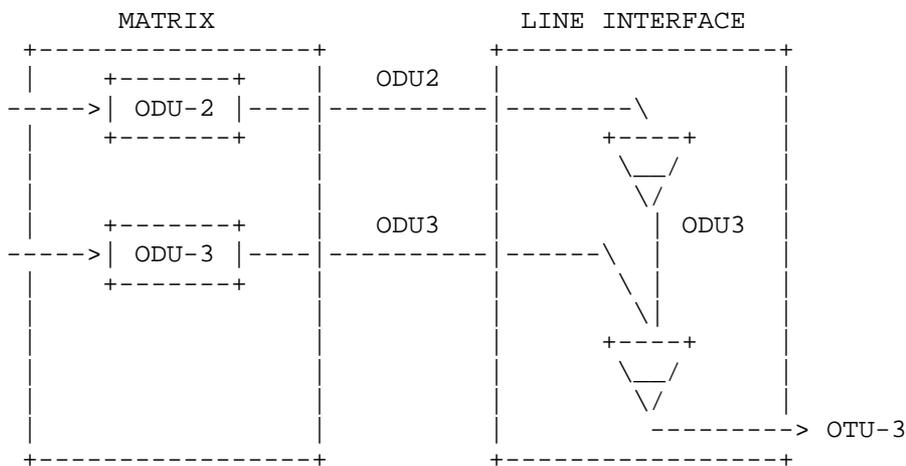


Figure 4: Switching and Terminating capabilities

The figure in the example shows a line interface able to:

- Multiplex an ODU2 coming from the switching matrix into an ODU3 and map it into an OTU3
- Map an ODU3 coming from the switching matrix into an OTU3

In this case the interface bandwidth advertised is ODU2 with switching capability and ODU3 with both switching and terminating capabilities.

This piece of information needs to be advertised together with the related unreserved bandwidth and signal type. As a consequence signaling must have the possibility to setup an LSP allowing the

local selection of resources consistent with the limitations considered during the path computation.

In figures 6 and 7 there are two examples of the need of termination/switching capability differentiation. In both examples all nodes are supposed to support single-stage capability. The figure 6 addresses a scenario in which a failure on link B-C forces node A to calculate another ODU2 LSP path carrying ODU0 service along the nodes B-E-D. Being D a single stage capable node, it is able to extract ODU0 service only from ODU2 interface. Node A has to know that from E to D exists an available OTU2 link from which node D can extract the ODU0 service. This information is required in order to avoid that the OTU3 link is considered in the path computation.

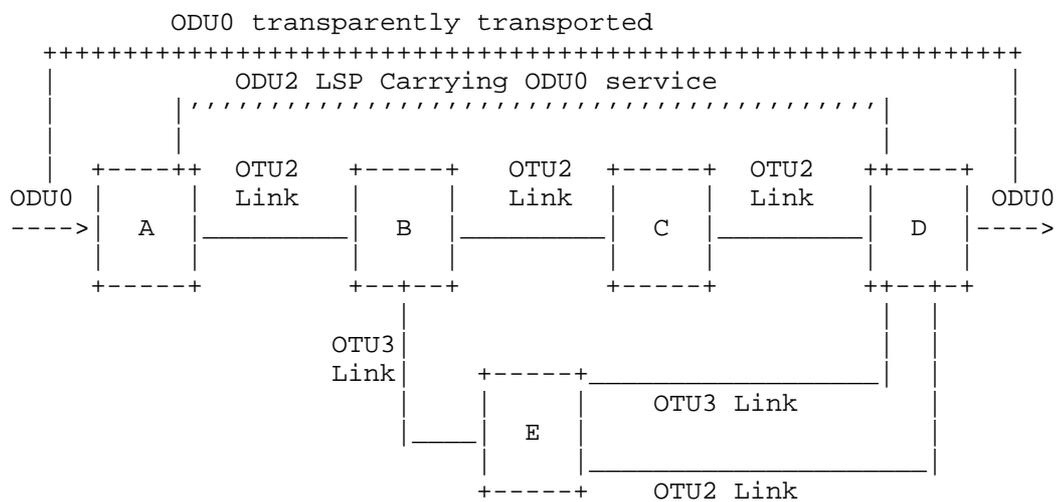


Figure 5: Switching and Terminating capabilities - Example 1

Figure 7 addresses the scenario in which the restoration of the ODU2 LSP (ABCD) is required. The two bundled component links between B and E could be used, but the ODU2 over the OTU2 component link can only be terminated and not switched. This implies that it cannot be used to restore the ODU2 LSP (ABCD). However such ODU2 unreserved bandwidth must be advertised since it can be used for a different ODU2 LSP terminating on E, e.g. (FBE). Node A has to know that the ODU2 capability on the OTU2 link can only be terminated and that the restoration of (ABCD) can only be performed using the ODU2 bandwidth available on the OTU3 link.

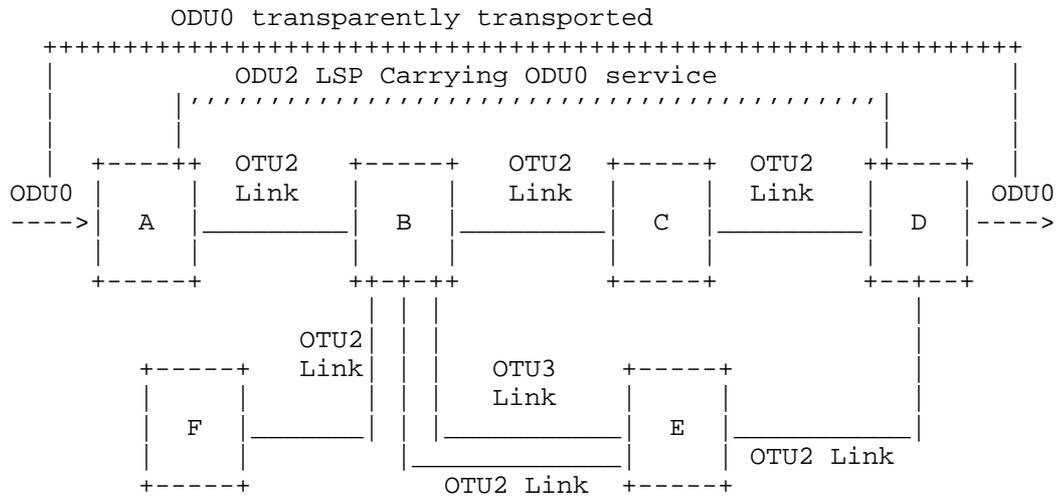


Figure 6: Switching and Terminating capabilities - Example 2

4.8. Priority Support

The IETF foresees that up to eight priorities must be supported and that all of them have to be advertised independently on the number of priorities supported by the implementation. Considering that the advertisement of all the different supported signal types will originate large LSAs, it is advised to advertise only the information related to the really supported priorities.

4.9. Multi-stage multiplexing

With reference to the [OTN-FWK], introduction of multi-stage multiplexing implies the advertisement of cascaded adaptation capabilities together with the matrix access constraints. The structure defined by IETF for the advertisement of adaptation capabilities is ISCD/IACD as in [RFC4202] and [RFC5339]. Modifications to ISCD/IACD, if needed, have to be addressed in the related encoding documents.

4.10. Generalized Label

The ODUk label format defined in [RFC4328] could be updated to support new signal types defined in [G709-V3] but would hardly be further enhanced to support possible new signal types.

Furthermore such label format may have scalability issues due to the

high number of labels needed when signaling large LSPs. For example, when an ODU3 is mapped into an ODU4 with 1.25G tributary slots, it would require the utilization of thirty-one labels ($31*4*8=992$ bits) to be allocated while an ODUflex into an ODU4 may need up to eighty labels ($80*4*8=2560$ bits).

A new flexible and scalable ODUk label format needs to be defined.

5. Security Considerations

TBD

6. IANA Considerations

TBD

7. Contributors

Jonathan Sadler, Tellabs

EMail: jonathan.sadler@tellabs.com

8. Acknowledgements

The authors would like to thank Eve Varma and Sergio Lanzone for their precious collaboration and review.

9. References

9.1. Normative References

[HIER-BIS]

K.Shiomoto, A.Farrel, "Procedure for Dynamically Signaled Hierarchical Label Switched Paths", work in progress draft-ietf-lsp-hierarchy-bis-08, February 2010.

[OTN-OSPF]

D.Ceccarelli, D.Caviglia, F.Zhang, D.Li, Y.Xu, P.Grandi, S.Belotti, "Traffic Engineering Extensions to OSPF for Generalized MPLS (GMPLS) Control of Evolutive G.709 OTN Networks", work in progress draft-cceccarelli-ccamp-gmpls-ospf-g709-03, August 2010.

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC3630] Katz, D., Kompella, K., and D. Yeung, "Traffic Engineering (TE) Extensions to OSPF Version 2", RFC 3630, September 2003.
- [RFC4202] Kompella, K. and Y. Rekhter, "Routing Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4202, October 2005.
- [RFC4203] Kompella, K. and Y. Rekhter, "OSPF Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4203, October 2005.
- [RFC4328] Papadimitriou, D., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Extensions for G.709 Optical Transport Networks Control", RFC 4328, January 2006.
- [RFC5250] Berger, L., Bryskin, I., Zinin, A., and R. Coltun, "The OSPF Opaque LSA Option", RFC 5250, July 2008.
- [RFC5339] Le Roux, JL. and D. Papadimitriou, "Evaluation of Existing GMPLS Protocols against Multi-Layer and Multi-Region Networks (MLN/MRN)", RFC 5339, September 2008.

9.2. Informative References

- [G.709-v1] ITU-T, "Interface for the Optical Transport Network (OTN)", G.709 Recommendation (and Amendment 1), February 2001.
- [G.709-v2] ITU-T, "Interface for the Optical Transport Network (OTN)", G.709 Recommendation (and Amendment 1), March 2003.
- [G.709-v3] ITU-T, "Rec G.709, version 3", approved by ITU-T on December 2009.

[G.872-am2]

ITU-T, "Amendment 2 of G.872 Architecture of optical transport networks for consent", consented by ITU-T on June 2010.

[OTN-FWK]

F.Zhang, D.Li, H.Li, S.Belotti, "Framework for GMPLS and PCE Control of G.709 Optical Transport Networks", work in progress draft-ietf-ccamp-gmpls-g709-framework-00, April 2010.

Authors' Addresses

Sergio Belotti
Alcatel-Lucent
Via Trento, 30
Vimercate
Italy

Email: sergio.belotti@alcatel-lucent.com

Pietro Vittorio Grandi
Alcatel-Lucent
Via Trento, 30
Vimercate
Italy

Email: pietro_vittorio.grandi@alcatel-lucent.com

Daniele Ceccarelli
Ericsson
Via A. Negrone 1/A
Genova - Sestri Ponente
Italy

Email: daniele.ceccarelli@ericsson.com

Diego Caviglia
Ericsson
Via A. Negrone 1/A
Genova - Sestri Ponente
Italy

Email: diego.caviglia@ericsson.com

Fatai Zhang
Huawei Technologies
F3-5-B R&D Center, Huawei Base
Shenzhen 518129 P.R.China Bantian, Longgang District
Phone: +86-755-28972912

Email: zhangfatai@huawei.com

Dan Li
Huawei Technologies
F3-5-B R&D Center, Huawei Base
Shenzhen 518129 P.R.China Bantian, Longgang District
Phone: +86-755-28973237

Email: danli@huawei.com

CCAMP Working Group
Internet-Draft
Intended status: Standards Track
Expires: September 15, 2011

D. Ceccarelli
D. Caviglia
Ericsson
F. Zhang
D. Li
Huawei Technologies
Y. Xu
CATR
S. Belotti
P. Grandi
Alcatel-Lucent
R. Rao
Infinera Corporation
J. Drake
Juniper
March 14, 2011

Traffic Engineering Extensions to OSPF for Generalized MPLS (GMPLS)
Control of Evolving G.709 OTN Networks
draft-ceccarelli-ccamp-gmpls-ospf-g709-05

Abstract

The recent revision of ITU-T Recommendation G.709 [G709-V3] has introduced new fixed and flexible ODU containers, enabling optimized support for an increasingly abundant service mix.

This document describes OSPF routing protocol extensions to support Generalized MPLS (GMPLS) control of all currently defined ODU containers, in support of both sub-lambda and lambda level routing granularity.

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on September 15, 2011.

Copyright Notice

Copyright (c) 2011 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1.	Introduction	4
1.1.	Terminology	4
2.	OSPF Extensions	4
2.1.	ISCD extensions	5
2.1.1.	Unreserved Bandwidth sub-sub-TLV	5
2.2.	IMCD - Interface Multiplexing Capability Descriptor	6
2.2.1.	IMCD Bandwidth sub-sub-TLV	7
3.	Procedures	10
3.1.	ODUk advertisement	10
3.2.	ODUj advertisement	10
3.3.	Link Bundling	10
4.	Optimization considerations	11
4.1.	Efficient priorities advertisement	11
4.2.	Efficient bandwidth encoding	12
5.	Example	13
6.	Compatibility	13
7.	Security Considerations	13
8.	IANA Considerations	13
9.	Contributors	13
10.	Acknowledgements	15
11.	References	15
11.1.	Normative References	15
11.2.	Informative References	16
	Authors' Addresses	16

1. Introduction

G.709 OTN [G709-V3] includes new fixed and flexible ODU containers, two types of Tributary Slots (i.e., 1.25Gbps and 2.5Gbps), and supports various multiplexing relationships (e.g., ODUj multiplexed into ODUk (j<k)), two different tributary slots for ODUk (K=1, 2, 3) and ODUflex service type, which is being standardized in ITU-T. In order to present this information in the routing process, this document provides OTN technology specific encoding for OSPF-TE.

For a short overview of OTN evolution and implications of OTN requirements on GMPLS routing please refer to [OTN-FWK]. The information model and an evaluation against the current solution are provided in [OTN-INFO].

The routing information for Optical Channel Layer (OCh) (i.e., wavelength) is out of the scope of this document. Please refer to [WSON-Frame] for further information.

1.1. Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

2. OSPF Extensions

In terms of GMPLS based OTN networks, each OTUk can be viewed as a component link, and each component link can carry one or more types of ODUj (j<k).

Each TE LSA can carry a top-level link TLV with several nested sub-TLVs to describe different attributes of a TE link. Two top-level TLVs are defined in [RFC 3630]. (1) The Router Address TLV (referred to as the Node TLV) and (2) the TE link TLV. One or more sub-TLVs can be nested into the two top-level TLVs. The sub-TLV set for the two top-level TLVs are also defined in [RFC 3630] and [RFC 4203].

As discussed in [OTN-FWK] and [OTN-INFO], the OSPF-TE must be extended so to be able to advertise the termination and switching capabilities related to each different ODUj and ODUk and the advertisement of related multiplexing capabilities. This document defines:

- New Switching Capability and Encoding Type values for the ISCD with related new sub-sub-TLVs

- A new Link type sub-TLV called IMCD with related sub-sub-TLVs

In the following we will use ODUj to indicate a service type that is multiplexed into an higher order ODU and ODUk to indicate the layer mapped into the OTUk. Moreover ODUj(S) and ODUk(S) are used to indicate ODUj and ODUk with switching capability only, ODUj(T) and ODUk(T) to indicate ODUj and ODUk with terminating capability only and ODUj(T,S) and ODUk(T,S) to indicate ODUj and ODUk that can be both switched or terminated. Moreover the ODUj->ODUk format is used to indicate the ODUj into ODUk multiplexing capability.

The advertisement of available bandwidth, max LSP bandwidth and multiplexing capabilities is performed as follows:

- ODUk(S) advertised in the ISCD
- ODUk(T) advertised in the IMCD (Interface Multiplexing Capability Descriptor)
- ODUk(T,S) advertised both in the ISCD and IMCD
- ODUj(*) and related multiplexing hierarchy advertised in the IMCD

The IMCD and new sub-sub-TLVs format are illustrated in the following sections.

2.1. ISCD extensions

This document defines a new Switching Capability value

Value	Type
-----	-----
101	OTN-TDM

while the values of the Encoding Type field are the ones defined in [RFC4328].

2.1.1. Unreserved Bandwidth sub-sub-TLV

The Unreserved bandwidth sub-sub-TLV is included into the SCSI (Switching Capability Specific Information) of the ISCD. It is used for the advertisement of ODUk(S) unreserved bandwidth. Please note that there is no need to advertise MAX LSP bandwidth within the ISCD because the only container with variable bandwidth (ODUflex) can be

an ODUj only. The format of the Unreserved Bandwidth sub-sub-TLV is shown in the following figure.

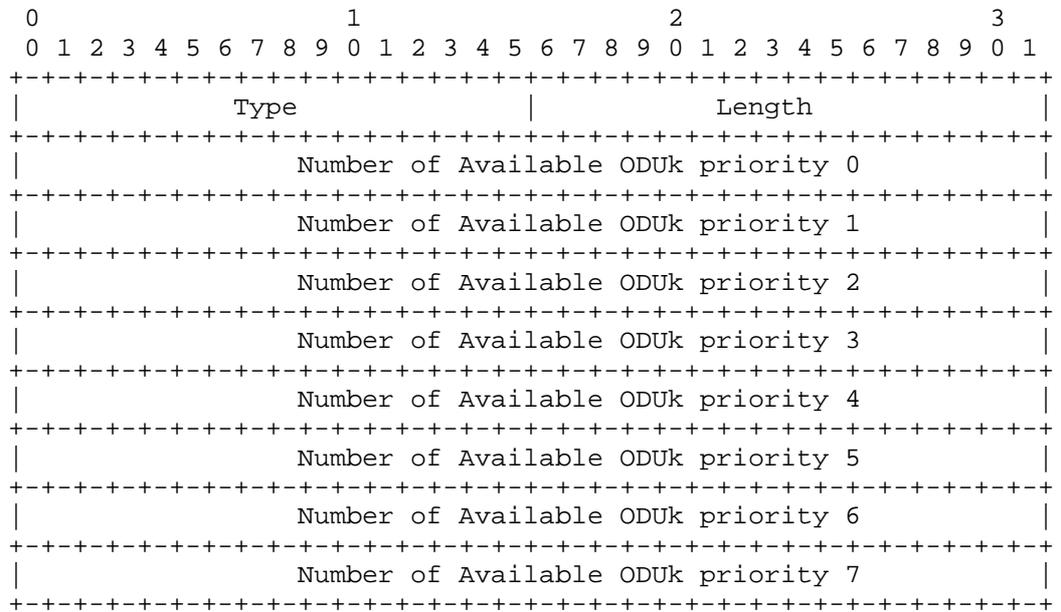


Figure 1: Unreserved Bandwidth sub-sub-TLV format

- Type: Type = 1 indicates Unreserved Bandwidth sub-sub-TLV. i.e. advertising Unreserved Bandwidth for ODUk containers.
- Lengths: Expressed in Bytes and aligned to 32bits.
- Number of Available ODUk at Priority Px: Indicates the number of Available ODUk al Priority Px that can be Switched in the advertised TE Link.

2.2. IMCD - Interface Multiplexing Capability Descriptor

The Interface Multiplexing Capability Descriptor (IMCD) is a new Link type sub-TLV (Type TBA by IANA) and is used for the advertisement of:

- ODUk Termination Unreserved Bandwidth
- ODUj Switching and Termination Unreserved Bandwidth with related muxing hierarchy

- ODUj Switching and Termination MAX LSP Bandwidth with related muxing hierarchy

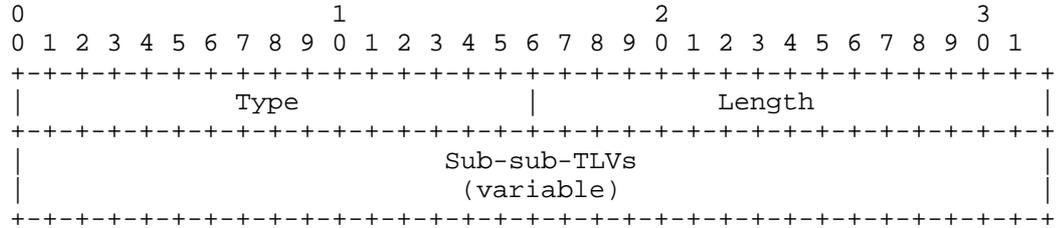


Figure 2: IMCD sub-TLV format

- Type: To be assigned by IANA.
- Length: Expressed in Bytes and aligned to 32bits.
- Sub-sub-TLVs: The body of the IMCD can include a variable number of sub-sub-TLVs.

2.2.1. IMCD Bandwidth sub-sub-TLV

This document defines three types of IMCD Unreserved Bandwidth sub-sub-TLVs:

- Type = 1, advertising the Unreserved Bandwidth of fixed bandwidth containers (e.g. ODU2,ODU3)
- Type = 2, advertising the Unreserved Bandwidth of variable bandwidth containers (e.g. ODUFlex)
- Type = 3, advertising the MAX LSP Bandwidth of variable bandwidth containers (e.g. ODUFlex)

The format is shown in figure below:

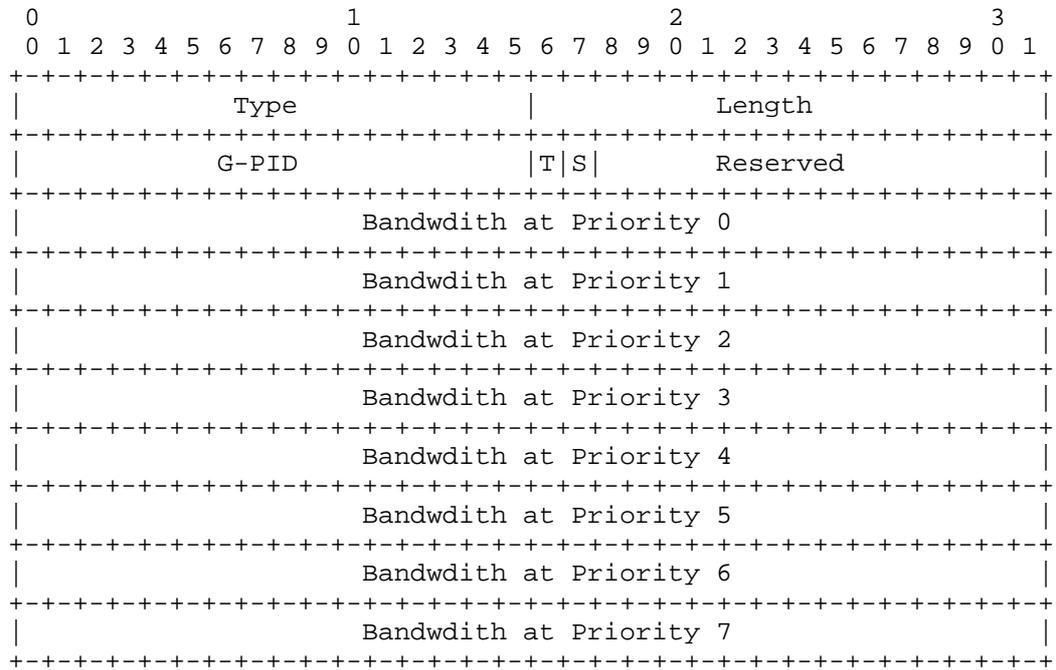


Figure 3: IMCD Bandwidth sub-sub-TLV format

The rest of the sub-sub-TLV fields is defined as follows:

- Length: Expressed in Bytes and aligned to 32bits.
- G-PID: Defines new values in addition to those already defined in RFC3471] and identifies the muxing hierarchy supported by a component link.

Value	G-PID
-----	-----
100	ODU1
101	ODU2
102	ODU3
103	ODU4
104	ODU1->ODU0
105	ODU2->ODU0
106	ODU2->ODU1
107	ODU2->ODU1->ODU0
108	ODU2->ODUflex
109	ODU3->ODU0
110	ODU3->ODU1
111	ODU3->ODU1->ODU0
112	ODU3->ODU2
113	ODU3->ODU2->ODU0
114	ODU3->ODU2->ODU1
115	ODU3->ODU2->ODU1->ODU0
116	ODU3->ODU2->ODUflex
117	ODU3->ODUflex
118	ODU3->ODU2e
119	ODU4->ODU0
120	ODU4->ODU1
121	ODU4->ODU1->ODU0
122	ODU4->ODU2
123	ODU4->ODU2->ODU0
124	ODU4->ODU2->ODU1
125	ODU4->ODU2->ODU1->ODU0
126	ODU4->ODU2->ODUflex
127	ODU4->ODU3
128	ODU4->ODU3->ODU0
129	ODU4->ODU3->ODU1
130	ODU4->ODU3->ODU1->ODU0
131	ODU4->ODU3->ODU2
132	ODU4->ODU3->ODU2->ODU0
133	ODU4->ODU3->ODU2->ODU1
134	ODU4->ODU3->ODU2->ODU1->ODU0
135	ODU4->ODU3->ODU2->ODUflex
136	ODU4->ODU3->ODUflex
137	ODU4->ODU3->ODU2e
138	ODU4->ODUflex
139	ODU4->ODU2e

- Flags: T,S flags are used to indicate Termination and Switching capabilities of the ODUj containers and MUST be set to 0 and ignored in case of ODUk.

- Unreserved Bandwidth: Indicates the Unreserved bandwidth of the container being advertised. It MUST be expressed in Number of Available containers in case of fixed containers (i.e. Type=1) and in IEEE floating point in case of variable bandwidth containers (i.e. Type=2).

3. Procedures

3.1. ODUk advertisement

The advertisement of ODUk is performed via ISCD, IMCD or both, depending on the terminating and switching capabilities of the given ODUk. In case of ODUk(S), its unreserved bandwidth MUST be advertised by means of the Unreserved Bandwidth sub-sub-TLV included into the ISCD. One ISCD for each ODUk(S) is advertised.

On the other hand, an ODUk(T) MUST be advertised via the Bandwidth sub-sub-TLV included into the IMCD. Multiple ODUk(T) MAY be advertised withing the same IMCD.

In the case of ODUk(T,S), the advertisement of such ODUk MUST be present both in the ISCD and the IMCD.

3.2. ODUj advertisement

The advertisement of ODUj MUST be performed via IMCD only and its terminating and switching capabilities are specified by the flags (T and S) of the Bandwidth sub-sub-TLV.

Unreserved and MAX LSP bandwidth are advertised by means of different types of the Bandwdith sub-sub-TLV as shown in Section 2.2.1.

The advertisement of ODUj(S) is not performed via ISCD because the ISCD does not provide the means for distinguishing between ODUj and ODUk and this would prevent the bundling of interfaces at different line rates.

3.3. Link Bundling

It is possible to bundle different interfaces with different line rates, muxing hierarchies and termination/switching capabilities except the case in which the end nodes of a TE link have ODUk at the same line rate but different terminating/switching capabilities or muxing hierarchies.

An example of this exemption is shown in figure below:

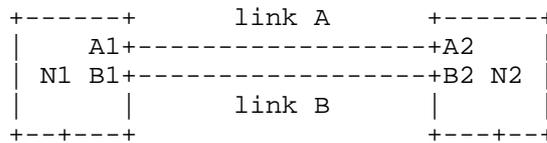


Figure 4: Bundling not allowed

In case link A has interface A1 supporting ODUk (T) and A2 supporting ODUk(S) and link B with interface B1 supporting the same ODUk(S) and B2 supporting ODUk(T), the bundling of the two links in a single TE link would give the false information of ODUk(S) availability at the ends of the TE link. Hence, link A and link B cannot be bundled into the same TE link.

4. Optimization considerations

Optimization considerations are extremely important not only under the scalability point of view but also considering the requirement that an LSA (Link State Advertisement) cannot be fragmented into multiple IP packets. Considering that in a conforming Ethernet environment the Frame_MTU is 1500 bytes, the amount of available bandwidth for the LSA payload is 1432 byte. (1500 byte - (IP header 20bytes + OSPF header 28bytes + LSA header 20 bytes)). IP packets fragmentation is not suggested in IPv4-IPv6 as it has a big impact on computation efficiency and CPU processing time.

4.1. Efficient priorities advertisement

Actual GMPLS definition foresees the advertisement of all the eight possible priorities. This is an inefficient approach in terms of bandwidth utilization in those cases where not all the priorities are supported. A possible enhancement consists on inserting an 8 bits bitmask identifying the supported priorities being advertised.

The bitmask can be applied to the Unreserved Bandwidth sub-sub-TLV of the ISCD and to the Bandwidth sub-sub-TLV of the IMCD. The following figure shows an example of bitmask application to the Bandwidth sub-sub-TLV in the advertisement of the MAX LSP bandwidth of a given service type:

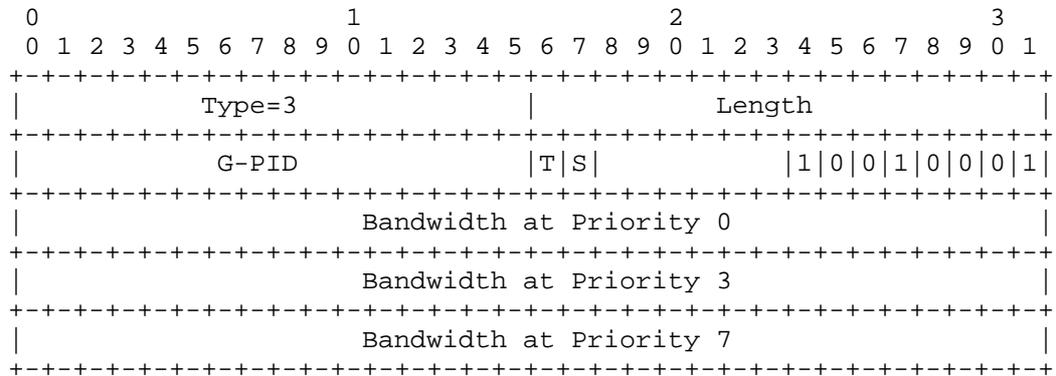


Figure 5: Efficient priorities advertisement

Only priorities 0,3 and 7 are supported and hence advertised. In this simple example the amount of bytes saved is 20, but in a scenario with traffic cards supporting a high number of service types and muxing hierarchies, the amount of saved bandwidth is meaningful.

4.2. Efficient bandwidth encoding

When a fixed bandwidth service type is advertised, the number of available service types is used as measurement units. This can be easily advertised via a 16 bits field instead of 32 bits (needed for IEEE floating point encoding). When the number of supported priorities is odd, padding to multiples of 32 bits is required. The following figure shows an example of Unreserved Bandwidth advertisement via Bandwidth sub-sub-TLV with 3 priorities supported and padding.

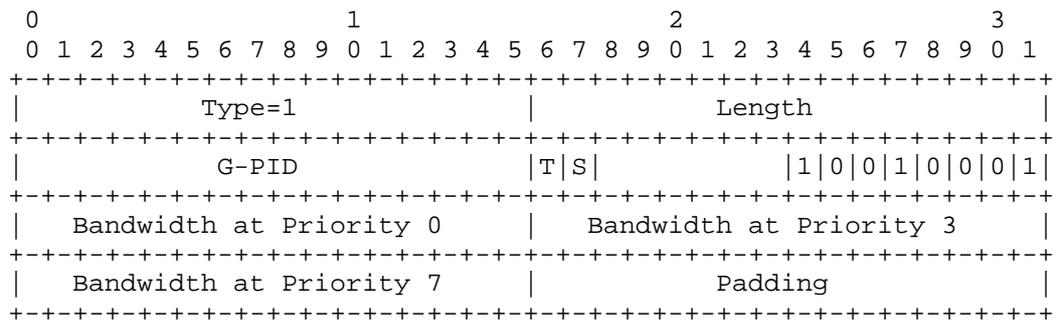


Figure 6: Efficient bandwidth encoding

5. Example

TBD

6. Compatibility

Backwards compatibility with implementations based on [RFC4328] can be achieved advertising the [RFC4328] based ISCDs in addition to the ISCD defined in this document.

7. Security Considerations

This document specifies the contents of Opaque LSAs in OSPFv2. As Opaque LSAs are not used for SPF computation or normal routing, the extensions specified here have no direct effect on IP routing. Tampering with GMPLS TE LSAs may have an effect on the underlying transport (optical and/or SONET-SDH) network. [RFC3630] suggests mechanisms such as [RFC2154] to protect the transmission of this information, and those or other mechanisms should be used to secure and/or authenticate the information carried in the Opaque LSAs.

8. IANA Considerations

TBD

9. Contributors

Xiaobing Zi, Huawei Technologies

Email: zixiaobing@huawei.com

Francesco Fondelli, Ericsson

Email: francesco.fondelli@ericsson.com

Marco Corsi, Altran Italia

EMail: marco.corsi@altran.it

Eve Varma, Alcatel-Lucent

EMail: eve.varma@alcatel-lucent.com

Jonathan Sadler, Tellabs

EMail: jonathan.sadler@tellabs.com

Lyndon Ong, Ciena

EMail: lyong@ciena.com

Ashok Kunjidhpatham

akunjidhpatham@infinera.com

Snigdho Bardalai

sbardalai@infinera.com

Khuzema Pithewan

kpithewan@infinera.com

Steve Balls

Steve.Balls@metaswitch.com

Xihua Fu

fu.xihua@zte.com.cn

10. Acknowledgements

The authors would like to thank Eric Gray for his precious comments and advices.

11. References

11.1. Normative References

- [MLN-EXT] D.Papadimitriou, M.Vigoureux, K.Shiomoto, D.Brungard, J.Le Roux, "Generalized Multi-Protocol Extensions for Multi-Layer and Multi-Region Network (MLN/MRN)", February 2010.
- [OTN-FWK] F.Zhang, D.Li, H.Li, S.Belotti, D.Ceccarelli, "Framework for GMPLS and PCE Control of G.709 Optical Transport networks, work in progress draft-ietf-ccamp-gmpls-g709-framework-02", July 2010.
- [OTN-INFO] S.Belotti, P.Grandi, D.Ceccarelli, D.Caviglia, F.Zhang, D.Li, "Information model for G.709 Optical Transport Networks (OTN), work in progress draft-bddg-ccamp-otn-g709-info-model-01", October 2010.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC2154] Murphy, S., Badger, M., and B. Wellington, "OSPF with Digital Signatures", RFC 2154, June 1997.
- [RFC2370] Coltun, R., "The OSPF Opaque LSA Option", RFC 2370, July 1998.
- [RFC3630] Katz, D., Kompella, K., and D. Yeung, "Traffic Engineering (TE) Extensions to OSPF Version 2", RFC 3630, September 2003.
- [RFC4201] Kompella, K., Rekhter, Y., and L. Berger, "Link Bundling in MPLS Traffic Engineering (TE)", RFC 4201, October 2005.
- [RFC4202] Kompella, K. and Y. Rekhter, "Routing Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4202, October 2005.
- [RFC4203] Kompella, K. and Y. Rekhter, "OSPF Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4203, October 2005.

- [RFC5250] Berger, L., Bryskin, I., Zinin, A., and R. Coltun, "The OSPF Opaque LSA Option", RFC 5250, July 2008.
- [RFC5339] Le Roux, JL. and D. Papadimitriou, "Evaluation of Existing GMPLS Protocols against Multi-Layer and Multi-Region Networks (MLN/MRN)", RFC 5339, September 2008.

11.2. Informative References

- [G.709] ITU-T, "Interface for the Optical Transport Network (OTN)", G.709 Recommendation (and Amendment 1), February 2001.
- [G.709-v3] ITU-T, "Draft revised G.709, version 3", consented by ITU-T on Oct 2009.
- [Gsup43] ITU-T, "Proposed revision of G.sup43 (for agreement)", December 2008.

Authors' Addresses

Daniele Ceccarelli
Ericsson
Via A. Negrone 1/A
Genova - Sestri Ponente
Italy

Email: daniele.ceccarelli@ericsson.com

Diego Caviglia
Ericsson
Via A. Negrone 1/A
Genova - Sestri Ponente
Italy

Email: diego.caviglia@ericsson.com

Fatai Zhang
Huawei Technologies
F3-5-B R&D Center, Huawei Base
Shenzhen 518129 P.R.China Bantian, Longgang District
Phone: +86-755-28972912

Email: zhangfatai@huawei.com

Dan Li
Huawei Technologies
F3-5-B R&D Center, Huawei Base
Shenzhen 518129 P.R.China Bantian, Longgang District
Phone: +86-755-28973237

Email: danli@huawei.com

Yunbin Xu
CATR
11 Yue Tan Nan Jie
Beijing
P.R.China

Email: xuyunbin@mail.ritt.com.cn

Sergio Belotti
Alcatel-Lucent
Via Trento, 30
Vimercate
Italy

Email: sergio.belotti@alcatel-lucent.com

Pietro Vittorio Grandi
Alcatel-Lucent
Via Trento, 30
Vimercate
Italy

Email: pietro_vittorio.grandi@alcatel-lucent.com

Rajan Rao
Infinera Corporation
169, Java Drive
Sunnyvale, CA-94089
USA

Email: rrao@infinera.com

John E Drake
Juniper

Email: jdrake@juniper.net

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: April 28, 2011

X. Fu
Q. Wang
Y. Bao
ZTE Corporation
R. Jing
X. Huo
China Telecom
October 25, 2010

RSVP-TE Signaling Extension for Explicit Control of LSP Boundary in A
GMPLS-Based Multi-Region and Multi-Layer Networks (MRN/MLN)
draft-fuxh-ccamp-boundary-explicit-control-ext-01

Abstract

[RFC5212] defines a Multi-Region and Multi-Layer Networks (MRN/MLN). [RFC4206] introduces a region boundary determination algorithm and a Hierarchy LSP (H-LSP) creation method. However, in some scenarios, some attributes have to be attached with the boundary nodes in order to explicit control the hierarchy LSP creation. This document extends GMPLS signaling protocol for the requirement of explicit control the hierarchy LSP creation.

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on April 28, 2011.

Copyright Notice

Copyright (c) 2010 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents

(<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	3
1.1. Conventions Used In This Document	3
2. Requirement of Explicit Control of Hierarchy LSP Creation . . .	3
2.1. Selection of Server Layer/Sub-Layer	3
2.2. Selection/Creation of FA-LSP based on characteristics of server layer	4
2.3. Configuration of Multi Stages Multiplexing Hierarchy . . .	5
3. Explicit Route Boundary Object (ERBO)	6
3.1. Server Layer/Sub-Layer Attributes TLV	8
3.2. Multiplexing Hierarchy Attribute TLV	9
3.3. Latency Attribute TLV	10
4. Signaling Procedure	11
5. Security Considerations	11
6. IANA Considerations	12
7. References	12
7.1. Normative References	12
7.2. Informative References	12
Authors' Addresses	13

1. Introduction

[RFC5212] defines a Multi-Region and Multi-Layer Networks (MRN/MLN). [RFC4206] introduces a region boundary determination algorithm and a Hierarchy LSP (H-LSP) creation method. However, in some scenarios, some attributes have to be attached with the boundary nodes in order to explicitly control the hierarchy LSP creation. This document extends GMPLS signaling protocol for the requirement of explicit control the hierarchy LSP creation.

1.1. Conventions Used In This Document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

2. Requirement of Explicit Control of Hierarchy LSP Creation

2.1. Selection of Server Layer/Sub-Layer

[RFC4206] describes a region boundary determination algorithm and a hierarchical LSP creation method. This region boundary determination algorithm and LSP creation method are well applied to Multi-Region Network. However it isn't fully applied to Multi-Layer Network. In the following figure, three LSPs belong to the same TDM region and different latyers, but the sub-layer boundary node could not determine which lower layer should be triggered according to the region boundary determination algorithm defined in [RFC4206]. Thus the higher layer (VC4 in figure 1) signaling can't trigger the lower layer (STM-N in figure 1) LSP creation. It needs to explicitly describe which sub-layer should be triggered in the signaling message.

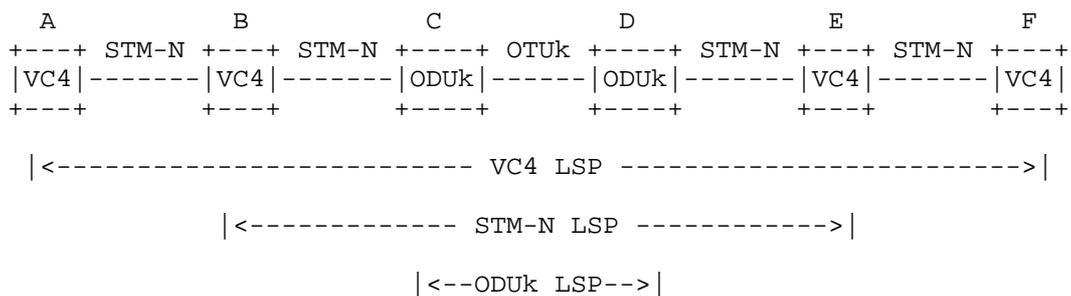


Figure 1: Example of Server Layer/Sub-Layer Selection

2.2. Selection/Creation of FA-LSP based on characteristics of server layer

ITU-T G.800 defines Composite Link. Individual component links in a composite link may be supported by different transport technologies such as OTN, MPLS-TP or SDH/SONET. Even if the transport technology implementing the component links is identical, the characteristics (e.g., latency) of the component links may differ. Operator may prefer its traffic to be transported over a specific transport technology server layer. Further more, operator may prefer its traffic to be transported over a specific transport technology component link with some specific characteristics (e.g., latency). So it desires to explicitly control the component link selection based on the attributes (e.g., switching capability and latency) attached with the boundary nodes during the signaling.

Latency is a key requirement for service provider. Restoration and/or protection can impact "provisioned" latency. The key driver for this is stock/commodity trading applications that use data base mirroring. A few delicacy can impact a transaction. Therefore latency and latency SLA is one of the key parameters that these "high value" customers use to select a private pipe line provider. So it desires to explicitly convey latency SLA to the boundary nodes where the hierarchy LSP will be triggered.

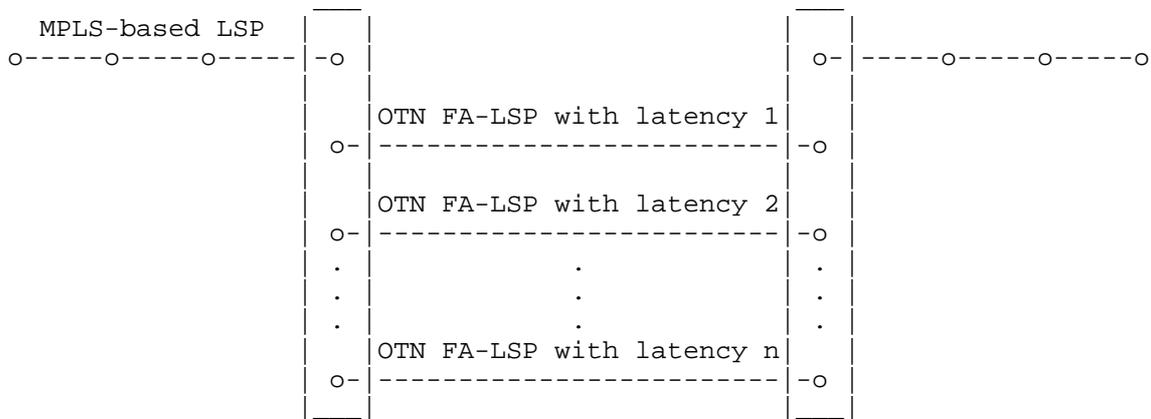


Figure 2: Example of FA-LSP Selection/Creation based on Latency

In Figure 2, a LSP traffic is over a composite link whose component links with different latency characteristic are supported by OTN. In order to meet the latency SLA, it needs to explicitly limit the

latency between boundary nodes to create an OTN tunnel.

2.3. Configuration of Multi Stages Multipelxing Hierarchy

In Figure 3, node B and C in the OTN network are connected to 2.5G TS network by two OTU3 link. They can support flexible multi stages multiplexing hierarchies. There are two multi stages multiplexing hierarchies for ODU0 being mapped into OTU3 link in B and C of Figure 1 (i.e., ODU0-ODU1-ODU3 and ODU0-ODU2-ODU3). So path computation entity has to determine which kind of multi stages multiplexing hierarchies should be used for the end-to-end ODU0 service and the type of tunnel (FA-LSP). In Figure 3, if path computation entity select the ODU0-ODU2-ODU3 multi stages multiplexing hierarch in Node B and C for one end-to-end ODU0 service from A to Z, there has to be an ODU2 tunnel between B and C. The selection of multi stages multiplexing hierarchies is based on the operator policy and the equipment capability. How to select the multiplexing hierarchies is the internal behavior of path computation entity.

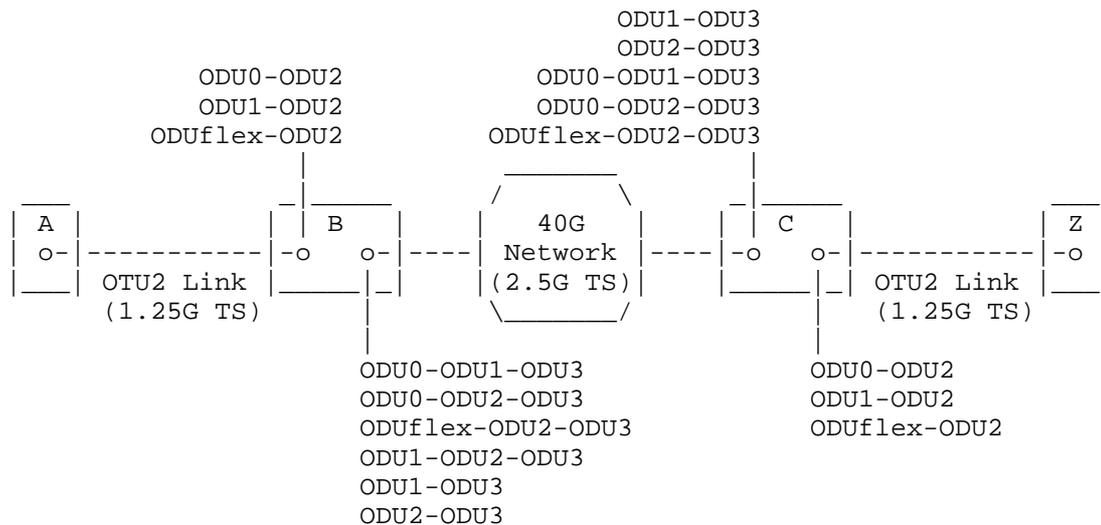


Figure 3 Example of Multi-Stages Multiplexing Hierarchy Selection

If path computation entity select the ODU0-ODU2-ODU3 for ODU0 being mapped into OTU3 Link, the multi stages multiplexing hierarchy has to be carried in signaling message to node B and C. After B receives the signaling message, it will triggered a creation of and ODU2 FA-LSP base on [RFC4206] and the selection of multi stages multiplexing hierarchy. Node B and C must config this kind of multi stages multiplexing hierarchy (i.e., ODU0-ODU2-ODU3) to its data plane. So

data plane can multiplex and demultiplex the ODU0 signal from/to ODU3 for a special end-to-end ODU0 service in terms of the control plane's configuration.

In Figure 4, the switching capability (e.g., TDM), switching granularity (i.e., ODU3) and multi stages multiplexing hierarchy (ODU0-ODU1-ODU3-ODU4) must be specified during signaling. Because the switching capability (TDM) and switching granularity (ODU3) information is not enough for data plane to know ODU0 is mapped into ODU3 tunnel by ODU0-ODU1-ODU3 then ODU4. In order to explicit specify multi stages multiplexing hierarchy, the switching capability, switching granularity and multi stages multiplexing hierarchy (ODU0-ODU1-ODU3) must be carried in the signaling message.

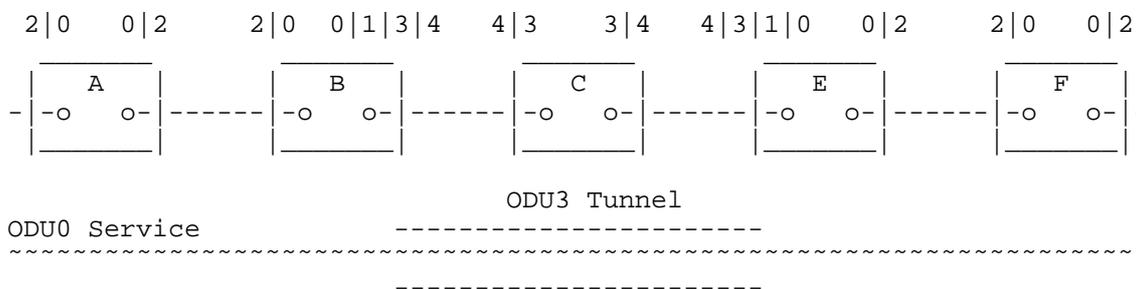


Figure 4 Example of Multi-Stages Multiplexing Hierarchy Selection

3. Explicit Route Boundary Object (ERBO)

In order to explicitly control hierarchy LSP creation, this document introduce a new object (ERBO- Explicit Route Boundary Object) carried in RSVP-TE message. The format of ERBO object is the same as ERO. The ERBO including the region boundaries information and some specific attributes (e.g., latency) can be carried in Path message. One pairs or multiple pairs of nodes within the ERBO can belong to the same layer or different layers.

This document introduce a new sub-object (BOUNDARY_ATTRIBUTES) carry the attributes of the associated hop specified in the ERBO. It allows the specification and reporting of attributes relevant to a particular hop of the signaled LSP. It follows an IPv4 or IPv6 prefix or unnumbered Interface ID sub-object in ERBO. A list of attribute TLV can be inserted into ERBO.

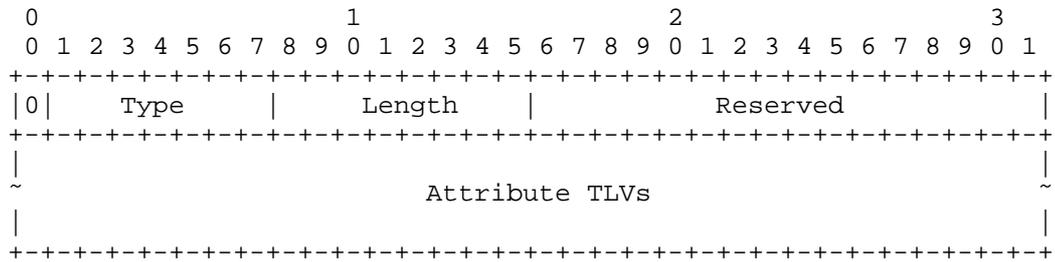


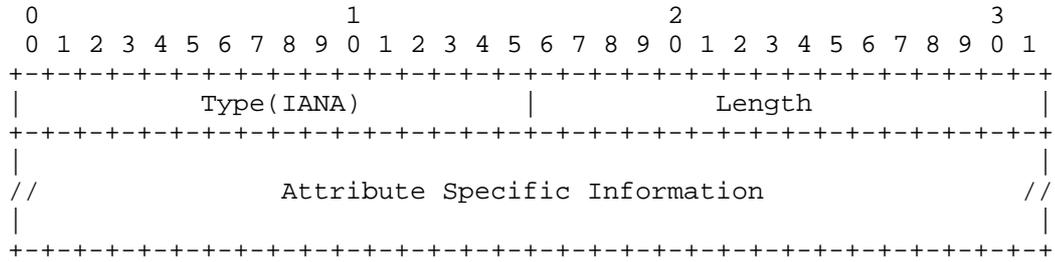
Figure 5 Format of BOUNDARY_ATTRIBUTES

- This field indicates different attribute TLV sub-objects.
- The total length of the sub-object in bytes, including the Type and Length fields. The value of this field is always a multiple of 4.
- Attribute TLVs: This field carries different TLV according to the Type filed.

A list of attributes TLV can be inserted into ERBO. These attributes may represent the following information. It can be further extended to carry other specific requirement in the future.

- Server Layer (e.g., PSC, L2SC, TDM, LSC, FSC) or Sub-Layer (e.g., VC4, VC11, VC4-4c, VC4-16c, VC4-64c, ODU0, ODU1, ODU2, ODU3, ODU4) used for boundary node to trigger one specific corresponding server layer or Sub-Layer FA-LSP creation. The region boundary node may support multiple interface switching capabilities and multiple switching granularities. It is very useful to indicate which server layer and/or sub-layer to be used at the region boundary node.
- Multiplexing hierarchy (e.g., ODU0-ODU1-ODU3-ODU4) used for boundary node to configure it to the data plane and trigger one specific corresponding tunnel creation.
- Server Layer and/or Sub-Layer's LSP Latency SLA (e.g., minimum latency value, maximum latency value, average latency value and latency variation value). Boundary node select a FA or create a FA-LSP based on the latency limitation.

The format of the Attributes TLV is as follows:

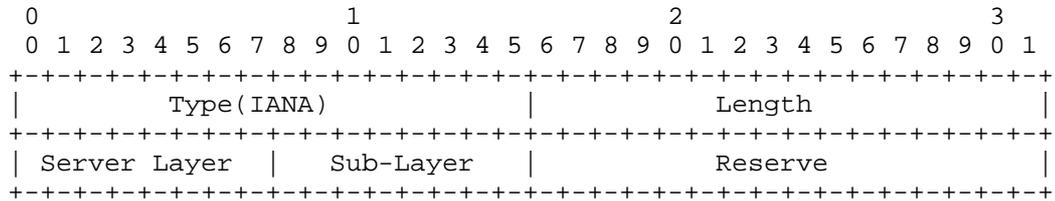


The following types are supported.

Type	Information
TBD	server layer/sub-layer
TBD	server layer/sub-layer characteristics (e.g., latency)
TBD	multi stage multiplexing hierarchy

3.1. Server Layer/Sub-Layer Attributes TLV

Switching capabilities and switching granularities of the region boundary can be carried in Attribute TLV. With these information carried in the RSVP-TE path message, the region boundary node can directly trigger one corresponding server layer or sub-layer FA-LSP creation which is defined in the Attribute TLV. The format of the Attribute TLV is shown below.

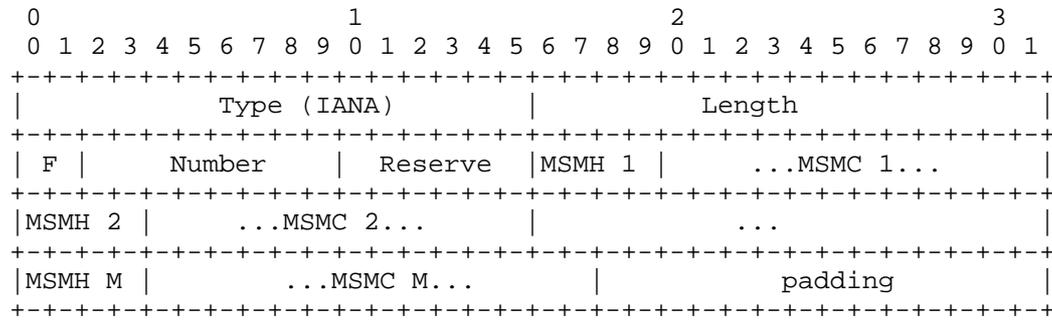


- o Type: indicates different values of Attribute TLV.
- o Length: indicates the total length of this Attribute TLV value.
- o Server Layer: Indicates which corresponding server layer should be triggered by the boundary node. The value of server layer is the same as the switching capability [RFC3471].
- o Sub-Layer: If there are several sub-layers within one server layer, it can further indicates which sub-layer should be triggered by the boundary node.

- * SDH/SONET: VC4, VC11, VC12, VC4-4c, VC4-16c, VC4-64c.
- * OTN: ODU0, ODU1, ODU2, ODU3, ODU2e, ODU4, and so on

3.2. Multiplexing Hierarchy Attribute TLV

Multiplexing Hierarchy Attribute TLV indicates the multiplexing hierarchies (e.g., ODU0-ODU2-ODU3) used for boundary node to configure it to the data plane and trigger one specific corresponding tunnel creation. The type of this sub-TLV will be assigned by IANA, and length is eight octets. The value field of this sub-TLV contains multi stages multiplexing hierachies constraint information of the link port.



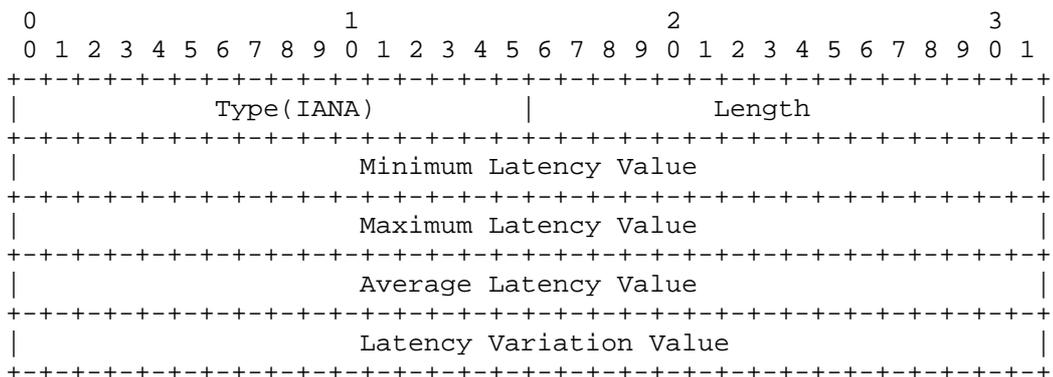
- o F (2 bits): Indicates the multi stages multiplexing hierarchies are included or excluded.
 - * 0 - Inclusive Multiplexing Hierarchies:Indicates that the object/TLV contains one or more multi stages multiplexing hierarchies which can be supported.
 - * 1 - Exclusive Multiplexing Hierarchies:Indicates that the object/TLV contains one or more multi stages multiplexing hierarchies which can't be supported.
- o Number (8 bits): Indicates the total number of multi stages multiplexing hierarchies which are supported or prohibited by the link port.
- o Reserve (8 bits): for future use.
- o (MSMH 1, MSMC 1), (MSMH 2, MSMC 2), ... ,(MSMH M, MSMC M): Indicates each multi stages multiplexing capability detailed information.

- * MSMH 1, MSMH2, ... , MSMH M (4 bits): Indicates the numbers of Multi Stages Multiplexing Hierarchies (MSMH).
 - + MSMH = 1: It indicates ODU_i is mapped into ODU_k (k > i) by single stage multiplexing (e.g., ODU0-ODU3).
 - + MSMH > 1: It indicates ODU_i is mapped into ODU_k (k > i) by multi stages multiplexing (e.g., ODU0-ODU1-ODU3).
- * MSMC 1, MSMC 2, ... ,MSMC M: Indicates the detailed information of multi stages multiplexing capability. The length of Multi Stages Multiplexing Capability (MSMC) information depends on the multi stages multiplexing hierarchies (MSMH). The length of MSMC is (MSMH+1) * 4. Each ODU_k (k=1, 2, 3, 4, 2e, flex) is indicated by 4 bits. Following is the Signal Type for G.709 Amendment 3.

Value	Type
0000	ODU0
0001	ODU1
0010	ODU2
0011	ODU3
0100	ODU4
0101	ODU2e
0110	ODUflex
7-15	Reserved (for future use)

- o The padding is used to make the Multi Stages Multiplexing Capability Descriptor sub-TLV 32-bits aligned.

3.3. Latency Attribute TLV



- Minimum Latency Value: a minimum value indicates the latency performance parameters which server layer/sub-layer LSP must meet.
- Maximum Latency Value: a maximum value indicates the latency performance parameters which server layer/sub-layer LSP must meet.
- Average Latency Value: a average value indicates the latency performance parameters which server layer/sub-layer LSP must meet.
- Latency Variation Value: a variation value indicates the latency performance parameters which server layer/sub-layer LSP must meet.

4. Signaling Procedure

In order to signal an end-to-end LSP across multi layer, the LSP source node sends the RSVP-TE PATH message with ERO which indicates LSP route and ERBO which indicates the LSP route boundary. When a interim node receives a PATH message, it will check ERBO to see if it is the layer boundary node. If a interim node isn't a layer boundary, it will process the PATH message as the normal one of single layer LSP. If a interim node finds its address is in ERBO, it is a layer boundary node. So it will directly extract another boundary egress node and other detail Attribute TLV information (e.g., Latency) from ERBO. If it is necessary, it will also extract the server layer/sub-layer routing information from ERO based on a pair of boundary node. Then the layer boundary node holds the PATH message and selects or creates a server layer/sub-layer LSP based on the detailed information of Attribute TLV (e.g., Latency) carried in ERBO.

On reception of a Path message containing BOUNDARY_ATTRIBUTES whose type of Attributes TLV is Multi States Multiplexing Hierarchy Sub-TLV, The interim node checks the local data plane capability to see if this kind of multi stages multiplexing/demultiplexing hierarchy is acceptable on specific interface. As there is an acceptable kind of multi stages multiplexing/demultiplexing, it must determine an ODUk tunnel must be created between a pair of boundary node. The kind of multi stages multiplexing/demultiplexing hierarchy must be configured into the data plane.

5. Security Considerations

TBD

6. IANA Considerations

TBD

7. References

7.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, December 2001.
- [RFC3471] Berger, L., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Functional Description", RFC 3471, January 2003.
- [RFC3473] Berger, L., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions", RFC 3473, January 2003.
- [RFC4203] Kompella, K. and Y. Rekhter, "OSPF Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4203, 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.
- [RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, August 2006.
- [RFC5212] Shiomoto, K., Papadimitriou, D., Le Roux, JL., Vigoureux, M., and D. Brungard, "Requirements for GMPLS-Based Multi-Region and Multi-Layer Networks (MRN/MLN)", RFC 5212, July 2008.
- [RFC5440] Vasseur, JP. and JL. Le Roux, "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, March 2009.

7.2. Informative References

- [I-D.ietf-ccamp-gmpls-mln-extensions]
Papadimitriou, D., Vigoureux, M., Shiomoto, K., Brungard,

D., and J. Roux, "Generalized Multi-Protocol Label Switching (GMPLS) Protocol Extensions for Multi-Layer and Multi-Region Networks (MLN/MRN)", draft-ietf-ccamp-gmpls-mln-extensions-12 (work in progress), February 2010.

[I-D.ietf-rtgwg-cl-requirement]

Ning, S., Malis, A., McDysan, D., Yong, L., JOUNAY, F., and Y. Kamite, "Requirements for MPLS Over a Composite Link", draft-ietf-rtgwg-cl-requirement-00 (work in progress), February 2010.

Authors' Addresses

Xihua Fu
ZTE Corporation
West District, ZTE Plaza, No.10, Tangyan South Road, Gaoxin District
Xi An 710065
P.R.China

Phone: +8613798412242
Email: fu.xihua@zte.com.cn
URI: <http://www.zte.com.cn/>

Qilei Wang
ZTE Corporation
No.68 ZiJingHua Road, Yuhuatai District
Nanjing 210012
P.R.China

Phone: +8613585171890
Email: wang.qilei@zte.com.cn
URI: <http://www.zte.com.cn/>

Yuanlin Bao
ZTE Corporation
5/F, R.D. Building 3, ZTE Industrial Park, Liuxian Road
Shenzhen 518055
P.R.China

Phone: +86 755 26773731
Email: bao.yuanlin@zte.com.cn
URI: <http://www.zte.com.cn/>

Ruiquan Jing
China Telecom

Email: jingrq@ctbri.com.cn

Xiaoli Huo
China Telecom

Email: huoxl@ctbri.com.cn

Internet Engineering Task Force
Internet-Draft
Intended status: Standards Track
Expires: August 22, 2011

GMG. G.Galimberti, Ed.
Cisco
RK. R.Kunze, Ed.
Deutsche Telekom
February 18, 2011

A SNMP MIB to manage the optical parameters characteristic of a DWDM
Black-Link
draft-galimbe-kunze-black-link-mib-00

Abstract

This memo defines a portion of the Management Information Base (MIB) used by Simple Network Management Protocol (SNMP) in TCP/IP- based internets. In particular, it defines objects for managing Optical Interfaces associated with Wavelength Division Multiplexing (WDM) systems or characterized by the Optical Transport Network (OTN) in accordance with the Black-Link approach defined in ITU-T Recommendation G.698. [ITU.G698.2]

The MIB module defined in this memo can be used for Optical Parameters monitoring and/or configuration of such optical interface.

Copyright Notice

Copyright (c) 2011 IETF Trust and the persons identified as the document authors. All rights reserved.

Note to RFC Editor re: [TEMPLATE TODO] markers

Note to RFC Editor: When a document is developed using this template, the editor of the document should replace or remove all the places marked [TEMPLATE TODO] before submitting the document. If there are still [TEMPLATE TODO] markers, please send the document back to the editor.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months

and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on August 22, 2011.

Copyright Notice

Copyright (c) 2011 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

- 1. Introduction 4
- 2. The Internet-Standard Management Framework 5
- 3. Conventions 5
- 4. Overview 5
 - 4.1. Optical Parameters Description 7
 - 4.1.1. General 7
 - 4.1.2. Parameters at Ss 8
 - 4.1.3. Optical path from point Ss to Rs 9
 - 4.1.4. Interface at point Rs 10
 - 4.1.5. Alarms and Threshold definition 10
 - 4.1.6. Performance Monitoring (PM) description 12
 - 4.1.7. Generic Parameter description 13
 - 4.2. Use of ifTable 14
- 5. Structure of the MIB Module 14
 - 5.1. The optIfOTMn group 15
 - 5.1.1. optIfOTMnTable 15
 - 5.2. The optIfOTSn groups 15
 - 5.2.1. optIfOTSn Configuration group 15
 - 5.3. The [TEMPLATE TODO] Subtree 15
 - 5.4. The Notifications Subtree 15
- 6. Object Definitions 15
- 7. Relationship to Other MIB Modules 17
 - 7.1. Relationship to the [TEMPLATE TODO] MIB 17
 - 7.2. MIB modules required for IMPORTS 17
- 8. Definitions 17
- 9. Security Considerations 17
- 10. IANA Considerations 18
- 11. Contributors 19
- 12. References 21
 - 12.1. Normative References 21
 - 12.2. Informative References 22
- Appendix A. Change Log 22
- Appendix B. Open Issues 23

1. Introduction

This memo defines a portion of the Management Information Base (MIB) used by Simple Network Management Protocol (SNMP) in TCP/IP- based internets. In particular, it defines objects for managing Optical Interfaces associated with Wavelength Division Multiplexing (WDM) systems or characterized by the Optical Transport Network (OTN) in accordance with the Black-Link approach defined in G.698.2 [ITU.G698.2]

Black Link approach allows supporting an optical transmitter/receiver pair of one vendor to inject a DWDM channel and run it over an optical network composed of amplifiers, filters, add-drop multiplexers from a different vendor. Whereas the standardization of black link for 2.5 and 10G is settled for 40G and 100G interfaces and Black Link extensions are still in progress. For carrier network deployments, interoperability is a key requirement. Today it is state-of-the-art to interconnect IP Routers from different vendors and WDM transport systems using short-reach, grey interfaces. Applying the Black Link (BL) concept, routers now get directly connected to each via transport interfaces which must be interoperable to each other.

The G.698.2 [ITU.G698.2] provides optical parameter values for physical layer interfaces of Dense Wavelength Division Multiplexing (DWDM) systems primarily intended for metro applications which include optical amplifiers. Applications are defined using optical interface parameters at the single-channel connection points between optical transmitters and the optical multiplexer, as well as between optical receivers and the optical demultiplexer in the DWDM system. This Recommendation uses a methodology which does not specify the details of the optical link, e.g. the maximum fibre length, explicitly. The Recommendation currently includes unidirectional DWDM applications at 2.5 and 10 Gbit/s with 100 GHz channel frequency spacing and may be extended to 40 and 100 Gbit/s channels with a lower channel frequency spacing.

The Building a SNMP MIB describing the optical parameters defined in G.698 [ITU.G698.2] allow the different vendors and operator to retrieve, provision and exchange information related to Optical Networks in a standardized way. This ensures interworking in case of using optical interfaces from different vendors at the end of the link. Decoupling DWDM layer from the optical layer The Optical Parameters and their values characterize the features and the performances of the Network optical components and allow a reliable network design in case of Multivendor Optical Networks.

Although RFC 3591 [RFC3591] describe and define the SNMP MIB of a

number of key optical parameters, alarms and Performance Monitoring, a more complete description of optical parameters and processes can be found in the ITU-T Recommendations. Appendix A of this document provides an overview about the extensive ITU-T documentation in this area. The same considerations can be applied to the RFC 4054 [RFC4054]

2. The Internet-Standard Management Framework

For a detailed overview of the documents that describe the current Internet-Standard Management Framework, please refer to section 7 of RFC 3410 [RFC3410].

Managed objects are accessed via a virtual information store, termed the Management Information Base or MIB. MIB objects are generally accessed through the Simple Network Management Protocol (SNMP). Objects in the MIB are defined using the mechanisms defined in the Structure of Management Information (SMI). This memo specifies a MIB module that is compliant to the SMIV2, which is described in STD 58, RFC 2578 [RFC2578], STD 58, RFC 2579 [RFC2579] and STD 58, RFC 2580 [RFC2580].

3. Conventions

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

4. Overview

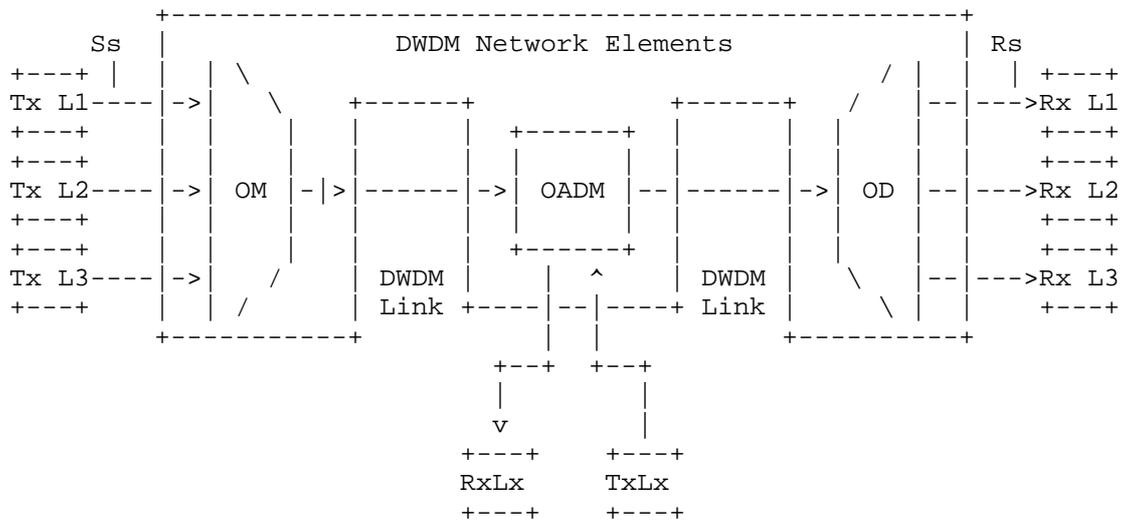
In this document, the term OTN (Optical Transport Network) system is used to describe devices that are compliant with the requirements specified in the ITU-T Recommendations G.872 [ITU.G872], G.709 [ITU.G709], G.798 [ITU.G798], G.874 [ITU.G874], and G.874.1 [ITU.G874.1] while refer to [ITU.G698.2] for the Black Link and DWDM parameter description.

The optical objects will be managed using the MIB II ifTable and ifStackTable. Additional tables will also be supported to monitor layer specific status and provide performance monitoring data. In the tables, some entries are required for OTN systems only. A Configuration (Config) table, Current Performance Monitoring (PM) table, and Interval PM table will be maintained for the OTSn, OMSn, OChGroup, and OCh layers on a source and sink trail termination basis. These tables will be linked to the ifTable by using the ifIndex that is associated with that layer.

An Alarm (Aalarm) table will be maintained for the OTSn, OMSn,

OChGroup, and OCh layers on a source and sink trail termination basis. These tables will be linked to the ifTable by using the ifIndex that is associated with that layer.

Figure ADD-REFERENCE shows a set of reference points, for the linear "black-link" approach, for single-channel connection (Ss and Rs) between transmitters (Tx) and receivers (Rx). Here the DWDM network elements include an OM and an OD (which are used as a pair with the opposing element), one or more optical amplifiers and may also include one or more OADMs.



Ss = reference point at the DWDM network element tributary output
 Rs = reference point at the DWDM network element tributary input
 Lx = Lambda x
 OM = Optical Mux
 OD = Optical Demux
 OADM = Optical Add Drop Mux

from Fig. 5.1/G.698.2

Figure 1: Linear Black Link

G.698.2 [ITU.G698.2] defines also Ring Black Link configurations [Fig. 5.2/G.698.2] and Bidirectional Black Link configurations [Fig. 5.3/G.698.2]

These objects are used when the particular media being used to realize an interface is an Optical Transport interface. At present,

this applies to these values of the ifType variable in the Internet-standard MIB:

opticalChannel (195), opticalChannelGroup (219), opticalTransport (196).

The definitions contained herein are based on the OTN specifications in ITU-T G.872 [ITU.G872], G.709 [ITU.G709], G.798 [ITU.G798], G.874 [ITU.G874], and G.874.1 [ITU.G874.1].

4.1. Optical Parameters Description

The terminology used in this document describes the optical parameters, the states and the Alarms at the points Ss, Rs and DWDM depicted in fig.1. The terms are defined in ITU-T Recommendations G.698.2 [ITU.G698.2]. Those definitions are made to increase the readability of the document.

4.1.1. General

Minimum channel spacing:

This is the minimum nominal difference in frequency between two adjacent channels (G).

Bit rate/line coding of optical tributary signals:

Optical tributary signal class NRZ 2.5G or NRZ 10G nominally 2.4 Gbit/s to nominally 10.71 Gbit/s. 40Gbit/s and 100Gbit/s are under definition (G, S).

Channel Modulation Format:

This parameter indicate what kind of modulation format is used at Ss (G).

FEC Coding:

This parameter indicate what Forward Error Correction (FEC) code is used at Ss and Rs (G, S).

Wavelength Range (see G.694.1): [ITU.G694.1]

This parameter indicate minimum and maximum wavelength spectrum (G) in a definite wavelength Band (L, C and S).

Wavelength Value (see G.694.1):

This parameter indicates the wavelength value that Ss and Rs will be set to work (G, S).

Vendor Transceiver Class:

Other than specifying all the Transceiver parameter, it might be convenient for the vendors to summarize a set of parameters in a single proprietary parameter: the Class of transceiver. The Transceiver classification will be based on the Vendor Name and the main TX and RX parameters (i.e. Trunk Mode, Framing, Bit rate, Trunk Type, Channel Band, Channel Grid, Modulation Format, etc.). If this parameter is used, the MIB parameters specifying the Transceiver characteristics may not be significant and the vendor will be responsible to specify the Class contents and values. The Vendor can publish the parameters of its Classes or declare to be compatible with published Classes.(G) Optional for compliance.

4.1.2. Parameters at Ss

Maximum and minimum mean channel output power:

The mean launched power at Ss is the average power of a pseudo-random data sequence coupled into the DWDM link It is defined the thange (Max and Min) of the parameter (G, S)

Minimum and maximum central frequency:

The central frequency is the nominal single-channel frequency on which the digital coded information of the particular optical channel is modulated by use of the NRZ line code. The central frequencies of all channels within an application lie on the frequency grid for the minimum channel spacing of the application given in ITU-T Rec. G.694.1. This parameter give the Maximum and minimum frequency interval the channel must be modulated (G)

Maximum spectral excursion:

This is the maximum acceptable difference between the nominal central frequency of the channel and the minus 15 dB points of the transmitter spectrum furthest from the nominal central frequency measured at point Ss. (G)

Maximum transmitter (residual) dispersion OSNR penalty (B.3/G.959.1) [ITU.G959.1]

Lowest OSNR at Ss with worst case (residual) dispersion. Lowest OSNR at Ss with no dispersion (G)

Electrical Signal Framing:

This is the indication of what framing (GE, Sonet/SDH, OTN) the Ss and Rs ports are set (G, S)

4.1.3. Optical path from point Ss to Rs

Maximum and minimum (residual) chromatic dispersion:

These parameters define the maximum and minimum value of the optical path "end to end chromatic dispersion" that the system shall be able to tolerate. (G)

Minimum optical return loss at Ss:

This parameter defines minimum optical return loss of the cable plant at the source reference point (Ss), including any connectors (G)

Maximum discrete reflectance between SS and RS:

Optical reflectance is defined to be the ratio of the reflected optical power present at a point, to the optical power incident to that point. Control of reflections is discussed extensively in ITU-T Rec. G.957 (G)

Maximum differential group delay:

Differential group delay (DGD) is the time difference between the fractions of a pulse that are transmitted in the two principal states of polarization of an optical signal. For distances greater than several kilometres, and assuming random (strong) polarization mode coupling, DGD in a fibre can be statistically modelled as having a Maxwellian distribution. (G)

Maximum polarisation dependent loss:

The polarisation dependent loss (PDL) is the difference (in dB) between the maximum and minimum values of the channel insertion loss (or gain) of the black-link from point SS to RS due to a variation of the state of polarization (SOP) over all SOPs. (G)

Maximum inter-channel crosstalk:

Inter-channel crosstalk is defined as the ratio of total power in all of the disturbing channels to that in the wanted channel, where the wanted and disturbing channels are at different wavelengths. The parameter specifies the isolation of a link conforming to the "black-link" approach such that under the worst-case operating conditions the inter-channel crosstalk at any reference point RS is less than the maximum inter-channel crosstalk value (G)

Maximum interferometric crosstalk:

This parameter places a requirement on the isolation of a link conforming to the "black-link" approach such that under the worst case operating conditions the interferometric crosstalk at any reference point RS is less than the maximum interferometric crosstalk value. (G)

Maximum optical path OSNR penalty:

The optical path OSNR penalty is defined as the difference between the Lowest OSNR at Rs and Lowest OSNR at Ss (G)

4.1.4. Interface at point Rs

Maximum and minimum mean input power:

The maximum and minimum values of the average received power at point Rs. (G)

Minimum optical signal-to-noise ratio (OSNR):

The minimum optical signal-to-noise ratio (OSNR) is the minimum value of the ratio of the signal power in the wanted channel to the highest noise power density in the range of the central frequency plus and minus the maximum spectral excursion (G)

Receiver OSNR tolerance:

The receiver OSNR tolerance is defined as the minimum value of OSNR at point Rs that can be tolerated while maintaining the maximum BER of the application. (G)

Minimum maximum Chromatic Dispersion (CD) :

This parameter define the CD range a Receiver (Rs) can tolerate in order to decode the received signal (G)

Maximum Polarization Mode Dispersion (PMD) :

This parameter define the maximum PMD value a Receiver (Rs) can tolerate in order to decode the received signal (G)

4.1.5. Alarms and Threshold definition

This section describes the Alarms and the Thresholds at Ss and Rs points according to ITU-T Recommendations G.872 [ITU.G872], G.709 [ITU.G709], G.798 [ITU.G798], G.874 [ITU.G874], and G.874.1 [ITU.G874.1]. The SNMP MIB of the above list is already defined and specified by the RFC3591

OTN alarms defined in RFC3591:

Threshold Crossing Alert (TCA Alarm)

LOW-TXPOWER

HIGH-TXPOWER

LOW-RXPOWER

HIGH-RXPOWER

OTUk-LOF or more generic LOF

Backward Defect Indication (BDI)

Trace Identifier Mismatch (tim)

Signal Degrade (sd)

Server Signal Failure (SSF)

Alarm Indication Signal (AIS)

Loss of Multiframe (lom)

OTN Thresholds (for TCA) defined in RFC3591

LOW-TXPOWER

HIGH-TXPOWER

LOW-RXPOWER

HIGH-RXPOWER

The list below reports the new Alarms and Thresholds not managed in RFC3591

Laser Bias Current:

This parameter report the Bias current of the Laser Transmitter (G)

Laser Bias Current Threshold:

This parameter is to set the Bias current Threshold of the Laser Transmitter used ri rise the related Alarm (G, S)

Forward Defect Indication (FDI):

This parameter indicates a notification to the receiver that a failure occurred in the network (G)

Backward Error Indication (BEI):

This parameter indicates the number of Errors occurred in the opposite line direction (G)

4.1.6. Performance Monitoring (PM) description

This section describes the Performance Monitoring parameters at Ss and Rs points (Near -End and Far-End) according to ITU-T Recommendations G.826 [ITU.G826], G.8201 [ITU.G8201], G.709 [ITU.G709], G.798 [ITU.G798], G.874 [ITU.G874], and G.874.1 [ITU.G874.1].

Failure Counts (fc) :

Number of Failures occurred in an observation period (G)

Errored Seconds (es) :

It is a one-second period in which one or more bits are in error or during which Loss of Signal (LOS) or Alarm Indication Signal (AIS) is detected (G)

Severely Errored Seconds (ses) :

It is a one-second period which has a bit-error ratio = 1×10^E minus 3 or during which Loss of Signal (LOS) or Alarm Indication Signal (AIS) is detected (G)

Unavailable Seconds (uas) :

A period of unavailable time begins at the onset of ten consecutive SES events. These ten seconds are considered to be part of unavailable time. A new period of available time begins at the onset of ten consecutive non-SES events. These ten seconds are considered to be part of available time (G)

Background Block Errors (bbe) :

An errored block not occurring as part of an SES(G)

Error Seconds Ratio (esr) :

The ratio of ES in available time to total seconds in available time during a fixed measurement interval(G)

Severely Errored Seconds Ratio (sesr) :

The ratio of SES in available time to total seconds in available time during a fixed measurement interval(G)

Background Block Errored Seconds Ratio (bber) :

The ratio of Background Block Errors (BBE) to total blocks in available time during a fixed measurement interval. The count of total blocks excludes all blocks during SESs.(G)

FEC corrected Bit Error (FECcorrErr):

The number of bits corrected by the FEC are counted over one second (G)

FEC un-corrected Bit Error :

The number of bits un-corrected by the FEC are counted over one second (G)

Pre-FEC Bit Error :

The number of Errored bits at receiving side before the FEC function counted over one second (G)

OTN Valid Intervals :

The number of contiguous 15 minute intervals for which valid OTN performance monitoring data is available for the particular interface (G)

FEC Valid Intervals :

The number of contiguous 15 minute intervals for which valid FEC PM data is available for the particular interface.(G)

4.1.7. Generic Parameter description

This section describes the Generic Parameters at Ss and Rs points according to ITU-T Recommendations G.872 [ITU.G872], G.709 [ITU.G709], G.798 [ITU.G798], G.874 [ITU.G874], and G.874.1 [ITU.G874.1].

Interface Admin Status :

The Administrative Status of an Interface: Up/Down - In Service/Out of Service (can be Automatic in Service) (G/S)

Interface Operational Status :

The Operational Status of an Interface: Up/Down - In Service/Out of Service (G)

Loopbacks :

The Interface loopbacks used for maintenance purposes, they are Terminal or Line (may be with send AIS)(G/S)

Pre-FEC BER (Mantissa + Exponent) :

Bit Error Rate at the Rs interface before error correction (G/S)

Q factor :

(G)

Q margin :

(G)

4.2. Use of ifTable

This section specifies how the MIB II interfaces group, as defined in RFC 2863 [RFC2863], is used for optical interfaces. As described in the RFC 3591 figure 1 [RFC3591] Only the ifGeneralInformationGroup will be supported for the ifTable and the ifStackTable to maintain the relationship between the various layers. The OTN layers are managed in the ifTable using IfEntries that correlate to the layers depicted in Figure 1. For example, a DWDM device with an Optical Network Node Interface (ONNI) will have an Optical Transmission Section (OTS) physical layer, an Optical Multiplex Section (OMS) layer (transports multiple optical channels), and an Optical Channel (OCh) layer. There is a one to one relationship between the OMS and OTS layers. The OMS layer has fixed connectivity via the OTS and thus no connectivity flexibility at the OMS layer is supported. This draft extend the RFC 3591 [RFC3591] as far as the OMSn and OTSn are concerned. The sections 2.5 and 2.6 of RFC 3591 [RFC3591] must be considered as a reference for the ifStackTable use and Optical Network Terminology.

5. Structure of the MIB Module

The managed Optical Networking interface objects are arranged into the following groups of tables:

The optIfOTMn group handles the OTM information structure of an optical interface.

optIfOTMnTable

The optIfPerfMon group handles the current 15-minute and 24-hour interval elapsed time, as well as the number of 15-minute intervals for all layers

optIfPerfMonIntervalTable

The optIfOTSn groups handle the configuration and performance monitoring information for OTS layers.

optIfOTSnConfigTable

optIfOTSnSinkCurrentTable

optIfOTSnSinkIntervalTable

optIfOTSnSinkCurDayTable

optIfOTSnSinkPrevDayTable

optIfOTSnSrcCurrentTable

optIfOTSnSrcIntervalTable

optIfOTSnSrcCurDayTable

optIfOTSnSrcPrevDayTable

5.1. The optIfOTMn group

5.1.1. optIfOTMnTable

This table contains the OTM structure information of an optical interface.

5.2. The optIfOTSn groups

5.2.1. optIfOTSn Configuration group

5.2.1.1. optIfOTSn Configuration Table

This table contains information on configuration of optIfOTSn interfaces, in addition to the information on such interfaces contained in the ifTable.

5.3. The [TEMPLATE TODO] Subtree

5.4. The Notifications Subtree

6. Object Definitions

```
OPT-IF-MIB DEFINITIONS ::= BEGIN
```

```
IMPORTS
```

```
    MODULE-IDENTITY, OBJECT-TYPE, Gauge32, Integer32,
        Unsigned32, transmission
        FROM SNMPv2-SMI
    TEXTUAL-CONVENTION, RowPointer, RowStatus, TruthValue
        FROM SNMPv2-TC
    SnmpAdminString
        FROM SNMP-FRAMEWORK-MIB
    MODULE-COMPLIANCE, OBJECT-GROUP
        FROM SNMPv2-CONF
    ifIndex
        FROM IF-MIB;
```

```
-- This is the MIB module for the OTN Interface objects.

optIfMibModule MODULE-IDENTITY
  LAST-UPDATED "200308130000Z"
  ORGANIZATION "IETF AToM MIB Working Group"
  CONTACT-INFO
    "WG charter:
     http://www.ietf.org/html.charters/atommib-charter.html

    Mailing Lists:
     General Discussion: atommib@research.telcordia.com
     To Subscribe: atommib-request@research.telcordia.com
    Editor: Hing-Kam Lam
    Postal: Lucent Technologies, Room 4C-616
           101 Crawfords Corner Road
           Holmdel, NJ 07733
           Tel: +1 732 949 8338
           Email: hklam@lucent.com"
  DESCRIPTION
    "The MIB module to describe pre-OTN and OTN interfaces.

    Copyright (C) The Internet Society (2003). This version
    of this MIB module is part of RFC 3591; see the RFC
    itself for full legal notices."
  REVISION "200308130000Z"
  DESCRIPTION
    "Initial version, published as RFC 3591."
  ::= { transmission 133 }
```

```
OptIfBitRateK ::= TEXTUAL-CONVENTION
  STATUS current
  DESCRIPTION
    "Indicates the index 'k' that is used to
    represent a supported bit rate and the different
    versions of OPUk, ODUk and OTUk.
    Allowed values of k are defined in ITU-T G.709.
    Currently allowed values in G.709 are:
     k=1 represents an approximate bit rate of 2.5 Gbit/s,
     k=2 represents an approximate bit rate of 10 Gbit/s,
     k=3 represents an approximate bit rate of 40 Gbit/s."
  SYNTAX Integer32
```

```
optIfOTMnBitRates OBJECT-TYPE
  SYNTAX BITS { bitRateK1(0), bitRateK2(1), bitRateK3(2) }
  MAX-ACCESS read-only
  STATUS current
  DESCRIPTION
    "This attribute is a bit map representing the bit
    rate or set of bit rates supported on the interface.
    The meaning of each bit position is as follows:
      bitRateK1(0) is set if the 2.5 Gbit/s rate is supported
      bitRateK2(1) is set if the 10 Gbit/s rate is supported
      bitRateK3(2) is set if the 40 Gbit/s rate is supported
    Note that each bit position corresponds to one possible
    value of the type OptIfBitRateK.
    The default value of this attribute is system specific."
 ::= { optIfOTMnEntry 3 }
```

7. Relationship to Other MIB Modules

7.1. Relationship to the [TEMPLATE TODO] MIB

7.2. MIB modules required for IMPORTS

8. Definitions

[TEMPLATE TODO]: put your valid MIB module here.
A list of tools that can help automate the process of
checking MIB definitions can be found at
<http://www.ops.ietf.org/mib-review-tools.html>

9. Security Considerations

There are a number of management objects defined in this MIB module with a MAX-ACCESS clause of read-write and/or read-create. Such objects may be considered sensitive or vulnerable in some network environments. The support for SET operations in a non-secure environment without proper protection can have a negative effect on network operations. These are the tables and objects and their sensitivity/vulnerability:

o

There are no management objects defined in this MIB module that have a MAX-ACCESS clause of read-write and/or read-create. So, if this MIB module is implemented correctly, then there is no risk that an intruder can alter or create any management objects of this MIB module via direct SNMP SET operations.

Some of the readable objects in this MIB module (i.e., objects with a MAX-ACCESS other than not-accessible) may be considered sensitive or vulnerable in some network environments. It is thus important to control even GET and/or NOTIFY access to these objects and possibly to even encrypt the values of these objects when sending them over the network via SNMP. These are the tables and objects and their sensitivity/vulnerability:

- o
- o

SNMP versions prior to SNMPv3 did not include adequate security. Even if the network itself is secure (for example by using IPsec), even then, there is no control as to who on the secure network is allowed to access and GET/SET (read/change/create/delete) the objects in this MIB module.

It is RECOMMENDED that implementers consider the security features as provided by the SNMPv3 framework (see [RFC3410], section 8), including full support for the SNMPv3 cryptographic mechanisms (for authentication and privacy).

Further, deployment of SNMP versions prior to SNMPv3 is NOT RECOMMENDED. Instead, it is RECOMMENDED to deploy SNMPv3 and to enable cryptographic security. It is then a customer/operator responsibility to ensure that the SNMP entity giving access to an instance of this MIB module is properly configured to give access to the objects only to those principals (users) that have legitimate rights to indeed GET or SET (change/create/delete) them.

10. IANA Considerations

Option #1:

The MIB module in this document uses the following IANA-assigned OBJECT IDENTIFIER values recorded in the SMI Numbers registry:

Descriptor	OBJECT IDENTIFIER value
-----	-----
sampleMIB	{ mib-2 XXX }

Option #2:

Editor's Note (to be removed prior to publication): the IANA is requested to assign a value for "XXX" under the 'mib-2' subtree and

to record the assignment in the SMI Numbers registry. When the assignment has been made, the RFC Editor is asked to replace "XXX" (here and in the MIB module) with the assigned value and to remove this note.

Note well: prior to official assignment by the IANA, an internet draft MUST use placeholders (such as "XXX" above) rather than actual numbers. See RFC4181 Section 4.5 for an example of how this is done in an internet draft MIB module.

Option #3:

This memo includes no request to IANA.

11. Contributors

Arnold Mattheus
Deutsche Telekom
Darmstadt
Germany
Phone +49xxxxxxxxxxx
email arnold.Mattheus@telekom.de

Manuel Paul
Deutsche Telekom
Berlin
Germany
phone +49xxxxxxxxxxx
email Manuel.Paul@telekom.de

Frank Luennemann
T-Com TE14
Germany
phone +49xxxxxxxxxxx
email Frank.Luennemann@telekom.de

Najam Saquib
Cisco
Ludwig-Erhard-Strasse 3
ESCHBORN, HESSEN 65760
GERMANY
phone +49 619 6773 9041
email nasaquib@cisco.com

Walid Wakim
Cisco
9501 Technology Blvd
ROSEMONT, ILLINOIS 60018
UNITED STATES
phone +1 847 678 5681
email wwakim@cisco.com

Ori Gerstel
Cisco
32 HaMelacha St., (HaSharon Bldg)
SOUTH NETANYA, HAMERKAZ 42504
ISRAEL
phone +972 9 864 6292
email ogerstel@cisco.com

12. References

12.1. Normative References

- [RFC2863] McCloghrie, K. and F. Kastenholz, "The Interfaces Group MIB", RFC 2863, June 2000.

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

- [RFC2578] McCloghrie, K., Ed., Perkins, D., Ed., and J. Schoenwaelder, Ed., "Structure of Management Information Version 2 (SMIV2)", STD 58, RFC 2578, April 1999.

- [RFC2579] McCloghrie, K., Ed., Perkins, D., Ed., and J. Schoenwaelder, Ed., "Textual Conventions for SMIV2", STD 58, RFC 2579, April 1999.

- [RFC2580] McCloghrie, K., Perkins, D., and J. Schoenwaelder, "Conformance Statements for SMIV2", STD 58, RFC 2580, April 1999.

- [RFC3591] Lam, H-K., Stewart, M., and A. Huynh, "Definitions of Managed Objects for the Optical Interface Type", RFC 3591, September 2003.

- [ITU.G698.2] International Telecommunications Union, "Amplified multichannel dense wavelength division multiplexing applications with single channel optical interfaces", ITU-T Recommendation G.698.2, November 2009.

- [ITU.G709] International Telecommunications Union, "Interface for the Optical Transport Network (OTN)", ITU-T Recommendation G.709, March 2003.

- [ITU.G872] International Telecommunications Union, "Architecture of optical transport networks", ITU-T Recommendation G.872, November 2001.

- [ITU.G798] International Telecommunications Union, "Characteristics of optical transport network hierarchy equipment functional blocks", ITU-T Recommendation G.798, October 2010.

- [ITU.G874] International Telecommunications Union, "Management aspects of optical transport network elements", ITU-T Recommendation G.874, July 2010.

- [ITU.G874.1] International Telecommunications Union, "Optical

transport network (OTN): Protocol-neutral management information model for the network element view", ITU-T Recommendation G.874.1, January 2002.

- [ITU.G959.1] International Telecommunications Union, "Optical transport network physical layer interfaces", ITU-T Recommendation G.959.1, November 2009.
- [ITU.G826] International Telecommunications Union, "End-to-end error performance parameters and objectives for international, constant bit-rate digital paths and connections", ITU-T Recommendation G.826, November 2009.
- [ITU.G8201] International Telecommunications Union, "Error performance parameters and objectives for multi-operator international paths within the Optical Transport Network (OTN)", ITU-T Recommendation G.8201, September 2003.
- [ITU.G694.1] International Telecommunications Union, "Spectral grids for WDM applications: DWDM frequency grid", ITU-T Recommendation G.694.1, June 2002.

12.2. Informative References

- [RFC3410] Case, J., Mundy, R., Partain, D., and B. Stewart, "Introduction and Applicability Statements for Internet-Standard Management Framework", RFC 3410, December 2002.
- [RFC2629] Rose, M., "Writing I-Ds and RFCs using XML", RFC 2629, June 1999.
- [RFC4181] Heard, C., "Guidelines for Authors and Reviewers of MIB Documents", BCP 111, RFC 4181, September 2005.
- [RFC4054] Strand, J. and A. Chiu, "Impairments and Other Constraints on Optical Layer Routing", RFC 4054, May 2005.

Appendix A. Change Log

This optional section should be removed before the internet draft is submitted to the IESG for publication as an RFC.

Note to RFC Editor: please remove this appendix before publication as an RFC.

Appendix B. Open Issues

Note to RFC Editor: please remove this appendix before publication as an RFC.

Authors' Addresses

Gabriele Galimberti (editor)
Cisco
Via Philips,12
20052 - Monza
Italy

Phone: +390392091462
EMail: ggalimbe@cisco.com

Ruediger Kunze (editor)
Deutsche Telekom
Dddd, xx
Berlin
Germany

Phone: +49xxxxxxxxxx
EMail: RKunze@telekom.de

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: September 8, 2011

O. Gonzalez de Dios, Ed.
Telefonica
G. Bernini
Nextworks
G. Zervas
Univ of Essex
M. Basham
Intune Networks
March 7, 2011

Framework for GMPLS and path computation support of sub-wavelength
switching optical networks
draft-gonzalezdedios-subwavelength-framework-00

Abstract

This document discusses the framework for enhancements to the GMPLS architecture to control sub-wavelength switching optical networks. Sub-wavelengths refer to the time-shared utilization of a single wavelength by optical bursts, packets or slots.

Sub-wavelength technologies are the base of new cost-effective network architectures. In particular, they are suited for metro areas, to cope with high traffic volumes as well as more demanding requirements of networked applications.

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on September 8, 2011.

Copyright Notice

Copyright (c) 2011 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

- 1. Introduction 3
- 2. Sub-wavelength optical networks 3
 - 2.1. Sub-wavelength switching research and applicability scenarios 5
- 3. Sub-wavelength network resource control 5
 - 3.1. Control functions and time-scales 5
 - 3.2. Network resource modeling 7
- 4. GMPLS implications 9
 - 4.1. Impact on GMPLS signaling 9
 - 4.1.1. Sub-wavelength resources and labels 9
 - 4.1.2. Sub-wavelength traffic specification 10
 - 4.2. Impact on GMPLS routing 11
 - 4.2.1. Sub-wavelength network resource availability advertisement 11
- 5. Route computation and sub-wavelength resource assignment scenarios 12
 - 5.1. Centralized PCE and centralized sub-wavelength resource assignment 13
 - 5.2. Centralized PCE and distributed sub-wavelength resource assignment 13
 - 5.3. Distributed PCE and distributed sub-wavelength resource assignment 13
- 6. Security Considerations 14
- 7. IANA Considerations 14
- 8. Contributing Authors 14
- 9. Acknowledgements 15
- 10. References 15
 - 10.1. Normative References 15
 - 10.2. Informative References 16
- Authors' Addresses 18

1. Introduction

A broad range of emerging services and applications drive the evolutionary trend of traffic growth at metro-regional networks with more and more demands for high bandwidth. However, recent measures and forecasts by network operators show that the expected traffic flows in the metro area for the next 10 years will occupy just a fraction of a wavelength (typically bearing 10 Gbps throughputs); consequently, the deployment of sub-wavelength statistically multiplexed networks can highly enhance the resource utilization [MAINS]. Moreover, the emergence of network centric services is also requiring more and more short-lived connections (in the range of secs, mins) with Quality of Service (QoS) guarantees: such services include Video on Demand (VoD), Storage Area Network (SAN) and a number of Cloud services.

In this network scenario, the metro-regional network should provide a contention-free sub-wavelength data transport environment with fast time-to-service delivery (few msec), low end-to-end delay and multiple levels of guaranteed QoS.

Nevertheless, an effective and sub-wavelength enabled supervising upper control layer might be needed to control the end-to-end resource reservation and routing across multiple sub-wavelength technologies. The IETF GMPLS is currently the most efficient solution for managing the physical core tunneling technologies of Internet and Telecom service providers. The natively generalized control approach enabled by GMPLS on the underlying Transport Plane allows also handling multiple switching technologies under a single Control Plane instance (MRN/MLN). The objective of this document is to define the framework for enhancements and extensions to the GMPLS protocols and procedures, to allow the automatic control of sub-wavelength optical switches.

2. Sub-wavelength optical networks

During the last few years there have been considerable developments on sub-wavelength optical networks for metro/core regions. Sub-wavelength optical networks incorporate the optical time domain in addition to the wavelength/frequency and space domains that are dealt with the wavelength switched optical networks (WSON). Such networks allow for time-shared use of individual or multiple wavelengths of a transparent optical network infrastructure, and multiple label switched paths (LSPs) can be transparently switched over the same wavelength of any link. This is possible due to the dynamic access and switch of transparent sub-wavelength data-sets such as optical time-slices/packet/bursts/flows.

The use of sub-wavelength networks is further motivated by the need to support the various granularity requirements (as in SONET/SDH, OTN) on optically transparent switching technologies either fixed or arbitrary. This allows for optical bypassing of statistically multiplexed sub-wavelength data-sets and as such reduces the use of O/E/O conversions and data processing at every node. This can potentially enhance optical network utilization and reduce the transport delays. Such network system has the ability to create, transport and switch tailored data-sets at sub-wavelength granularities to match network service requirements.

The main enablers of sub-wavelength optical networks are the fast tunable lasers [FTL] and fast optical switching elements [Fast-Switches]. Fast tunable lasers that span across the ITU-T C-Band can deliver nanoseconds wavelength-tuning time. Fast tunable lasers have been used in ingress and bypass nodes. In the ingress node they enable time-shared tunable transmission [Tunable-Trans], whereas in bypass nodes time-shared optical switching [Tunable-Switch] and time-shared wavelength conversion [Lambda-Convert]. More specifically, tunable ingress nodes can map optical data-sets (e.g. slots/packets/bursts/flows) to particular wavelength(s) according to their destination address. Bypass nodes use fast tunable lasers together with other optical devices (e.g. semiconductor optical amplifiers, SOAs) to convert incoming wavelength(s) for contention purposes. Combination of fast and transparent optical switches, MUX/DEMUX, optical wavelength routers (e.g. NxN arrayed waveguide gratings, AWGs), fast EDFAs (fast transient response time) and bursty receivers have been used to demonstrate sub-wavelength transparent optical networks [OPS-Network1][OPS-Network2].

In addition, Optical Packet Switch and Transport is a new networking platform that collapses layers 0 to 2 inside a ring network, by using ultra-fast tunable laser transmitters. The tunable transmitters act as both transmitters and switches simultaneously, which collapses the layers under one control system. The ring supports wavelength routing scheme to address packet flows based on wavelength selective switch, which acts as the address.[OPST-Network3].

Such optical technologies have enabled a variety of optical switching techniques. These include optical slotted and un-slotted optical packet/label/burst [OPS-OLS-OBS], as well as optical flow switching [OFS]. Such switching techniques have been designed to support both connection-oriented and connection-less services and as such different quality of service (QoS) guarantees. Furthermore, effort has been made to deliver interoperation between optical burst and circuit switched networks [CP-OBS-OCS] as well as a common control plane solution for optical packet and circuit switched networks [CP-OPS-OCS].

Considering the evolution, progress and variety of sub-wavelength optical networks it is important to address standardization aspects that would specify a transport-agnostic GMPLS control plane able to control and provision different sub-wavelength optical transport networks on a generic way. This would deliver interoperability between different sub-wavelength transport networks but also between sub-wavelength and WSON networks.

2.1. Sub-wavelength switching research and applicability scenarios

Some positioning statements by networks operators are provided in this section to describe their main objectives, activities and interests towards the deployment of sub-wavelength switching technologies.

The main drivers for sub-wavelength switching technologies are:

- o Increased capacity and scalability for intensive bandwidth consuming applications (e.g. 3D video, high definition videoconference, etc)
- o Operational simplicity by integrating optical (DWDM) and packet switching technologies (e.g. Ethernet, MPLS) in a single networking platform
- o Dynamic high capacity data transfers between distributed servers would enable an optimized planning of both IT (e.g. storage and computation) and network resources for cloud services

Both network operators and vendors will be benefitiated form the definition of common GMPLS extensions for any kind of sub-wavelength technology enabling end-to-end resource reservation and routing across different technological domains (e.g. sub-wavelength, WSON, etc).

3. Sub-wavelength network resource control

3.1. Control functions and time-scales

The time-scale of sub-wavelength optical networks control is broad and can be structured in three different levels. The first and coarser time-scale represents the duration of an LSP, which could be long-lived (e.g. days, hours) or short-lived (e.g. minutes, minutes, hours, days and can be directly controlled by GMPLS. The LSP durations decreasing towards seconds may have an impact on both GMPLS signaling and routing procedures, in terms of provisioning success, signaling overheads and TE topology accuracy.

The second level of time-scale is the optical frame in a repeating cycle. The duration (i.e. microseconds, milliseconds) and frame structure is controlled by sub-wavelength optical transport plane and is used to accommodate any fixed (i.e. optical time-slots) or flexible (i.e. optical packets/bursts, time-slices) data-sets. LSP durations in this context correspond to multiple frames. Framing the time on a cycle manner can aid towards the provisioning of the sub-wavelength data-sets.

The finest level of time-scale is the time-slice. It represents a fixed or flexible time proportion of a frame that corresponds to the amount of data-set (e.g. time-slot/time-slice, packet, burst, flow) and in turn bandwidth that can be individually switched and transported over the sub-wavelength optical network. The allocation and assignment of time-slices is controlled by the sub-wavelength optical transport plane. In this context, an LSP could be associated with a number of fixed or arbitrary sized time-slices per frame and the allocation of these time-sliced resources is described in section 6.

Consequently, the control of sub-wavelength network resources can be effectively performed just through the tight interworking of two different control layers: the GMPLS and the specific sub-wavelength optical transport control functions. This vertical cooperation can follow two different models, i.e. the overlay-style and the augmented, which depends on the procedures for information exchange and resource provisioning, and the information contents.

The overlay-style interworking is based on an independent control by GMPLS and sub-wavelength optical transport network. Main concepts of this architectural model are the following:

- o The TED maintained by the GMPLS (either distributed or centralized).
- o The sub-wavelength optical transport control functions, which manage the sub-wavelength scheduling database.
- o Using this model only the concept of sub-wavelength data over WSON is feasible. In that case a LSP is first established and then sub-wavelength flows use it to transport data across the network. Only a single LSP can be established per wavelength (no statistical multiplexing is feasible) and as such resource utilization and service flexibility is limited to grooming at the ingress node.

The advantage of such solution is the minimum extensions required for the GMPLS protocols. However, the major disadvantage of such

interworking approach is the lack of statistical multiplexing capabilities (i.e. multiple sub-wavelength LSPs per wavelength).

The augmented model allows for TE information exchange from sub-wavelength optical transport control system and GMPLS CP. Main concepts are:

- o Deployment of two types of TED:
 - * the GMPLS TED holds and maintains aggregated/abstracted sub-wavelength information (e.g. total time-slices - number or duration - per wavelength)
 - * the sub-wavelength optical transport TED maintains detailed information (e.g. accurate time availability representation of each frame per wavelength).
- o Route selection and time resource assignment is coordinated among both GMPLS and sub-wavelength transport control functions in a centralized or distributed style:
 - * the GMPLS is able to assign the possible route(s) and/or wavelength(s) based on the abstracted resource information to follow standard procedures.
 - * in a subsequent step, based on the calculated routes and/or wavelengths, the sub-wavelength transport control can assign the time-slices with the frame structure.
- o The LSP is provisioned by GMPLS and the individual access and switching of each time-slice is guaranteed by sub-wavelength optical transport control functions (e.g. data mapping to optical time-slices at ingress nodes, time-slice switching at bypass nodes).
- o Due to this collaborative task more than one LSP per wavelength is feasible and, thus, sub-wavelength statistical multiplexing is feasible. Also guaranteed contention-free network services can be delivered due to pre-established LSPs.

3.2. Network resource modeling

The vertical cooperation among the GMPLS control plane and the sub-wavelength transport plane control functions can be achieved under the condition of a certain level of GMPLS awareness of the sub-wavelength network resource availabilities. Therefore, a proper modeling of these new switching resources is needed in terms of Traffic Engineering parameters.

As previously stated, the operational time-scale of sub-wavelength optical networks is highly dynamic in comparison with the standard GMPLS operation time (i.e. signaling and routing procedures). This results in an impact of the potential fast variations of sub-wavelength resource availabilities at GMPLS control plane. For a proper GMPLS operation and control traffic balance, the frequency of subwavelength resource updates needs to be limited, though minimizing the potential contention on resource (i.e. wavelengths) due to subsequent inaccurate TE topology representation. Specific aggregation procedures need to be performed by the sub-wavelength optical network transport plane control functions to let the GMPLS maintain a summarized knowledge of sub-wavelength network resource availabilities, coherent and compatible with its operation time-scale. For a given sub-wavelength optical network link, the GMPLS control plane should be aware of the free capacity in each wavelength, to allow the time-shared use of the single wavelengths.

Such an aggregated and summarized description of sub-wavelength network resources would enable, on the one hand, the exchange of sub-wavelength TE routing information, and, on the other hand, the signaling and configuration of multiple LSPs sharing - where possible - the same wavelengths. The resource contention avoidance as well as the sub-wavelength switching configuration would be left to the sub-wavelength optical network transport plane control functions. This further control action will occur along the end-to-end path provisioned by the GMPLS control plane.

Since the sub-wavelength enabled GMPLS control plane is responsible for the end-to-end resource reservation and routing across multiple sub-wavelength technologies, the network resource modeling should be valid for any kind of sub-wavelength technology (i.e. optical packets, bursts, flows, etc.). To this purpose, a new Switching Type should be defined to model the sub-wavelength network resources:

- o Sub-Wavelength Switching Capability: to indicate the switching performed on a link, which supports the time-shared use of a wavelength. The SWSC would group all the specific implementations of sub-wavelength switching.

Since the Sub-Wavelength Switching Capability would be an intermediate switching type between TDM (value: 100) and LSC (value: 150), its value should be chosen in the <100,150> range to preserve the switching capability ordering and LSP region definitions specified in [RFC4206].

The identification of the specific sub-wavelength technologies should be performed defining new LSP Encoding Types (i.e. one for each switching technology), to univocally identify the links able to carry

and switch a signal encoded in a sub-wavelength format rather than another.

As a result, a given sub-wavelength optical network link should be described at least by the following parameters:

- o List of allowed/available wavelengths, e.g. described through the "Wavelength Label" format specified in [LAMBDA-LABELS]
- o For each wavelength, a sub-wavelength TE parameter accounting the free wavelength capacity

4. GMPLS implications

The GMPLS architecture [RFC3945] is designed to provide automatic provisioning of connections with traffic engineering, traffic survivability (i.e. protections, restorations), and automatic resource discovery and management. The GMPLS specifications are fully agnostic of specific deployment models and transport environments. Specific procedures have been defined to control transport networks as diverse as SDH/SONET [RFC3946], OTNs incorporating G.709 encapsulation [RFC4328], and Ethernet[RFC5828].

The sub-wavelength optical networks expose switching granularities and capabilities not natively supported by GMPLS. The following sub-sections provides a description of the impact of sub-wavelength switching granularity support on the GMPLS signaling and routing control functions, identifying a set of requirements to be evaluated for extensions of the current GMPLS protocol suite.

4.1. Impact on GMPLS signaling

Current GMPLS signaling procedures does not support the provisioning of sub-wavelength optical LSPs, where a single wavelength in a link can be shared among multiple LSPs. Two GMPLS signaling aspects are mainly affected by the introduction of sub-wavelength switching granularity: the identification of the sub-wavelength labels, and the characterization of the sub-wavelength data traffic.

4.1.1. Sub-wavelength resources and labels

An LSP signaled in a sub-wavelength optical network will reserve hop-by-hop the sub-wavelength resources. Current GMPLS signaling procedures does not support the identification of such fine-grained transport network resources. This means that a new type of label, i.e. a sub-wavelength label, should be defined to identify the sub-wavelength resources to be reserved in the transport plane for a

given LSP.

Different formats and encodings of the sub-wavelength label should be supported, depending on the specific sub-wavelength technologies controlled by the GMPLS.

Depending on how the sub-wavelength network resources are assigned along the LSP route, the sub-wavelength label would be processed in different ways.

When the sub-wavelength network resources assignment adopts the centralized model, a sub-wavelength label should be provided for each hop in the ERO of the LSP to be signaled. Therefore, the resources to be reserved along the LSP route would be selected and assigned before the signaling of the LSP, and the reservation would be performed hop-by-hop in the ERO processing during the LSP setup phase.

On the other hand, when the sub-wavelength network resources assignment adopts the distributed model (see section 5.2 and 5.3), the selection of the resources would not be performed before the LSP signaling and the ERO would not contain any sub-wavelength label. One or more sub-wavelength labels might be signaled in the Suggested Label object [RFC3473] to provide, if needed, the downstream node with the upstream node's label preference.

4.1.2. Sub-wavelength traffic specification

GMPLS signaling allows the inclusion of technology-specific parameters during the LSP setup, as described in [RFC3471][RFC3471]. In particular, when an LSP has to be established in a sub-wavelength optical network domain, a dedicated traffic profiling should be defined to describe the traffic characteristics of the sub-wavelength data flow, and identify network specific performance parameters (e.g. based on the sub-wavelength control parameters, such as burst durations, blockings, delays, etc.)

In GMPLS RSVP-TE [RFC3473], the SENDER_TSPEC object is used to describe the traffic parameters for the LSP being established, and allows for the inclusion of technology specific parameters. Therefore, a specific traffic profiling should be used in the sub-wavelength optical networks context, for two main purposes:

- o Identification of the requested LSP switching granularity, to distinguish among the different sub-wavelength technologies
- o Identification of the sub-wavelength traffic requirements and characteristics for the LSP to be signaled in the sub-wavelength

optical network domain

In particular, due to the short-lived fast-paced nature of the sub-wavelength data flows, the sub-wavelength traffic characteristics for the LSP to be established should be described, at least, in terms of bandwidth and QoS (i.e. delay, jitter, etc.) requirements, such as:

- o Bandwidth information: to specify the bandwidth requested for the reservation of the LSP, e.g. indicating average and peak bandwidths associated to the sub-wavelength data flow traffic to reserve.
- o Delay information: to specify the end-to-end delay requirements for a burst of traffic to be transmitted across the sub-wavelength optical network from source to destination (e.g. in terms of average and maximum delays).
- o Jitter information: to specify the maximum acceptable variation of latency for the bursts of traffic transmitted in the sub-wavelength optical network.

4.2. Impact on GMPLS routing

When an LSP has to be installed in a sub-wavelength optical network, the path computation process should find a suitable route for the requested connection. The selection of the end-to-end route (i.e. hops and links) should be performed at the GMPLS layer, while the sub-wavelength network resources assignment should be carried out by the sub-wavelength transport plane control functions. This means that the GMPLS routing protocol should be extended to advertise some aggregated sub-wavelength information to represent the actual availabilities in the transport plane, and allow the selection of the optimal end-to-end route.

4.2.1. Sub-wavelength network resource availability advertisement

GMPLS routing [RFC4202][RFC4203] defines the Interface Switching Capability Descriptor to advertise switching capabilities and encoding formats supported by a given link. According to what stated in section 4.2, in a sub-wavelength optical network scenario the Interface Switching Capability Descriptor should support:

- o The new Sub-Wavelength Switching Capability defined to describe the time-shared use of a wavelength
- o The new LSP Encoding Types defined to identify the different sub-wavelength encoding formats

Moreover, per [RFC4203], the Interface Switching Capability Descriptor advertises also aggregated link information at the bandwidth level (i.e. Maximum LSP Bandwidth). It specifies the maximum bandwidth that a single LSP can reserve on that interface.

However, rather than operating at the bandwidth level, the sub-wavelength enabled GMPLS should operate at least on a wavelength basis, and preferably on a sub-wavelength basis. Indeed, the distribution of wavelength and sub-wavelength availabilities is the key element to enable both GMPLS and path computation support of sub-wavelength switching optical networks.

GMPLS routing extensions in support of Wavelength Switching Optical Networks (WSON) are currently under study, and the usage of the Available Labels sub-TLV to advertise the available wavelengths in a given link has been proposed in [WSON-Routing]. On the other hand, dedicated GMPLS routing extensions must be defined to advertise, for each wavelength, the TE parameters accounting the free wavelength capacity, according to the network resource modeling detailed in section 4.2. The distribution of the sub-wavelength availabilities should be used in addition to the Available Labels sub-TLV to further detail the time-shared usage of the single wavelengths, introducing a finer granularity.

5. Route computation and sub-wavelength resource assignment scenarios

Based on the interworking models between GMPLS and the sub-wavelength switching control layer, three approaches for route computation can exist:

- o Centralized in both control planes
- o Centralized in GMPLS and distributed in the sub-wavelength switching control layer
- o Distributed in both control planes

A key functional element in all the scenarios is the Path Computation Element (PCE), which may be centralized or distributed, and perform the computation of a set of potential routes between the source and destination sub-wavelength capable nodes, matching the specified service and end-to-end traffic parameters. For this purpose, the PCE should store a summarized view of the sub-wavelength network topology, detailed in terms of nodes, TE links, the related wavelengths and aggregated slots availabilities. This TE information may be dynamically updated in different ways, e.g. through IGP routing protocols, like OSPF-TE, properly extended.

The PCE may interact with a sub-wavelength resource assignment entity, which operates in the sub-wavelength control layer and stores a detailed view of the slot/time-period utilization on all the links of the sub-wavelength network, and provides the logical representation of synchronized frames per link. The SLAE assigns the requested bandwidth on one of the potential routes, and may additionally guarantee both slot/time-period continuity constraint and wavelength-continuity constraint where needed.

Specific aspects of the different architectural scenarios regarding the association of PCE and sub-wavelength resource assignment are described below.

5.1. Centralized PCE and centralized sub-wavelength resource assignment

This case consists of GMPLS TED with complete network view of aggregated sub-wavelength information for route selection and complete view of sub-wavelength availability (e.g. time-slice(s) per wavelength per link). The process of both is concurrent, performed at the same location, and considers the assignment of route and sub-wavelength resource assignment for every link of an end-to-end LSP. The sub-wavelength assignment occurs before the actual LSP provisioning.

5.2. Centralized PCE and distributed sub-wavelength resource assignment

This case differs from the first one on the aspect of distributed sub-wavelength assignment based on initial route calculations. As such, after the calculation of K possible paths (e.g. shortest paths) the sub-wavelength assignment is attempted on a hop-by-hop basis at sub-wavelength optical transport level. The sub-wavelength TED information on each node might consist of either only neighbor's link information or complete network information.

5.3. Distributed PCE and distributed sub-wavelength resource assignment

In that case both route and sub-wavelength assignment happens in a distributed manner. As such each node calculates the possible next hops at GMPLS level and the sub-wavelength assignment uses such info to assign the available time-slices. Again the aggregate TED might consist of complete network resource availability information and sub-wavelength TED might consist of either neighbor or complete network information. In case of complete sub-wavelength TED information at each node the success ratio on provisioning LSPs might increase at the expense of increased routing information exchange at sub-wavelength transport control system.

6. Security Considerations

This document does not introduce new security issues; the considerations in [RFC3471], [RFC3473] and [RFC3945] apply.

GMPLS control of sub-wavelength switching assumes that users and devices attached to UNIs may behave maliciously, negligently, or incorrectly. Intra-provider control traffic is trusted to not be malicious. In general, these requirements are no different from the security requirements for operating any GMPLS network. Access to the trusted network will only occur through the protocols defined for the UNI or NNI or through protected management interfaces.

When in-band GMPLS signaling is used for the control plane the security of the control plane and the data plane may affect each other. When out-of-band GMPLS signaling is used for the control plane the data plane security is decoupled from the control plane and therefore the security of the data plane has less impact on overall security.

For a more comprehensive discussion on GMPLS security please refer to [RFC5920].

7. IANA Considerations

This document introduces the following requests for IANA action:

- o Assign a new Switching Type: "Sub-Wavelength" (suggested value TBD) in the GMPLS Signaling Parameters / Switching Types registry.
- o Assign new LSP Encoding Types for the different sub-wavelength switching technologies":
- o New error codes in the RSVP Parameters / Error Codes and Globally-Defined Error Value Sub-Codes registry:

8. Contributing Authors

Juan Pedro Fernandez-Palacios Gimenez
Telefonica Investigacion y Desarrollo
C/Ramon de la Cruz
Madrid, 28006
Spain

Phone: +34 91 3379037
Email: jpfpg@tid.es

Gino Carrozzo
Nextworks
via Turati 43/45
Pisa
Italy

Phone:
Email: g.carrozzo@nextworks.it

Dimitra Simeonidou
University of Essex
Wivenhoe Park
Colchester, Essex
U.K.

Phone:
Email: dsimeo@essex.ac.uk

9. Acknowledgements

This work has been partially supported by the EC through the IST STREP project MAINS (INFISO-ICT-247706).

10. References

10.1. Normative References

- [RFC3471] Berger, L., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Functional Description", RFC 3471, January 2003.
- [RFC3473] Berger, L., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions", RFC 3473, January 2003.
- [RFC3945] Mannie, E., "Generalized Multi-Protocol Label Switching (GMPLS) Architecture", RFC 3945, October 2004.
- [RFC3946] Mannie, E. and D. Papadimitriou, "Generalized Multi-Protocol Label Switching (GMPLS) Extensions for Synchronous Optical Network (SONET) and Synchronous Digital Hierarchy (SDH) Control", RFC 3946, October 2004.
- [RFC4202] Kompella, K. and Y. Rekhter, "Routing Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4202, October 2005.

- [RFC4203] Kompella, K. and Y. Rekhter, "OSPF Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4203, 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.
- [RFC4328] Papadimitriou, D., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Extensions for G.709 Optical Transport Networks Control", RFC 4328, January 2006.
- [RFC5828] Fedyk, D., Berger, L., and L. Andersson, "Generalized Multiprotocol Label Switching (GMPLS) Ethernet Label Switching Architecture and Framework", RFC 5828, March 2010.
- [RFC5920] Fang, L., "Security Framework for MPLS and GMPLS Networks", RFC 5920, July 2010.

10.2. Informative References

- [CP-OBS-OCS] Hong, X., "Testbed of OBS/GMPLS interworking, WOBS 2009", 2009.
- [CP-OPS-OCS] Miyazawa, T., "Experimental Performance Evaluation of Control Mechanisms for Integrated Optical Packet- and Circuit-Switched Networks, IEEE GLOBECOM, FutNet05.1, pp.360-360, Miami, USA", 2010.
- [FTL] Simsarian, J., "Fast tunable lasers for optical routers and networks, CLEO/QELS 2006", 2006.
- [Fast-Switches] Zervas, E., "Multi-Granular Optical Cross-Connect: Design, Analysis and Demonstration IEEE Journal of Optical Communications and Networking, Vol. 1, Issue 1, pp. 69-84", 2009.
- [Lambda-Convert] Lal, V., "Monolithic Wavelength Converters for High-Speed Packet-Switched Optical Networks JSTQE, 13, PP. 49-57", 2007.
- [OFS] Weichenberg, G. and V. Chan, "Design and Analysis of Optically Flow Switched Networks, IEEE/OSA Journal on

Optical Communications and Networking", Aug 2009.

[OPS-Network1]

Chiaroni, D., "Demonstration of the Interconnection of Two Optical Packet Rings with a Hybrid Optoelectronic Packet Router, PD3.5, ECOC", 2010.

[OPS-Network2]

Furukawa, H., "First Development of Integrated Optical Packet and Circuit Switching Node for New-Generation Networks, We.8.A.4, ECOC", 2010.

[OPS-Network3]

Fernandez-Palacios, J., "IP Offloading over Multi-granular Photonic Switching Technologies, Mo.2.D.6, ECOC", 2010.

[OPS-OLS-OBS]

Yoo, S., "Optical packet and burst switching technologies for the future photonic Internet, J. Lightwave Technol., vol., no. 12, pp. 4468 4492", Dec 2006.

[Tunable-Switch]

Klonidis, D., "Fast and Widely Tunable Optical Packet Switching Scheme based on Tunable Laser and Dual-Pump Four-Wave Mixing, IEEE PTL, vol. 16, no. 5", May 2004.

[Tunable-Trans]

Dunne, J., "Optical Packet Switch and Transport: A New Metro Platform to Reduce Costs and Power by 50% to 75% While Simultaneously Increasing Deterministic Performance, WOBS 2009", 2009.

[WSON-Routing]

Zhang, F., Bernstein, G., and Y. Xu, "OSPF Extensions in Support of Routing and Wavelength Assignment (RWA) in Wavelength Switched Optical Networks (WSONs), draft-zhang-ccamp-rwa-wson-routing-ospf-03.txt, March 2010.", 2010.

[lambda-label]

Otani, T. and D. Li, "Generalized Labels for Lambda-Switching Capable Label Switching Routers, work in progress: draft-ietf-ccamp-gmpls-g-694-lambda-labels-11.txt", 2011.

Authors' Addresses

Oscar Gonzalez de Dios (editor)
Telefonica
Ramon de la Cruz, 82-84
Madrid, 28006
Spain

Phone: +34 913374013
Email: ogondio@tid.es

Giacomo Bernini
Nextworks
via Turati 43/45
Pisa
Italy

Phone:
Email: g.bernini@nextworks.it

Georgios Zervas
Univ of Essex
Wivenhoe Park
Colchester, Essex
U.K.

Phone:
Email: gzerva@essex.ac.uk

Mark Basham
Intune Networks
via Turati 43/45
Colchester
U.K.

Phone:
Email: mark.basham@intunenetworks.com

Internet Draft
Updates: 2205, 3209, 3473
Category: Standards Track
Expiration Date: September 14, 2011

Lou Berger (LabN)
Francois Le Faucheur (Cisco)
Ashok Narayanan (Cisco)

March 14, 2011

Usage of The RSVP Association Object

draft-ietf-ccamp-assoc-info-01.txt

Abstract

The RSVP ASSOCIATION object was defined in the context of GMPLS (Generalized Multi-Protocol Label Switching) controlled label switched paths (LSPs). In this context, the object is used to associate recovery LSPs with the LSP they are protecting. This object also has broader applicability as a mechanism to associate RSVP state, and this document defines how the ASSOCIATION object can be more generally applied. The document also reviews how the association is to be provided in the context of GMPLS recovery. No new new procedures or mechanisms are defined with respect to GMPLS recovery. This document also defines extended ASSOCIATION objects which can be used in the context of Transport Profile of Multiprotocol Label Switching (MPLS-TP).

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at <http://www.ietf.org/lid-abstracts.html>

The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>

This Internet-Draft will expire on September 14, 2011

Copyright and License Notice

Copyright (c) 2011 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1	Introduction	3
1.1	Conventions Used In This Document	4
2	Background	4
2.1	LSP Association	4
2.2	End-to-End Recovery LSP Association	6
2.3	Segment Recovery LSP Association	8
2.4	Resource Sharing LSP Association	9
3	Association of GMPLS Recovery LSPs	10
4	Non-GMPLS Recovery Usage	11
4.1	Upstream Initiated Association	11
4.1.1	Path Message Format	12
4.1.2	Path Message Processing	12
4.2	Downstream Initiated Association	13
4.2.1	Resv Message Format	14
4.2.2	Resv Message Processing	14
4.3	Association Types	15
4.3.1	Resource Sharing Association Type	15
5	IPv4 and IPv6 Extended ASSOCIATION Objects	16
5.1	IPv4 and IPv6 Extended ASSOCIATION Object Format	17
5.2	Processing	18
6	Security Considerations	20
7	IANA Considerations	20
7.1	IPv4 and IPv6 Extended ASSOCIATION Objects	20
7.2	Resource Sharing Association Type	21
8	Acknowledgments	21
9	References	21
9.1	Normative References	21
9.2	Informative References	22
10	Authors' Addresses	23

1. Introduction

End-to-end and segment recovery are defined for GMPLS (Generalized Multi-Protocol Label Switching) controlled label switched paths (LSPs) in [RFC4872] and [RFC4873] respectively. Both definitions use the ASSOCIATION object to associate recovery LSPs with the LSP they are protecting. This document provides additional narrative on how such associations are to be identified. In the context of GMPLS recovery, this document does not define any new procedures or mechanisms and is strictly informative in nature.

In addition to the narrative, this document also explicitly expands the possible usage of the ASSOCIATION object in other contexts. In Section 4, this document reviews how association should be made in the case where the object is carried in a Path message and defines usage with Resv messages. This section also discusses usage of the ASSOCIATION object outside the context of GMPLS LSPs.

Some examples of non-LSP association in order to enable resource sharing are:

- o Voice Call-Waiting:
A bidirectional voice call between two endpoints A and B is signaled using two separate unidirectional RSVP reservations for the flows A->B and B->A. If endpoint A wishes to put the A-B call on hold and join a separate A-C call, it is desirable that network resources on common links be shared between the A-B and A-C calls. The B->A and C->A subflows of the call can share resources using existing RSVP sharing mechanisms, but only if they use the same destination IP addresses and ports. However, there is no way in RSVP today to share the resources between the A->B and A->C subflows of the call since by definition the RSVP reservations for these subflows must have different IP addresses in the SESSION objects.
- o Voice Shared Line:
A single number that rings multiple endpoints (which may be geographically diverse), such as phone lines on a manager's desk and their assistant. A VoIP system that models these calls as multiple P2P unicast pre-ring reservations would result in significantly over-counting bandwidth on shared links, since today unicast reservations to different endpoints cannot share bandwidth.
- o Symmetric NAT:
RSVP permits sharing of resources between multiple flows addressed to the same destination D, even from different senders S1 and S2. However, if D is behind a NAT operating in symmetric mode [RFC5389], it is possible that the destination port of the flows S1->D and S2->D may be different outside the NAT. In this case, these flows cannot share resources using RSVP today, since

the SESSION objects for these two flows outside the NAT would have different ports.

Section 5 of this document defines the extended ASSOCIATION objects which can be used in the context of Transport Profile of Multiprotocol Label Switching (MPLS-TP). Although, the scope of the extended ASSOCIATION objects is not limited to MPLS-TP.

1.1. Conventions Used In This Document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

2. Background

This section reviews the definition of LSP association in the contexts of end-to-end and segment recovery as defined in [RFC4872] and [RFC4873]. This section merely reiterates what has been defined, if differences exist between this text and [RFC4872] or [RFC4873], the earlier RFCs provide the authoritative text.

2.1. LSP Association

[RFC4872] introduces the concept and mechanisms to support the association of one LSP to another LSP across different RSVP-TE sessions. Such association is enabled via the introduction of the ASSOCIATION object. The ASSOCIATION object is defined in Section 16 of [RFC4872]. It is explicitly defined as having both general application and specific use within the context of recovery. End-to-end recovery usage is defined in [RFC4872] and is covered in Section 2.2. Segment recovery usage is defined in [RFC4873] and is covered in Section 2.3. Resource sharing LSP association is also defined in [RFC4873], while strictly speaking such association is beyond the scope of this document, for completeness it is covered in Section 2.4. The remainder of this section covers generic usage of the ASSOCIATION object.

In general, LSP association using the ASSOCIATION object can take place based on the values carried in the ASSOCIATION object. This means that association between LSPs can take place independent from and across different sessions. This is a significant enhancement from the association of LSPs that is possible in base MPLS [RFC3209] and GMPLS [RFC3473].

When using ASSOCIATION object, LSP association is always initiated by an upstream node that inserts appropriate ASSOCIATION objects in the Path message of LSPs that are to be associated. Downstream nodes

then correlate LSPs based on received ASSOCIATION objects. Multiple types of LSP association is supported by the ASSOCIATION object, and downstream correlation is made based on the type.

[RFC4872] defines C-Types 1 and 2 of the ASSOCIATION object. Both objects have essentially the same semantics, only differing in the type of address carried (IPv4 and IPv6). The defined objects carry multiple fields. The fields, taken together, enable the identification of which LSPs are association with one another. The [RFC4872] defined fields are:

- o Association Type:
This field identifies the usage, or application, of the association object. The currently defined values are Recovery [RFC4872] and Resource Sharing [RFC4873]. This field also scopes the interpretation of the object. In other words, the type field is included when matching LSPs (i.e., the type fields must match), and the way associations are identified may be type dependent.
- o Association Source:
This field is used to provide global scope (within the address space) to the identified association. There are no specific rules in the general case for which address should be used by a node creating an ASSOCIATION object beyond that the address is "associated to the node that originated the association", see [RFC4872].
- o Association ID:
This field provides an "identifier" that further scopes an association. Again, this field is combined with the other ASSOCIATION object fields to support identification of associated LSPs. The generic definition does not provide any specific rules on how matching is to be done, so such rules are governed by the Association Type. Note that the definition permits the association of an arbitrary number of LSPs.

As defined, the ASSOCIATION object may only be carried in a Path message, so LSP association takes place based on Path state. The definition permits one or more objects to be present. The support for multiple objects enables an LSP to be associated with other LSPs in more than one way at a time. For example, an LSP may carry one ASSOCIATION object to associate the LSP with another LSP for end-to-end recovery, and at the same time carry a second ASSOCIATION object to associate the LSP with another LSP for segment recovery, and at the same time carry a third ASSOCIATION object to associate the LSP with yet another LSP for resource sharing.

2.2. End-to-End Recovery LSP Association

The association of LSPs in support of end-to-end LSP recovery is defined in Section 16.2 of [RFC4872]. There are also several additional related conformance statements (i.e., use of [RFC2119] defined key words) in Sections 7.3, 8.3, 9.3, 11.1. When analyzing the definition, as with any Standards Track RFC, it is critical to note and differentiate which statements are made using [RFC2119] defined key words, which relate to conformance, and which statements are made without such key words, which are only informative in nature.

As defined in Section 16.2, end-to-end recovery related LSP association may take place in two distinct forms:

- a. Between multiple (one or more) working LSPs and a single shared (associated) recovery LSP. This form essentially matches the shared 1:N ($N \geq 1$) recovery type described in the other sections of [RFC4872].
- b. Between a single working LSP and multiple (one or more) recovery LSPs. This form essentially matches all other recovery types described in [RFC4872].

Both forms share the same Association Type (Recovery) and the same Association Source (the working LSP's tunnel sender address). They also share the same definition of the Association ID, which is (quoting [RFC4872]):

"The Association ID MUST be set to the LSP ID of the LSP being protected by this LSP or the LSP protecting this LSP. If unknown, this value is set to its own signaled LSP ID value (default). Also, the value of the Association ID MAY change during the lifetime of the LSP."

The interpretation of the above is fairly straightforward. The Association ID carries one of 3 values:

- The LSP ID of the LSP being protected.
- The LSP ID of the LSP protecting an LSP.
- In the case where the matching LSP is not yet known (i.e., initiated), the LSP ID value of the LSP itself.

The text also explicitly allows for changing the Association ID during the lifetime of an LSP. But this is only an option, and is neither required (i.e., "MUST") nor recommended (i.e., "SHOULD"). It should be noted that the document does not describe when such a change should be initiated, or the procedures for such a change. Clearly care needs to be taken when changing the Association ID to ensure that the old association is not lost during the transition to a new association.

The text does not preclude, and it is therefore assumed, that one or more ASSOCIATION objects may also be added to an LSP that was originated without any ASSOCIATION objects. Again this is a case that is not explicitly discussed in [RFC4872].

From the above, this means that the following combinations may occur:

- Case 1. When the ASSOCIATION object of the LSP being protected is initialized before the ASSOCIATION objects of any recovery LSPs are initialized, the Association ID in the LSP being protected and any recovery LSPs will carry the same value and this value will be the LSP ID value of the LSP being protected.
- Case 2. When the ASSOCIATION object of a recovery LSP is initialized before the ASSOCIATION object of any protected LSP is initialized, the Association ID in the recovery LSP and any LSPs being protected by that LSP will carry the same value and this value will be the LSP ID value of the recovery LSP.
- Case 3. When the ASSOCIATION objects of both the LSP being protected and the recovery LSP are concurrently initialized, the value of the Association ID carried in the LSP being protected is the LSP ID value of the recovery LSP, and the value of the Association ID carried in the recovery LSP is the LSP ID value of the LSP being protected. As this case can only be applied to LSPs with matching tunnel sender addresses, the scope of this case is limited to end-to-end recovery. Note that this is implicit in [RFC4872] as its scope is limited to end-to-end recovery.

In practical terms, case 2 will only occur when using the shared 1:N (N >= 1) end-to-end recovery type and case 1 will occur with all other end-to-end recovery types. Case 3 is allowed, and it is subject to interpretation how often it will occur. Some believe that this case is the common case and, furthermore, that working and recovery LSPs will often first be initiated without any ASSOCIATION objects and then case 3 objects will be added once the LSPs are established. Others believe that case 3 will rarely if ever occur. Such perspectives have little impact on interoperability as a [RFC4872] compliant implementation needs to properly handle (identify associations for) all three cases.

It is important to note that Section 16.2 of [RFC4872] provides no further requirements on how or when the Association ID value is to be selected. The other sections of the document do provide further narrative and 3 additional requirements. In general, the narrative highlights case 3 identified above but does not preclude the other cases. The 3 additional requirements are, by [RFC4872] Section

number:

- o Section 7.3 -- "The Association ID MUST be set by default to the LSP ID of the protected LSP corresponding to $N = 1$."

When considering this statement together with the 3 cases enumerated above, it can be seen that this statement clarifies which LSP ID value should be used when a single shared protection LSP is established simultaneously with (case 3), or after (case 2), more than one LSP to be protected.

- o Section 8.3 -- "Secondary protecting LSPs are signaled by setting in the new PROTECTION object the S bit and the P bit to 1, and in the ASSOCIATION object, the Association ID to the associated primary working LSP ID, which MUST be known before signaling of the secondary LSP."

This requirement clarifies that the Rerouting without Extra-Traffic type of recovery is required to follow either case 1 or 3, but not 2, as enumerated above.

- o Section 9.3 -- "Secondary protecting LSPs are signaled by setting in the new PROTECTION object the S bit and the P bit to 1, and in the ASSOCIATION object, the Association ID to the associated primary working LSP ID, which MUST be known before signaling of the secondary LSP."

This requirement clarifies that the Shared-Mesh Restoration type of recovery is required to follow either case 1 or 3, but not 2, as enumerated above.

- o Section 11.1 -- "In both cases, the Association ID of the ASSOCIATION object MUST be set to the LSP ID value of the signaled LSP."

This requirement clarifies that when using the LSP Rerouting type of recovery is required to follow either case 1 or 3, but not 2, as enumerated above.

2.3. Segment Recovery LSP Association

GMPLS segment recovery is defined in [RFC4873]. Segment recovery reuses the LSP association mechanisms, including the Association Type field value, defined in [RFC4872]. The primary text to this effect in [RFC4873] is:

3.2.1. Recovery Type Processing

Recovery type processing procedures are the same as those defined in [RFC4872], but processing and identification occur

with respect to segment recovery LSPs. Note that this means that multiple ASSOCIATION objects of type recovery may be present on an LSP.

This statement means that case 2 as enumerated above is to be followed and furthermore that Association Source is set to the tunnel sender address of the segment recovery LSPs. The explicit exclusion of case 3 is not listed as its non-applicability was considered obvious to the informed reader. (Perhaps having this exclusion explicitly identified would have obviated the need for this document.)

2.4. Resource Sharing LSP Association

Section 3.2.2 of [RFC4873] defines an additional type of LSP association which is used for "Resource Sharing". Resource sharing enables the sharing of resources across LSPs with different SESSION objects. Without this object only sharing across LSPs with a shared SESSION object was possible, see [RFC3209].

Resource sharing is indicated using a new Association Type value. As the Association Type field value is not the same as is used in Recovery LSP association, the semantics used for the association of LSPs using an ASSOCIATION object containing the new type differs from Recovery LSP association.

Section 3.2.2 of [RFC4873] states the following rules for the construction of an ASSOCIATION object in support of resource sharing LSP association:

- o The Association Type value is set to "Resource Sharing".
- o Association Source is set to the originating node's router address.
- o The Association ID is set to a value that uniquely identifies the set of LSPs to be associated.

The setting of the Association ID value to the working LSP's LSP ID value is mentioned, but using the "MAY" key word. Per [RFC2119], this translates to the use of LSP ID value as being completely optional and that the choice of Association ID is truly up to the originating node.

Additionally, the identical ASSOCIATION object is used for all LSPs that should be associated using Resource Sharing. This differs from recovery LSP association where it is possible for the LSPs to carry different Association ID fields and still be associated (see case 3 in Section 2.2).

3. Association of GMPLS Recovery LSPs

The previous section reviews the construction of an ASSOCIATION object, including the selection of the value used in the Association ID field, as defined in [RFC4872] and [RFC4873]. This section reviews how a downstream receiver identifies that one LSP is associated within another LSP based on ASSOCIATION objects. Note that this section in no way modifies the normative definitions of end-to-end and segment recovery, see [RFC4872] or [RFC4873].

As the ASSOCIATION object is only carried in Path messages, such identification only takes place based on Path state. In order to support the identification of the recovery type association between LSPs, a downstream receiver needs to be able to handle all three cases identified in Section 2.2. Cases 1 and 2 are simple as the associated LSPs will carry the identical ASSOCIATION object. This is also always true for resource sharing type LSP association, see Section 2.4. Case 3 is more complicated as it is possible for the LSPs to carry different Association ID fields and still be associated. The receiver also needs to allow for changes in the set of ASSOCIATION objects included in an LSP.

Based on the [RFC4872] and [RFC4873] definitions related to the ASSOCIATION object, the following behavior can be followed to ensure that a receiver always properly identifies the association between LSPs:

- o Covering cases 1 and 2 and resource sharing type LSP association:

For ASSOCIATION objects with the Association Type field values of "Recovery" (1) and "Resource Sharing" (2), the association between LSPs is identified by comparing all fields of each of the ASSOCIATION objects carried in the Path messages associated with each LSP. An association is deemed to exist when the same values are carried in all fields of an ASSOCIATION object carried in each LSP's Path message. As more than one association may exist (e.g., in support of different association types or end-to-end and segment recovery), all carried ASSOCIATION objects need to be examined.

- o Covering case 3:

Any ASSOCIATION object with the Association Type field value of "Recovery" (1) that does not yield an association in the prior comparison needs to be checked to see if a case 3 association is indicated. As this case only applies to end-to-end recovery, the first step is to locate any other LSPs with the identical SESSION object fields and the identical tunnel sender address fields as the LSP carrying the ASSOCIATION object. If such LSPs exist, a case 3 association is identified by comparing the value of the Association ID field with the LSP ID field of the other LSP. If

the values are identical, then an end-to-end recovery association exists. As this behavior only applies to end-to-end recovery, this check need only be performed at the egress.

No additional behavior is needed in order to support changes in the set of ASSOCIATION objects included in an LSP, as long as the change represents either a new association or a change in identifiers made as described in Section 2.2.

4. Non-GMPLS Recovery Usage

While the ASSOCIATION object, [RFC4872], is defined in the context of GMPLS Recovery, the object can have wider application. [RFC4872] defines the object to be used to "associate LSPs with each other", and then defines an Association Type field to identify the type of association being identified. It also defines that the Association Type field is to be considered when determining association, i.e., there may be type-specific association rules. As discussed above, this is the case for Recovery type association objects. The text above, notably the text related to resource sharing types, can also be used as the foundation for a generic method for associating LSPs when there is no type-specific association defined.

The remainder of this section defines the general rules to be followed when processing ASSOCIATION objects. Object usage in both Path and Resv messages is discussed. The usage applies equally to GMPLS LSPs [RFC3473], MPLS LSPs [RFC3209] and non-LSP RSVP sessions [RFC2205], [RFC2207], [RFC3175] and [RFC4860]. As described below, association is always done based on matching either Path state or Resv state, but not Path state to Resv State. This section applies to the ASSOCIATION objects defined in [RFC4872].

4.1. Upstream Initiated Association

Upstream initiated association is represented in ASSOCIATION objects carried in Path messages and can be used to associate RSVP Path state across MPLS Tunnels / RSVP sessions. (Note, per [RFC3209] an MPLS tunnel is represented by a RSVP SESSION object, and multiple LSPs may be represented within a single tunnel.) Cross-session association based on Path state is defined in [RFC4872]. This definition is extended by this section, which defined generic association rules and usage for non-LSP uses. This section does not modify processing required to support [RFC4872] and [RFC4873], which is reviewed above in Section 3.

4.1.1. Path Message Format

This section provides the Backus-Naur Form (BNF), see [RFC5511], for Path messages containing ASSOCIATION objects. BNF is provided for both MPLS and for non-LSP session usage. Unmodified RSVP message formats and some optional objects are not listed.

The format for MPLS and GMPLS sessions is unmodified from [RFC4872], and can be represented based on the BNF in [RFC3209] as:

```
<Path Message> ::= <Common Header> [ <INTEGRITY> ]
                   <SESSION> <RSVP_HOP>
                   <TIME_VALUES>
                   [ <EXPLICIT_ROUTE> ]
                   <LABEL_REQUEST>
                   [ <SESSION_ATTRIBUTE> ]
                   [ <ASSOCIATION> ... ]
                   [ <POLICY_DATA> ... ]
                   <sender descriptor>
```

The format for non-LSP sessions as based on the BNF in [RFC2205] is:

```
<Path Message> ::= <Common Header> [ <INTEGRITY> ]
                   <SESSION> <RSVP_HOP>
                   <TIME_VALUES>
                   [ <ASSOCIATION> ... ]
                   [ <POLICY_DATA> ... ]
                   [ <sender descriptor> ]
```

In general, relative ordering of ASSOCIATION objects with respect to each other as well as with respect to other objects is not significant. Relative ordering of ASSOCIATION objects of the same type SHOULD be preserved by transit nodes. Association type specific ordering requirements MAY be defined in the future.

4.1.2. Path Message Processing

This section is based on the processing rules described in [RFC4872] and [RFC4873], which is reviewed above. These procedures apply equally to GMPLS LSPs, MPLS LSPs and non-LSP session state.

A node that wishes to allow downstream nodes to associate Path state across RSVP sessions MUST include an ASSOCIATION object in the outgoing Path messages corresponding to the RSVP sessions to be associated. In the absence of Association Type-specific rules for identifying association, the included ASSOCIATION objects MUST be identical. When there is an Association Type-specific definition of association rules, the definition SHOULD allow for association based on identical ASSOCIATION objects. This document does not define any Association Type-specific rules. (See Section 3 for a discussion of

an example of Association Type-specific rules which are derived from [RFC4872].)

When creating an ASSOCIATION object, the originator MUST format the object as defined in Section 16.1 of [RFC4872]. The originator MUST set the Association Type field based on the type of association being identified. The Association ID field MUST be set to a value that uniquely identifies the sessions to be associated within the context of the Association Source field. The Association Source field MUST be set to a unique address assigned to the node originating the association.

A downstream node can identify an upstream initiated association by performing the following checks. When a node receives a Path message it MUST check each ASSOCIATION object received in the Path message to see if it contains an Association Type field value supported by the node. For each ASSOCIATION object containing a supported association type, the node MUST then check to see if the object matches an ASSOCIATION object received in any other Path message. To perform this matching, a node MUST examine the Path state of all other sessions and compare the fields contained in the newly received ASSOCIATION object with the fields contained in the Path state's ASSOCIATION objects. An association is deemed to exist when the same values are carried in all fields of the ASSOCIATION objects being compared. Processing once an association is identified is type specific and is outside the scope of this document.

Note that as more than one association may exist, all ASSOCIATION objects carried in a received Path message which have supported association types MUST be compared against all Path state.

Unless there are type-specific processing rules, downstream nodes MUST forward all ASSOCIATION objects received in a Path message with any corresponding outgoing Path messages.

4.2. Downstream Initiated Association

Downstream initiated association is represented in ASSOCIATION objects carried in Resv messages and can be used to associate RSVP Resv state across MPLS Tunnels / RSVP sessions. Cross-session association based on Path state is defined in [RFC4872]. This section defines cross-session association based on Resv state. This section places no additional requirements on implementations supporting [RFC4872] and [RFC4873].

4.2.1. Resv Message Format

This section provides the Backus-Naur Form (BNF), see [RFC5511], for Resv messages containing ASSOCIATION objects. BNF is provided for both MPLS and for non-LSP session usage. Unmodified RSVP message formats and some optional objects are not listed.

The format for MPLS, GMPLS and non-LSP sessions are identical, and is represented based on the BNF in [RFC2205] and [RFC3209]:

```
<Resv Message> ::= <Common Header> [ <INTEGRITY> ]
                   <SESSION> <RSVP_HOP>
                   <TIME_VALUES>
                   [ <RESV_CONFIRM> ] [ <SCOPE> ]
                   [ <ASSOCIATION> ... ]
                   [ <POLICY_DATA> ... ]
                   <STYLE> <flow descriptor list>
```

Relative ordering of ASSOCIATION objects with respect to each other as well as with respect to other objects is not currently significant. Relative ordering of ASSOCIATION objects of the same type MUST be preserved by transit nodes. Association type specific ordering requirements MAY be defined in the future.

4.2.2. Resv Message Processing

This section apply equally to GMPLS LSPs, MPLS LSPs and non-LSP session state.

A node that wishes to allow upstream nodes to associate Resv state across RSVP sessions MUST include an ASSOCIATION object in the outgoing Resv messages corresponding to the RSVP sessions to be associated. In the absence of Association Type-specific rules for identifying association, the included ASSOCIATION objects MUST be identical. When there is an Association Type-specific definition of association rules, the definition SHOULD allow for association based on identical ASSOCIATION objects. This document does not define any Association Type-specific rules.

When creating an ASSOCIATION object, the originator MUST format the object as defined in Section 16.1 of [RFC4872]. The originator MUST set the Association Type field based on the type of association being identified. The Association ID field MUST be set to a value that uniquely identifies the sessions to be associated within the context of the Association Source field. The Association Source field MUST be set to a unique address assigned to the node originating the association.

An upstream node can identify a downstream initiated association by performing the following checks. When a node receives a Resv message

it MUST check each ASSOCIATION object received in the Resv message to see if it contains an Association Type field value supported by the node. For each ASSOCIATION object containing a supported association type, the node MUST then check to see if the object matches an ASSOCIATION object received in any other Resv message. To perform this matching, a node MUST examine the Resv state of all other sessions and compare the fields contained in the newly received ASSOCIATION object with the fields contained in the Resv state's ASSOCIATION objects. An association is deemed to exist when the same values are carried in all fields of the ASSOCIATION objects being compared. Processing once an association is identified is type specific and is outside the scope of this document.

Note that as more than one association may exist, all ASSOCIATION objects with support Association Types carried in a received Resv message MUST be compared against all Resv state.

Unless there are type-specific processing rules, upstream nodes MUST forward all ASSOCIATION objects received in a Resv message with any corresponding outgoing Resv messages.

4.3. Association Types

Two association types are currently defined: recovery and resource sharing. Recovery type association is only applicable within the context of recovery, [RFC4872] and [RFC4873]. Resource sharing is generally useful and its general use is defined in this section.

4.3.1. Resource Sharing Association Type

The resource sharing association type was defined in [RFC4873] and was defined within the context of GMPLS and upstream initiated association. This section presents a definition of the resource sharing association that allows for its use with any RSVP session type and in both Path and Resv messages. This definition is consistent with the definition of the resource sharing association type in [RFC4873] and no changes are required by this section in order to support [RFC4873]. The Resource Sharing Association Type MUST be supported by any implementation compliant with this document.

The Resource Sharing Association Type is used to enable resource sharing across RSVP sessions. Per [RFC4873], Resource Sharing uses the Association Type field value of 2. ASSOCIATION objects with an Association Type with the value Resource Sharing MAY be carried in Path and Resv messages. Association for the Resource Sharing type MUST follow the procedures defined in Section 4.1.2 for upstream (Path message) initiated association and Section 4.2.1 for downstream (Resv message) initiated association. There are no type-specific association rules, processing rules, or ordering requirements. Note

that as is always the case with association as enabled by this document, no associations are made across Path and Resv state.

Once an association is identified, resources SHOULD be shared across the identified sessions. Resource sharing is discussed in general in [RFC2205] and within the context of LSPs in [RFC3209].

5. IPv4 and IPv6 Extended ASSOCIATION Objects

[RFC4872] defines the IPv4 ASSOCIATION object and the IPv6 ASSOCIATION object. As defined, these objects each contain an Association Source field and a 16-bit Association ID field. The combination of the Association Source and the Association ID uniquely identifies the association. Because the association-ID field is a 16-bit field, an association source can allocate up to 65536 different associations and no more. There are scenarios where this number is insufficient. (For example where the association identification is best known and identified by a fairly centralized entity, which therefore may be involved in a large number of associations.)

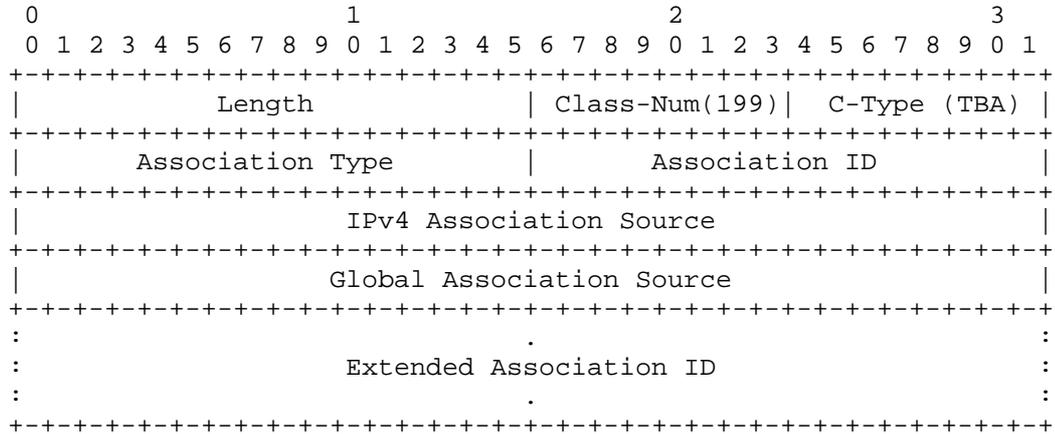
Furthermore, per [TP-IDENTIFIERS], MPLS-TP LSPs can be identified in two forms that cannot be supported using the existing ASSOCIATION objects. The first form is a global identifier and the second uses an ITU Carrier Code (ICC). The [TP-IDENTIFIERS] defined "global identifier", or Global_ID, is based on [RFC5003] and includes the operator's Autonomous System Number (ASN). [TP-IDENTIFIERS] identifies the ICC as "a string of one to six characters, each character being either alphabetic (i.e. A-Z) or numeric (i.e. 0-9) characters. Alphabetic characters in the ICC SHOULD be represented with upper case letters."

This sections defines new ASSOCIATION objects to support extended identification in order to address the limitations described above. Specifically, the IPv4 Extended ASSOCIATION object and IPv6 Extended ASSOCIATION object are defined below. Both new objects include the fields necessary to enable identification of a larger number of associations, as well as MPLS-TP required identification.

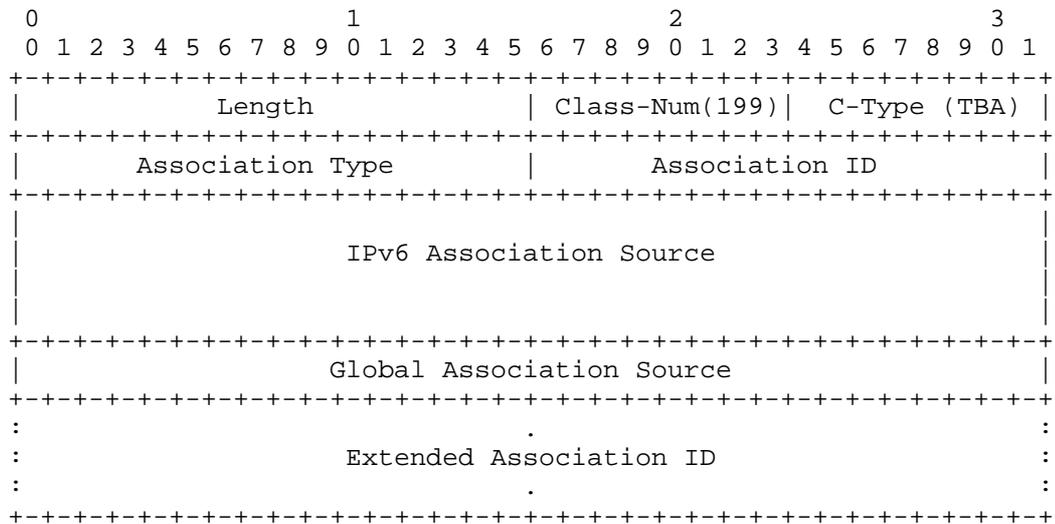
The IPv4 Extended ASSOCIATION object and IPv6 Extended ASSOCIATION object SHOULD be supported by an implementation compliant with this document. The processing rules for the IPv4 and IPv6 Extended ASSOCIATION object are described below, and are based on the rules for the IPv4 and IPv6 ASSOCIATION objects as described above.

5.1. IPv4 and IPv6 Extended ASSOCIATION Object Format

The IPv4 Extended ASSOCIATION object (Class-Num of the form 11bbbbbb with value = 199, C-Type = TBA) has the format:



The IPv6 Extended ASSOCIATION object (Class-Num of the form 11bbbbbb with value = 199, C-Type = TBA) has the format:



Association Type: 16 bits

Same as for IPv4 and IPv6 ASSOCIATION objects, see [RFC4872].

Association ID: 16 bits

Same as for IPv4 and IPv6 ASSOCIATION objects, see [RFC4872].

Association Source: 4 or 16 bytes

Same as for IPv4 and IPv6 ASSOCIATION objects, see [RFC4872].

Global Association Source: 4 bytes

This field contains a value that is unique to the provider, i.e., a global identifier. This field MAY contain the 2-octet or 4-octet value of the provider's Autonomous System Number (ASN). It is expected that the global identifier will be derived from the globally unique ASN of the autonomous system hosting the Association Source. The special value of zero (0) indicates that no global identifier is present. Note that a Global Association Source of zero SHOULD be limited to entities contained within a single operator.

If the Global Association Source field value is derived from a 2-octet AS number, then the two high-order octets of this 4-octet field MUST be set to zero.

Please note that, as stated in [TP-IDENTIFIERS], the use of the provider's ASN as a global identifier DOES NOT have anything at all to do with the use of the ASN in protocols such as BGP.

This field is based on the definition of Global_ID defined in [RFC5003] and used by [TP-IDENTIFIERS].

Extended Association ID: variable, 4-byte aligned

This field contains data that is additional information to support unique identification. The length and contents of this field is determined by the Association Source. This field MAY be omitted, i.e., have a zero length. This field MUST be padded with zeros (0s) to ensure 32-bit alignment.

5.2. Processing

The processing of a IPv4 or IPv6 Extended ASSOCIATION object MUST be identical to the processing of a IPv4 or IPv6 ASSOCIATION object as described above in Section 4 except as extended by this section. This section applies to both upstream-initiated (Path message) and downstream-initiated (Resv message) association.

The following are the modified procedures for Extended ASSOCIATION object processing:

- o When creating an Extended ASSOCIATION object, the originator MUST format the object as defined in this document.

- o The originator MUST set the Association Type, Association ID and Association Source fields as described in Section 4.
- o When ASN-based global identification of the Association Source is desired, the originator MUST set the Global Association Source field. When ASN-based global identification is not desired, the originator MUST set the Global Association Source field to zero (0).
- o The Extended ASSOCIATION object originator MAY include the Extended Association ID field. The field is included based on local policy. The field MUST be included when the Association ID field is insufficient to uniquely identify association within the scope of the source of the association. When included, this field MUST be set to a value that, when taken together with the other fields in the object, uniquely identifies the sessions to be associated.

When used in support of ICC identified (MPLS-TP) LSPs, this field MUST be at least eight (8) bytes long, and MAY be longer; the first six (6) bytes MUST be set to the ICC as defined in Section 3.2 of [TP-IDENTIFIERS] and the next two bytes MUST be set to zero (0). For non-ICC identified MPLS-TP LSPs, this field MUST either be omitted, or MUST have the first 6 bytes set to all zeros (0s).

- o The object Length field is set based on the length of the Extended Association ID field. When the Extended Association ID field is omitted, the object Length field MUST be set to 16 or 28 for the IPv4 and IPv6 ASSOCIATION objects, respectively. When the Extended Association ID field is present, the object Length field MUST be set to indicate the additional bytes carried in the Extended Association ID field, including pad bytes.

Note: per [RFC2205], the object Length field is set to the total object length in bytes, and is always a multiple of 4, and at least 4.

Identification of association is not modified by this section. It is important to note that Section 4 defines association identification based on ASSOCIATION object matching, and that such matching is based on the comparison of all fields in a ASSOCIATION object (unless type-specific comparison rules are defined). This applies equally to ASSOCIATION objects and Extended ASSOCIATION objects.

6. Security Considerations

A portion of this document reviews procedures defined in [RFC4872] and [RFC4873] and does not define any new procedures. As such, no new security considerations are introduced in this portion.

Section 4 defines broader usage of the ASSOCIATION object, but does not fundamentally expand on the association function that was previously defined in [RFC4872] and [RFC4873]. Section 5 increases the number of bits that are carried in an ASSOCIATION object (by 32), and similarly does not expand on the association function that was previously defined. This broader definition does allow for additional information to be conveyed, but this information is not fundamentally different from the information that is already carried in RSVP. Therefore there are no new risks or security considerations introduced by this document.

For a general discussion on MPLS and GMPLS related security issues, see the MPLS/GMPLS security framework [RFC5920].

7. IANA Considerations

IANA is requested to administer assignment of new values for namespaces defined in this document and summarized in this section.

7.1. IPv4 and IPv6 Extended ASSOCIATION Objects

Upon approval of this document, IANA will make the assignment of two new C-Types (which are defined in section 5.1) for the existing ASSOCIATION object in the "Class Names, Class Numbers, and Class Types" section of the "Resource Reservation Protocol (RSVP) Parameters" registry located at <http://www.iana.org/assignments/rsvp-parameters>:

199 ASSOCIATION [RFC4872]

Class Types or C-Types

3	Type 3 IPv4 Extended Association	[this document]
4	Type 4 IPv6 Extended Association	[this document]

7.2. Resource Sharing Association Type

This document also broadens the potential usage of the Resource Sharing Association Type defined in [RFC4873]. As such, IANA is requested to change the Reference of the Resource Sharing Association Type included in the associate registry. This document also directs IANA to correct the duplicate usage of '(R)' in this Registry. In particular, the Association Type registry found at <http://www.iana.org/assignments/gmpls-sig-parameters/> should be updated as follows:

OLD:		
2	Resource Sharing (R)	[RFC4873]
NEW		
2	Resource Sharing (S)	[RFC4873][this-document]

There are no other IANA considerations introduced by this document.

8. Acknowledgments

Sections 2 and 3 of this document formalizes the explanation provided in an e-mail to the working group authored by Adrian Farrel, see [AF-EMAIL]. This portion of the document was written in response to questions raised in the CCAMP working group by Nic Neate <nhn@dataconnection.com>. Valuable comments and input was also received from Dimitri Papadimitriou.

We thank Subha Dhesikan for her contribution to the early work on sharing of resources across RSVP reservations.

9. References

9.1. Normative References

- [RFC2205] Braden, R., Zhang, L., Berson, S., Herzog, S. and S. Jamin, "Resource ReSerVation Protocol (RSVP) -- Version 1, Functional Specification", RFC 2205, September 1997.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC4872] Lang, J., Rekhter, Y., and Papadimitriou, D., "RSVP-TE Extensions in Support of End-to-End Generalized Multi-Protocol Label Switching (GMPLS) Recovery", RFC 4872, May 2007.

- [RFC4873] Berger, L., Bryskin, I., Papadimitriou, D., Farrel, A., "GMPLS Segment Recovery", RFC 4873, May 2007.
- [RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, December 2001.
- [RFC3473] Berger, L., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions", RFC 3473, January 2003.
- [RFC5511] Farrel, A., "Routing Backus-Naur Form (RBNF): A Syntax Used to Form Encoding Rules in Various Routing Protocol Specifications", RFC 5511, April 2009

9.2. Informative References

- [AF-EMAIL] Farrel, A. "Re: Clearing up your misunderstanding of the Association ID", CCAMP working group mailing list, <http://www.ietf.org/mail-archive/web/ccamp/current/msg00644.html>, November 18, 2008.
- [RFC2207] Berger., L., O'Malley., T., "RSVP Extensions for IPSEC RSVP Extensions for IPSEC Data Flows", RFC 2207, September 1997.
- [RFC3175] Baker, F., Iturralde, C., Le, F., Davie, B., "Aggregation of RSVP for IPv4 and IPv6 Reservations", RFC 3175, September 2001.
- [RFC4860] Le, F., Davie, B., Bose, P., Christou, C., Davenport, M., "Generic Aggregate Resource ReSerVation Protocol (RSVP) Reservations", RFC 4860, May 2007.
- [RFC5003] Metz, C., Martini, L., Balus, F., Sugimoto, J., "Attachment Individual Identifier (AII) Types for Aggregation", RFC 5003, September 2007.
- [RFC5389] Rosenberg, J., Mahy, R., Matthews, P., Wing, D., "Session Traversal Utilities for NAT (STUN)", RFC 5389, October 2008.
- [RFC5920] Fang, L., et al, "Security Framework for MPLS and GMPLS Networks", work in progress, RFC 5920, July 2010.
- [TP-IDENTIFIERS] Bocci, M., Swallow, G., Gray, E., "MPLS-TP Identifiers", work in progress, draft-ietf-mpls-tp-identifiers.

10. Authors' Addresses

Lou Berger
LabN Consulting, L.L.C.
Phone: +1-301-468-9228
Email: lberger@labn.net

Francois Le Faucheur
Cisco Systems
Greenside, 400 Avenue de Roumanille
Sophia Antipolis 06410
France
Email: flefauch@cisco.com

Ashok Narayanan
Cisco Systems
300 Beaver Brook Road
Boxborough, MA 01719
United States
Email: ashokn@cisco.com

Generated on: Mon, Mar 14, 2011 7:36:53 AM

Network Working Group
Internet Draft
Intended status: Standards Track
Expires: June 2011

G. Bernstein
Grotto Networking
Y. Lee
D. Li
Huawei
W. Imajuku
NTT

December 1, 2010

General Network Element Constraint Encoding for GMPLS Controlled
Networks

draft-ietf-ccamp-general-constraint-encode-04.txt

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at <http://www.ietf.org/ietf/lid-abstracts.txt>

The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>

This Internet-Draft will expire on June 1, 2011.

Copyright Notice

Copyright (c) 2010 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Abstract

Generalized Multiprotocol Label Switching can be used to control a wide variety of technologies. In some of these technologies network elements and links may impose additional routing constraints such as asymmetric switch connectivity, non-local label assignment, and label range limitations on links.

This document provides efficient, protocol-agnostic encodings for general information elements representing connectivity and label constraints as well as label availability. It is intended that protocol-specific documents will reference this memo to describe how information is carried for specific uses.

Conventions used in this document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC-2119 [RFC2119].

Table of Contents

1. Introduction.....	3
1.1. Node Switching Asymmetry Constraints.....	3
1.2. Non-Local Label Assignment Constraints.....	4
1.3. Change Log.....	5
2. Encoding.....	5
2.1. Link Set Field.....	5
2.2. Label Set Field.....	7
2.2.1. Inclusive/Exclusive Label Lists.....	8
2.2.2. Inclusive/Exclusive Label Ranges.....	9
2.2.3. Bitmap Label Set.....	9
2.3. Available Labels Sub-TLV.....	10

2.4. Shared Backup Labels Sub-TLV.....	11
2.5. Connectivity Matrix Sub-TLV.....	11
2.6. Port Label Restriction sub-TLV.....	12
2.6.1. SIMPLE_LABEL.....	13
2.6.2. CHANNEL_COUNT.....	14
2.6.3. LABEL_RANGE1.....	14
2.6.4. SIMPLE_LABEL & CHANNEL_COUNT.....	15
2.6.5. Link Label Exclusivity.....	15
3. Security Considerations.....	15
4. IANA Considerations.....	16
5. Acknowledgments.....	16
APPENDIX A: Encoding Examples.....	17
A.1. Link Set Field.....	17
A.2. Label Set Field.....	17
A.3. Connectivity Matrix Sub-TLV.....	18
A.4. Connectivity Matrix with Bi-directional Symmetry.....	21
6. References.....	24
6.1. Normative References.....	24
6.2. Informative References.....	24
7. Contributors.....	25
Authors' Addresses.....	26
Intellectual Property Statement.....	27
Disclaimer of Validity.....	27

1. Introduction

Some data plane technologies that wish to make use of a GMPLS control plane contain additional constraints on switching capability and label assignment. In addition, some of these technologies must perform non-local label assignment based on the nature of the technology, e.g., wavelength continuity constraint in WSON [WSON-Frame]. Such constraints can lead to the requirement for link by link label availability in path computation and label assignment.

This document provides efficient encodings of information needed by the routing and label assignment process in technologies such as WSON and are potentially applicable to a wider range of technologies. Such encodings can be used to extend GMPLS signaling and routing protocols. In addition these encodings could be used by other mechanisms to convey this same information to a path computation element (PCE).

1.1. Node Switching Asymmetry Constraints

For some network elements the ability of a signal or packet on a particular ingress port to reach a particular egress port may be

limited. In addition, in some network elements the connectivity between some ingress ports and egress ports may be fixed, e.g., a simple multiplexer. To take into account such constraints during path computation we model this aspect of a network element via a connectivity matrix.

The connectivity matrix (ConnectivityMatrix) represents either the potential connectivity matrix for asymmetric switches or fixed connectivity for an asymmetric device such as a multiplexer. Note that this matrix does not represent any particular internal blocking behavior but indicates which ingress ports and labels (e.g., wavelengths) could possibly be connected to a particular output port. Representing internal state dependent blocking for a node is beyond the scope of this document and due to its highly implementation dependent nature would most likely not be subject to standardization in the future. The connectivity matrix is a conceptual M by N matrix representing the potential switched or fixed connectivity, where M represents the number of ingress ports and N the number of egress ports.

1.2. Non-Local Label Assignment Constraints

If the nature of the equipment involved in a network results in a requirement for non-local label assignment we can have constraints based on limits imposed by the ports themselves and those that are implied by the current label usage. Note that constraints such as these only become important when label assignment has a non-local character. For example in MPLS an LSR may have a limited range of labels available for use on an egress port and a set of labels already in use on that port and hence unavailable for use. This information, however, does not need to be shared unless there is some limitation on the LSR's label swapping ability. For example if a TDM node lacks the ability to perform time-slot interchange or a WSON lacks the ability to perform wavelength conversion then the label assignment process is not local to a single node and it may be advantageous to share the label assignment constraint information for use in path computation.

Port label restrictions (PortLabelRestriction) model the label restrictions that the network element (node) and link may impose on a port. These restrictions tell us what labels may or may not be used on a link and are intended to be relatively static. More dynamic information is contained in the information on available labels. Port label restrictions are specified relative to the port in general or

to a specific connectivity matrix for increased modeling flexibility. Reference [Switch] gives an example where both switch and fixed connectivity matrices are used and both types of constraints occur on the same port.

1.3. Change Log

Changes from 03 version:

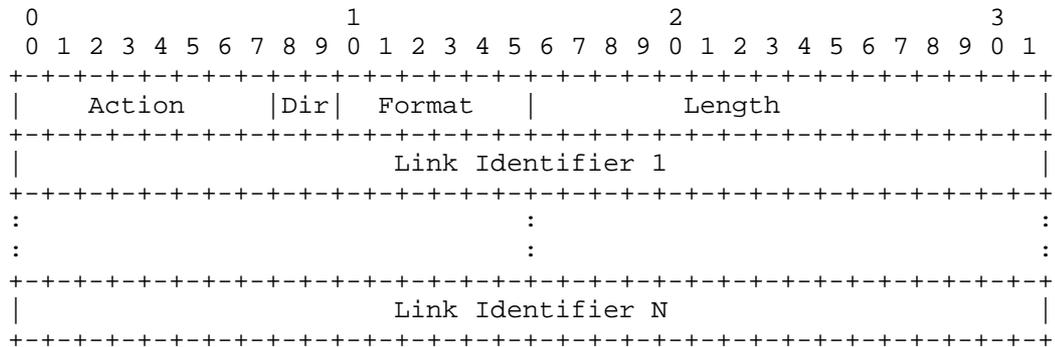
- (a) Removed informational BNF from section 1.
- (b) Removed section on "Extension Encoding Usage Recommendations"

2. Encoding

A type-length-value (TLV) encoding of the general connectivity and label restrictions and availability extensions is given in this section. This encoding is designed to be suitable for use in the GMPLS routing protocols OSPF [RFC4203] and IS-IS [RFC5307] and in the PCE protocol PCEP [PCEP]. Note that the information distributed in [RFC4203] and [RFC5307] is arranged via the nesting of sub-TLVs within TLVs and this document makes use of such constructs. First, however we define two general purpose fields that will be used repeatedly in the subsequent TLVs.

2.1. Link Set Field

We will frequently need to describe properties of groups of links. To do so efficiently we can make use of a link set concept similar to the label set concept of [RFC3471]. This Link Set Field is used in the <ConnectivityMatrix> sub-TLV, which is defined in Section 2.5. The information carried in a Link Set is defined by:



Action: 8 bits

0 - Inclusive List

Indicates that one or more link identifiers are included in the Link Set. Each identifies a separate link that is part of the set.

1 - Inclusive Range

Indicates that the Link Set defines a range of links. It contains two link identifiers. The first identifier indicates the start of the range (inclusive). The second identifier indicates the end of the range (inclusive). All links with numeric values between the bounds are considered to be part of the set. A value of zero in either position indicates that there is no bound on the corresponding portion of the range. Note that the Action field can be set to 0x02(Inclusive Range) only when unnumbered link identifier is used.

Dir: Directionality of the Link Set (2 bits)

0 -- bidirectional

1 -- ingress

2 -- egress

For example in optical networks we think in terms of unidirectional as well as bidirectional links. For example, label restrictions or connectivity may be different for an ingress port, than for its "companion" egress port if one exists. Note that "interfaces" such as those discussed in the Interfaces MIB [RFC2863] are assumed to be

bidirectional. This also applies to the links advertised in various link state routing protocols.

Format: The format of the link identifier (6 bits)

0 -- Link Local Identifier

Indicates that the links in the Link Set are identified by link local identifiers. All link local identifiers are supplied in the context of the advertising node.

1 -- Local Interface IPv4 Address

2 -- Local Interface IPv6 Address

Indicates that the links in the Link Set are identified by Local Interface IP Address. All Local Interface IP Address are supplied in the context of the advertising node.

Others TBD.

Note that all link identifiers in the same list must be of the same type.

Length: 16 bits

This field indicates the total length in bytes of the Link Set field.

Link Identifier: length is dependent on the link format

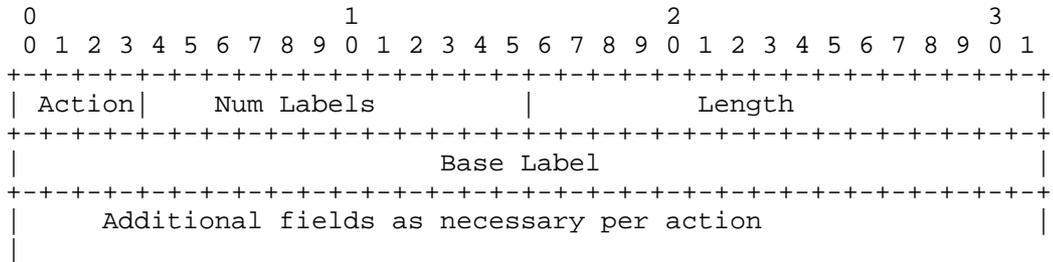
The link identifier represents the port which is being described either for connectivity or label restrictions. This can be the link local identifier of [RFC4202], GMPLS routing, [RFC4203] GMPLS OSPF routing, and [RFC5307] IS-IS GMPLS routing. The use of the link local identifier format can result in more compact encodings when the assignments are done in a reasonable fashion.

2.2. Label Set Field

Label Set Field is used within the <AvailableLabels> sub-TLV or the <SharedBackupLabels> sub-TLV, which is defined in Section 2.3. and 2.4. , respectively.

The general format for a label set is given below. This format uses the Action concept from [RFC3471] with an additional Action to define

a "bit map" type of label set. The second 32 bit field is a base label used as a starting point in many of the specific formats.



Action:

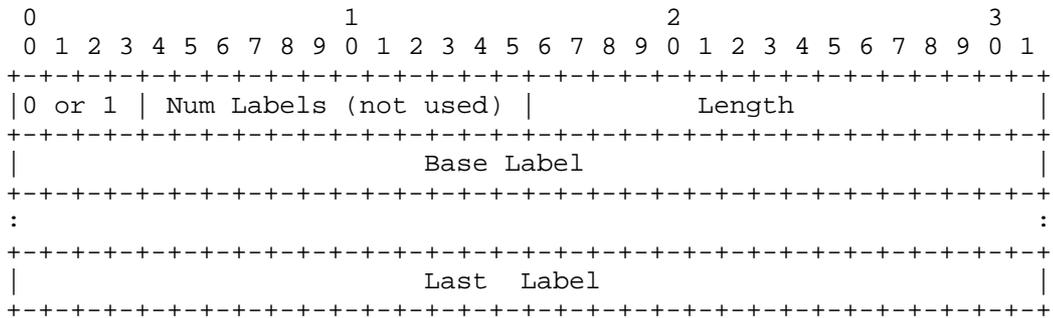
- 0 - Inclusive List
- 1 - Exclusive List
- 2 - Inclusive Range
- 3 - Exclusive Range
- 4 - Bitmap Set

Num Labels is only meaningful for Action value of 4 (Bitmap Set). It indicates the number of labels represented by the bit map. See more detail in section 3.2.3.

Length is the length in bytes of the entire field.

2.2.1. Inclusive/Exclusive Label Lists

In the case of the inclusive/exclusive lists the wavelength set format is given by:

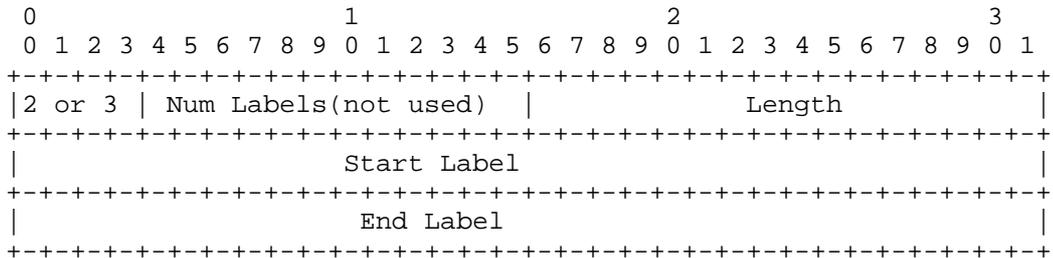


Where:

Num Labels is not used in this particular format since the Length parameter is sufficient to determine the number of labels in the list.

2.2.2. Inclusive/Exclusive Label Ranges

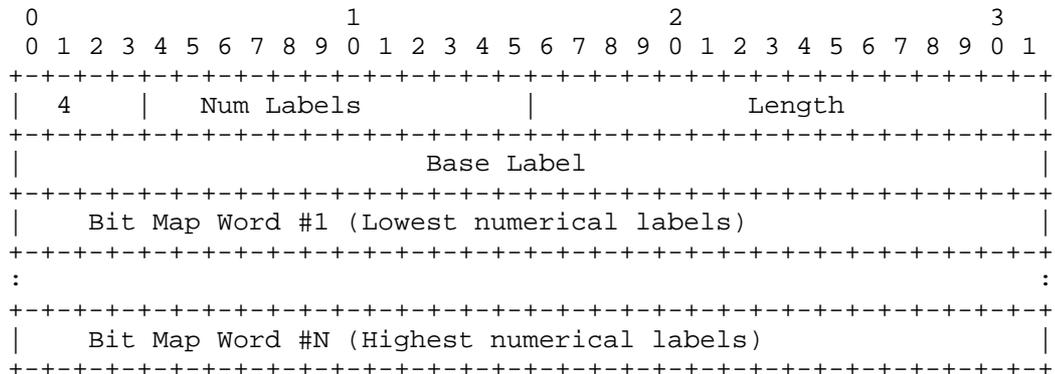
In the case of inclusive/exclusive ranges the label set format is given by:



Note that the start and end label must in some sense "compatible" in the technology being used.

2.2.3. Bitmap Label Set

In the case of Action = 4, the bitmap the label set format is given by:

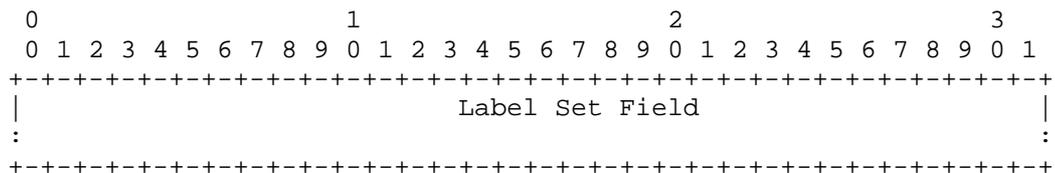


Where Num Labels in this case tells us the number of labels represented by the bit map. Each bit in the bit map represents a particular label with a value of 1/0 indicating whether the label is in the set or not. Bit position zero represents the lowest label and corresponds to the base label, while each succeeding bit position represents the next label logically above the previous.

The size of the bit map is Num Label bits, but the bit map is padded out to a full multiple of 32 bits so that the TLV is a multiple of four bytes. Bits that do not represent labels (i.e., those in positions (Num Labels) and beyond SHOULD be set to zero and MUST be ignored.

2.3. Available Labels Sub-TLV

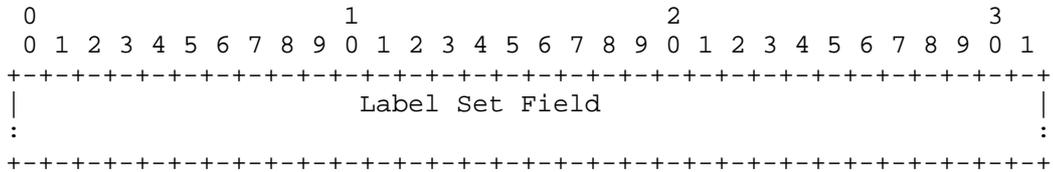
To indicate the labels available for use on a link the Available Labels sub-TLV consists of a single variable length label set field as follows:



Note that Label Set Field is defined in Section 3.2.

2.4. Shared Backup Labels Sub-TLV

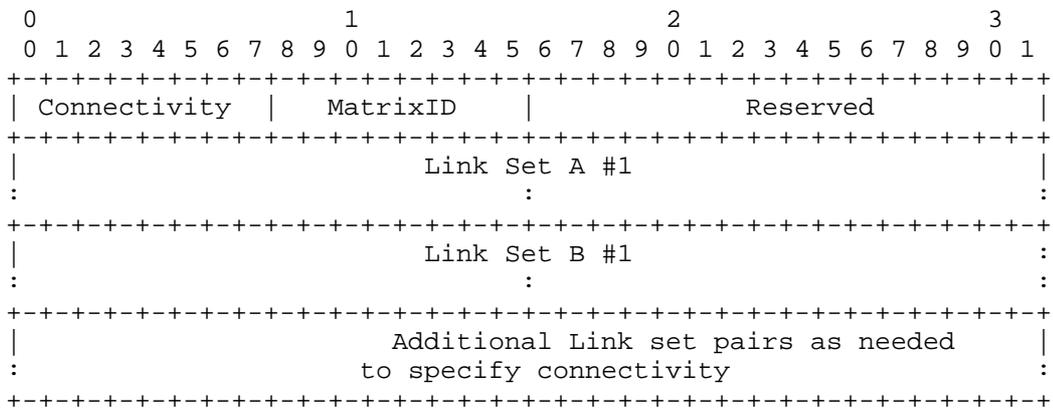
To indicate the labels available for shared backup use on a link the Shared Backup Labels sub-TLV consists of a single variable length label set field as follows:



2.5. Connectivity Matrix Sub-TLV

The Connectivity Matrix represents how ingress ports are connected to egress ports for network elements. The switch and fixed connectivity matrices can be compactly represented in terms of a minimal list of ingress and egress port set pairs that have mutual connectivity. As described in [Switch] such a minimal list representation leads naturally to a graph representation for path computation purposes that involves the fewest additional nodes and links.

A TLV encoding of this list of link set pairs is:



Where

Connectivity is the device type.

0 -- the device is fixed

1 -- the device is switched(e.g., ROADM/OXC)

MatrixID represents the ID of the connectivity matrix and is an 8 bit integer. The value of 0xFF is reserved for use with port wavelength constraints and should not be used to identify a connectivity matrix.

Link Set A #1 and Link Set B #1 together represent a pair of link sets. There are two permitted combinations for the link set field parameter "dir" for Link Set A and B pairs:

- o Link Set A dir=ingress, Link Set B dir=egress

The meaning of the pair of link sets A and B in this case is that any signal that ingresses a link in set A can be potentially switched out of an egress link in set B.

- o Link Set A dir=bidirectional, Link Set B dir=bidirectional

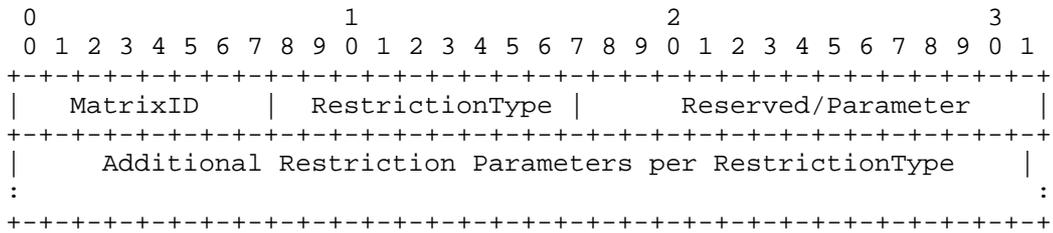
The meaning of the pair of link sets A and B in this case is that any signal that ingresses on the links in set A can potentially egress on a link in set B, and any ingress signal on the links in set B can potentially egress on a link in set A.

See Appendix A for both types of encodings as applied to a ROADM example.

2.6. Port Label Restriction sub-TLV

Port Label Restriction tells us what labels may or may not be used on a link.

The port label restriction of section 1.2. can be encoded as a sub-TLV as follows. More than one of these sub-TLVs may be needed to fully specify a complex port constraint. When more than one of these sub-TLVs are present the resulting restriction is the intersection of the restrictions expressed in each sub-TLV. To indicate that a restriction applies to the port in general and not to a specific connectivity matrix use the reserved value of 0xFF for the MatrixID.



Where:

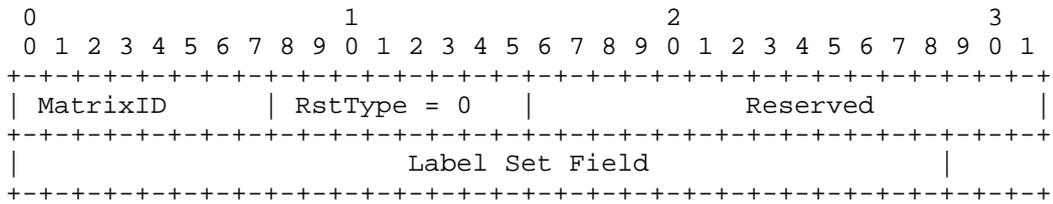
MatrixID: either is the value in the corresponding Connectivity Matrix sub-TLV or takes the value 0xFF to indicate the restriction applies to the port regardless of any Connectivity Matrix.

RestrictionType can take the following values and meanings:

- 0: SIMPLE_LABEL (Simple label selective restriction)
- 1: CHANNEL_COUNT (Channel count restriction)
- 2: LABEL_RANGE1 (Label range device with a movable center label and width)
- 3: SIMPLE_LABEL & CHANNEL_COUNT (Combination of SIMPLE_LABEL and CHANNEL_COUNT restriction. The accompanying label set and channel count indicate labels permitted on the port and the maximum number of channels that can be simultaneously used on the port)
- 4: LINK_LABEL_EXCLUSIVITY (A label may be used at most once amongst a set of specified ports)

2.6.1. SIMPLE_LABEL

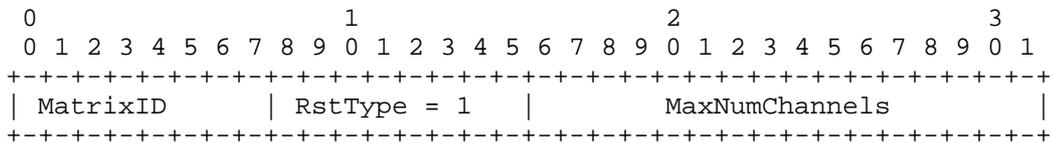
In the case of the SIMPLE_LABEL the GeneralPortRestrictions (or MatrixSpecificRestrictions) format is given by:



In this case the accompanying label set indicates the labels permitted on the port.

2.6.2. CHANNEL_COUNT

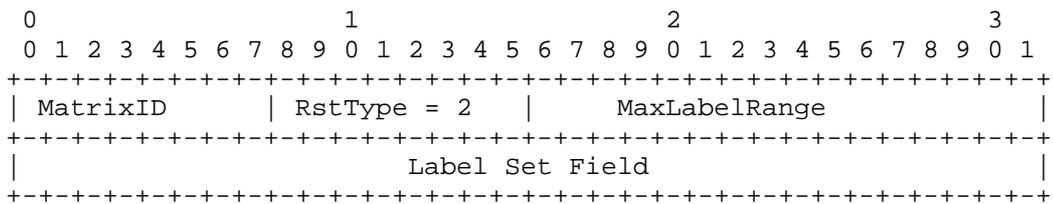
In the case of the CHANNEL_COUNT the format is given by:



In this case the accompanying MaxNumChannels indicates the maximum number of channels (labels) that can be simultaneously used on the port/matrix.

2.6.3. LABEL_RANGE1

In the case of the LABEL_RANGE1 the GeneralPortRestrictions (or MatrixSpecificRestrictions) format is given by:

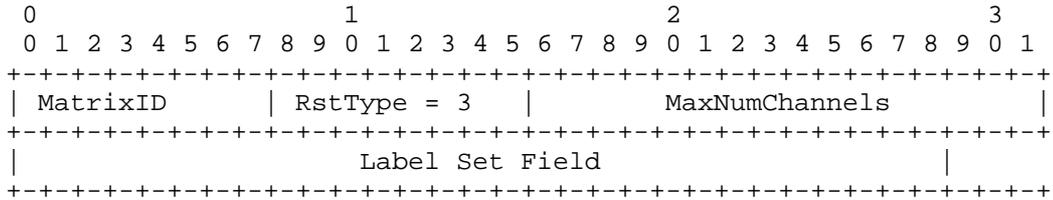


In this case the accompanying MaxLabelRange indicates the maximum range of the labels. The corresponding label set is used to indicate the overall label range. Specific center label information can be obtained from dynamic label in use information. It is assumed that

both center label and range tuning can be done without causing faults to existing signals.

2.6.4. SIMPLE_LABEL & CHANNEL_COUNT

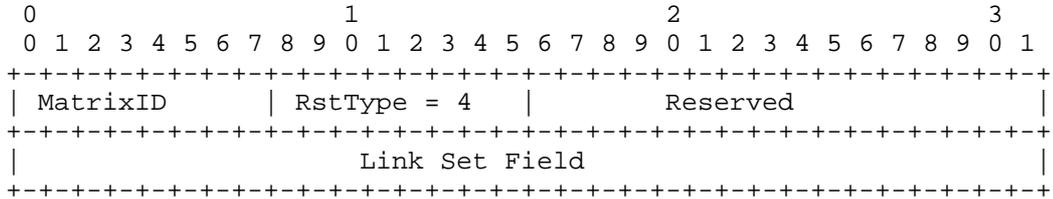
In the case of the SIMPLE_LABEL & CHANNEL_COUNT the format is given by:



In this case the accompanying label set and MaxNumChannels indicate labels permitted on the port and the maximum number of labels that can be simultaneously used on the port.

2.6.5. Link Label Exclusivity

In the case of the SIMPLE_LABEL & CHANNEL_COUNT the format is given by:



In this case the accompanying port set indicate that a label may be used at most once among the ports in the link set field.

3. Security Considerations

This document defines protocol-independent encodings for WSON information and does not introduce any security issues.

However, other documents that make use of these encodings within protocol extensions need to consider the issues and risks associated with, inspection, interception, modification, or spoofing of any of this information. It is expected that any such documents will describe the necessary security measures to provide adequate protection.

4. IANA Considerations

TBD. Once our approach is finalized we may need identifiers for the various TLVs and sub-TLVs.

5. Acknowledgments

This document was prepared using 2-Word-v2.0.template.dot.

APPENDIX A: Encoding Examples

Here we give examples of the general encoding extensions applied to some simple ROADM network elements and links.

A.1. Link Set Field

Suppose that we wish to describe a set of ingress ports that are have link local identifiers number 3 through 42. In the link set field we set the Action = 1 to denote an inclusive range; the Dir = 1 to denote ingress links; and, the Format = 0 to denote link local identifiers. In particular we have:

```

+-----+
| Action=1      | 0 1|0 0 0 0 0 0|                               Length = 12      |
+-----+-----+
|                               Link Local Identifier = #3                               |
+-----+-----+
|                               Link Local Identifier = #42                               |
+-----+-----+

```

A.2. Label Set Field

Example:

A 40 channel C-Band DWDM system with 100GHz spacing with lowest frequency 192.0THz (1561.4nm) and highest frequency 195.9THz (1530.3nm). These frequencies correspond to n = -11, and n = 28 respectively. Now suppose the following channels are available:

Frequency (THz)	n Value	bit map position
192.0	-11	0
192.5	-6	5
193.1	0	11
193.9	8	19
194.0	9	20
195.2	21	32
195.8	27	38

With the Grid value set to indicate an ITU-T G.694.1 DWDM grid, C.S. set to indicate 100GHz this lambda bit map set would then be encoded as follows:

```

0                               1                               2                               3
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-----+-----+-----+-----+-----+-----+-----+-----+
| 4 | Num Wavelengths = 40 | Length = 16 bytes |
+-----+-----+-----+-----+-----+-----+-----+-----+
|Grid | C.S. | Reserved | n for lowest frequency = -11 |
+-----+-----+-----+-----+-----+-----+-----+-----+
|1 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 |
+-----+-----+-----+-----+-----+-----+-----+-----+
|1 0 0 0 0 0 1 0 | Not used in 40 Channel system (all zeros) |
+-----+-----+-----+-----+-----+-----+-----+-----+

```

To encode this same set as an inclusive list we would have:

```

0                               1                               2                               3
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-----+-----+-----+-----+-----+-----+-----+-----+
| 0 | Num Wavelengths = 40 | Length = 20 bytes |
+-----+-----+-----+-----+-----+-----+-----+-----+
|Grid | C.S. | Reserved | n for lowest frequency = -11 |
+-----+-----+-----+-----+-----+-----+-----+-----+
|Grid | C.S. | Reserved | n for lowest frequency = -6 |
+-----+-----+-----+-----+-----+-----+-----+-----+
|Grid | C.S. | Reserved | n for lowest frequency = -0 |
+-----+-----+-----+-----+-----+-----+-----+-----+
|Grid | C.S. | Reserved | n for lowest frequency = 8 |
+-----+-----+-----+-----+-----+-----+-----+-----+
|Grid | C.S. | Reserved | n for lowest frequency = 9 |
+-----+-----+-----+-----+-----+-----+-----+-----+
|Grid | C.S. | Reserved | n for lowest frequency = 21 |
+-----+-----+-----+-----+-----+-----+-----+-----+
|Grid | C.S. | Reserved | n for lowest frequency = 27 |
+-----+-----+-----+-----+-----+-----+-----+-----+

```

A.3. Connectivity Matrix Sub-TLV

Example:

Suppose we have a typical 2-degree 40 channel ROADM. In addition to its two line side ports it has 80 add and 80 drop ports. The picture below illustrates how a typical 2-degree ROADM system that works with bi-directional fiber pairs is a highly asymmetrical system composed of two unidirectional ROADM subsystems.


```

      0           1           2           3
      0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|   Conn = 1   |   MatrixID   |   Reserved   |1
+-----+-----+-----+-----+-----+-----+-----+
                        Note: adds to line
+-----+-----+-----+-----+-----+-----+-----+-----+
| Action=1     |0 1|0 0 0 0 0 0|           Length = 12   |2
+-----+-----+-----+-----+-----+-----+-----+
|                               Link Local Identifier = #3   |3
+-----+-----+-----+-----+-----+-----+-----+
|                               Link Local Identifier = #42  |4
+-----+-----+-----+-----+-----+-----+-----+
| Action=0     |1 0|0 0 0 0 0 0|           Length = 8     |5
+-----+-----+-----+-----+-----+-----+-----+
|                               Link Local Identifier = #1    |6
+-----+-----+-----+-----+-----+-----+-----+
                        Note: line to drops
+-----+-----+-----+-----+-----+-----+-----+-----+
| Action=0     |0 1|0 0 0 0 0 0|           Length = 8     |7
+-----+-----+-----+-----+-----+-----+-----+
|                               Link Local Identifier = #2    |8
+-----+-----+-----+-----+-----+-----+-----+
| Action=1     |1 0|0 0 0 0 0 0|           Length = 12   |9
+-----+-----+-----+-----+-----+-----+-----+
|                               Link Local Identifier = #3    |10
+-----+-----+-----+-----+-----+-----+-----+
|                               Link Local Identifier = #42  |11
+-----+-----+-----+-----+-----+-----+-----+
                        Note: line to line
+-----+-----+-----+-----+-----+-----+-----+-----+
| Action=0     |0 1|0 0 0 0 0 0|           Length = 8     |12
+-----+-----+-----+-----+-----+-----+-----+
|                               Link Local Identifier = #2    |13
+-----+-----+-----+-----+-----+-----+-----+
| Action=0     |1 0|0 0 0 0 0 0|           Length = 8     |14
+-----+-----+-----+-----+-----+-----+-----+
|                               Link Local Identifier = #1    |15
+-----+-----+-----+-----+-----+-----+-----+
                        Note: adds to line
+-----+-----+-----+-----+-----+-----+-----+-----+
| Action=1     |0 1|0 0 0 0 0 0|           Length = 12   |16
+-----+-----+-----+-----+-----+-----+-----+
|                               Link Local Identifier = #43  |17
+-----+-----+-----+-----+-----+-----+-----+

```

```

|                               Link Local Identifier = #82                               |18
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| Action=0      |1 0|0 0 0 0 0 0|                               Length = 8           |19
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|                               Link Local Identifier = #2                               |20
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|                               Note: line to drops                                     |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| Action=0      |0 1|0 0 0 0 0 0|                               Length = 8           |21
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|                               Link Local Identifier = #1                               |22
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| Action=1      |1 0|0 0 0 0 0 0|                               Length = 12          |23
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|                               Link Local Identifier = #43                               |24
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|                               Link Local Identifier = #82                               |25
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|                               Note: line to line                                     |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| Action=0      |0 1|0 0 0 0 0 0|                               Length = 8           |26
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|                               Link Local Identifier = #1                               |27
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| Action=0      |1 0|0 0 0 0 0 0|                               Length = 8           |28
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|                               Link Local Identifier = #2                               |30
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

```

A.4. Connectivity Matrix with Bi-directional Symmetry

If one has the ability to renumber the ports of the previous example as shown in the next figure then we can take advantage of the bi-directional symmetry and use bi-directional encoding of the connectivity matrix. Note that we set dir=bidirectional in the link set fields.


```

      0           1           2           3
      0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| Conn = 1 | MatrixID | Reserved |1
+-----+-----+-----+-----+-----+-----+-----+-----+
| Add/Drops #3-42 to Line side #1
+-----+-----+-----+-----+-----+-----+-----+-----+
| Action=1 |0 0|0 0 0 0 0 0| Length = 12 |2
+-----+-----+-----+-----+-----+-----+-----+-----+
| Link Local Identifier = #3 |3
+-----+-----+-----+-----+-----+-----+-----+-----+
| Link Local Identifier = #42 |4
+-----+-----+-----+-----+-----+-----+-----+-----+
| Action=0 |0 0|0 0 0 0 0 0| Length = 8 |5
+-----+-----+-----+-----+-----+-----+-----+-----+
| Link Local Identifier = #1 |6
+-----+-----+-----+-----+-----+-----+-----+-----+
| Note: line #2 to add/drops #43-82
+-----+-----+-----+-----+-----+-----+-----+-----+
| Action=0 |0 0|0 0 0 0 0 0| Length = 8 |7
+-----+-----+-----+-----+-----+-----+-----+-----+
| Link Local Identifier = #2 |8
+-----+-----+-----+-----+-----+-----+-----+-----+
| Action=1 |0 0|0 0 0 0 0 0| Length = 12 |9
+-----+-----+-----+-----+-----+-----+-----+-----+
| Link Local Identifier = #43 |10
+-----+-----+-----+-----+-----+-----+-----+-----+
| Link Local Identifier = #82 |11
+-----+-----+-----+-----+-----+-----+-----+-----+
| Note: line to line
+-----+-----+-----+-----+-----+-----+-----+-----+
| Action=0 |0 0|0 0 0 0 0 0| Length = 8 |12
+-----+-----+-----+-----+-----+-----+-----+-----+
| Link Local Identifier = #1 |13
+-----+-----+-----+-----+-----+-----+-----+-----+
| Action=0 |0 0|0 0 0 0 0 0| Length = 8 |14
+-----+-----+-----+-----+-----+-----+-----+-----+
| Link Local Identifier = #2 |15
+-----+-----+-----+-----+-----+-----+-----+-----+

```

6. References

6.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC2863] McCloghrie, K. and F. Kastenholz, "The Interfaces Group MIB", RFC 2863, June 2000.
- [RFC3471] Berger, L., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Functional Description", RFC 3471, January 2003.
- [G.694.1] ITU-T Recommendation G.694.1, "Spectral grids for WDM applications: DWDM frequency grid", June, 2002.
- [RFC4202] Kompella, K., Ed., and Y. Rekhter, Ed., "Routing Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4202, October 2005
- [RFC4203] Kompella, K., Ed., and Y. Rekhter, Ed., "OSPF Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4203, October 2005.

6.2. Informative References

- [G.694.1] ITU-T Recommendation G.694.1, Spectral grids for WDM applications: DWDM frequency grid, June 2002.
- [G.694.2] ITU-T Recommendation G.694.2, Spectral grids for WDM applications: CWDM wavelength grid, December 2003.
- [RFC5307] Kompella, K., Ed., and Y. Rekhter, Ed., "IS-IS Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 5307, October 2008.

[Switch] G. Bernstein, Y. Lee, A. Gavler, J. Martensson, " Modeling WDM Wavelength Switching Systems for Use in GMPLS and Automated Path Computation", Journal of Optical Communications and Networking, vol. 1, June, 2009, pp. 187-195.

[PCEP] Vasseur, JP., Ed. and JL. Le Roux, Ed., "Path Computation Element (PCE) communication Protocol (PCEP) - Version 1", RFC5440.

7. Contributors

Diego Caviglia
Ericsson
Via A. Negrone 1/A 16153
Genoa Italy

Phone: +39 010 600 3736
Email: diego.caviglia@(marconi.com, ericsson.com)

Anders Gavler
Acreo AB
Electrum 236
SE - 164 40 Kista Sweden

Email: Anders.Gavler@acreo.se

Jonas Martensson
Acreo AB
Electrum 236
SE - 164 40 Kista, Sweden

Email: Jonas.Martensson@acreo.se

Itaru Nishioka
NEC Corp.
1753 Simonumabe, Nakahara-ku, Kawasaki, Kanagawa 211-8666
Japan

Phone: +81 44 396 3287
Email: i-nishioka@cb.jp.nec.com

Authors' Addresses

Greg M. Bernstein (ed.)
Grotto Networking
Fremont California, USA

Phone: (510) 573-2237
Email: gregb@grotto-networking.com

Young Lee (ed.)
Huawei Technologies
1700 Alma Drive, Suite 100
Plano, TX 75075
USA

Phone: (972) 509-5599 (x2240)
Email: ylee@huawei.com

Dan Li
Huawei Technologies Co., Ltd.
F3-5-B R&D Center, Huawei Base,
Bantian, Longgang District
Shenzhen 518129 P.R.China

Phone: +86-755-28973237
Email: danli@huawei.com

Wataru Imajuku
NTT Network Innovation Labs
1-1 Hikari-no-oka, Yokosuka, Kanagawa
Japan

Phone: +81-(46) 859-4315
Email: imajuku.wataru@lab.ntt.co.jp

Jianrui Han
Huawei Technologies Co., Ltd.
F3-5-B R&D Center, Huawei Base,
Bantian, Longgang District
Shenzhen 518129 P.R.China

Phone: +86-755-28972916
Email: hanjianrui@huawei.com

Intellectual Property Statement

The IETF Trust takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in any IETF Document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights.

Copies of Intellectual Property disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a 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 any standard or specification contained in an IETF Document. Please address the information to the IETF at ietf-ipr@ietf.org.

Disclaimer of Validity

All IETF Documents and the information contained therein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY, THE IETF TRUST AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION THEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Acknowledgment

Funding for the RFC Editor function is currently provided by the Internet Society.

Network Working Group
Internet Draft
Category: Informational

Fatai Zhang
Dan Li
Huawei
Han Li
CMCC
S. Belotti
Alcatel-Lucent
D. Ceccarelli
Ericsson
March 11, 2011

Expires: September 11, 2011

Framework for GMPLS and PCE Control of
G.709 Optical Transport Networks

draft-ietf-ccamp-gmpls-g709-framework-04.txt

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at <http://www.ietf.org/ietf/lid-abstracts.txt>.

The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>.

This Internet-Draft will expire on September 11, 2011.

Abstract

This document provides a framework to allow the development of protocol extensions to support Generalized Multi-Protocol Label Switching (GMPLS) and Path Computation Element (PCE) control of

Optical Transport Networks (OTN) as specified in ITU-T Recommendation G.709 as consented in October 2009.

Table of Contents

1. Introduction	2
2. Terminology	3
3. G.709 Optical Transport Network (OTN)	4
3.1. OTN Layer Network	4
3.1.1. Client signal mapping	5
3.1.2. Multiplexing ODUj onto Links	7
3.1.2.1. Structure of MSI information	8
4. Connection management in OTN	9
4.1. Connection management of the ODU	10
5. GMPLS/PCE Implications	12
5.1. Implications for LSP Hierarchy with GMPLS TE	12
5.2. Implications for GMPLS Signaling	13
5.3. Implications for GMPLS Routing	16
5.4. Implications for Link Management Protocol (LMP)	18
5.5. Implications for Path Computation Elements	19
6. Data Plane Backward Compatibility Considerations	19
7. Security Considerations	20
8. IANA Considerations	20
9. Acknowledgments	20
10. References	21
10.1. Normative References	21
10.2. Informative References	22
11. Authors' Addresses	23
12. Contributors	24
APPENDIX A: ODU connection examples	25

1. Introduction

OTN has become a mainstream layer 1 technology for the transport network. Operators want to introduce control plane capabilities based on Generalized Multi-Protocol Label Switching (GMPLS) to OTN networks, to realize the benefits associated with a high-function control plane (e.g., improved network resiliency, resource usage efficiency, etc.).

GMPLS extends MPLS to encompass time division multiplexing (TDM) networks (e.g., SONET/SDH, PDH, and G.709 sub-lambda), lambda switching optical networks, and spatial switching (e.g., incoming port or fiber to outgoing port or fiber). The GMPLS architecture is provided in [RFC3945], signaling function and Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) extensions are described in

[RFC3471] and [RFC3473], routing and OSPF extensions are described in [RFC4202] and [RFC4203], and the Link Management Protocol (LMP) is described in [RFC4204].

The GMPLS protocol suite including provision [RFC4328] provides the mechanisms for basic GMPLS control of OTN networks based on the 2001 revision of the G.709 specification [G709-V1]. Later revisions of the G.709 specification, including [G709-V3], have included some new features; for example, various multiplexing structures, two types of TSs (i.e., 1.25Gbps and 2.5Gbps), and extension of the Optical Data Unit (ODU) ODUj definition to include the ODUFlex function.

This document reviews relevant aspects of OTN technology evolution that affect the GMPLS control plane protocols and examines why and how to update the mechanisms described in [RFC4328]. This document additionally provides a framework for the GMPLS control of OTN networks and includes a discussion of the implication for the use of the Path Computation Element (PCE) [RFC4655]. No additional Switching Type and LSP Encoding Type are required to support the control of the evolved OTN, because the Switching Type and LSP Encoding Type defined in [RFC4328] are still applicable.

For the purposes of the control plane the OTN can be considered as being comprised of ODU and wavelength (OCh) layers. This document focuses on the control of the ODU layer, with control of the wavelength layer considered out of the scope. Please refer to [WSON-Frame] for further information about the wavelength layer.

2. Terminology

OTN: Optical Transport Network

ODU: Optical Channel Data Unit

OTU: Optical channel transport unit

OMS: Optical multiplex section

MSI: Multiplex Structure Identifier

TPN: Tributary Port Number

LO ODU: Lower Order ODU. The LO ODU_j (j can be 0, 1, 2, 2e, 3, 4, flex.) represents the container transporting a client of the OTN that is either directly mapped into an OTU_k (k = j) or multiplexed into a server HO ODU_k (k > j) container.

HO ODU: Higher Order ODU. The HO ODU_k (k can be 1, 2, 2e, 3, 4.) represents the entity transporting a multiplex of LO ODU_j tributary signals in its OPU_k area.

ODUflex: Flexible ODU. A flexible ODU_k can have any bit rate and a bit rate tolerance up to +/-100 ppm.

3. G.709 Optical Transport Network (OTN)

This section provides an informative overview of those aspects of the OTN impacting control plane protocols. This overview is based on the ITU-T Recommendations that contain the normative definition of the OTN. Technical details regarding OTN architecture and interfaces are provided in the relevant ITU-T Recommendations.

Specifically, [G872-2001] and [G872Am2] describe the functional architecture of optical transport networks providing optical signal transmission, multiplexing, routing, supervision, performance assessment, and network survivability. [G709-V1] defines the interfaces of the optical transport network to be used within and between subnetworks of the optical network. With the evolution and deployment of OTN technology many new features have been specified in ITU-T recommendations, including for example, new ODU₀, ODU_{2e}, ODU₄ and ODUflex containers as described in [G709-V3].

3.1. OTN Layer Network

The simplified signal hierarchy of OTN is shown in Figure 1, which illustrates the layers that are of interest to the control plane. Other layers below OCh (e.g. Optical Transmission Section - OTS) are not included in this Figure. The full signal hierarchy is provided in [G709-V3].

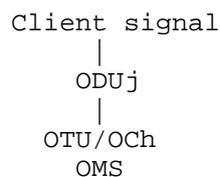


Figure 1 - Basic OTN signal hierarchy

Client signals are mapped into ODU_j containers. These ODU_j containers are multiplexed onto the OTU/OCh. The individual OTU/OCh signals are combined in the Optical Multiplex Section (OMS) using WDM multiplexing, and this aggregated signal provides the link between the nodes.

3.1.1. Client signal mapping

The client signals are mapped into a Low Order (LO) ODU_j. Appendix A gives more information about LO ODU.

The current values of *j* defined in [G709-V3] are: 0, 1, 2, 2e, 3, 4, Flex. The approximate bit rates of these signals are defined in [G709-V3] and are reproduced in Tables 1 and 2.

ODU Type	ODU nominal bit rate
ODU0	1 244 160 kbits/s
ODU1	239/238 x 2 488 320 kbit/s
ODU2	239/237 x 9 953 280 kbit/s
ODU3	239/236 x 39 813 120 kbit/s
ODU4	239/227 x 99 532 800 kbit/s
ODU2e	239/237 x 10 312 500 kbit/s
ODUflex for CBR Client signals	239/238 x client signal bit rate
ODUflex for GFP-F Mapped client signal	Configured bit rate

Table 1 - ODU types and bit rates

NOTE - The nominal ODU_k rates are approximately: 2 498 775.126 kbit/s (ODU1), 10 037 273.924 kbit/s (ODU2), 40 319 218.983 kbit/s (ODU3), 104 794 445.815 kbit/s (ODU4) and 10 399 525.316 kbit/s (ODU2e).

ODU Type	ODU bit-rate tolerance
ODU0	+ - 20 ppm
ODU1	+ - 20 ppm
ODU2	+ - 20 ppm
ODU3	+ - 20 ppm
ODU4	+ - 20 ppm
ODU2e	+ - 100 ppm
ODUflex for CBR Client signals	+ - 100 ppm
ODUflex for GFP-F Mapped client signal	+ - 100 ppm

Table 2 - ODU types and tolerance

One of two options is for mapping client signals into ODUflex depending on the client signal type:

- Circuit clients are proportionally wrapped. Thus the bit rate and tolerance are defined by the client signal.
- Packet clients are mapped using the Generic Framing Procedure (GFP). [G709-V3] recommends that the bit rate should be set to an integer multiplier of the High Order (HO) Optical Channel Physical Unit (OPU) OPUk TS rate, the tolerance should be +/-100ppm, and the bit rate should be determined by the node that performs the mapping.

[Editors' Note: As outcome of ITU SG15/q11 expert meeting held in Vimercate in September 2010 it was decided that a resizable ODUflex(GFP) occupies the same number of TS on every link of the path (independently of the High Order (HO) OPUk TS rate). Please see WD07 and the meeting report of this meeting for more information.

The authors will update the above text related to Packet client mapping as soon as new version of G.709 will be updated accordingly with expert meeting decision reported here.]

3.1.2. Multiplexing ODUj onto Links

The links between the switching nodes are provided by one or more wavelengths. Each wavelength carries one OCh, which carries one OTU, which carries one ODU. Since all of these signals have a 1:1:1 relationship, we only refer to the OTU for clarity. The ODUs are mapped into the TS of the OPUk. Note that in the case where $j=k$ the ODUj is mapped into the OTU/OCh without multiplexing.

The initial versions of G.709 [G709-V1] only provided a single TS granularity, nominally 2.5Gb/s. [G709-V3], approved in 2009, added an additional TS granularity, nominally 1.25Gb/s. The number and type of TSs provided by each of the currently identified OTUk is provided below:

	2.5Gb/s	1.25Gb/s	Nominal Bit rate
OTU1	1	2	2.5Gb/s
OTU2	4	8	10Gb/s
OTU3	16	32	40Gb/s
OTU4	--	80	100Gb/s

To maintain backwards compatibility while providing the ability to interconnect nodes that support 1.25Gb/s TS at one end of a link and 2.5Gb/s TS at the other, the 'new' equipment will fall back to the use of a 2.5Gb/s TS if connected to legacy equipment. This information is carried in band by the payload type.

The actual bit rate of the TS in an OTUk depends on the value of k. Thus the number of TS occupied by an ODUj may vary depending on the values of j and k. For example an ODU2e uses 9 TS in an OTU3 but only 8 in an OTU4. Examples of the number of TS used for various cases are provided below:

- ODU0 into ODU1, ODU2, ODU3 or ODU4 multiplexing with 1,25Gbps TS granularity
 - o ODU0 occupies 1 of the 2, 8, 32 or 80 TS for ODU1, ODU2, ODU3 or ODU4
- ODU1 into ODU2, ODU3 or ODU4 multiplexing with 1,25Gbps TS granularity
 - o ODU1 occupies 2 of the 8, 32 or 80 TS for ODU2, ODU3 or ODU4
- ODU1 into ODU2, ODU3 multiplexing with 2.5Gbps TS granularity
 - o ODU1 occupies 1 of the 4 or 16 TS for ODU2 or ODU3

- ODU2 into ODU3 or ODU4 multiplexing with 1.25Gbps TS granularity
 - o ODU2 occupies 8 of the 32 or 80 TS for ODU3 or ODU4
- ODU2 into ODU3 multiplexing with 2.5Gbps TS granularity
 - o ODU2 occupies 4 of the 16 TS for ODU3
- ODU3 into ODU4 multiplexing with 1.25Gbps TS granularity
 - o ODU3 occupies 31 of the 80 TS for ODU4
- ODUflex into ODU2, ODU3 or ODU4 multiplexing with 1.25Gbps TS granularity
 - o ODUflex occupies n of the 8, 32 or 80 TS for ODU2, ODU3 or ODU4 (n <= Total TS numbers of ODUk)
- ODU2e into ODU3 or ODU4 multiplexing with 1.25Gbps TS granularity
 - o ODU2e occupies 9 of the 32 TS for ODU3 or 8 of the 80 TS for ODU4

In general the mapping of an ODU_j (including ODUflex) into the OTU_k TSs is determined locally, and it can also be explicitly controlled by a specific entity (e.g., head end, NMS) through Explicit Label Control [RFC3473].

3.1.2.1. Structure of MSI information

When multiplexing an ODU_j into a HO ODU_k (k>j), G.709 specifies the information that has to be transported in-band in order to allow for correct demultiplexing. This information, known as Multiplex Structure Information (MSI), is transported in the OPU_k overhead and is local to each link. In case of bidirectional paths the association between TPN and TS MUST be the same in both directions.

The MSI information is organized as a set of entries, with one entry for each HO ODU_j TS. The information carried by each entry is:

- Payload Type: the type of the transported payload.
- Tributary Port Number (TPN): the port number of the ODU_j transported by the HO ODU_k. The TPN is the same for all the TSs assigned to the transport of the same ODU_j instance.

For example, an ODU2 carried by a HO ODU3 is described by 4 entries in the OPU3 overhead when the TS size is 2.5 Gbit/s, and by 8 entries when the TS size is 1.25 Gbit/s.

On each node and on every link, two MSI values have to be provisioned:

- The TxMSI information inserted in OPU (e.g., OPU3) overhead by the source of the HO ODUk trail.
- The expectedMSI information that is used to check the acceptedMSI information. The acceptedMSI information is the MSI valued received in-band, after a 3 frames integration.

The sink of the HO ODU trail checks the complete content of the acceptedMSI information (against the expectedMSI). If the acceptedMSI is different from the expectedMSI, then the traffic is dropped and a payload mismatch alarm is generated.

Provisioning of TPN can be performed either by network management system or control plane. In the last case, control plane is also responsible for negotiating the provisioned values on a link by link base.

4. Connection management in OTN

OTN-based connection management is concerned with controlling the connectivity of ODU paths and optical channels (OCh). This document focuses on the connection management of ODU paths. The management of OCh paths is described in [WSON-FRAME].

While [G872-2001] considered the ODU as a set of layers in the same way as SDH has been modeled, recent ITU-T OTN architecture progress [G872-Am2] includes an agreement to model the ODU as a single layer network with the bit rate as a parameter of links and connections. This allows the links and nodes to be viewed in a single topology as a common set of resources that are available to provide ODU_j connections independent of the value of j. Note that when the bit rate of ODU_j is less than the server bit rate, ODU_j connections are supported by HO-ODU (which has a one-to-one relationship with the OTU).

From an ITU-T perspective, the ODU connection topology is represented by that of the OTU link layer, which has the same topology as that of the OCh layer (independent of whether the OTU supports HO-ODU, where multiplexing is utilized, or LO-ODU in the case of direct mapping).

Thus, the OTU and OCh layers should be visible in a single topological representation of the network, and from a logical perspective, the OTU and OCh may be considered as the same logical, switchable entity.

Note that the OTU link layer topology may be provided via various infrastructure alternatives, including point-to-point optical connections, flexible optical connections fully in the optical domain, flexible optical connections involving hybrid sub-lambda/lambda nodes involving 3R, etc.

The document will be updated to maintain consistency with G.872 progress when it is consented for publication.

4.1. Connection management of the ODU

LO ODU_j can be either mapped into the OTU_k signal ($j = k$), or multiplexed with other LO ODU_js into an OTU_k ($j < k$), and the OTU_k is mapped into an OCh. See Appendix A for more information.

From the perspective of control plane, there are two kinds of network topology to be considered.

(1) ODU layer

In this case, the ODU links are presented between adjacent OTN nodes, which is illustrated in Figure 2. In this layer there are ODU links with a variety of TSs available, and nodes that are ODXCs. Lo ODU connections can be setup based on the network topology.

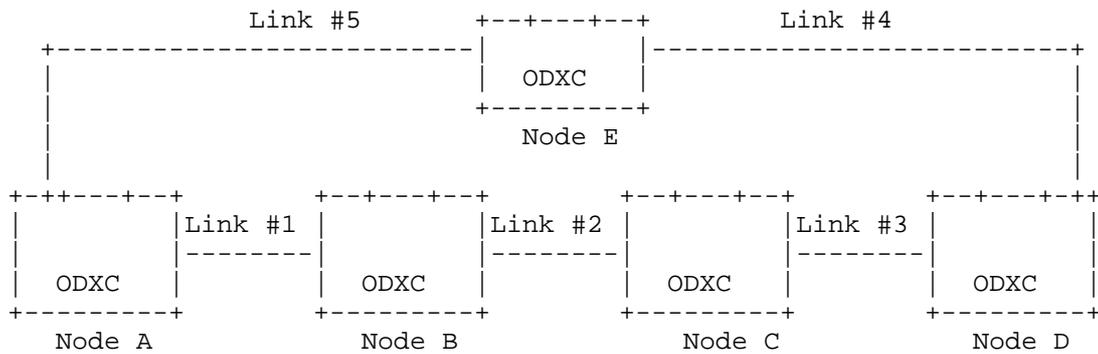


Figure 2 - Example Topology for LO ODU connection management

If an ODU_j connection is requested between Node C and Node E routing/path computation must select a path that has the required number of TS available and that offers the lowest cost. Signaling is then invoked to set up the path and to provide the information (e.g., selected TS) required by each transit node to allow the configuration of the ODU_j to OTU_k mapping ($j = k$) or multiplexing ($j < k$), and demapping ($j = k$) or demultiplexing ($j < k$).

(2) ODU layer with OCh switching capability

In this case, the OTN nodes interconnect with wavelength switched node (e.g., ROADM,OXC) that are capable of OCh switching, which is illustrated in Figure 3 and Figure 4. There are ODU layer and OCh layer, so it is simply a MLN. OCh connections may be created on demand, which is described in section 5.1.

In this case, an operator may choose to allow the underlined OCh layer to be visible to the ODU routing/path computation process in which case the topology would be as shown in Figure 4. In Figure 3 below, instead, a cloud representing OCH capable switching nodes is represented. In Figure 3, the operator choice is to hide the real RWA network topology.

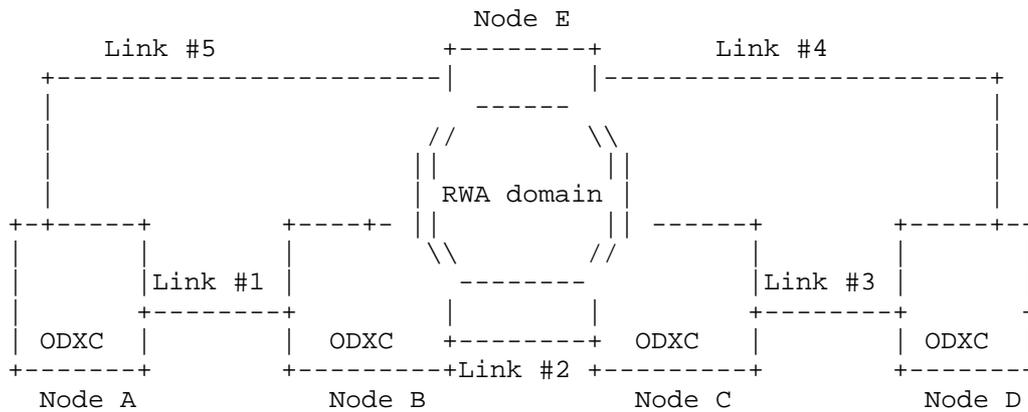


Figure 3 - RWA Hidden Topology for LO ODU connection management

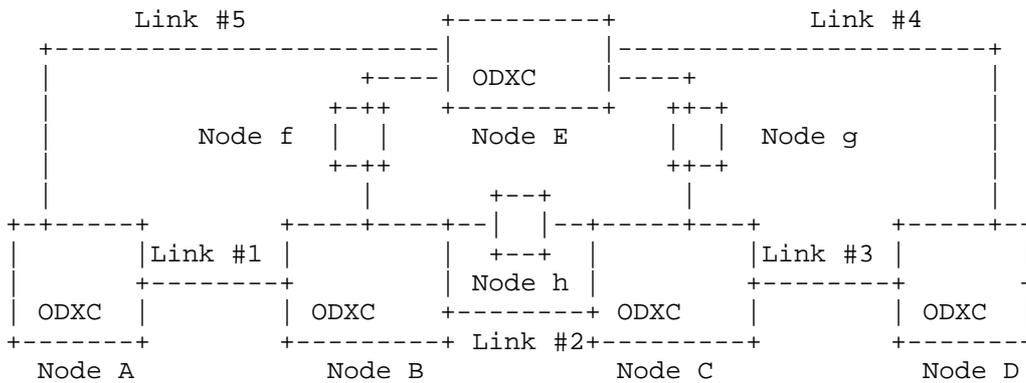


Figure 4 - RWA Visible Topology for LO ODUj connection management

In Figure 4, the cloud of previous figure is substitute by the real topology. The nodes f, g, h are nodes with OCH switching capability.

In the examples (i.e., Figure 3 and Figure 4), we have considered the case in which LO-ODUj connections are supported by OCH connection, and the case in which the supporting underlying connection can be also made by a combination of HO-ODU/OCh connections.

In this case, the ODU routing/path selection process will request an HO-ODU/OCh connection between node C and node E from the RWA domain. The connection will appear at ODU level as a Forwarding Adjacency, which will be used to create the ODU connection.

5. GMPLS/PCE Implications

The purpose of this section is to provide a set of requirements to be evaluated for extensions of the current GMPLS protocol suite and the PCE applications and protocols to encompass OTN enhancements and connection management.

5.1. Implications for LSP Hierarchy with GMPLS TE

The path computation for ODU connection request is based on the topology of ODU layer, including OCh layer visibility.

The OTN path computation can be divided into two layers. One layer is OCh/OTUk, the other is ODUj. [RFC4206] and [RFC6107] define the mechanisms to accomplish creating the hierarchy of LSPs. The LSP management of multiple layers in OTN can follow the procedures defined in [RFC4206], [RFC6107] and related MLN drafts.

As discussed in section 4, the route path computation for OCh is in the scope of WSON [WSON-Frame]. Therefore, this document only considers ODU layer for ODU connection request.

LSP hierarchy could be applied within the ODU layers. One of the typical scenarios for ODU layer hierarchy is to maintain compatibility with introducing new [G709-V3] services (e.g., ODU0, ODUflex) into a legacy network configuration (containing [G709-V1] or [G709-V2] OTN equipment). In this scenario, it may be needed to consider introducing hierarchical multiplexing capability in specific network transition scenarios. One method for enabling multiplexing hierarchy is by introducing dedicated boards in a few specific places in the network and tunneling these new services through [G709-V1] or [G709-V2] containers (ODU1, ODU2, ODU3), thus postponing the need to upgrade every network element to [G709-V3] capabilities.

In such case, one ODUj connection can be nested into another ODUk (j<k) connection, which forms the LSP hierarchy in ODU layer. The creation of the outer ODUk connection can be triggered via network planning, or by the signaling of the inner ODUj connection. For the former case, the outer ODUk connection can be created in advance based on network planning. For the latter case, the multi-layer network signaling described in [RFC4206], [RFC6107] and [RFC6001] (including related modifications, if needed) are relevant to create the ODU connections with multiplexing hierarchy. In both cases, the outer ODUk connection is advertised as a Forwarding Adjacency (FA).

5.2. Implications for GMPLS Signaling

The signaling function and Resource reSerVation Protocol-Traffic Engineering (RSVP-TE) extensions are described in [RFC3471] and [RFC3473]. For OTN-specific control, [RFC4328] defines signaling extensions to support G.709 Optical Transport Networks Control as defined in [G709-V1].

As described in Section 3, [G709-V3] introduced some new features that include the ODU0, ODU2e, ODU4 and ODUflex containers. The mechanisms defined in [RFC4328] do not support such new OTN features, and protocol extensions will be necessary to allow them to be controlled by a GMPLS control plane.

[RFC4328] defines the LSP Encoding Type, the Switching Type and the Generalized Protocol Identifier (Generalized-PID) constituting the common part of the Generalized Label Request. The G.709 Traffic Parameters are also defined in [RFC4328]. The following signaling aspects should be considered additionally since [RFC4328] was published:

- Support for specifying the new signal types and the related traffic information

THE traffic parameters should be extended in signaling message to support the new optical Channel Data Unit (ODUj) including:

- ODU0
- ODU2e
- ODU4
- ODUflex

For ODUflex, since it has a variable bandwidth/bit rate BR and a bit rate tolerance T, the (node local) mapping process must be aware of the bit rate and tolerance of the ODUj being multiplexed in order to select the correct number of TS and the fixed/variable stuffing bytes. Therefore, bit rate and bit rate tolerance should also be carried in the Traffic Parameter in the signaling of connection setup request.

For other ODU signal types, the bit rates and tolerances of them are fixed and can be deduced from the signal types.

- Support for LSP setup using different Tributary Slot granularity

New label should be defined to identify the type of TS (i.e., the 2.5 Gbps TS granularity and the new 1.25 Gbps TS granularity).

- Support for LSP setup of new ODUk/ODUflex containers with related mapping and multiplexing capabilities

New label should be defined to carry the exact TS allocation information related to the extended mapping and multiplexing hierarchy (For example, ODU0 into ODU2 multiplexing (with 1,25Gbps TS granularity)), in order to setting up the ODU connection.

- Support for Tributary Port Number allocation and negotiation

Tributary Port Number needs to be configured as part of the MSI information (See more information in Section 3.1.2.1). A new

extension object has to be defined to carry TPN information if control plane is used to configure MSI information.

- Support for ODU Virtual Concatenation (VCAT) and Link Capacity Adjustment Scheme (LCAS)

GMPLS signaling should support the creation of Virtual Concatenation of ODUk signal with $k=1, 2, 3$. The signaling should also support the control of dynamic capacity changing of a VCAT container using LCAS ([G.7042]). [VCAT] has a clear description of VCAT and LCAS control in SONET/SDH and OTN networks.

- Support for constraint signaling

How an ODUk connection service is transported within an operator network is governed by operator policy. For example, the ODUk connection service might be transported over an ODUk path over an OTUk section, with the path and section being at the same rate as that of the connection service. In this case, an entire lambda of capacity is consumed in transporting the ODUk connection service. On the other hand, the operator might leverage sub-lambda multiplexing capabilities in the network to improve infrastructure efficiencies within any given networking domain. In this case, ODUk multiplexing may be performed prior to transport over various rate ODU servers over associated OTU sections.

The identification of constraints and associated encoding in the signaling for differentiating full lambda LSP or sub lambda LSP is for further study.

- Support for Control of Hitless Adjustment of ODUflex (GFP)

[G.HAO] has been created in ITU-T to specify hitless adjustment of ODUflex (GFP) (HAO) that is used to increase or decrease the bandwidth of an ODUflex (GFP) that is transported in an OTN network.

The procedure of ODUflex (GFP) adjustment requires the participation of every node along the path. Therefore, it is recommended to use the control plane signaling to initiate the adjustment procedure in order to avoid the manual configuration at each node along the path.

Since the [G.HAO] is being developed currently, the control of HAO is for further study.

All the extensions above should consider the extensibility to match future evolvement of OTN.

5.3. Implications for GMPLS Routing

The path computation process should select a suitable route for a ODU_j connection request. In order to compute the lowest cost path it must evaluate the available bandwidth on each candidate link. The routing protocol should be extended to convey some information to represent ODU TE topology.

GMPLS Routing [RFC4202] defines Interface Switching Capability Descriptor of TDM which can be used for ODU. However, some other issues should also be considered which are discussed below.

Interface Switching Capability Descriptors present a new constraint for LSP path computation. [RFC4203] defines the switching capability and related Maximum LSP Bandwidth and the Switching Capability specific information. When the Switching Capability field is TDM the Switching Capability specific information field includes Minimum LSP Bandwidth, an indication whether the interface supports Standard or Arbitrary SONET/SDH, and padding. So routing protocol should be extended when TDM is ODU type to support representation of ODU switching information, especially the following requirements should be considered:

- Support for carrying the link multiplexing capability

As discussed in section 3.1.2, many different types of ODU_j can be multiplexed into the same OTU_k. For example, both ODU₀ and ODU₁ may be multiplexed into ODU₂. An OTU link may support one or more types of ODU_j signals. The routing protocol should be capable of carrying this multiplexing capability.

- Support any ODU and ODUflex

The bit rate (i.e., bandwidth) of TS is dependent on the TS granularity and the signal type of the link. For example, the bandwidth of a 1.25G TS in an OTU₂ is about 1.249409620 Gbps, while the bandwidth of a 1.25G TS in an OTU₃ is about 1.254703729 Gbps.

One LO ODU may need different number of TSSs when multiplexed into different HO ODUs. For example, for ODU_{2e}, 9 TSSs are needed when multiplexed into an ODU₃, while only 8 TSSs are needed when

multiplexed into an ODU4. For ODUflex, the total number of TSs to be reserved in a HO ODU equals the maximum of [bandwidth of ODUflex / bandwidth of TS of the HO ODU].

Therefore, the routing protocol must be capable of carrying the necessary and sufficient link bandwidth information for performing accurate route computation for any of the fixed rate ODUs as well as ODUflex.

- Support for differentiating between terminating and switching capability

Due to internal constraints and/or limitations, the type of signal being advertised by an interface could be just switched (i.e. forwarded to switching matrix without multiplexing/demultiplexing actions), just terminated (demuxed) or both of them. The capability advertised by an interface needs further distinction in order to separate termination and switching capabilities.

Therefore, to allow the required flexibility, the routing protocol should clearly distinguish the terminating and switching capability.

- Support different priorities for resource reservation

How many priorities levels should be supported depends on the operator's policy. Therefore, the routing protocol should be capable of supporting either no priorities or up to 8 priority levels as defined in [RFC4202].

- Support link bundling

Link bundling can improve routing scalability by reducing the amount of TE links that has to be handled by routing protocol. The routing protocol must be capable of supporting bundling multiple OTU links, at the same or different line rates, between a pair of nodes as a TE link. Note that link bundling is optional and is implementation dependent.

- Support for Control of Hitless Adjustment of ODUflex (GFP)

As described in Section 5.2, the routing requirements for supporting hitless adjustment of ODUflex (GFP) (HAO) are for further study.

As mentioned in Section 5.1, one method of enabling multiplexing hierarchy is via usage of dedicated boards to allow tunneling of new services through legacy ODU1, ODU2, ODU3 containers. Such dedicated boards may have some constraints with respect to switching matrix access; detection and representation of such constraints is for further study.

5.4. Implications for Link Management Protocol (LMP)

As discussed in section 5.3, Path computation needs to know the interface switching capability of links. The switching capability of two ends of the link may be different, so the link capability of two ends should be correlated.

The Link Management Protocol (LMP) [RFC4204] provides a control plane protocol for exchanging and correlating link capabilities.

It is not necessary to use LMP to correlate link-end capabilities if the information is available from another source such as management configuration or automatic discovery/negotiation within the data plane.

Note that LO ODU type information can be, in principle, discovered by routing. Since in certain case, routing is not present (e.g. UNI case) we need to extend link management protocol capabilities to cover this aspect. In case of routing presence, the discovering procedure by LMP could also be optional.

- Correlating the granularity of the TS

As discussed in section 3.1.2, the two ends of a link may support different TS granularity. In order to allow interconnection the node with 1.25Gb/s granularity must fall back to 2.5Gb/s granularity.

Therefore, it is necessary for the two ends of a link to correlate the granularity of the TS. This ensures to allocate the TS over the TE link correctly.

- Correlating the supported LO ODU signal types and multiplexing hierarchy capability

Many new ODU signal types have been introduced in [G709-V3], such as ODU0, ODU4, ODU2e and ODUFlex. It is possible that equipment does not support all the LO ODU signal types introduced by those new standards or drafts. Furthermore, since multiplexing hierarchy is not allowed before [G709-V3], it is possible that

only one end of an ODU link can support multiplexing hierarchy capability, or the two ends of the link support different multiplexing hierarchy capabilities (e.g., one end of the link supports ODU0 into ODU1 into ODU3 multiplexing while the other end supports ODU0 into ODU2 into ODU3 multiplexing).

For the control and management consideration, it is necessary for the two ends of an HO ODU link to correlate which types of LO ODU can be supported and what multiplexing hierarchy capabilities can be provided by the other end.

5.5. Implications for Path Computation Elements

[PCE-APS] describes the requirements for GMPLS applications of PCE in order to establish GMPLS LSP. PCE needs to consider the GMPLS TE attributes appropriately once a PCC or another PCE requests a path computation. The TE attributes which can be contained in the path calculation request message from the PCC or the PCE defined in [RFC5440] includes switching capability, encoding type, signal type, etc.

As described in section 5.2.1, new signal types and new signals with variable bandwidth information need to be carried in the extended signaling message of path setup. For the same consideration, PCECP also has a desire to be extended to carry the new signal type and related variable bandwidth information when a PCC requests a path computation.

6. Data Plane Backward Compatibility Considerations

The node supporting 1.25Gbps TS can interwork with the other nodes that supporting 2.5Gbps TS by combining Specific TSs together in data plane. The control plane MUST support this TS combination.

Take Figure 5 as an example. Assume that there is an ODU2 link between node A and B, where node A only supports the 2.5Gbps TS while node B supports the 1.25Gbps TS. In this case, the TS#i and TS#i+4 (where $i \leq 4$) of node B are combined together. When creating an ODU1 service in this ODU2 link, node B reserves the TS#i and TS#i+4 with the granularity of 1.25Gbps. But in the label sent from B to A, it is indicated that the TS#i with the granularity of 2.5Gbps is reserved.

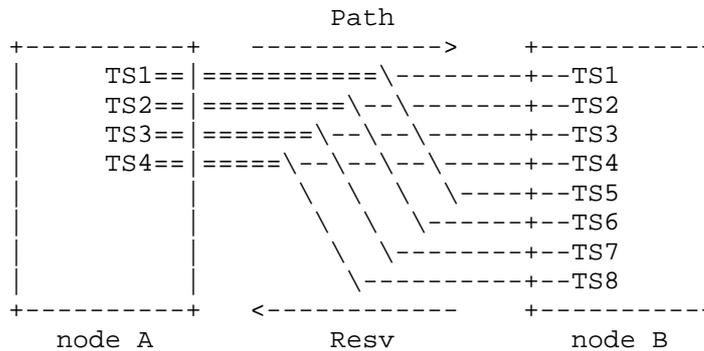


Figure 5 - Interworking between 1.25Gbps TS and 2.5Gbps TS

In the contrary direction, when receiving a label from node A indicating that the TS#i with the granularity of 2.5Gbps is reserved, node B will reserved the TS#i and TS#i+4 with the granularity of 1.25Gbps in its control plane.

7. Security Considerations

The use of control plane protocols for signaling, routing, and path computation opens an OTN to security threats through attacks on those protocols. The data plane technology for an OTN does not introduce any specific vulnerabilities, and so the control plane may be secured using the mechanisms defined for the protocols discussed.

For further details of the specific security measures refer to the documents that define the protocols ([RFC3473], [RFC4203], [RFC4205], [RFC4204], and [RFC5440]). [GMPLS-SEC] provides an overview of security vulnerabilities and protection mechanisms for the GMPLS control plane.

8. IANA Considerations

This document makes not requests for IANA action.

9. Acknowledgments

We would like to thank Maarten Vissers for his review and useful comments.

10. References

10.1. Normative References

- [RFC4328] D. Papadimitriou, Ed. "Generalized Multi-Protocol LabelSwitching (GMPLS) Signaling Extensions for G.709 Optical Transport Networks Control", RFC 4328, Jan 2006.
- [RFC3471] Berger, L., Editor, "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Functional Description", RFC 3471, January 2003.
- [RFC3473] L. Berger, Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions", RFC 3473, January 2003.
- [RFC4201] K. Kompella, Y. Rekhter, Ed., "Link Bundling in MPLS Traffic Engineering (TE)", RFC 4201, October 2005.
- [RFC4202] K. Kompella, Y. Rekhter, Ed., "Routing Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4202, October 2005.
- [RFC4203] K. Kompella, Y. Rekhter, Ed., "OSPF Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4203, October 2005.
- [RFC4205] K. Kompella, Y. Rekhter, Ed., "Intermediate System to Intermediate System (IS-IS) Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4205, October 2005.
- [RFC4204] Lang, J., Ed., "Link Management Protocol (LMP)", RFC 4204, October 2005.
- [RFC4206] K. Kompella, Y. Rekhter, Ed., " Label Switched Paths (LSP) Hierarchy with Generalized Multi-Protocol Label Switching (GMPLS) Traffic Engineering (TE)", RFC 4206, October 2005.
- [RFC6107] K. Shiomoto, A. Farrel, "Procedures for Dynamically Signaled Hierarchical Label Switched Paths", RFC6107, February 2011.

- [RFC6001] Dimitri Papadimitriou et al, "Generalized Multi-Protocol Label Switching (GMPLS) Protocol Extensions for Multi-Layer and Multi-Region Networks (MLN/MRN)", RFC6001, February 21, 2010.
- [RFC5440] JP. Vasseur, JL. Le Roux, Ed., " Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, March 2009.
- [VCAT] G. Bernstein et al, "Operating Virtual Concatenation (VCAT) and the Link Capacity Adjustment Scheme (LCAS) with Generalized Multi-Protocol Label Switching (GMPLS)", draft-ietf-ccamp-gmpls-vcat-lcas-11.txt, March 9, 2011.
- [G709-V3] ITU-T, "Interfaces for the Optical Transport Network (OTN)", G.709 Recommendation, December 2009.

10.2. Informative References

- [G709-V1] ITU-T, "Interface for the Optical Transport Network (OTN)," G.709 Recommendation and Amendment1, November 2001.
- [G709-V2] ITU-T, "Interface for the Optical Transport Network (OTN)," G.709 Recommendation, March 2003.
- [G7042] ITU-T G.7042/Y.1305, "Link capacity adjustment scheme (LCAS) for virtual concatenated signals", March 2006.
- [G872-2001] ITU-T, "Architecture of optical transport networks", November 2001 (11 2001).
- [G872-Am2] Draft Amendment 2, ITU-T, "Architecture of optical transport networks".
- [G.HAO] TD 382 (WP3/15), 31 May - 11 June 2010, Q15 Plenary Meeting in Geneva, Initial draft G.hao "Hitless Adjustment of ODUflex (HAO)".
- [HZang00] H. Zang, J. Jue and B. Mukherjee, "A review of routing and wavelength assignment approaches for wavelength-routed optical WDM networks", Optical Networks Magazine, January 2000.

[WSON-FRAME] Y. Lee, G. Bernstein, W. Imajuku, "Framework for GMPLS and PCE Control of Wavelength Switched Optical Networks (WSON)", draft-ietf-ccamp-rwa-wson-framework, work in progress.

[PCE-APS] Tomohiro Otani, Kenichi Ogaki, Diego Caviglia, and Fatai Zhang, "Requirements for GMPLS applications of PCE", draft-ietf-pce-gmpls-aps-req-01.txt, July 2009.

[GMPLS-SEC] Fang, L., Ed., "Security Framework for MPLS and GMPLS Networks", Work in Progress, October 2009.

11. Authors' Addresses

Fatai Zhang
Huawei Technologies
F3-5-B R&D Center, Huawei Base
Bantian, Longgang District
Shenzhen 518129 P.R.China

Phone: +86-755-28972912
Email: zhangfatai@huawei.com

Dan Li
Huawei Technologies Co., Ltd.
F3-5-B R&D Center, Huawei Base
Bantian, Longgang District
Shenzhen 518129 P.R.China

Phone: +86-755-28973237
Email: huawei.danli@huawei.com

Han Li
China Mobile Communications Corporation
53 A Xibianmennei Ave. Xuanwu District
Beijing 100053 P.R. China

Phone: +86-10-66006688
Email: lihan@chinamobile.com

Sergio Belotti
Alcatel-Lucent
Optics CTO
Via Trento 30 20059 Vimercate (Milano) Italy
+39 039 6863033

Email: sergio.belotti@alcatel-lucent.it

Daniele Ceccarelli
Ericsson
Via A. Negrone 1/A
Genova - Sestri Ponente
Italy
Email: daniele.ceccarelli@ericsson.com

12. Contributors

Jianrui Han
Huawei Technologies Co., Ltd.
F3-5-B R&D Center, Huawei Base
Bantian, Longgang District
Shenzhen 518129 P.R.China

Phone: +86-755-28972913
Email: hanjianrui@huawei.com

Malcolm Betts
Huawei Technologies Co., Ltd.

Email: malcolm.betts@huawei.com

Pietro Grandi
Alcatel-Lucent
Optics CTO
Via Trento 30 20059 Vimercate (Milano) Italy
+39 039 6864930

Email: pietro_vittorio.grandi@alcatel-lucent.it

Eve Varma
Alcatel-Lucent
1A-261, 600-700 Mountain Av
PO Box 636
Murray Hill, NJ 07974-0636
USA
Email: eve.varma@alcatel-lucent.com

APPENDIX A: ODU connection examples

This appendix provides a description of ODU terminology and connection examples. This section is not normative, and is just intended to facilitate understanding.

In order to transmit a client signal, an ODU connection must first be created. From the perspective of [G709-V3] and [G872-Am2], some types of ODUs (i.e., ODU1, ODU2, ODU3, ODU4) may assume either a client or server role within the context of a particular networking domain:

(1) An ODU_j client that is mapped into an OTU_k server. For example, if a STM-16 signal is encapsulated into ODU1, and then the ODU1 is mapped into OTU1, the ODU1 is a LO ODU (from a multiplexing perspective).

(2) An ODU_j client that is mapped into an ODU_k ($j < k$) server occupying several TSs. For example, if ODU1 is multiplexed into ODU2, and ODU2 is mapped into OTU2, the ODU1 is a LO ODU and the ODU2 is a HO ODU (from a multiplexing perspective).

Thus, a LO ODU_j represents the container transporting a client of the OTN that is either directly mapped into an OTU_k ($k = j$) or multiplexed into a server HO ODU_k ($k > j$) container. Consequently, the HO ODU_k represents the entity transporting a multiplex of LO ODU_j tributary signals in its OPU_k area.

In the case of LO ODU_j mapped into an OTU_k ($k = j$) directly, Figure 6 give an example of this kind of LO ODU connection.

In Figure 6, The LO ODU_j is switched at the intermediate ODXC node. OCh and OTU_k are associated with each other. From the viewpoint of connection management, the management of OTU_k is similar with OCh. LO ODU_j and OCh/OTU_k have client/server relationships.

For example, one LO ODU1 connection can be setup between Node A and Node C. This LO ODU1 connection is to be supported by OCh/OTU1

connections, which are to be set up between Node A and Node B and between Node B and Node C. LO ODU1 can be mapped into OTU1 at Node A, demapped from it in Node B, switched at Node B, and then mapped into the next OTU1 and demapped from this OTU1 at Node C.

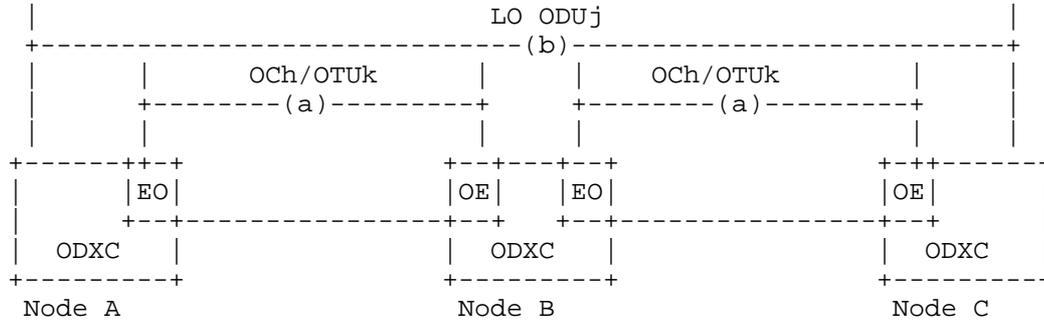


Figure 6 - Connection of LO ODUj (1)

In the case of LO ODUj multiplexing into HO ODUk, Figure 7 gives an example of this kind of LO ODU connection.

In Figure 7, OCh, OTUk, HO ODUk are associated with each other. The LO ODUj is multiplexed/de-multiplexed into/from the HO ODU at each ODXC node and switched at each ODXC node (i.e. trib port to line port, line card to line port, line port to trib port). From the viewpoint of connection management, the management of these HO ODUk and OTUk are similar to OCh. LO ODUj and OCh/OTUk/HO ODUk have client/server relationships. When a LO ODU connection is setup, it will be using the existing HO ODUk (/OTUk/OCh) connections which have been set up. Those HO ODUk connections provide LO ODU links, of which the LO ODU connection manager requests a link connection to support the LO ODU connection.

For example, one HO ODU2 (/OTU2/OCh) connection can be setup between Node A and Node B, another HO ODU3 (/OTU3/OCh) connection can be setup between Node B and Node C. LO ODU1 can be generated at Node A, switched to one of the 10G line ports and multiplexed into a HO ODU2 at Node A, demultiplexed from the HO ODU2 at Node B, switched at Node B to one of the 40G line ports and multiplexed into HO ODU3 at Node B, demultiplexed from HO ODU3 at Node C and switched to its LO ODU1 terminating port at Node C.

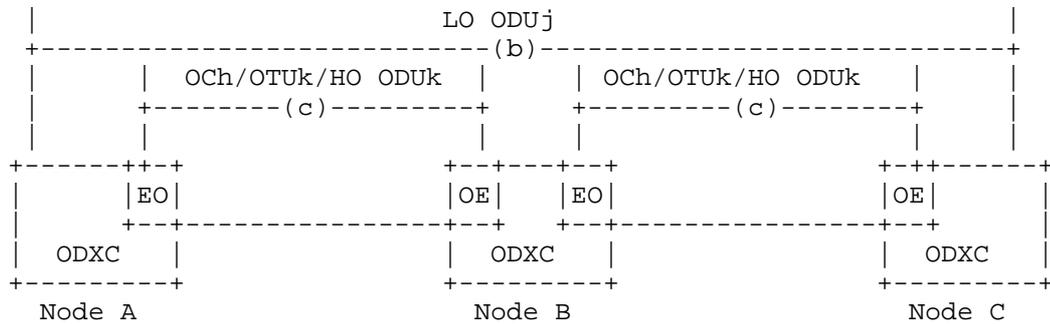


Figure 7 - Connection of LO ODUj (2)

Intellectual Property

The IETF Trust takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in any IETF Document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights.

Copies of Intellectual Property disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a 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 any standard or specification contained in an IETF Document. Please address the information to the IETF at ietf-ipr@ietf.org.

The definitive version of an IETF Document is that published by, or under the auspices of, the IETF. Versions of IETF Documents that are published by third parties, including those that are translated into other languages, should not be considered to be definitive versions of IETF Documents. The definitive version of these Legal Provisions is that published by, or under the auspices of, the IETF. Versions of these Legal Provisions that are published by third parties, including

those that are translated into other languages, should not be considered to be definitive versions of these Legal Provisions.

For the avoidance of doubt, each Contributor to the IETF Standards Process licenses each Contribution that he or she makes as part of the IETF Standards Process to the IETF Trust pursuant to the provisions of RFC 5378. No language to the contrary, or terms, conditions or rights that differ from or are inconsistent with the rights and licenses granted under RFC 5378, shall have any effect and shall be null and void, whether published or posted by such Contributor, or included with or in such Contribution.

Disclaimer of Validity

All IETF Documents and the information contained therein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY, THE IETF TRUST AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION THEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Copyright Notice

Copyright (c) 2010 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Network Working Group
Internet Draft
Updates: RFC4204
Category: Standards Track

Dan Li
Huawei
D. Ceccarelli
Ericsson

Expires: September 2011

March 14, 2011

Behavior Negotiation in The Link Management Protocol

draft-ietf-ccamp-lmp-behavior-negotiation-02.txt

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at <http://www.ietf.org/ietf/lid-abstracts.txt>

The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>

This Internet-Draft will expire on September 13, 2011.

Copyright Notice

Copyright (c) 2010 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in

Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Abstract

The Link Management Protocol (LMP) is used to coordinate the properties, use, and faults of data links in Generalized Multiprotocol Label Switching (GMPLS) networks. Various proposals have been advanced to provide extensions to the base LMP specification. This document defines an extension to negotiate capabilities and support for those extensions, and provides a generic procedure for LMP implementations that do not recognize or do not support any one of these extensions.

Conventions used in this document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

Table of Contents

1. Introduction	2
2. LMP Behavior Negotiation Procedure.....	3
3. Backward Compatibility.....	5
4. Security Considerations.....	6
5. IANA Considerations	6
5.1. New LMP Class Type.....	6
5.2. New Capabilities Registry.....	7
6. Contributors	7
7. Acknowledgments	8
8. References	8
8.1. Normative References.....	8
8.2. Informative References.....	8
9. Authors' Addresses	9

1. Introduction

The Link Management Protocol (LMP) [RFC4204] has been successfully deployed in Generalized Multiprotocol Label Switching (GMPLS)-controlled networks.

New LMP behaviors and protocol extensions have been introduced in a number of IETF documents as set out later in this section. It is likely that future extensions will be made to support additional functions.

In the network, if one GMPLS Label Switch Router (LSR) supports a new behavior or protocol extension, but its peer LSR does not, it is necessary to have a protocol mechanism for resolving issues that may arise. It is also beneficial to have a protocol mechanism to discover the capabilities of peer LSRs so that the right protocol extensions can be selected and the correct features enabled. There are no such procedures defined in the base LMP specification [RFC4204], so this document defines how to handle LMP extensions both at legacy LSRs and at upgraded LSRs that would communicate with legacy LSRs.

In [RFC4204], the basic behaviors have been defined around the use of the standard LMP messages, which include Config, Hello, Verify, Test, LinkSummary, and ChannelStatus. Per [RFC4204], these behaviors MUST be supported when LMP is implemented, and the message types from 1 to 20 have been assigned by IANA for these messages.

In [RFC4207], the SONET/SDH technology-specific behavior and information for LMP is defined. The Trace behavior is added to LMP, and the message types from 21 to 31 were assigned by IANA for the messages that provide the TRACE function. The Trace function has been extended for the support of OTNs (Optical Transport Networks) in [LMP-TEST].

In [RFC4209], extensions to LMP are defined to allow it to be used between a peer node and an adjacent Optical Line System (OLS). The LMP object class type and sub-object class name have been extended to support DWDM behavior.

In [RFC5818], the data channel consistency check behavior is defined, and the message types from 32 to 34 have been assigned by IANA for messages that provide this behavior.

It is likely that future extensions to LMP for other functions or technologies will require the definition of further LMP messages.

This document describes the behavior negotiation procedure to make sure both LSRs at the ends of each link understand the LMP messages that they exchange.

2. LMP Behavior Negotiation Procedure

The Config message is used in the control channel negotiation phase of LMP [RFC4204]. The LMP behavior negotiation procedure is defined in this document as an addition to this phase.

The Config message is defined in Section 12.3.1 of [RFC4204] and carries the <CONFIG> object (class name 6) as defined in Section 13.6 of [RFC4204].

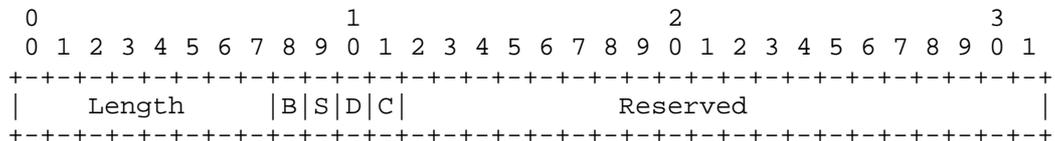
Two class types have been defined:

- C-Type = 1, HelloConfig, defined in [RFC4204]
- C-Type = 2, LMP_WDM_CONFIG, defined in [RFC4209]

This document defines a third C-Type with value 3 (TBD by IANA) to report and negotiate currently defined LMP mechanisms and behaviors, and to allow future LMP extensions to be reported and negotiated.

- C-Type = 3, BEHAVIOR_CONFIG

The format of the new type of CONFIG Class is defined as follows:



Length: 8 bits

This field indicates the total length of the objects expressed in multiples of 4 bytes.

Flags:

B: 1 bit

This bit indicates support for the basic behaviors defined in [RFC4204].

S: 1 bit

This bit indicates support for the Trace behavior of SONET/SDH technology-specific defined in [RFC4207].

D: 1 bit

This bit indicates support for the DWDM behavior defined in [RFC4209].

C: 1 bit

This bit indicates support for the data channel consistency check behavior defined in [RFC5818].

Further bits may be defined in future documents.

The Reserved field MUST be sent as zero and MUST NOT be ignored on receipt. This allows the detection of unsupported or unknown LMP behaviors when new bits are allocated to indicate further capabilities and are sent as one.

Upon receiving a bit set related to an unsupported or unknown behavior, a ConfigNack message MUST be sent with a <CONFIG> object, the BEHAVIOR_CONFIG C-Type representing the supported LMP behaviors. An LSR receiving such a ConfigNack SHOULD select a supported set of capabilities and send a further Config message, or MAY raise an alert to the management system (or log an error) and stop trying to perform LMP communications with its neighbor.

Note that multiple <CONFIG> objects (each with a different Class Type) MAY be present on a Config message in which case all of the objects SHOULD be processed, but see the note on backward compatibility in the next section. However, if more than one <CONFIG> object with the same Class Type is present on a Config message, the message SHOULD be rejected.

3. Backward Compatibility

An LSR that receives a Config message containing a <CONFIG> object with a C-Type that it does not recognize should respond with a ConfigNack message as described in [RFC4204]. Thus, legacy LMP nodes that do not support the BEHAVIOR_CONFIG C-Type defined in this document will respond with a ConfigNack message.

Note that [RFC4204] does not describe how multiple <CONFIG> objects with different C-Types should be processed. Thus it is possible that a legacy node receiving a BEHAVIOR_CONFIG object on a Config message that also includes a HelloConfig or LMP_WDM_CONFIG object might react as follows:

- Reject the message because of the unknown BEHAVIOR_CONFIG object as described above.

- Reject the message because of multiple <CONFIG> objects. This achieves the same effective result.
- Ignore the second <CONFIG> object. This would result in the BEHAVIOR_CONFIG object being unprocessed and also not rejected.

An LSR that receives a ConfigNack message rejecting a Config message that it sent containing the BEHAVIOR CONFIG C-Type because that object variant is not supported by its peer MUST NOT draw any conclusions about the level of support at the peer for LMP options described by bits B, S, D, and C. Instead, the LSR MUST revert to current practices of configuration or discovery through attempts to exercise the options.

However, as future documents are published describing new LMP features, and those documents require support of the BEHAVIOR CONFIG C-Type, an LSR that receives a ConfigNack message rejecting a Config message that it sent containing the BEHAVIOR CONFIG C-Type because that object variant is not supported by its peer SHOULD conclude that the additional options it wants to use are not supported by the peer.

4. Security Considerations

[RFC4204] describes how LMP messages between peers can be secured, and these measures are equally applicable to messages carrying the new <CONFIG> object defined in this document.

The operation of the procedures described in this document does not of itself constitute a security risk since they do not cause any change in network state. It would be possible, if the messages were intercepted or spoofed to cause bogus alerts in the management plane, or to cause LMP peers to consider that they could or could not operate protocol extensions, and so the use of the LMP security measures are RECOMMENDED.

5. IANA Considerations

5.1. New LMP Class Type

IANA maintains the "Link Management Protocol (LMP)" registry which has a subregistry called "LMP Object Class name space and Class type (C-Type)".

IANA is requested to make an assignment from this registry as follows:

6 CONFIG [RFC4204]

CONFIG Object Class type name space:

C-Type	Description	Reference
-----	-----	-----
3	BEHAVIOR_CONFIG	[This.I-D]

5.2. New Capabilities Registry

IANA is requested to create a new subregistry of the "Link Management Protocol (LMP)" registry to track the Behaviour Configuration bits defined in Section 2 of this document. It is suggested that this registry be called "LMP Behaviour Configuration Flags".

Allocations from this registry are by Standards Action.

Bits in this registry are numbered from zero as the most significant bit (transmitted first). The number of bits that can be present is limited by the length field of the <CONFIG> object which gives rise to $(255 \times 32) - 8 = 8152$. IANA is strongly recommended to allocate new bits with the lowest available unused number.

The registry is initially populated as follows:

Bit Number	Bit Name	Meaning	Reference
-----	-----	-----	-----
0	B	Basic LMP behavior support	[This.ID]
1	S	SONET/SDH Test support	[This.ID]
2	D	DWDM support	[This.ID]
3	C	Data Channel consistency check support	[This.ID]

6. Contributors

Diego Caviglia
Ericsson
Via A. Negrone 1/A 16153
Genoa Italy
Phone: +39 010 600 3736
Email: diego.caviglia@ericsson.com

7. Acknowledgments

Thanks to Adrian Farrel and Lou Berger for their useful comments.

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.
- [RFC4204] J. Lang, Ed., "Link Management Protocol (LMP)", RFC 4204, October 2005.
- [RFC4207] J. Lang, Ed., "Synchronous Optical Network (SONET)/ Synchronous Digital Hierarchy (SDH) Encoding for Link Management Protocol (LMP) Test Messages", RFC 4207, October 2005.
- [RFC4209] A. Fredette, Ed., "Link Management Protocol (LMP) for Dense Wavelength Division Multiplexing (DWDM) Optical Line Systems", RFC 4209, October 2005.
- [RFC5818] D. Li, Ed., "Data Channel Status Confirmation Extensions for the Link Management Protocol", RFC 5818, April 2010.

8.2. Informative References

- [LMP TEST] D. Ceccarelli, Ed., "Link Management Protocol (LMP) Test Messages Extensions for Evolutive Optical Transport Networks (OTN)" draft-ceccarelli-ccamp-gmpls-g709-lmp-test-02.txt, May, 2010.

9. Authors' Addresses

Dan Li
Huawei Technologies
F3-5-B R&D Center, Huawei Industrial Base,
Shenzhen 518129 China
Phone: +86 755-289-70230
Email: danli@huawei.com

Daniele Ceccarelli
Ericsson
Via A. Negrone 1/A
Genova - Sestri Ponente
Italy
Email: daniele.ceccarelli@ericsson.com

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: September 15, 2011

A. Takacs
Ericsson
D. Fedyk
Alcatel-Lucent
J. He
Huawei
March 14, 2011

GMPLS RSVP-TE extensions for OAM Configuration
draft-ietf-ccamp-oam-configuration-fwk-05

Abstract

OAM is an integral part of transport connections, hence it is required that OAM functions are activated/deactivated in sync with connection commissioning/decommissioning; avoiding spurious alarms and ensuring consistent operation. In certain technologies OAM entities are inherently established once the connection is set up, while other technologies require extra configuration to establish and configure OAM entities. This document specifies extensions to RSVP-TE to support the establishment and configuration of OAM entities along with LSP signaling.

Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on September 15, 2011.

Copyright Notice

Copyright (c) 2011 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1.	Introduction	4
2.	Requirements	6
3.	RSVP-TE based OAM Configuration	9
3.1.	Establishment of OAM Entities and Functions	9
3.2.	Adjustment of OAM Parameters	11
3.3.	Deleting OAM Entities	11
4.	RSVP-TE Extensions	13
4.1.	LSP Attributes Flags	13
4.2.	OAM Configuration TLV	14
4.2.1.	OAM Function Flags Sub-TLV	15
4.2.2.	Technology Specific sub-TLVs	16
4.3.	Administrative Status Information	16
4.4.	Handling OAM Configuration Errors	16
4.5.	Considerations on Point-to-Multipoint OAM Configuration	17
5.	IANA Considerations	19
6.	Security Considerations	20
7.	Acknowledgements	21
8.	References	22
8.1.	Normative References	22
8.2.	Informative References	22
	Authors' Addresses	24

1. Introduction

GMPLS is designed as an out-of-band control plane supporting dynamic connection provisioning for any suitable data plane technology; including spatial switching (e.g., incoming port or fiber to outgoing port or fiber), wavelength-division multiplexing (e.g., DWDM), time-division multiplexing (e.g., SONET/SDH, G.709), and lately Ethernet Provider Backbone Bridging -- Traffic Engineering (PBB-TE) and MPLS Transport Profile (MPLS-TP). In most of these technologies there are Operations and Management (OAM) functions employed to monitor the health and performance of the connections and to trigger data plane (DP) recovery mechanisms. Similarly to connections, OAM functions follow general principles but also have some technology specific characteristics.

OAM is an integral part of transport connections, hence it is required that OAM functions are activated/deactivated in sync with connection commissioning/decommissioning; avoiding spurious alarms and ensuring consistent operation. In certain technologies OAM entities are inherently established once the connection is set up, while other technologies require extra configuration to establish and configure OAM entities. In some situations the use of OAM functions, like those of Fault- (FM) and Performance Management (PM), may be optional confirming to actual network management policies. Hence the network operator must be able to choose which kind of OAM functions to apply to specific connections and with what parameters the selected OAM functions should be configured and operated. To achieve this objective OAM entities and specific functions must be selectively configurable.

In general, it is required that the management plane and control plane connection establishment mechanisms are synchronized with OAM establishment and activation. In particular, if the GMPLS control plane is employed it is desirable to bind OAM setup and configuration to connection establishment signaling to avoid two separate management/configuration steps (connection setup followed by OAM configuration) which increases delay, processing and more importantly may be prone to misconfiguration errors. Once OAM entities are setup and configured, pro-active as well as on-demand OAM functions can be activated via the management plane. On the other hand, it should be possible to activate/deactivate pro-active OAM functions via the GMPLS control plane as well.

This document describes requirements on OAM configuration and control via RSVP-TE, and specifies extensions to the RSVP-TE protocol providing a framework to configure and control OAM entities along with the capability to carry technology specific information. Extensions can be grouped into generic elements that are applicable

to any OAM solution and technology specific elements that provide additional configuration parameters, only needed for a specific OAM technology. This document specifies the technology agnostic elements, which alone can be used to establish and control OAM entities in the case no technology specific information is needed, and specifies the way additional technology specific OAM parameters are provided.

This document addresses end-to-end OAM configuration, that is, the setup of OAM entities bound to an end-to-end LSP, and configuration and control of OAM functions running end-to-end in the LSP. Configuration of OAM entities for LSP segments and tandem connections are out of the scope of this document.

The mechanisms described in this document provide an additional option for bootstrapping OAM that is not intended to replace or deprecate the use of other technology specific OAM bootstrapping techniques; e.g., LSP Ping [RFC4379] for MPLS networks. The procedures specified in this document are intended only for use in environments where RSVP-TE signaling is already in use to set up the LSPs that are to be monitored using OAM.

2. Requirements

MPLS OAM requirements are described in [RFC4377], which provides requirements to create consistent OAM functionality for MPLS networks.

The following list is an excerpt of MPLS OAM requirements documented in [RFC4377]. Only a few requirements are discussed that bear a direct relevance to the discussion set forth in this document.

- o It is desired to support the automation of LSP defect detection. It is especially important in cases where large numbers of LSPs might be tested.
- o In particular some LSPs may require automated ingress-LSR to egress-LSR testing functionality, while others may not.
- o Mechanisms are required to coordinate network responses to defects. Such mechanisms may include alarm suppression, translating defect signals at technology boundaries, and synchronizing defect detection times by setting appropriately bounded detection timeframes.

MPLS-TP defines a profile of MPLS targeted at transport applications [RFC5921]. This profile specifies the specific MPLS characteristics and extensions required to meet transport requirements, including providing additional OAM, survivability and other maintenance functions not currently supported by MPLS. Specific OAM requirements for MPLS-TP are specified in [RFC5654] [RFC5860]. MPLS-TP poses requirements on the control plane to configure and control OAM entities:

- o From [RFC5860]: OAM functions MUST operate and be configurable even in the absence of a control plane. Conversely, it SHOULD be possible to configure as well as enable/disable the capability to operate OAM functions as part of connectivity management, and it SHOULD also be possible to configure as well as enable/disable the capability to operate OAM functions after connectivity has been established.
- o From [RFC5654]: The MPLS-TP control plane MUST support the configuration and modification of OAM maintenance points as well as the activation/ deactivation of OAM when the transport path or transport service is established or modified.

Ethernet Connectivity Fault Management (CFM) defines an adjunct connectivity monitoring OAM flow to check the liveness of Ethernet networks [IEEE-CFM]. With PBB-TE [IEEE-PBBTE] Ethernet networks

support explicitly-routed Ethernet connections. CFM can be used to track the liveness of PBB-TE connections and detect data plane failures. In IETF the GMPLS controlled Ethernet Label Switching (GELS) (see [RFC5828] and [RFC6060]) work extended the GMPLS control plane to support the establishment of PBB-TE data plane connections. Without control plane support separate management commands would be needed to configure and start CFM.

GMPLS based OAM configuration and control should be general to be applicable to a wide range of data plane technologies and OAM solutions. There are three typical data plane technologies used for transport application, which are wavelength based such as WSON, TDM based such as SDH/SONET, packet based such as MPLS-TP [RFC5921] and Ethernet PBB-TE [IEEE-PBBTE]. In all these data planes, the operator MUST be able to configure and control the following OAM functions.

- o It MUST be possible to explicitly request the setup of OAM entities for the signaled LSP and provide specific information for the setup if this is required by the technology.
- o Control of alarms is important to avoid false alarm indications and reporting to the management system. It MUST be possible to enable/disable alarms generated by OAM functions. In some cases selective alarm control may be desirable when, for instance, the operator is only concerned about critical alarms thus the non-service affecting alarms should be inhibited.
- o When periodic messages are used for liveness check (continuity check) of LSPs it MUST be possible to set the frequency of messages allowing proper configuration for fulfilling the requirements of the service and/or meeting the detection time boundaries posed by possible congruent connectivity check operations of higher layer applications. For a network operator to be able to balance the trade-off in fast failure detection and overhead it is beneficial to configure the frequency of continuity check messages on a per LSP basis.
- o Pro-active Performance Monitoring (PM) functions are continuously collecting information about specific characteristics of the connection. For consistent measurement of Service Level Agreements (SLAs) measurement points must use common probing rate to avoid measurement errors.
- o The extensions MUST allow the operator to use only a minimal set of OAM configuration and control features if the data plane technology, the OAM solution or network management policy allows. The extensions must be reusable as much as reasonably possible. That is generic OAM parameters and data plane or OAM technology

specific parameters must be separated.

3. RSVP-TE based OAM Configuration

In general, two types of Maintenance Points (MPs) can be distinguished: Maintenance End Points (MEPs) and Maintenance Intermediate Points (MIPs). MEPs reside at the ends of an LSP and are capable of initiating and terminating OAM messages for Fault Management (FM) and Performance Monitoring (PM). MIPs on the other hand are located at transit nodes of an LSP and are capable of reacting to some OAM messages but otherwise do not initiate messages. Maintenance Entity (ME) refers to an association of MEPs and MIPs that are provisioned to monitor an LSP. The ME association is achieved by configuring MPs to belong to the same ME.

When an LSP is signaled, forwarding association is established between endpoints and transit nodes via label bindings. This association creates a context for the OAM entities monitoring the LSP. On top of this association OAM entities may be configured to unambiguously identify MPs and MEs.

In addition to MP and ME identification parameters pro-active OAM functions (e.g., Continuity Check (CC), Performance Monitoring) may have specific parameters requiring configuration as well. In particular, the frequency of periodic CC packets and the measurement interval for loss and delay measurements may need to be configured.

In some cases all the above parameters may be either derived from some existing information or pre-configured default values can be used. In the simplest case the control plane needs to provide information whether or not OAM entities need to be setup for the signaled LSP. If OAM entities are created signaling must provide means to activate/deactivate OAM message flows and associated alarms.

OAM identifiers as well as the configuration of OAM functions are technology specific, i.e., vary depending on the data plane technology and the chosen OAM solution. In addition, for any given data plane technology a set of OAM solutions may be applicable. The OAM configuration framework allows selecting a specific OAM solution to be used for the signaled LSP and provides technology specific TLVs to carry further detailed configuration information.

3.1. Establishment of OAM Entities and Functions

In order to avoid spurious alarms OAM functions must be setup and enabled in the appropriate order. When using the GMPLS control plane, establishment and enabling of OAM functions must be bound to RSVP-TE message exchanges.

An LSP may be signaled and established without OAM configuration

first, and OAM entities may be added later with a subsequent re-signaling of the LSP. Alternatively, the LSP may be setup with OAM entities right with the first signaling of the LSP. The below procedures apply to both cases.

Before the initiator first sends a Path messages with OAM Configuration information, it MUST establish and configure the corresponding OAM entities locally, however OAM source functions MUST NOT start sending any OAM messages. In the case of bidirectional connections, the initiator node MUST setup the OAM sink function to be prepared to receive OAM messages but MUST suppress any OAM alarms (e.g., due to missing or unidentified OAM messages). The Path message MUST be sent with the "OAM Alarms Enabled" ADMIN_STATUS flag cleared, i.e, data plane OAM alarms are suppressed.

When the Path message arrives at the receiver, the remote end MUST establish and configure OAM entities according to the OAM information provided in the Path message. If this is not possible a PathErr SHOULD be sent and neither the OAM entities nor the LSP SHOULD be established. If OAM entities are established successfully, the OAM sink function MUST be prepared to receive OAM messages but MUST not generate any OAM alarms (e.g., due to missing or unidentified OAM messages). In the case of bidirectional connections, an OAM source function MUST be setup and, according to the requested configuration, the OAM source function MUST start sending OAM messages. Then a Resv message is sent back, including the OAM Configuration TLV that corresponds to the actually established and configured OAM entities and functions. Depending on the OAM technology, some elements of the OAM Configuration TLV MAY be updated/changed; i.e., if the remote end is not supporting a certain OAM configuration it may suggest an alternative setting, which may or may not be accepted by the initiator of the Path message. If it is accepted, the initiator will reconfigure its OAM functions according to the information received in the Resv message. If the alternate setting is not acceptable a ResvErr may be sent tearing down the LSP. Details of this operation are technology specific and should be described in accompanying technology specific documents.

When the initiating side receives the Resv message it completes any pending OAM configuration and enables the OAM source function to send OAM messages.

After this round, OAM entities are established and configured for the LSP and OAM messages are already exchanged. OAM alarms can now be enabled. The initiator, while still keeping OAM alarms disabled sends a Path message with "OAM Alarms Enabled" ADMIN_STATUS flag set. The receiving node enables the OAM alarms after processing the Path message. The initiator enables OAM alarms after it receives the Resv

message. Data plane OAM is now fully functional.

3.2. Adjustment of OAM Parameters

There may be a need to change the parameters of an already established and configured OAM function during the lifetime of the LSP. To do so the LSP needs to be re-signaled with the updated parameters. OAM parameters influence the content and timing of OAM messages and identify the way OAM defects and alarms are derived and generated. Hence, to avoid spurious alarms, it is important that both sides, OAM sink and source, are updated in a synchronized way. First, the alarms of the OAM sink function should be suppressed and only then should expected OAM parameters be adjusted. Subsequently, the parameters of the OAM source function can be updated. Finally, the alarms of the OAM sink side can be enabled again.

In accordance with the above operation, the LSP MUST first be re-signaled with "OAM Alarms Enabled" ADMIN_STATUS flag cleared and including the updated OAM Configuration TLV corresponding to the new parameter settings. The initiator MUST keep its OAM sink and source functions running unmodified, but it MUST suppress OAM alarms after the updated Path message is sent. The receiver MUST first disable all OAM alarms, then update the OAM parameters according to the information in the Path message and reply with a Resv message acknowledging the changes by including the OAM Configuration TLV. Note that the receiving side has the possibility to adjust the requested OAM configuration parameters and reply with an updated OAM Configuration TLV in the Resv message, reflecting the actually configured values. However, in order to avoid an extensive negotiation phase, in the case of adjusting already configured OAM functions, the receiving side SHOULD NOT update the parameters requested in the Path message to an extent that would provide lower performance than what has been configured previously.

The initiator MUST only update its OAM sink and source functions after it received the Resv message. After this Path/Resv message exchange (in both unidirectional and bidirectional LSP cases) the OAM parameters are updated and OAM is running according to the new parameter settings. However OAM alarms are still disabled. A subsequent Path/Resv message exchange with "OAM Alarms Enabled" ADMIN_STATUS flag set is needed to enable OAM alarms again.

3.3. Deleting OAM Entities

In some cases it may be useful to remove some or all OAM entities and functions from an LSP without actually tearing down the connection.

To avoid any spurious alarm, first the LSP SHOULD be re-signaled with

"OAM Alarms Enabled" ADMIN_STATUS flag cleared but unchanged OAM configuration. Subsequently, the LSP is re-signaled with "OAM MEP Entities desired" and "OAM MIP Entities desired" LSP ATTRIBUTES flags cleared, and without the OAM Configuration TLV, this MUST result in the deletion of all OAM entities associated with the LSP. All control and data plane resources in use by the OAM entities and functions SHOULD be freed up. Alternatively, if only some OAM functions need to be removed, the LSP is re-signalled with the updated OAM Configuration TLV. Changes between the contents of the previously signalled OAM Configuration TLV and the currently received TLV represent which functions SHOULD be removed/added.

First, OAM source functions SHOULD be deleted and only after that SHOULD the associated OAM sink functions be removed, this will ensure that OAM messages do not leak outside the LSP. To this end the initiator, before sending the Path message, SHOULD remove the OAM source, hence terminating the OAM message flow associated to the downstream direction. In the case of a bidirectional connection, it SHOULD leave in place the OAM sink functions associated to the upstream direction. The remote end, after receiving the Path message, SHOULD remove all associated OAM entities and functions and reply with a Resv message without an OAM Configuration TLV. The initiator completely removes OAM entities and functions after the Resv message arrived.

4. RSVP-TE Extensions

4.1. LSP Attributes Flags

In RSVP-TE the Flags field of the SESSION_ATTRIBUTE object is used to indicate options and attributes of the LSP. The Flags field has 8 bits and hence is limited to differentiate only 8 options. [RFC5420] defines new objects for RSVP-TE messages to allow the signaling of arbitrary attribute parameters making RSVP-TE easily extensible to support new applications. Furthermore, [RFC5420] allows options and attributes that do not need to be acted on by all Label Switched Routers (LSRs) along the path of the LSP. In particular, these options and attributes may apply only to key LSRs on the path such as the ingress LSR and egress LSR. Options and attributes can be signaled transparently, and only examined at those points that need to act on them. The LSP_ATTRIBUTES and the LSP_REQUIRED_ATTRIBUTES objects are defined in [RFC5420] to provide means to signal LSP attributes and options in the form of TLVs. Options and attributes signaled in the LSP_ATTRIBUTES object can be passed transparently through LSRs not supporting a particular option or attribute, while the contents of the LSP_REQUIRED_ATTRIBUTES object must be examined and processed by each LSR. One TLV is defined in [RFC5420]: the Attributes Flags TLV.

One bit (IANA to assign): "OAM MEP entities desired" is allocated in the LSP Attributes Flags TLV. If the "OAM MEP entities desired" bit is set it is indicating that the establishment of OAM MEP entities are required at the endpoints of the signaled LSP. If the establishment of MEPs is not supported an error must be generated: "OAM Problem/MEP establishment not supported".

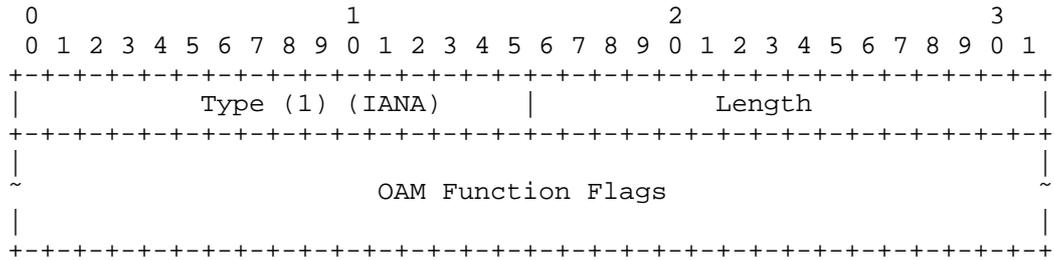
If the "OAM MEP entities desired" bit is set but additional parameters need also to be configured, an OAM Configuration TLV MAY be included in the LSP_ATTRIBUTES Object.

One bit (IANA to assign): "OAM MIP entities desired" is allocated in the LSP Attributes Flags TLV. This bit can only be set if the "OAM MEP entities desired" bit is set. If the "OAM MIP entities desired" bit is set in the LSP_ATTRIBUTES Flags TLV in the LSP_REQUIRED_ATTRIBUTES Object, it is indicating that the establishment of OAM MIP entities is required at every transit node of the signalled LSP. If the establishment of a MIP is not supported an error must be generated: "OAM Problem/MIP establishment not supported".

Note that there is a hierarchical dependency in between the OAM configuration elements. First, the "OAM MEP (and MIP) entities desired" flag needs to be set. Only when that is set MAY an "OAM Configuration TLV" be included in the LSP_ATTRIBUTES or LSP_REQUIRED_ATTRIBUTES Object. When this TLV is present, based on the "OAM Type" field, it MAY carry a technology specific OAM configuration sub-TLV. If this hierarchy is broken (e.g., "OAM MEP entities desired" flag is not set but an OAM Configuration TLV is present) an error MUST be generated: "OAM Problem/Configuration Error".

4.2.1. OAM Function Flags Sub-TLV

As the first sub-TLV the "OAM Function Flags sub-TLV" MUST be always included in the "OAM Configuration TLV". "OAM Function Flags" specifies which pro-active OAM functions (e.g., connectivity monitoring, loss and delay measurement) and which fault management signals MUST be established and configured. If the selected OAM Function(s) is(are) not supported, an error MUST be generated: "OAM Problem/Unsupported OAM Function".



OAM Function Flags is bitmap with extensible length based on the Length field of the TLV. Bits are numbered from left to right. IANA is requested to maintain the OAM Function Flags in the new "RSVP-TE OAM Configuration Registry". This document defines the following flags.

OAM Function Flag bit#	Description
0	Continuity Check (CC)
1	Connectivity Verification (CV)
2	Performance Monitoring/Loss (PM/Loss)
3	Performance Monitoring/Delay (PM/Delay)

4.2.2. Technology Specific sub-TLVs

One technology specific sub-TLV MAY be defined for each "OAM Type". This sub-TLV MUST contain any further OAM configuration information for that specific "OAM Type". The technology specific sub-TLV, when used, MUST be carried within the OAM Configuration TLV. IANA is requested to maintain the sub-TLV space in the new "RSVP-TE OAM Configuration Registry".

4.3. Administrative Status Information

Administrative Status Information is carried in the ADMIN_STATUS Object. The Administrative Status Information is described in [RFC3471], the ADMIN_STATUS Object is specified for RSVP-TE in [RFC3473].

Two bits are allocated for the administrative control of OAM monitoring. Two bits (IANA to assign) are allocated by this draft: the "OAM Flows Enabled" (M) and "OAM Alarms Enabled" (O) bits. When the "OAM Flows Enabled" bit is set, OAM packets are sent if it is cleared no OAM packets are emitted. When the "OAM Alarms Enabled" bit is set OAM triggered alarms are enabled and associated consequent actions are executed including the notification of the management system. When this bit is cleared, alarms are suppressed and no action is executed and the management system is not notified.

4.4. Handling OAM Configuration Errors

To handle OAM configuration errors a new Error Code (IANA to assign) "OAM Problem" is introduced. To refer to specific problems a set of Error Values is defined.

If a node does not support the establishment of OAM MEP or MIP entities it must use the error value (IANA to assign): "MEP establishment not supported" or "MIP establishment not supported" respectively in the PathErr message.

If a node does not support a specific OAM technology/solution it must use the error value (IANA to assign): "Unsupported OAM Type" in the PathErr message.

If a different technology specific OAM configuration TLV is included than what was specified in the OAM Type an error must be generated with error value: "OAM Type Mismatch" in the PathErr message.

There is a hierarchy in between the OAM configuration elements. If this hierarchy is broken the error value: "Configuration Error" must be used in the PathErr message.

If a node does not support a specific OAM Function it must use the error value: "Unsupported OAM Function" in the PathErr message.

4.5. Considerations on Point-to-Multipoint OAM Configuration

RSVP-TE extensions for the establishment of point-to-multipoint (P2MP) LSPs are specified in [RFC4875]. A P2MP LSP is comprised of multiple source-to-leaf (S2L) sub-LSPs. These S2L sub-LSPs are set up between the ingress and egress LSRs and are appropriately combined by the branch LSRs using RSVP semantics to result in a P2MP TE LSP. One Path message may signal one or multiple S2L sub-LSPs for a single P2MP LSP. Hence the S2L sub-LSPs belonging to a P2MP LSP can be signaled using one Path message or split across multiple Path messages.

P2MP OAM mechanisms are very specific to the data plane technology, hence in this document we only highlight basic operations for P2MP OAM configuration. We consider only the configuration of the root to leaves OAM flows of P2MP LSPs and as such aspects of any return path are outside the scope of our discussions. We also limit our consideration to cases where all leaves must successfully establish OAM entities in order a P2MP OAM is successfully established. In any case, the discussion set forth below provides only guidelines for P2MP OAM configuration, details SHOULD be specified in technology specific documents.

The root node may select if it uses a single Path message or multiple Path messages to setup the whole P2MP tree. In the case when multiple Path messages are used the root node is responsible also to keep the OAM Configuration information consistent in each of the sent Path messages, i.e., the same information MUST be included in all Path messages used to construct the multicast tree. Each branching node will propagate the Path message downstream on each of the branches, when constructing a Path message the OAM Configuration information MUST be copied unchanged from the received Path message, including the related ADMIN_STATUS bits, LSP Attribute Flags and the OAM Configuration TLV. The latter two also imply that the LSP_ATTRIBUTES and LSP_REQUIRED_ATTRIBUTES Object MUST be copied for the upstream Path message to the subsequent downstream Path messages.

Leaves MUST create and configure OAM sink functions according to the parameters received in the Path message, for P2MP OAM configuration there is no possibility for parameter negotiation on a per leaf basis. This is due to the fact that the only OAM source function, residing in the root of the tree, can only operate with a single configuration which must be obeyed by all leaves. If a leaf cannot accept the OAM parameters it MUST use the RRO Attributes sub-object [RFC5420] to notify the root of the problem. In particular, if the

OAM configuration was successful the leaf would set the "OAM MEP entities desired" flag in the RRO Attributes sub-object in the Resv message, while, if due to any reason, OAM entities could not be established the Resv message should be sent with the "OAM MEP entities desired" bit cleared in the RRO Attributes sub-object. Branching nodes should collect and merge the received RROs according to the procedures described in [RFC4875]. This way, the root when receiving the Resv message (or messages if multiple Path messages were used to setup the tree) will have a clear information on which of the leaves could the OAM sink functions be established. If all leaves established OAM entities successfully, the root can enable the OAM message flow. On the other hand, if at some leaves the establishment was unsuccessful additional actions will be needed before the OAM message flow can be enabled. Such action could be to setup two independent P2MP LSPs. One with OAM Configuration information towards leaves which successfully setup OAM. This can be done by pruning the leaves which failed to setup OAM of the previously signalled P2MP LSP. The other P2MP LSP could be constructed for leaves without OAM entities. What exact procedures are needed are technology specific and should be described in technology specific documents.

5. IANA Considerations

Two bits ("OAM Alarms Enabled" (O) and "OAM Flows Enabled" (M)) needs to be allocated in the ADMIN_STATUS Object.

Two bits ("OAM MEP entities desired" and "OAM MIP entities desired") needs to be allocated in the LSP Attributes Flags Registry.

This document specifies one new TLV to be carried in the LSP_ATTRIBUTES and LSP_REQUIRED_ATTRIBUTES objects in Path and Resv messages: OAM Configuration TLV.

One new Error Code: "OAM Problem" and a set of new values: "MEP establishment not supported", "MIP establishment not supported", "Unsupported OAM Type", "Configuration Error" and "Unsupported OAM Function" needs to be assigned.

IANA is requested to open a new registry: "RSVP-TE OAM Configuration Registry" that maintains the "OAM Type" code points, an associated sub-TLV space, and the allocations of "OAM Function Flags" within the OAM Configuration TLV.

6. Security Considerations

The signaling of OAM related parameters and the automatic establishment of OAM entities based on RSVP-TE messages adds a new aspect to the security considerations discussed in [RFC3473]. In particular, a network element could be overloaded, if a remote attacker could request liveliness monitoring, with frequent periodic messages, for a high number of LSPs, targeting a single network element. Such an attack can efficiently be prevented when mechanisms for message integrity and node authentication are deployed. Since the OAM configuration extensions rely on the hop-by-hop exchange of existing RSVP-TE messages, procedures specified for RSVP message security in [RFC2747] can be used to mitigate possible attacks.

For a more comprehensive discussion on GMPLS security please see the Security Framework for MPLS and GMPLS Networks [RFC5920]. Cryptography can be used to protect against many attacks described in [RFC5920].

7. Acknowledgements

The authors would like to thank Francesco Fondelli, Adrian Farrel, Loa Andersson, Eric Gray and Dimitri Papadimitriou for their useful comments.

8. References

8.1. Normative References

- [RFC3471] "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Functional Description", RFC 3471, January 2003.
- [RFC3473] "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions", RFC 3473, January 2003.
- [RFC5420] "Encoding of Attributes for Multiprotocol Label Switching (MPLS) Label Switched Path (LSP) Establishment Using Resource ReserVation Protocol-Traffic Engineering (RSVP-TE)", RFC 5420, February 2009.

8.2. Informative References

- [IEEE-CFM] "IEEE 802.lag, Draft Standard for Connectivity Fault Management", work in progress.
- [IEEE-PBBTE] "IEEE 802.1Qay Draft Standard for Provider Backbone Bridging Traffic Engineering", work in progress.
- [RFC2747] "RSVP Cryptographic Authentication", RFC 2747, January 2000.
- [RFC3469] "Framework for Multi-Protocol Label Switching (MPLS)-based Recovery", RFC 3469, February 2003.
- [RFC4377] "Operations and Management (OAM) Requirements for Multi-Protocol Label Switched (MPLS) Networks", RFC 4377, February 2006.
- [RFC4379] "Detecting Multi-Protocol Label Switched (MPLS) Data Plane Failures", RFC 4379, February 2006.
- [RFC4875] "Extensions to Resource Reservation Protocol - Traffic Engineering (RSVP-TE) for Point-to-Multipoint TE Label Switched Paths (LSPs)", RFC 4875, May 2007.
- [RFC5654] "Requirements of an MPLS Transport Profile", RFC 5654, September 2009.
- [RFC5828] "GMPLS Ethernet Label Switching Architecture and Framework", RFC 5828, March 2010.

- [RFC5860] "Requirements for OAM in MPLS Transport Networks", RFC 5860, May 2010.
- [RFC5920] "Security Framework for MPLS and GMPLS Networks", RFC 5920, July 2010.
- [RFC5921] "A Framework for MPLS in Transport Networks", RFC 5921, July 2010.
- [RFC6060] "Generalized Multiprotocol Label Switching (GMPLS) Control of Ethernet Provider Backbone Traffic Engineering (PBB-TE)", RFC 6060.

Authors' Addresses

Attila Takacs
Ericsson
Laborc u. 1.
Budapest, 1037
Hungary

Email: attila.takacs@ericsson.com

Don Fedyk
Alcatel-Lucent
Groton, MA 01450
USA

Email: donald.fedyk@alcatel-lucent.com

Jia He
Huawei

Email: hejia@huawei.com

Network Working Group
Internet Draft
Intended status: Informational
Expires: September 2011

Y. Lee
Huawei
G. Bernstein
Grotto Networking
D. Li
Huawei
W. Imajuku
NTT

March 14, 2011

Routing and Wavelength Assignment Information Model for Wavelength
Switched Optical Networks

draft-ietf-ccamp-rwa-info-11.txt

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at
<http://www.ietf.org/ietf/lid-abstracts.txt>

The list of Internet-Draft Shadow Directories can be accessed at
<http://www.ietf.org/shadow.html>

This Internet-Draft will expire on September 14, 2011.

Copyright Notice

Copyright (c) 2011 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Abstract

This document provides a model of information needed by the routing and wavelength assignment (RWA) process in wavelength switched optical networks (WSONs). The purpose of the information described in this model is to facilitate constrained optical path computation in WSONs. This model takes into account compatibility constraints between WSON signal attributes and network elements but does not include constraints due to optical impairments. Aspects of this information that may be of use to other technologies utilizing a GMPLS control plane are discussed.

Table of Contents

1. Introduction.....	3
1.1. Revision History.....	4
1.1.1. Changes from 01.....	4
1.1.2. Changes from 02.....	4
1.1.3. Changes from 03.....	4
1.1.4. Changes from 04.....	5
1.1.5. Changes from 05.....	5
1.1.6. Changes from 06.....	5
1.1.7. Changes from 07.....	5
1.1.8. Changes from 08.....	5
1.1.9. Changes from 09.....	5
1.1.10. Changes from 10.....	6
2. Terminology.....	6
3. Routing and Wavelength Assignment Information Model.....	6
3.1. Dynamic and Relatively Static Information.....	7
4. Node Information (General).....	7
4.1. Connectivity Matrix.....	8
4.2. Shared Risk Node Group.....	8
5. Node Information (WSON specific).....	9
5.1. Resource Accessibility/Availability.....	10

5.2. Resource Signal Constraints and Processing Capabilities..	14
5.3. Compatibility and Capability Details.....	15
5.3.1. Shared Ingress or Egress Indication.....	15
5.3.2. Modulation Type List.....	15
5.3.3. FEC Type List.....	15
5.3.4. Bit Rate Range List.....	15
5.3.5. Acceptable Client Signal List.....	16
5.3.6. Processing Capability List.....	16
6. Link Information (General).....	16
6.1. Administrative Group.....	17
6.2. Interface Switching Capability Descriptor.....	17
6.3. Link Protection Type (for this link).....	17
6.4. Shared Risk Link Group Information.....	17
6.5. Traffic Engineering Metric.....	17
6.6. Port Label (Wavelength) Restrictions.....	17
6.6.1. Port-Wavelength Exclusivity Example.....	19
7. Dynamic Components of the Information Model.....	20
7.1. Dynamic Link Information (General).....	21
7.2. Dynamic Node Information (WSON Specific).....	21
8. Security Considerations.....	21
9. IANA Considerations.....	22
10. Acknowledgments.....	22
11. References.....	23
11.1. Normative References.....	23
11.2. Informative References.....	24
12. Contributors.....	25
Author's Addresses.....	25
Intellectual Property Statement.....	26
Disclaimer of Validity.....	27

1. Introduction

The purpose of the following information model for WSONs is to facilitate constrained optical path computation and as such is not a general purpose network management information model. This constraint is frequently referred to as the "wavelength continuity" constraint, and the corresponding constrained optical path computation is known as the routing and wavelength assignment (RWA) problem. Hence the information model must provide sufficient topology and wavelength restriction and availability information to support this computation. More details on the RWA process and WSON subsystems and their properties can be found in [WSON-Frame]. The model defined here includes constraints between WSON signal attributes and network elements, but does not include optical impairments.

In addition to presenting an information model suitable for path computation in WSON, this document also highlights model aspects that

may have general applicability to other technologies utilizing a GMPLS control plane. The portion of the information model applicable to other technologies beyond WSON is referred to as "general" to distinguish it from the "WSON-specific" portion that is applicable only to WSON technology.

1.1. Revision History

1.1.1. Changes from 01

Added text on multiple fixed and switched connectivity matrices.

Added text on the relationship between SRNG and SRLG and encoding considerations.

Added clarifying text on the meaning and use of port/wavelength restrictions.

Added clarifying text on wavelength availability information and how to derive wavelengths currently in use.

1.1.2. Changes from 02

Integrated switched and fixed connectivity matrices into a single "connectivity matrix" model. Added numbering of matrices to allow for wavelength (time slot, label) dependence of the connectivity. Discussed general use of this node parameter beyond WSON.

Integrated switched and fixed port wavelength restrictions into a single port wavelength restriction of which there can be more than one and added a reference to the corresponding connectivity matrix if there is one. Also took into account port wavelength restrictions in the case of symmetric switches, developed a uniform model and specified how general label restrictions could be taken into account with this model.

Removed the Shared Risk Node Group parameter from the node info, but left explanation of how the same functionality can be achieved with existing GMPLS SRLG constructs.

Removed Maximum bandwidth per channel parameter from link information.

1.1.3. Changes from 03

Removed signal related text from section 3.2.4 as signal related information is deferred to a new signal compatibility draft.

Removed encoding specific text from Section 3.3.1 of version 03.

1.1.4. Changes from 04

Removed encoding specific text from Section 4.1.

Removed encoding specific text from Section 3.4.

1.1.5. Changes from 05

Renumbered sections for clarity.

Updated abstract and introduction to encompass signal compatibility/generalization.

Generalized Section on wavelength converter pools to include electro optical subsystems in general. This is where signal compatibility modeling was added.

1.1.6. Changes from 06

Simplified information model for WSON specifics, by combining similar fields and introducing simpler aggregate information elements.

1.1.7. Changes from 07

Added shared fiber connectivity to resource pool modeling. This includes information for determining wavelength collision on an internal fiber providing access to resource blocks.

1.1.8. Changes from 08

Added PORT_WAVELENGTH_EXCLUSIVITY in the RestrictionType parameter. Added section 6.6.1 that has an example of the port wavelength exclusivity constraint.

1.1.9. Changes from 09

Section 5: clarified the way that the resource pool is modeled from blocks of identical resources.

Section 5.1: grammar fixes. Removed reference to "academic" modeling pre-print. Clarified RBNF resource pool model details.

Section 5.2: Formatting fixes.

1.1.10. Changes from 10

Enhanced the explanation of shared fiber access to resources and updated Figure 2 to show a more general situation to be modeled.

Removed all 1st person idioms.

2. Terminology

CWDM: Coarse Wavelength Division Multiplexing.

DWDM: Dense Wavelength Division Multiplexing.

FOADM: Fixed Optical Add/Drop Multiplexer.

ROADM: Reconfigurable Optical Add/Drop Multiplexer. A reduced port count wavelength selective switching element featuring ingress and egress line side ports as well as add/drop side ports.

RWA: Routing and Wavelength Assignment.

Wavelength Conversion. The process of converting an information bearing optical signal centered at a given wavelength to one with "equivalent" content centered at a different wavelength. Wavelength conversion can be implemented via an optical-electronic-optical (OEO) process or via a strictly optical process.

WDM: Wavelength Division Multiplexing.

Wavelength Switched Optical Network (WSON): A WDM based optical network in which switching is performed selectively based on the center wavelength of an optical signal.

3. Routing and Wavelength Assignment Information Model

The following WSON RWA information model is grouped into four categories regardless of whether they stem from a switching subsystem or from a line subsystem:

- o Node Information
- o Link Information
- o Dynamic Node Information
- o Dynamic Link Information

Note that this is roughly the categorization used in [G.7715] section 7.

In the following, where applicable, the reduced Backus-Naur form (RBNF) syntax of [RBNF] is used to aid in defining the RWA information model.

3.1. Dynamic and Relatively Static Information

All the RWA information of concern in a WSON network is subject to change over time. Equipment can be upgraded; links may be placed in or out of service and the like. However, from the point of view of RWA computations there is a difference between information that can change with each successive connection establishment in the network and that information that is relatively static on the time scales of connection establishment. A key example of the former is link wavelength usage since this can change with connection setup/teardown and this information is a key input to the RWA process. Examples of relatively static information are the potential port connectivity of a WDM ROADM, and the channel spacing on a WDM link.

This document separates, where possible, dynamic and static information so that these can be kept separate in possible encodings and hence allowing for separate updates of these two types of information thereby reducing processing and traffic load caused by the timely distribution of the more dynamic RWA WSON information.

4. Node Information (General)

The node information described here contains the relatively static information related to a WSON node. This includes connectivity constraints amongst ports and wavelengths since WSON switches can exhibit asymmetric switching properties. Additional information could include properties of wavelength converters in the node if any are present. In [Switch] it was shown that the wavelength connectivity constraints for a large class of practical WSON devices can be modeled via switched and fixed connectivity matrices along with corresponding switched and fixed port constraints. These connectivity matrices are included with the node information while the switched and fixed port wavelength constraints are included with the link information.

Formally,

```
<Node_Information> ::= <Node_ID> [<ConnectivityMatrix>...]
```

Where the Node_ID would be an appropriate identifier for the node within the WSON RWA context.

Note that multiple connectivity matrices are allowed and hence can fully support the most general cases enumerated in [Switch].

4.1. Connectivity Matrix

The connectivity matrix (ConnectivityMatrix) represents either the potential connectivity matrix for asymmetric switches (e.g. ROADMs and such) or fixed connectivity for an asymmetric device such as a multiplexer. Note that this matrix does not represent any particular internal blocking behavior but indicates which ingress ports and wavelengths could possibly be connected to a particular output port. Representing internal state dependent blocking for a switch or ROADM is beyond the scope of this document and due to its highly implementation dependent nature would most likely not be subject to standardization in the future. The connectivity matrix is a conceptual M by N matrix representing the potential switched or fixed connectivity, where M represents the number of ingress ports and N the number of egress ports. This is a "conceptual" matrix since the matrix tends to exhibit structure that allows for very compact representations that are useful for both transmission and path computation [Encode].

Note that the connectivity matrix information element can be useful in any technology context where asymmetric switches are utilized.

ConnectivityMatrix ::= <MatrixID> <ConnType> <Matrix>

Where

<MatrixID> is a unique identifier for the matrix.

<ConnType> can be either 0 or 1 depending upon whether the connectivity is either fixed or potentially switched.

<Matrix> represents the fixed or switched connectivity in that $\text{Matrix}(i, j) = 0$ or 1 depending on whether ingress port i can connect to egress port j for one or more wavelengths.

4.2. Shared Risk Node Group

SRNG: Shared risk group for nodes. The concept of a shared risk link group was defined in [RFC4202]. This can be used to achieve a desired "amount" of link diversity. It is also desirable to have a similar capability to achieve various degrees of node diversity. This is

explained in [G.7715]. Typical risk groupings for nodes can include those nodes in the same building, within the same city, or geographic region.

Since the failure of a node implies the failure of all links associated with that node a sufficiently general shared risk link group (SRLG) encoding, such as that used in GMPLS routing extensions can explicitly incorporate SRNG information.

5. Node Information (WSON specific)

As discussed in [WSON-Frame] a WSON node may contain electro-optical subsystems such as regenerators, wavelength converters or entire switching subsystems. The model present here can be used in characterizing the accessibility and availability of limited resources such as regenerators or wavelength converters as well as WSON signal attribute constraints of electro-optical subsystems. As such this information element is fairly specific to WSON technologies.

A WSON node may include regenerators or wavelength converters arranged in a shared pool. As discussed in [WSON-Frame] this can include OEO based WDM switches as well. There are a number of different approaches used in the design of WDM switches containing regenerator or converter pools. However, from the point of view of path computation the following need to be known:

1. The nodes that support regeneration or wavelength conversion.
2. The accessibility and availability of a wavelength converter to convert from a given ingress wavelength on a particular ingress port to a desired egress wavelength on a particular egress port.
3. Limitations on the types of signals that can be converted and the conversions that can be performed.

For modeling purposes and encoding efficiency identical processing resources such as regenerators or wavelength converters with identical limitations, and processing and accessibility properties are grouped into "blocks". Such blocks can consist of a single resource, though grouping resources into blocks leads to more efficient encodings. The resource pool model is composed of one or more resource blocks where the accessibility to and from any resource within a block is the same.

This leads to the following formal high level model:

```
<Node_Information> ::= <Node_ID> [<ConnectivityMatrix>...]  
[<ResourcePool>]
```

Where

```
<ResourcePool> ::= <ResourceBlockInfo>...  
[<ResourceBlockAccessibility>...] [<ResourceWaveConstraints>...]  
[<RBPoolState>]
```

First the accessibility of resource blocks is addressed then their properties are discussed.

5.1. Resource Accessibility/Availability

A similar technique as used to model ROADMs and optical switches can be used to model regenerator/converter accessibility. This technique was generally discussed in [WSON-Frame] and consisted of a matrix to indicate possible connectivity along with wavelength constraints for links/ports. Since regenerators or wavelength converters may be considered a scarce resource it is desirable that the model include, if desired, the usage state (availability) of individual regenerators or converters in the pool. Models that incorporate more state to further reveal blocking conditions on ingress or egress to particular converters are for further study and not included here.

The three stage model is shown schematically in Figure 1 and Figure 2. The difference between the two figures is that Figure 1 assumes that each signal that can get to a resource block may do so, while in Figure 2 the access to sets of resource blocks is via a shared fiber which imposes its own wavelength collision constraint. The representation of Figure 1 can have more than one ingress to each resource block since each ingress represents a single wavelength signal, while in Figure 2 shows a single multiplexed WDM ingress or egress, e.g., a fiber, to/from each set of block.

This model assumes N ingress ports (fibers), P resource blocks containing one or more identical resources (e.g. wavelength converters), and M egress ports (fibers). Since not all ingress ports can necessarily reach each resource block, the model starts with a resource pool ingress matrix $RI(i,p) = \{0,1\}$ whether ingress port i can reach potentially reach resource block p .

Since not all wavelengths can necessarily reach all the resources or the resources may have limited input wavelength range the model has a set of relatively static ingress port constraints for each resource. In addition, if the access to a set of resource blocks is via a shared fiber (Figure 2) this would impose a dynamic wavelength

availability constraint on that shared fiber. The resource block ingress port constraint is modeled via a static wavelength set mechanism and the case of shared access to a set of blocks is modeled via a dynamic wavelength set mechanism.

Next a state vector $RA(j) = \{0, \dots, k\}$ is used to track the number of resources in resource block j in use. This is the only state kept in the resource pool model. This state is not necessary for modeling "fixed" transponder system or full OEO switches with WDM interfaces, i.e., systems where there is no sharing.

After that, a set of static resource egress wavelength constraints and possibly dynamic shared egress fiber constraints maybe used. The static constraints indicate what wavelengths a particular resource block can generate or are restricted to generating e.g., a fixed regenerator would be limited to a single lambda. The dynamic constraints would be used in the case where a single shared fiber is used to egress the resource block (Figure 2).

Finally, to complete the model, a resource pool egress matrix $RE(p,k) = \{0,1\}$ depending on whether the output from resource block p can reach egress port k , may be used.

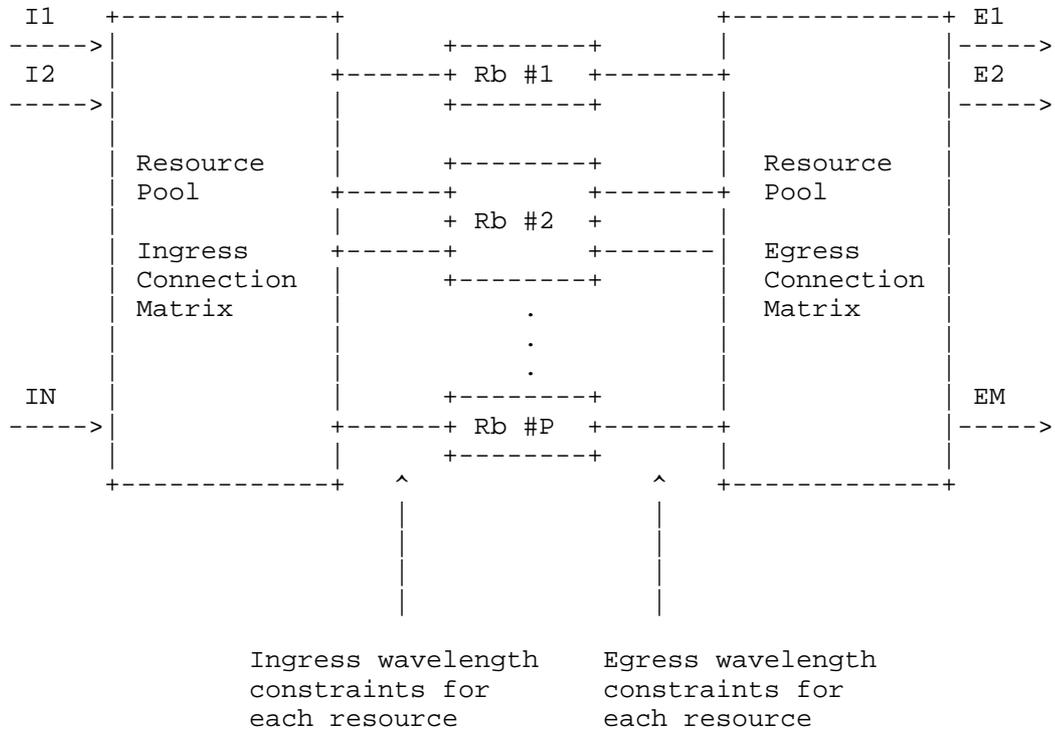


Figure 1 Schematic diagram of resource pool model.

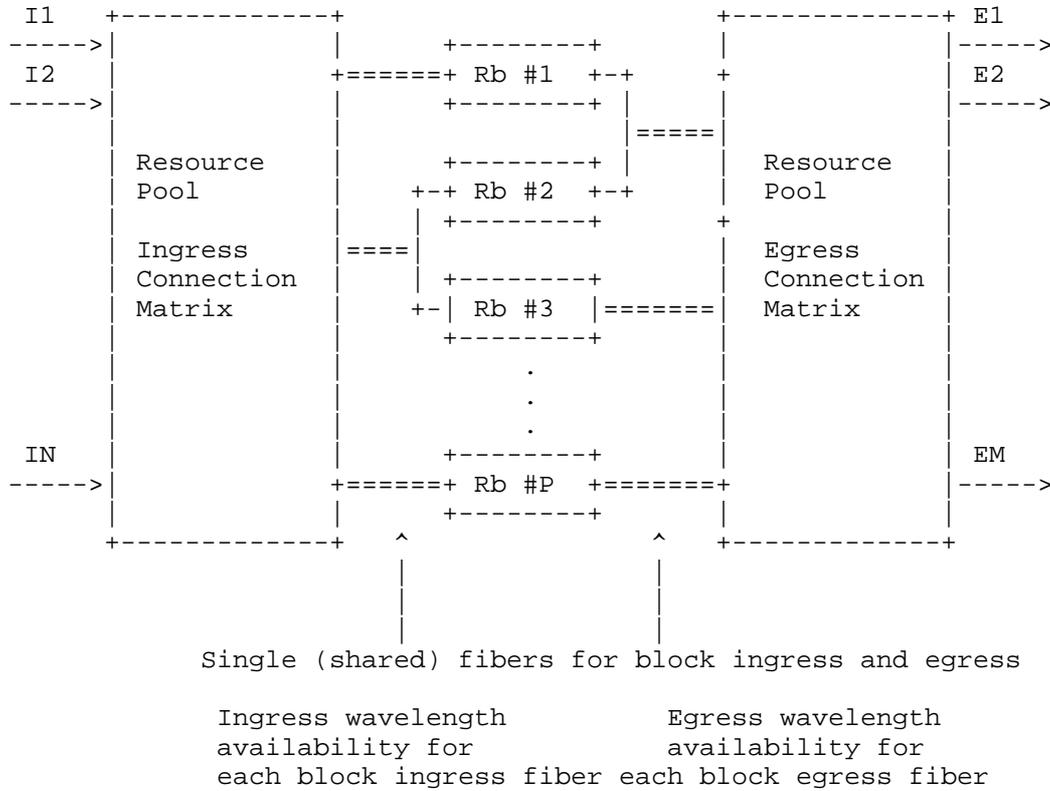


Figure 2 Schematic diagram of resource pool model with shared block accessibility.

Formally the model can be specified as:

```

<ResourceBlockAccessibility> ::= <PoolIngressMatrix>
<PoolEgressMatrix>

<ResourceWaveConstraints> ::= <IngressWaveConstraints>
<EgressWaveConstraints>

<RBPoolState>
::=( <ResourceBlockID> <NumResourcesInUse> <InAvailableWavelengths> <OutA
availableWavelengths> )...
  
```

Note that except for <RBPoolState> all the other components of <ResourcePool> are relatively static. Also the <InAvailableWavelengths> and <OutAvailableWavelengths> are only used in the cases of shared ingress or egress access to the particular block. See the resource block information in the next section to see how this is specified.

5.2. Resource Signal Constraints and Processing Capabilities

The wavelength conversion abilities of a resource (e.g. regenerator, wavelength converter) were modeled in the <EgressWaveConstraints> previously discussed. As discussed in [WSON-Frame] the constraints on an electro-optical resource can be modeled in terms of input constraints, processing capabilities, and output constraints:

```
<ResourceBlockInfo> ::= ([<ResourceSet>] <InputConstraints>
<ProcessingCapabilities> <OutputConstraints>)*
```

Where <ResourceSet> is a list of resource block identifiers with the same characteristics. If this set is missing the constraints are applied to the entire network element.

The <InputConstraints> are signal compatibility based constraints and/or shared access constraint indication. The details of these constraints are defined in section 5.3.

```
<InputConstraints> ::= <SharedIngress> <ModulationTypeList>
<FECTypeList> <BitRateRange> <ClientSignalList>
```

The <ProcessingCapabilities> are important operations that the resource (or network element) can perform on the signal. The details of these capabilities are defined in section 5.3.

```
<ProcessingCapabilities> ::= <NumResources>
<RegenerationCapabilities> <FaultPerfMon> <VendorSpecific>
```

The <OutputConstraints> are either restrictions on the properties of the signal leaving the block, options concerning the signal properties when leaving the resource or shared fiber egress constraint indication.

```
<OutputConstraints> ::= <SharedEgress> <ModulationTypeList>
<FECTypeList>
```

5.3. Compatibility and Capability Details

5.3.1. Shared Ingress or Egress Indication

As discussed in the previous section and shown in Figure 2 the ingress or egress access to a resource block may be via a shared fiber. The <SharedIngress> and <SharedEgress> elements are indicators for this condition with respect to the block being described.

5.3.2. Modulation Type List

Modulation type, also known as optical tributary signal class, comes in two distinct flavors: (i) ITU-T standardized types; (ii) vendor specific types. The permitted modulation type list can include any mixture of standardized and vendor specific types.

```
<modulation-list> ::=
  [<STANDARD_MODULATION> | <VENDOR_MODULATION>]...
```

Where the STANDARD_MODULATION object just represents one of the ITU-T standardized optical tributary signal class and the VENDOR_MODULATION object identifies one vendor specific modulation type.

5.3.3. FEC Type List

Some devices can handle more than one FEC type and hence a list is needed.

```
<fec-list> ::= [<FEC>]
```

Where the FEC object represents one of the ITU-T standardized FECs defined in [G.709], [G.707], [G.975.1] or a vendor-specific FEC.

5.3.4. Bit Rate Range List

Some devices can handle more than one particular bit rate range and hence a list is needed.

```
<rate-range-list> ::= [<rate-range>]...
```

```
<rate-range> ::= <START_RATE> <END_RATE>
```

Where the START_RATE object represents the lower end of the range and the END_RATE object represents the higher end of the range.

5.3.5. Acceptable Client Signal List

The list is simply:

```
<client-signal-list> ::= [<GPID>]...
```

Where the Generalized Protocol Identifiers (GPID) object represents one of the IETF standardized GPID values as defined in [RFC3471] and [RFC4328].

5.3.6. Processing Capability List

The ProcessingCapabilities were defined in Section 5.2 as follows:

```
<ProcessingCapabilities> ::= <NumResources>  
<RegenerationCapabilities> <FaultPerfMon> <VendorSpecific>
```

The processing capability list sub-TLV is a list of processing functions that the WSON network element (NE) can perform on the signal including:

1. Number of Resources within the block
2. Regeneration capability
3. Fault and performance monitoring
4. Vendor Specific capability

Note that the code points for Fault and performance monitoring and vendor specific capability are subject to further study.

6. Link Information (General)

MPLS-TE routing protocol extensions for OSPF and IS-IS [RFC3630], [RFC5305] along with GMPLS routing protocol extensions for OSPF and IS-IS [RFC4203, RFC5307] provide the bulk of the relatively static link information needed by the RWA process. However, WSON networks bring in additional link related constraints. These stem from WDM line system characterization, laser transmitter tuning restrictions, and switching subsystem port wavelength constraints, e.g., colored ROADMs drop ports.

In the following summarize both information from existing GMPLS route protocols and new information that maybe needed by the RWA process.

```
<LinkInfo> ::= <LinkID> [<AdministrativeGroup>] [<InterfaceCapDesc>]
 [<Protection>] [<SRLG>]... [<TrafficEngineeringMetric>]
 [<PortLabelRestriction>]
```

6.1. Administrative Group

AdministrativeGroup: Defined in [RFC3630]. Each set bit corresponds to one administrative group assigned to the interface. A link may belong to multiple groups. This is a configured quantity and can be used to influence routing decisions.

6.2. Interface Switching Capability Descriptor

InterfaceSwCapDesc: Defined in [RFC4202], lets us know the different switching capabilities on this GMPLS interface. In both [RFC4203] and [RFC5307] this information gets combined with the maximum LSP bandwidth that can be used on this link at eight different priority levels.

6.3. Link Protection Type (for this link)

Protection: Defined in [RFC4202] and implemented in [RFC4203, RFC5307]. Used to indicate what protection, if any, is guarding this link.

6.4. Shared Risk Link Group Information

SRLG: Defined in [RFC4202] and implemented in [RFC4203, RFC5307]. This allows for the grouping of links into shared risk groups, i.e., those links that are likely, for some reason, to fail at the same time.

6.5. Traffic Engineering Metric

TrafficEngineeringMetric: Defined in [RFC3630]. This allows for the definition of one additional link metric value for traffic engineering separate from the IP link state routing protocols link metric. Note that multiple "link metric values" could find use in optical networks, however it would be more useful to the RWA process to assign these specific meanings such as link mile metric, or probability of failure metric, etc...

6.6. Port Label (Wavelength) Restrictions

Port label (wavelength) restrictions (PortLabelRestriction) model the label (wavelength) restrictions that the link and various optical devices such as OXCs, ROADMs, and waveband multiplexers may impose on

a port. These restrictions tell us what wavelength may or may not be used on a link and are relatively static. This plays an important role in fully characterizing a WSON switching device [Switch]. Port wavelength restrictions are specified relative to the port in general or to a specific connectivity matrix (section 4.1. Reference [Switch] gives an example where both switch and fixed connectivity matrices are used and both types of constraints occur on the same port. Such restrictions could be applied generally to other label types in GMPLS by adding new kinds of restrictions.

```
<PortLabelRestriction> ::= [<GeneralPortRestrictions>...]  
[<MatrixSpecificRestrictions>...]  
  
<GeneralPortRestrictions> ::= <RestrictionType>  
[<RestrictionParameters>]  
  
<MatrixSpecificRestriction> ::= <MatrixID> <RestrictionType>  
[<RestrictionParameters>]  
  
<RestrictionParameters> ::= [<LabelSet>...] [<MaxNumChannels>]  
[<MaxWaveBandWidth>]
```

Where

MatrixID is the ID of the corresponding connectivity matrix (section 4.1.

The RestrictionType parameter is used to specify general port restrictions and matrix specific restrictions. It can take the following values and meanings:

SIMPLE_WAVELENGTH: Simple wavelength set restriction; The wavelength set parameter is required.

CHANNEL_COUNT: The number of channels is restricted to be less than or equal to the Max number of channels parameter (which is required).

PORT_WAVELENGTH_EXCLUSIVITY: A wavelength can be used at most once among a given set of ports. The set of ports is specified as a parameter to this constraint.

WAVEBAND1: Waveband device with a tunable center frequency and passband. This constraint is characterized by the MaxWaveBandWidth parameters which indicates the maximum width of the waveband in terms of channels. Note that an additional wavelength set can be used to

indicate the overall tuning range. Specific center frequency tuning information can be obtained from dynamic channel in use information. It is assumed that both center frequency and bandwidth (Q) tuning can be done without causing faults in existing signals.

Restriction specific parameters are used with one or more of the previously listed restriction types. The currently defined parameters are:

LabelSet is a conceptual set of labels (wavelengths).

MaxNumChannels is the maximum number of channels that can be simultaneously used (relative to either a port or a matrix).

MaxWaveBandWidth is the maximum width of a tunable waveband switching device.

PortSet is a conceptual set of ports.

For example, if the port is a "colored" drop port of a ROADM then there are two restrictions: (a) CHANNEL_COUNT, with MaxNumChannels = 1, and (b) SIMPLE_WAVELENGTH, with the wavelength set consisting of a single member corresponding to the frequency of the permitted wavelength. See [Switch] for a complete waveband example.

This information model for port wavelength (label) restrictions is fairly general in that it can be applied to ports that have label restrictions only or to ports that are part of an asymmetric switch and have label restrictions. In addition, the types of label restrictions that can be supported are extensible.

6.6.1. Port-Wavelength Exclusivity Example

Although there can be many different ROADM or switch architectures that can lead to the constraint where a lambda (label) maybe used at most once on a set of ports Figure 3 shows a ROADM architecture based on components known as a Wavelength Selective Switch (WSS)[OFC08]. This ROADM is composed of splitters, combiners, and WSSes. This ROADM has 11 egress ports, which are numbered in the diagram. Egress ports 1-8 are known as drop ports and are intended to support a single wavelength. Drop ports 1-4 egress from WSS #2, which is fed from WSS #1 via a single fiber. Due to this internal structure a constraint is placed on the egress ports 1-4 that a lambda can be only used once over the group of ports (assuming uni-cast and not multi-cast operation). Similarly the egress ports 5-8 have a similar constraint due to the internal structure.

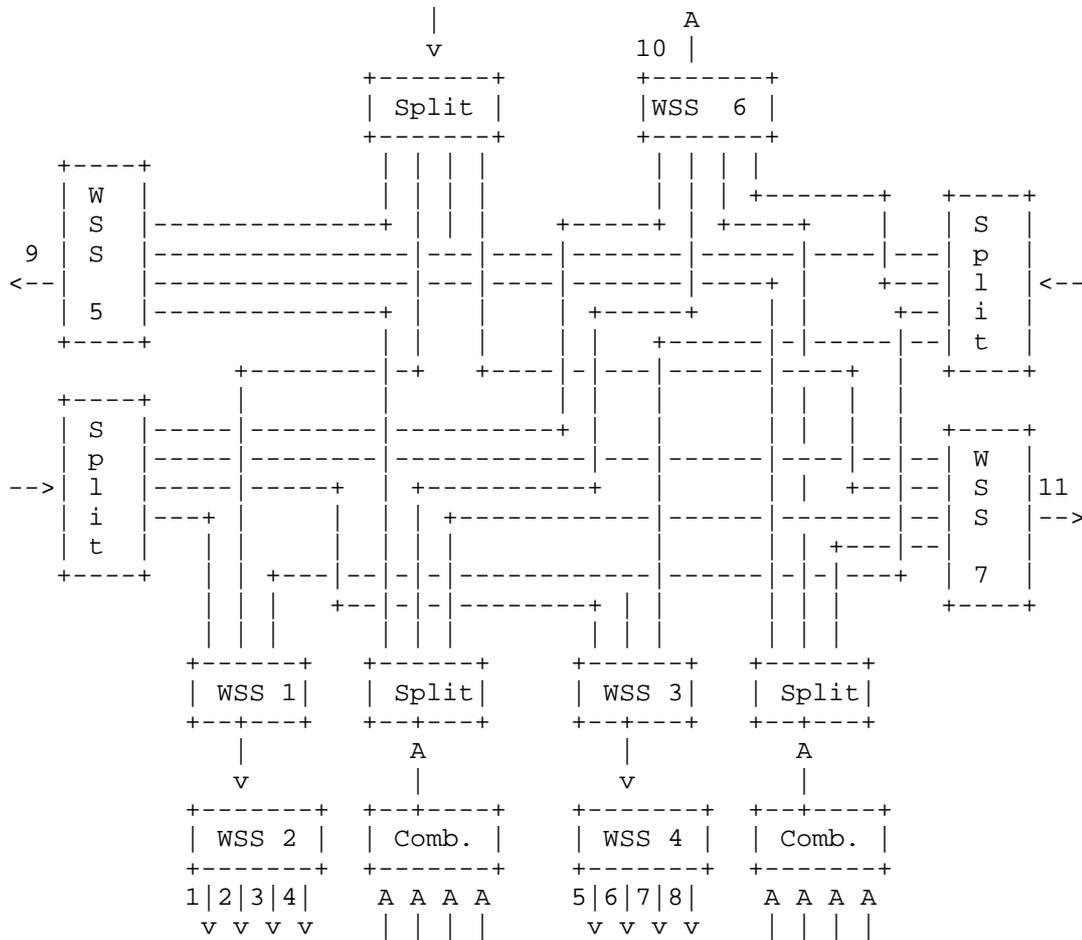


Figure 3 A ROADMs composed from splitter, combiners, and WSSs.

7. Dynamic Components of the Information Model

In the previously presented information model there are a limited number of information elements that are dynamic, i.e., subject to change with subsequent establishment and teardown of connections. Depending on the protocol used to convey this overall information model it may be possible to send this dynamic information separate from the relatively larger amount of static information needed to characterize WSON's and their network elements.

7.1. Dynamic Link Information (General)

For WSON links wavelength availability and wavelengths in use for shared backup purposes can be considered dynamic information and hence are grouped with the dynamic information in the following set:

```
<DynamicLinkInfo> ::= <LinkID> <AvailableLabels>  
[<SharedBackupLabels>]
```

AvailableLabels is a set of labels (wavelengths) currently available on the link. Given this information and the port wavelength restrictions one can also determine which wavelengths are currently in use. This parameter could potential be used with other technologies that GMPLS currently covers or may cover in the future.

SharedBackupLabels is a set of labels (wavelengths) currently used for shared backup protection on the link. An example usage of this information in a WSON setting is given in [Shared]. This parameter could potential be used with other technologies that GMPLS currently covers or may cover in the future.

7.2. Dynamic Node Information (WSON Specific)

Currently the only node information that can be considered dynamic is the resource pool state and can be isolated into a dynamic node information element as follows:

```
<DynamicNodeInfo> ::= <NodeID> [<ResourcePoolState>]
```

8. Security Considerations

This document discussed an information model for RWA computation in WSONs. Such a model is very similar from a security standpoint of the information that can be currently conveyed via GMPLS routing protocols. Such information includes network topology, link state and current utilization, and well as the capabilities of switches and routers within the network. As such this information should be protected from disclosure to unintended recipients. In addition, the intentional modification of this information can significantly affect network operations, particularly due to the large capacity of the optical infrastructure to be controlled.

9. IANA Considerations

This informational document does not make any requests for IANA action.

10. Acknowledgments

This document was prepared using 2-Word-v2.0.template.dot.

11. References

11.1. Normative References

- [Encode] G. Bernstein, Y. Lee, D. Li, W. Imajuku, "Routing and Wavelength Assignment Information Encoding for Wavelength Switched Optical Networks", work in progress: draft-ietf-ccamp-rwa-wson-encode.
- [G.707] ITU-T Recommendation G.707, Network node interface for the synchronous digital hierarchy (SDH), January 2007.
- [G.709] ITU-T Recommendation G.709, Interfaces for the Optical Transport Network(OTN), March 2003.
- [G.975.1] ITU-T Recommendation G.975.1, Forward error correction for high bit-rate DWDM submarine systems, February 2004.
- [RBNF] A. Farrel, "Reduced Backus-Naur Form (RBNF) A Syntax Used in Various Protocol Specifications", RFC 5511, April 2009.
- [RFC3471] Berger, L., Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Functional Description", RFC 3471, January 2003.
- [RFC3630] Katz, D., Kompella, K., and D. Yeung, "Traffic Engineering (TE) Extensions to OSPF Version 2", RFC 3630, September 2003.
- [RFC4202] Kompella, K., Ed., and Y. Rekhter, Ed., "Routing Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4202, October 2005
- [RFC4203] Kompella, K., Ed., and Y. Rekhter, Ed., "OSPF Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4203, October 2005.
- [RFC4328] Papadimitriou, D., Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Extensions for G.709 Optical Transport Networks Control", RFC 4328, January 2006.
- [RFC5305] Li, T. and H. Smit, "IS-IS Extensions for Traffic Engineering", RFC 5305, October 2008.

[RFC5307] Kompella, K., Ed., and Y. Rekhter, Ed., "IS-IS Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 5307, October 2008.

11.2. Informative References

[OFC08] P. Roorda and B. Collings, "Evolution to Colorless and Directionless ROADM Architectures," Optical Fiber communication/National Fiber Optic Engineers Conference, 2008. OFC/NFOEC 2008. Conference on, 2008, pp. 1-3.

[Shared] G. Bernstein, Y. Lee, "Shared Backup Mesh Protection in PCE-based WSON Networks", iPOP 2008, http://www.grotto-networking.com/wson/iPOP2008_WSON-shared-mesh-poster.pdf.

[Switch] G. Bernstein, Y. Lee, A. Gavler, J. Martensson, " Modeling WDM Wavelength Switching Systems for Use in GMPLS and Automated Path Computation", Journal of Optical Communications and Networking, vol. 1, June, 2009, pp. 187-195.

[G.Sup39] ITU-T Series G Supplement 39, Optical system design and engineering considerations, February 2006.

[WSON-Frame] Y. Lee, G. Bernstein, W. Imajuku, "Framework for GMPLS and PCE Control of Wavelength Switched Optical Networks", work in progress: draft-ietf-ccamp-rwa-wson-framework.

12. Contributors

Diego Caviglia
Ericsson
Via A. Negrone 1/A 16153
Genoa Italy

Phone: +39 010 600 3736
Email: diego.caviglia@(marconi.com, ericsson.com)

Anders Gavler
Acreo AB
Electrum 236
SE - 164 40 Kista Sweden

Email: Anders.Gavler@acreo.se

Jonas Martensson
Acreo AB
Electrum 236
SE - 164 40 Kista, Sweden

Email: Jonas.Martensson@acreo.se

Itaru Nishioka
NEC Corp.
1753 Simonumabe, Nakahara-ku, Kawasaki, Kanagawa 211-8666
Japan

Phone: +81 44 396 3287
Email: i-nishioka@cb.jp.nec.com

Lyndon Ong
Ciena
Email: lyong@ciena.com

Author's Addresses

Greg M. Bernstein (ed.)
Grotto Networking
Fremont California, USA

Phone: (510) 573-2237
Email: gregb@grotto-networking.com

Young Lee (ed.)
Huawei Technologies
1700 Alma Drive, Suite 100
Plano, TX 75075
USA

Phone: (972) 509-5599 (x2240)
Email: ylee@huawei.com

Dan Li
Huawei Technologies Co., Ltd.
F3-5-B R&D Center, Huawei Base,
Bantian, Longgang District
Shenzhen 518129 P.R.China

Phone: +86-755-28973237
Email: danli@huawei.com

Wataru Imajuku
NTT Network Innovation Labs
1-1 Hikari-no-oka, Yokosuka, Kanagawa
Japan

Phone: +81-(46) 859-4315
Email: imajuku.wataru@lab.ntt.co.jp

Intellectual Property Statement

The IETF Trust takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in any IETF Document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights.

Copies of Intellectual Property disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a 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 any standard or specification contained in an IETF Document. Please address the information to the IETF at ietf-ipr@ietf.org.

Disclaimer of Validity

All IETF Documents and the information contained therein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY, THE IETF TRUST AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION THEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Acknowledgment

Funding for the RFC Editor function is currently provided by the Internet Society.

Network Working Group
Internet Draft
Intended status: Standards Track
Expires: September 2011

G. Bernstein
Grotto Networking
Y. Lee
D. Li
Huawei
W. Imajuku
NTT

March 14, 2011

Routing and Wavelength Assignment Information Encoding for
Wavelength Switched Optical Networks

draft-ietf-ccamp-rwa-wson-encode-11.txt

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at <http://www.ietf.org/ietf/lid-abstracts.txt>

The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>

This Internet-Draft will expire on September 14, 2011.

Copyright Notice

Copyright (c) 2011 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Abstract

A wavelength switched optical network (WSON) requires that certain key information elements are made available to facilitate path computation and the establishment of label switching paths (LSPs). The information model described in "Routing and Wavelength Assignment Information for Wavelength Switched Optical Networks" shows what information is required at specific points in the WSON. Part of the WSON information model contains aspects that may be of general applicability to other technologies, while other parts are fairly specific to WSONs.

This document provides efficient, protocol-agnostic encodings for the WSON specific information elements. It is intended that protocol-specific documents will reference this memo to describe how information is carried for specific uses. Such encodings can be used to extend GMPLS signaling and routing protocols. In addition these encodings could be used by other mechanisms to convey this same information to a path computation element (PCE).

Conventions used in this document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC-2119 [RFC2119].

Table of Contents

1. Introduction.....	4
1.1. Revision History.....	4
1.1.1. Changes from 00 draft.....	4
1.1.2. Changes from 01 draft.....	5
1.1.3. Changes from 02 draft.....	5

1.1.4. Changes from 03 draft.....	5
1.1.5. Changes from 04 draft.....	5
1.1.6. Changes from 05 draft.....	5
1.1.7. Changes from 06 draft.....	5
1.1.8. Changes from 07 draft.....	5
1.1.9. Changes from 08 draft.....	6
1.1.10. Changes from 09 draft.....	6
1.1.11. Changes from 10 draft.....	6
2. Terminology.....	6
3. Resource Pool Accessibility/Availability.....	7
3.1. Resource Pool Accessibility Sub-TLV.....	8
3.2. Resource Block Wavelength Constraints Sub-TLV.....	10
3.3. Resource Pool State Sub-TLV.....	10
3.4. Block Shared Access Wavelength Availability sub-TLV.....	12
4. Resource Properties Encoding.....	13
4.1. Resource Block Information Sub-TLV.....	13
4.2. Input Modulation Format List Sub-Sub-TLV.....	14
4.2.1. Modulation Format Field.....	15
4.3. Input FEC Type List Sub-Sub-TLV.....	17
4.3.1. FEC Type Field.....	17
4.4. Input Bit Range List Sub-Sub-TLV.....	19
4.4.1. Bit Range Field.....	19
4.5. Input Client Signal List Sub-Sub-TLV.....	20
4.6. Processing Capability List Sub-Sub-TLV.....	21
4.6.1. Processing Capabilities Field.....	21
4.7. Output Modulation Format List Sub-Sub-TLV.....	23
4.8. Output FEC Type List Sub-Sub-TLV.....	23
5. Security Considerations.....	23
6. IANA Considerations.....	23
7. Acknowledgments.....	23
APPENDIX A: Encoding Examples.....	23
A.1. Wavelength Converter Accessibility Sub-TLV.....	23
A.2. Wavelength Conversion Range Sub-TLV.....	23
A.3. An OEO Switch with DWDM Optics.....	23
8. References.....	23
8.1. Normative References.....	23
8.2. Informative References.....	23
9. Contributors.....	23
Authors' Addresses.....	23
Intellectual Property Statement.....	23
Disclaimer of Validity.....	23

1. Introduction

A Wavelength Switched Optical Network (WSON) is a Wavelength Division Multiplexing (WDM) optical network in which switching is performed selectively based on the center wavelength of an optical signal.

[WSON-Frame] describes a framework for Generalized Multiprotocol Label Switching (GMPLS) and Path Computation Element (PCE) control of a WSON. Based on this framework, [WSON-Info] describes an information model that specifies what information is needed at various points in a WSON in order to compute paths and establish Label Switched Paths (LSPs).

This document provides efficient encodings of information needed by the routing and wavelength assignment (RWA) process in a WSON. Such encodings can be used to extend GMPLS signaling and routing protocols. In addition these encodings could be used by other mechanisms to convey this same information to a path computation element (PCE). Note that since these encodings are relatively efficient they can provide more accurate analysis of the control plane communications/processing load for WSONs looking to utilize a GMPLS control plane.

Note that encodings of information needed by the routing and label assignment process applicable to general networks beyond WSON are addressed in a separate document [Gen-Encode].

1.1. Revision History

1.1.1. Changes from 00 draft

Edits to make consistent with update to [Otani], i.e., removal of sign bit.

Clarification of TBD on connection matrix type and possibly numbering.

New sections for wavelength converter pool encoding: Wavelength Converter Set Sub-TLV, Wavelength Converter Accessibility Sub-TLV, Wavelength Conversion Range Sub-TLV, WC Usage State Sub-TLV.

Added optional wavelength converter pool TLVs to the composite node TLV.

1.1.2. Changes from 01 draft

The encoding examples have been moved to an appendix. Classified and corrected information elements as either reusable fields or sub-TLVs. Updated Port Wavelength Restriction sub-TLV. Added available wavelength and shared backup wavelength sub-TLVs. Changed the title and scope of section 6 to recommendations since the higher level TLVs that this encoding will be used in is somewhat protocol specific.

1.1.3. Changes from 02 draft

Removed inconsistent text concerning link local identifiers and the link set field.

Added E bit to the Wavelength Converter Set Field.

Added bidirectional connectivity matrix example. Added simple link set example. Edited examples for consistency.

1.1.4. Changes from 03 draft

Removed encodings for general concepts to [Gen-Encode].

Added in WSON signal compatibility and processing capability information encoding.

1.1.5. Changes from 04 draft

Added encodings to deal with access to resource blocks via shared fiber.

1.1.6. Changes from 05 draft

Revised the encoding for the "shared access" indicators to only use one bit each for ingress and egress.

1.1.7. Changes from 06 draft

Removed section on "WSON Encoding Usage Recommendations"

1.1.8. Changes from 07 draft

Section 3: Enhanced text to clarify relationship between pools, blocks and resources. Section 3.1, 3.2: Change title to clarify Pool-Block relationship. Section 3.3: clarify block-resource state.

Section 4: Deleted reference to previously removed RBNF element. Fixed TLV figures and descriptions for consistent sub-sub-TLV nomenclature.

1.1.9. Changes from 08 draft

Fixed ordering of fields in second half of sub-TLV example in Appendix A.1.

Clarifying edits in section 3 on pools, blocks, and resources.

1.1.10. Changes from 09 draft

Fixed the "Block Shared Access Wavelength Availability sub-TLV" of section 3.4 to use an "RB set field" rather than a single RB ID. Removed all 1st person idioms.

1.1.11. Changes from 10 draft

Removed remaining 1st person idioms. Updated IANA section.

2. Terminology

CWDM: Coarse Wavelength Division Multiplexing.

DWDM: Dense Wavelength Division Multiplexing.

FOADM: Fixed Optical Add/Drop Multiplexer.

ROADM: Reconfigurable Optical Add/Drop Multiplexer. A reduced port count wavelength selective switching element featuring ingress and egress line side ports as well as add/drop side ports.

RWA: Routing and Wavelength Assignment.

Wavelength Conversion. The process of converting an information bearing optical signal centered at a given wavelength to one with "equivalent" content centered at a different wavelength. Wavelength conversion can be implemented via an optical-electronic-optical (OEO) process or via a strictly optical process.

WDM: Wavelength Division Multiplexing.

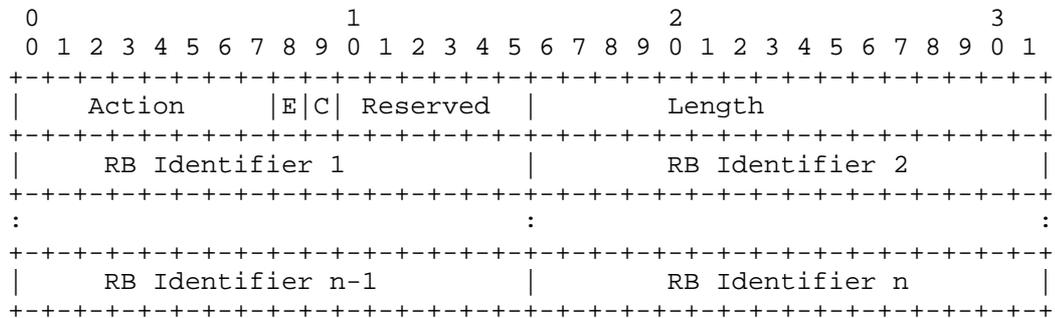
Wavelength Switched Optical Network (WSO): A WDM based optical network in which switching is performed selectively based on the center wavelength of an optical signal.

3. Resource Pool Accessibility/Availability

This section defines the sub-TLVs for dealing with accessibility and availability of resource blocks within a pool of resources. These include the ResourceBlockAccessibility, ResourceWaveConstraints, and RBPoolState sub-TLVs. All these sub-TLVs are concerned with sets of resources. As described in [WSO-Info] a resource pool is composed of blocks of resources with similar properties and accessibility characteristics.

In a WSON node that includes resource blocks (RB) denoting subsets of these blocks allows one to efficiently describe common properties the blocks and to describe the structure, if non-trivial, of the resource pool. The RB Set field is defined in a similar manner to the label set concept of [RFC3471].

The information carried in a RB set field is defined by:



Action: 8 bits

0 - Inclusive List

Indicates that the TLV contains one or more RB elements that are included in the list.

2 - Inclusive Range

Indicates that the TLV contains a range of RBs. The object/TLV contains two WC elements. The first element indicates the start of the range. The second element indicates the end of the range. A value of zero indicates that there is no bound on the corresponding portion of the range.

E (Even bit): Set to 0 denotes an odd number of RB identifiers in the list (last entry zero pad); Set to 1 denotes an even number of RB identifiers in the list (no zero padding).

C (Connectivity bit): Set to 0 to denote fixed (possibly multi-cast) connectivity; Set to 1 to denote potential (switched) connectivity. Used in resource pool accessibility sub-TLV. Ignored elsewhere.

Reserved: 6 bits

This field is reserved. It MUST be set to zero on transmission and MUST be ignored on receipt.

Length: 16 bits

The total length of this field in bytes.

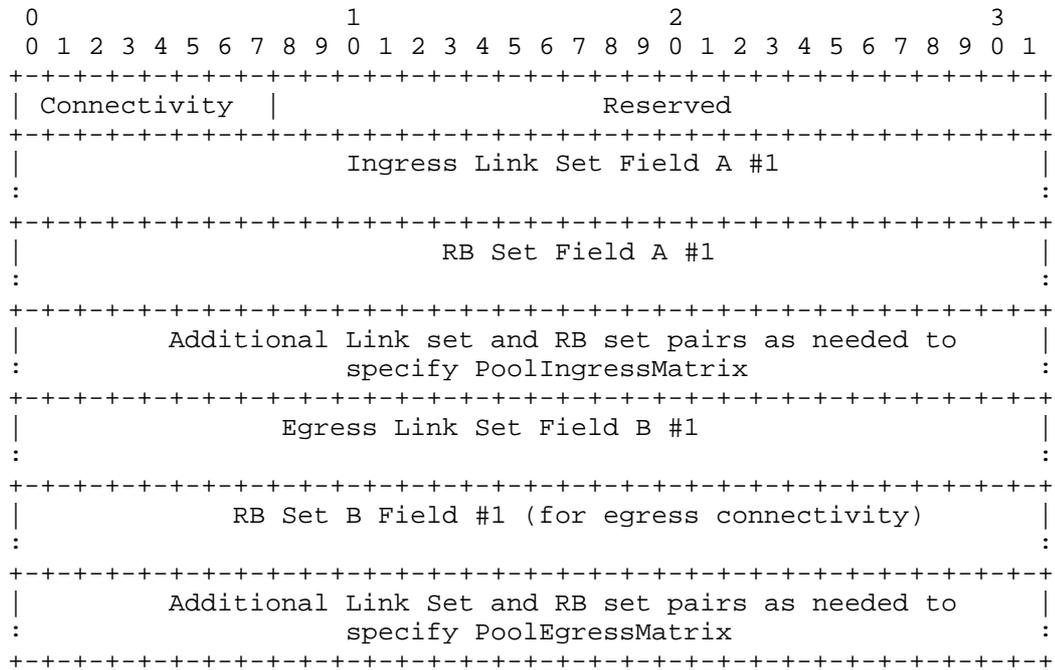
RB Identifier:

The RB identifier represents the ID of the resource block which is a 16 bit integer.

3.1. Resource Pool Accessibility Sub-TLV

This sub-TLV describes the structure of the resource pool in relation to the switching device. In particular it indicates the ability of an ingress port to reach a resource block and of a resource block to reach a particular egress port. This is the PoolIngressMatrix and PoolEgressMatrix of [WSON-Info].

The resource pool accessibility sub-TLV is defined by:



Where

Connectivity indicates how the ingress/egress ports connect to the resource blocks.

0 -- the device is fixed (e.g., a connected port must go through the resource block)

1 -- the device is switched (e.g., a port can be configured to go through a resource but isn't required)

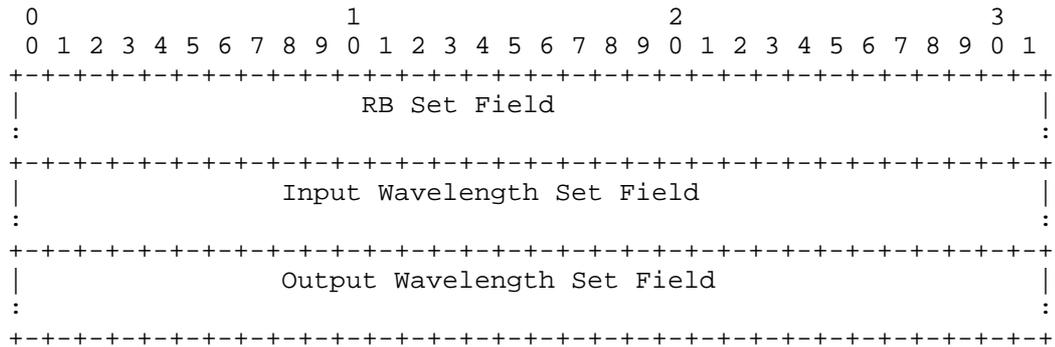
The Link Set Field is defined in [Gen-Encode].

Note that the direction parameter within the Link Set Field is used to indicate whether the link set is an ingress or egress link set, and the bidirectional value for this parameter is not permitted in this sub-TLV.

See Appendix A.1 for an illustration of this encoding.

3.2. Resource Block Wavelength Constraints Sub-TLV

Resources, such as wavelength converters, etc., may have a limited input or output wavelength ranges. Additionally, due to the structure of the optical system not all wavelengths can necessarily reach or leave all the resources. These properties are described by using one or more resource wavelength restrictions sub-TLVs as defined below:



RB Set Field:

A set of resource blocks (RBs) which have the same wavelength restrictions.

Input Wavelength Set Field:

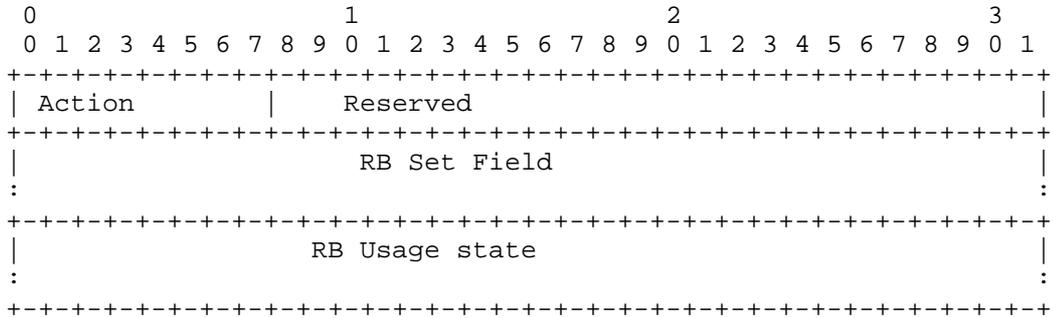
Indicates the wavelength input restrictions of the RBs in the corresponding RB set.

Output Wavelength Set Field:

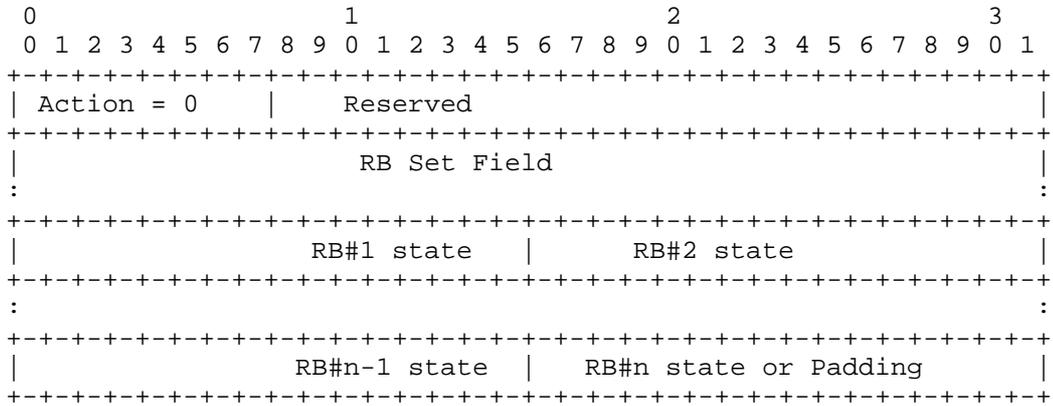
Indicates the wavelength output restrictions of RBs in the corresponding RB set.

3.3. Resource Pool State Sub-TLV

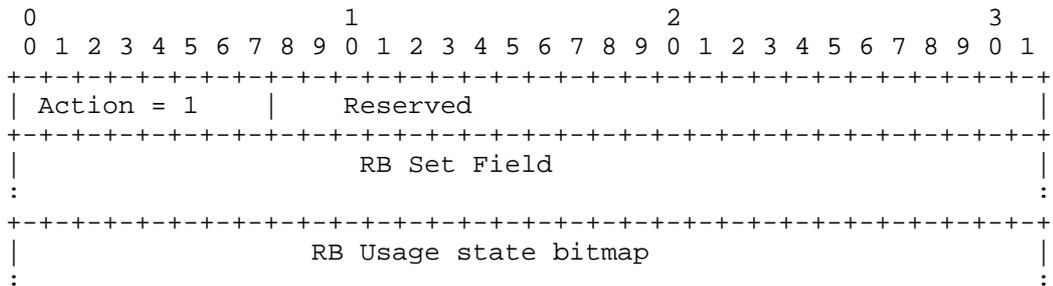
The state of the pool is given by the number of resources available in each block. The usage state of resources within a block is encoded as either a list of 16 bit integer values or a bit map indicating whether a single resource is available or in use. The bit map encoding is appropriate when resource blocks consist of a single resource. This information can be relatively dynamic, i.e., can change when a connection is established or torn down.

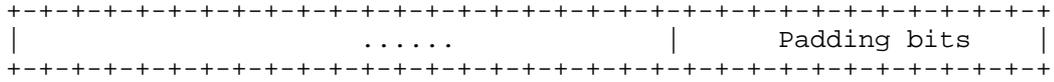


Where Action = 0 denotes a list of 16 bit integers and Action = 1 denotes a bit map. In both cases the elements of the RB Set field are in a one-to-one correspondence with the values in the usage RB usage state area.



Whether the last 16 bits is a wavelength converter (RB) state or padding is determined by the number of elements in the RB set field.





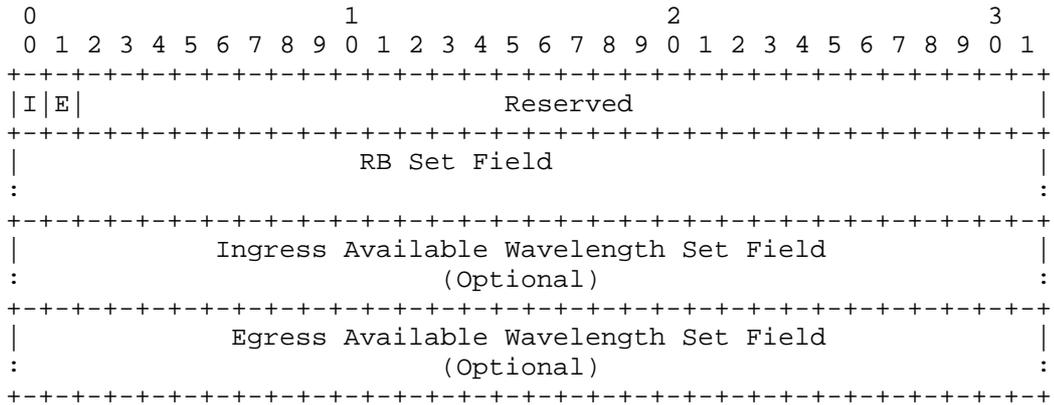
RB Usage state: Variable Length but must be a multiple of 4 bytes.

Each bit indicates the usage status of one RB with 0 indicating the RB is available and 1 indicating the RB is in used. The sequence of the bit map is ordered according to the RB Set field with this sub-TLV.

Padding bits: Variable Length

3.4. Block Shared Access Wavelength Availability sub-TLV

Resources blocks may be accessed via a shared fiber. If this is the case then wavelength availability on these shared fibers is needed to understand resource availability.



I bit:

Indicates whether the ingress available wavelength set field is included (1) or not (0).

E bit:

Indicates whether the egress available wavelength set field is included (1) or not (0).

RB Set Field:

A Resource Block set in which all the members share the same ingress or egress fiber or both.

Ingress Available Wavelength Set Field:

Indicates the wavelengths currently available (not being used) on the ingress fiber to this resource block.

Egress Available Wavelength Set Field:

Indicates the wavelengths currently available (not being used) on the egress fiber from this resource block.

4. Resource Properties Encoding

Within a WSON network element (NE) there may be resources with signal compatibility constraints. Such resources typically come in "blocks" which contain a group of identical and indistinguishable individual resources. These resource blocks may consist of regenerators, wavelength converters, etc... Such resource blocks may also constitute the network element as a whole as in the case of an electro optical switch. This section primarily focuses on the signal compatibility and processing properties of such a resource block, the accessibility aspects of a resource in a shared pool, except for the shared access indicators, were encoded in the previous section.

The fundamental properties of a resource block, such as a regenerator or wavelength converter, are:

- (a) Input constraints (shared ingress, modulation, FEC, bit rate, GPID)
- (b) Processing capabilities (number of resources in a block, regeneration, performance monitoring, vendor specific)
- (c) Output Constraints (shared egress, modulation, FEC)

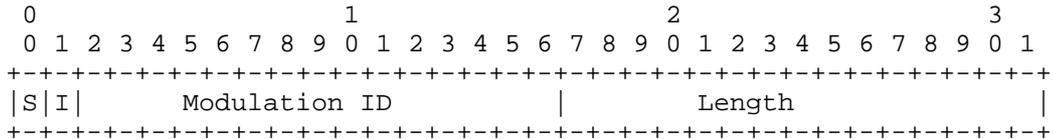
4.1. Resource Block Information Sub-TLV

Resource Block descriptor sub-TLVs are used to convey relatively static information about individual resource blocks including the

Value := A list of Modulation Format Fields

4.2.1. Modulation Format Field

Two different types of modulation format fields are defined: a standard modulation field and a vendor specific modulation field. Both start with the same 32 bit header shown below.

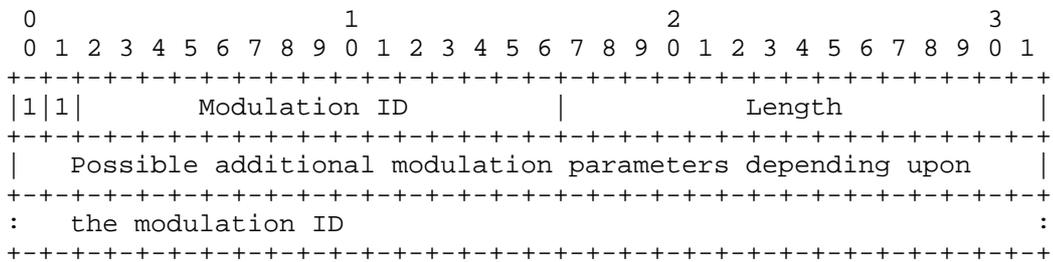


Where S bit set to 1 indicates a standardized modulation format and S bit set to 0 indicates a vendor specific modulation format. The length is the length in bytes of the entire modulation type field.

Where I bit set to 1 indicates it is an input modulation constraint and I bit set to 0 indicates it is an output modulation constraint.

Note that if an output modulation is not specified then it is implied that it is the same as the input modulation. In such case, no modulation conversion is performed.

The format for the standardized type for the input modulation is given by:



Modulation ID (S bit = 1); Input modulation (I bit = 1)

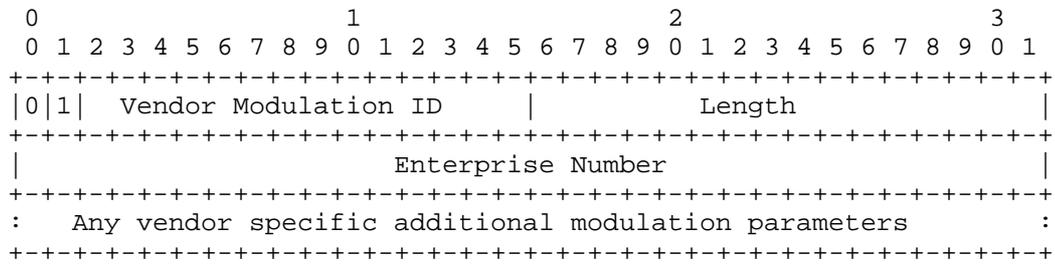
Takes on the following currently defined values:

- 0 Reserved

- 1 optical tributary signal class NRZ 1.25G
- 2 optical tributary signal class NRZ 2.5G
- 3 optical tributary signal class NRZ 10G
- 4 optical tributary signal class NRZ 40G
- 5 optical tributary signal class RZ 40G

Note that future modulation types may require additional parameters in their characterization.

The format for vendor specific modulation field (for input constraint) is given by:



Vendor Modulation ID

This is a vendor assigned identifier for the modulation type.

Enterprise Number

A unique identifier of an organization encoded as a 32-bit integer. Enterprise Numbers are assigned by IANA and managed through an IANA registry [RFC2578].

Vendor Specific Additional parameters

There can be potentially additional parameters characterizing the vendor specific modulation.

4.3. Input FEC Type List Sub-Sub-TLV

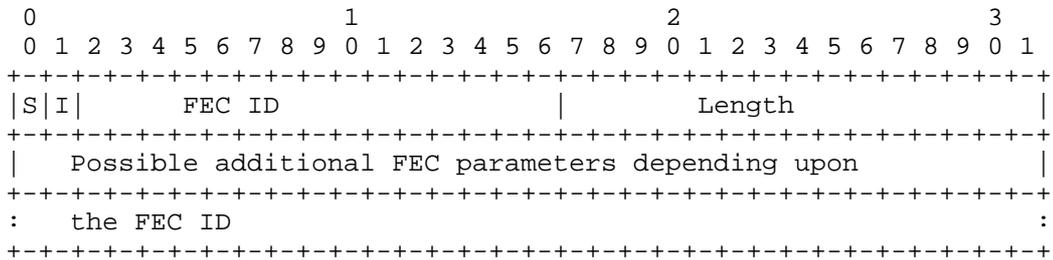
This sub-sub-TLV contains a list of acceptable FEC types.

Type := Input FEC Type field List

Value := A list of FEC type Fields

4.3.1. FEC Type Field

The FEC type Field may consist of two different formats of fields: a standard FEC field or a vendor specific FEC field. Both start with the same 32 bit header shown below.



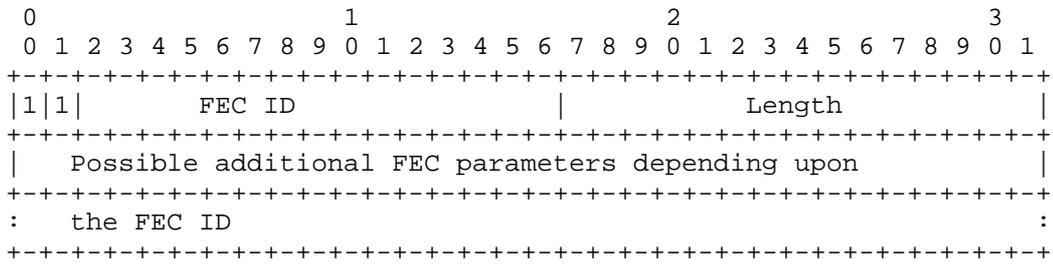
Where S bit set to 1 indicates a standardized FEC format and S bit set to 0 indicates a vendor specific FEC format. The length is the length in bytes of the entire FEC type field.

Where I bit set to 1 indicates it is an input FEC constraint and I bit set to 0 indicates it is an output FEC constraint.

Note that if an output FEC is not specified then it is implied that it is the same as the input FEC. In such case, no FEC conversion is performed.

The length is the length in bytes of the entire FEC type field.

The format for input standard FEC field is given by:

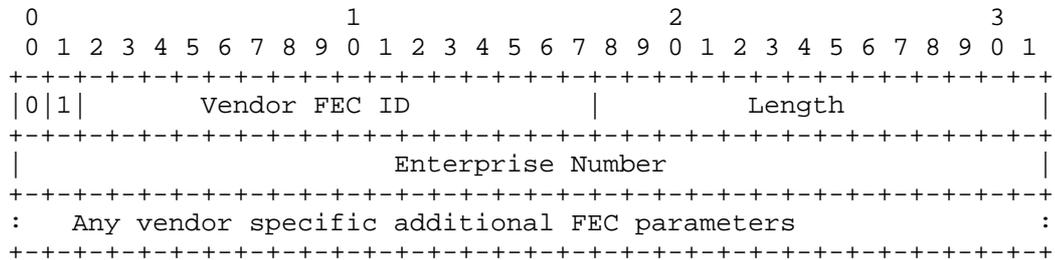


Takes on the following currently defined values for the standard FEC ID:

- 0 Reserved
- 1 G.709 RS FEC
- 2 G.709V compliant Ultra FEC
- 3 G.975.1 Concatenated FEC
(RS(255,239)/CSOC(n0/k0=7/6,J=8))
- 4 G.975.1 Concatenated FEC (BCH(3860,3824)/BCH(2040,1930))
- 5 G.975.1 Concatenated FEC (RS(1023,1007)/BCH(2407,1952))
- 6 G.975.1 Concatenated FEC (RS(1901,1855)/Extended Hamming
Product Code (512,502)X(510,500))
- 7 G.975.1 LDPC Code
- 8 G.975.1 Concatenated FEC (Two orthogonally concatenated
BCH codes)
- 9 G.975.1 RS(2720,2550)
- 10 G.975.1 Concatenated FEC (Two interleaved extended BCH
(1020,988) codes)

Where RS stands for Reed-Solomon and BCH for Bose-Chaudhuri-Hocquengham.

The format for input vendor-specific FEC field is given by:



Vendor FEC ID

This is a vendor assigned identifier for the FEC type.

Enterprise Number

A unique identifier of an organization encoded as a 32-bit integer. Enterprise Numbers are assigned by IANA and managed through an IANA registry [RFC2578].

Vendor Specific Additional FEC parameters

There can be potentially additional parameters characterizing the vendor specific FEC.

4.4. Input Bit Range List Sub-Sub-TLV

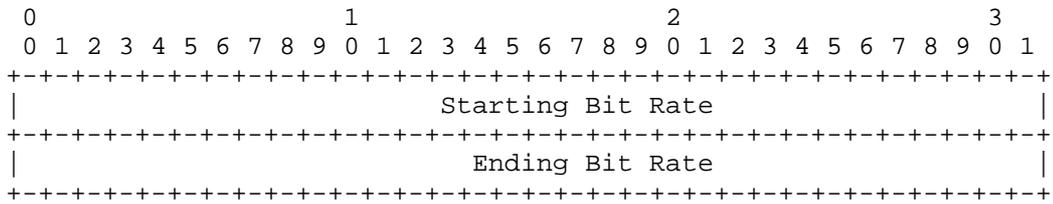
This sub-sub-TLV contains a list of acceptable input bit rate ranges.

Type := Input Bit Range List

Value := A list of Bit Range Fields

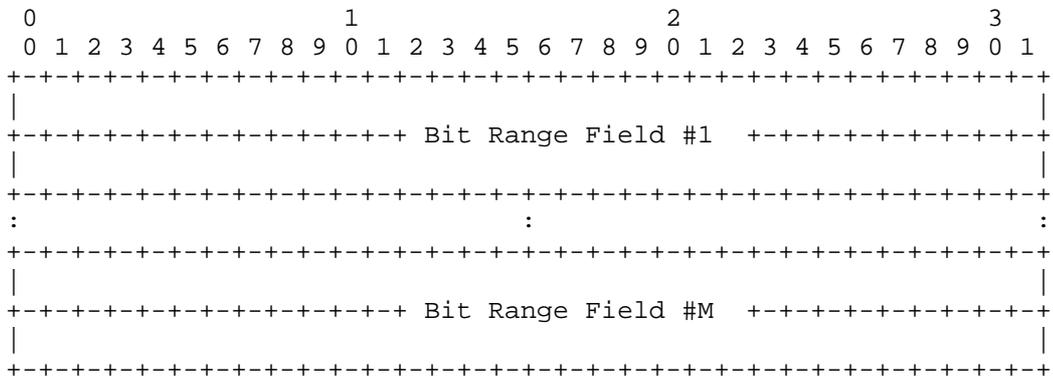
4.4.1. Bit Range Field

The bit rate range list sub-TLV makes use of the following bit rate range field:



The starting and ending bit rates are given as 32 bit IEEE floating point numbers in bits per second. Note that the starting bit rate is less than or equal to the ending bit rate.

The bit rate range list sub-TLV is then given by:



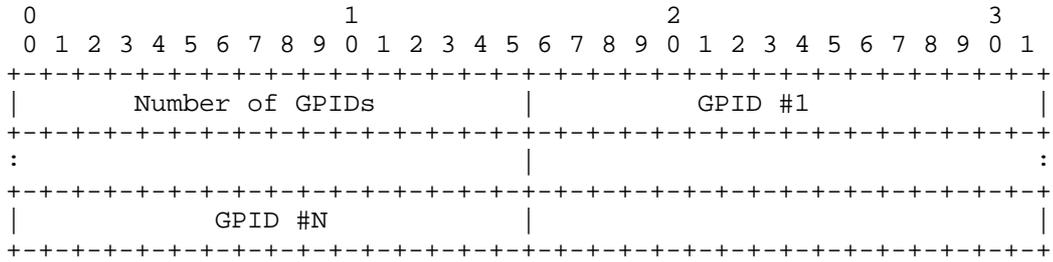
4.5. Input Client Signal List Sub-Sub-TLV

This sub-sub-TLV contains a list of acceptable input client signal types.

Type := Input Client Signal List

Value := A list of GPIDs

The acceptable client signal list sub-TLV is a list of Generalized Protocol Identifiers (GPIDs). GPIDs are assigned by IANA and many are defined in [RFC3471] and [RFC4328].



Where the number of GPIDs is an integer greater than or equal to one.

4.6. Processing Capability List Sub-Sub-TLV

This sub-sub-TLV contains a list of resource block processing capabilities.

Type := Processing Capabilities List

Value := A list of Processing Capabilities Fields

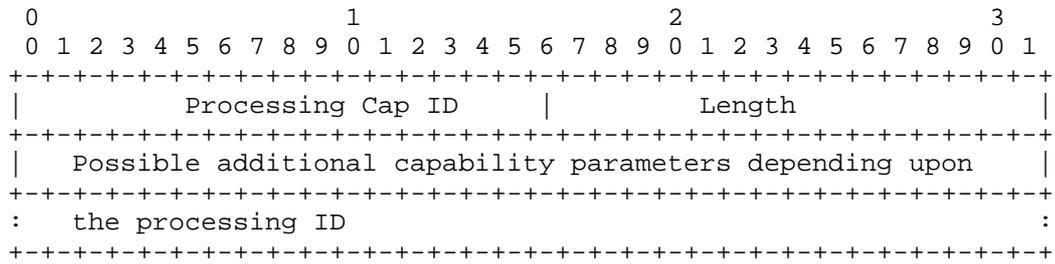
The processing capability list sub-TLV is a list of WSON network element (NE) that can perform signal processing functions including:

1. Number of Resources within the block
2. Regeneration capability
3. Fault and performance monitoring
4. Vendor Specific capability

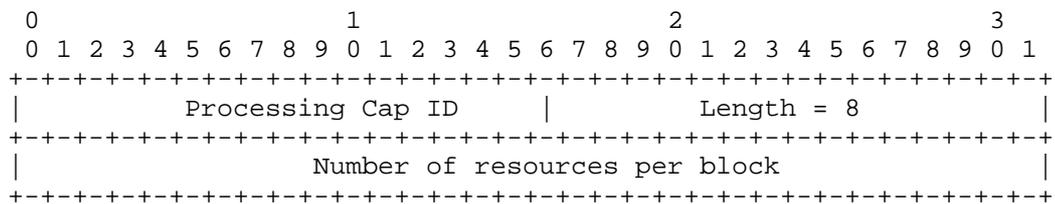
Note that the code points for Fault and performance monitoring and vendor specific capability are subject to further study.

4.6.1. Processing Capabilities Field

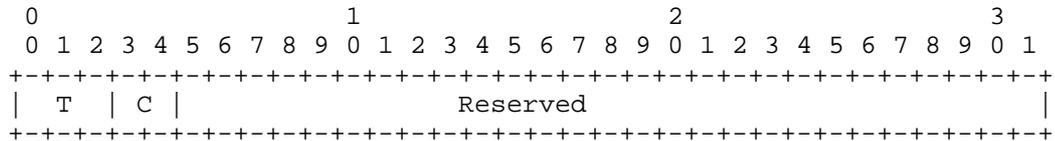
The processing capability field is then given by:



When the processing Cap ID is "number of resources" the format is simply:



When the processing Cap ID is "regeneration capability", the following additional capability parameters are provided in the sub-TLV:



Where T bit indicates the type of regenerator:

- T=0: Reserved
- T=1: 1R Regenerator
- T=2: 2R Regenerator
- T=3: 3R Regenerator

Where C bit indicates the capability of regenerator:

C=0: Reserved

C=1: Fixed Regeneration Point

C=2: Selective Regeneration Point

Note that when the capability of regenerator is indicated to be Selective Regeneration Pools, regeneration pool properties such as ingress and egress restrictions and availability need to be specified. This encoding is to be determined in the later revision.

4.7. Output Modulation Format List Sub-Sub-TLV

This sub-sub-TLV contains a list of available output modulation formats.

Type := Output Modulation Format List

Value := A list of Modulation Format Fields

4.8. Output FEC Type List Sub-Sub-TLV

This sub-sub-TLV contains a list of output FEC types.

Type := Output FEC Type field List

Value := A list of FEC type Fields

5. Security Considerations

This document defines protocol-independent encodings for WSON information and does not introduce any security issues.

However, other documents that make use of these encodings within protocol extensions need to consider the issues and risks associated with, inspection, interception, modification, or spoofing of any of this information. It is expected that any such documents will describe the necessary security measures to provide adequate protection.

6. IANA Considerations

This document provides general protocol independent information encodings. There is no IANA allocation request for the TLVs defined in this document. IANA allocation requests will be addressed in protocol specific documents based on the encodings defined here.

7. Acknowledgments

This document was prepared using 2-Word-v2.0.template.dot.

APPENDIX A: Encoding Examples

A.1. Wavelength Converter Accessibility Sub-TLV

Example:

Figure 1 shows a wavelength converter pool architecture know as "shared per fiber". In this case the ingress and egress pool matrices are simply:

$$WI = \begin{array}{|c|c|} \hline 1 & 1 \\ \hline 1 & 1 \\ \hline \end{array}, \quad WE = \begin{array}{|c|c|} \hline 1 & 0 \\ \hline 0 & 1 \\ \hline \end{array}$$

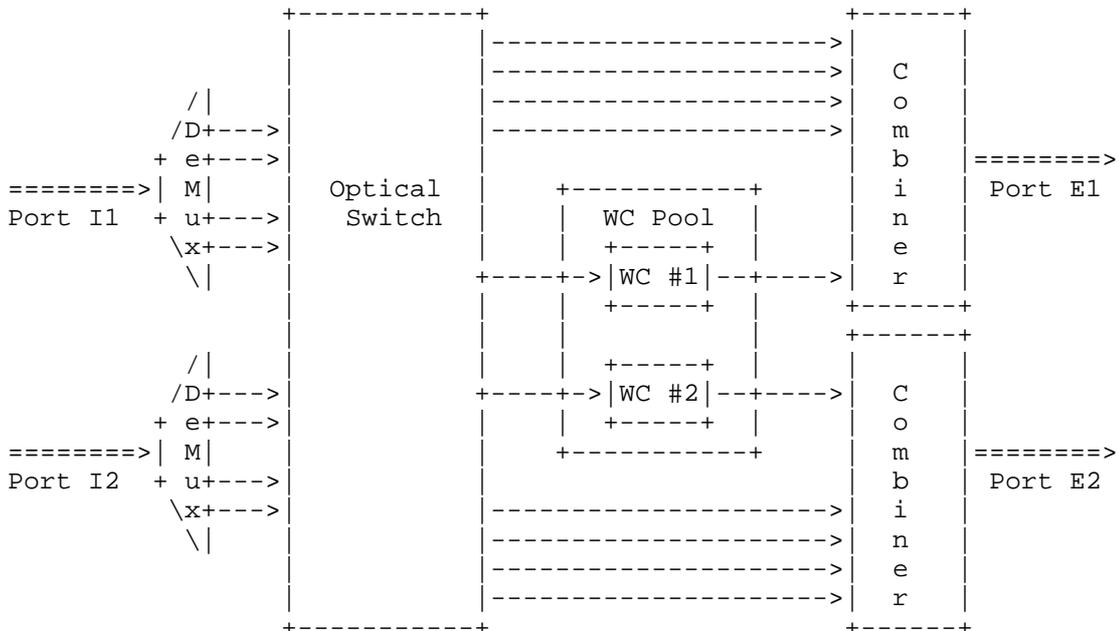


Figure 1 An optical switch featuring a shared per fiber wavelength converter pool architecture.

This wavelength converter pool can be encoded as follows:

```

0          1          2          3
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| Connectivity=1|                               Reserved |
|           Note: I1,I2 can connect to either WC1 or WC2
+-----+-----+-----+-----+-----+-----+-----+-----+
| Action=0     |0 1|0 0 0 0 0 0|                               Length = 12 |
+-----+-----+-----+-----+-----+-----+-----+-----+
|                               Link Local Identifier = #1 |
+-----+-----+-----+-----+-----+-----+-----+-----+
|                               Link Local Identifier = #2 |
+-----+-----+-----+-----+-----+-----+-----+-----+
| Action=0     |1| Reserved |                               Length = 8 |
+-----+-----+-----+-----+-----+-----+-----+-----+
|           RB ID = #1 |           RB ID = #2 |
+-----+-----+-----+-----+-----+-----+-----+-----+
|           Note: WC1 can only connect to E1
+-----+-----+-----+-----+-----+-----+-----+-----+
| Action=0     |1 0|0 0 0 0 0 0|                               Length = 8 |
+-----+-----+-----+-----+-----+-----+-----+-----+
|                               Link Local Identifier = #1 |
+-----+-----+-----+-----+-----+-----+-----+-----+
| Action=0     |0| Reserved |                               Length = 8 |
+-----+-----+-----+-----+-----+-----+-----+-----+
|           RB ID = #1 |           zero padding |
+-----+-----+-----+-----+-----+-----+-----+-----+
|           Note: WC2 can only connect to E2
+-----+-----+-----+-----+-----+-----+-----+-----+
| Action=0     |1 0|0 0 0 0 0 0|                               Length = 8 |
+-----+-----+-----+-----+-----+-----+-----+-----+
|                               Link Local Identifier = #2 |
+-----+-----+-----+-----+-----+-----+-----+-----+
| Action=0     |0|                               Length = 8 |
+-----+-----+-----+-----+-----+-----+-----+-----+
|           RB ID = #2 |           zero padding |
+-----+-----+-----+-----+-----+-----+-----+-----+

```

A.2. Wavelength Conversion Range Sub-TLV

Example:

This example, based on figure 1, shows how to represent the wavelength conversion range of wavelength converters. Suppose the

wavelength range of input and output of WC1 and WC2 are {L1, L2, L3, L4}:

```

0           1           2           3
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
Note: WC Set
+++++
| Action=0 |1| Reserved | Length = 8 |
+++++
| WC ID = #1 | WC ID = #2 |
+++++
Note: wavelength input range
+++++
| 2 | Num Wavelengths = 4 | Length = 8 |
+++++
|Grid | C.S. | Reserved | n for lowest frequency = 1 |
+++++
Note: wavelength output range
+++++
| 2 | Num Wavelengths = 4 | Length = 8 |
+++++
|Grid | C.S. | Reserved | n for lowest frequency = 1 |
+++++

```

A.3. An OEO Switch with DWDM Optics

Figure 2 shows an electronic switch fabric surrounded by DWDM optics. In this example the electronic fabric can handle either G.709 or SDH signals only (2.5 or 10 Gbps). To describe this node, the following information is needed:

```

<Node_Info> ::= <Node_ID>[Other GMPLS sub-
TLVs][<ConnectivityMatrix>...] [<ResourcePool>][<RBPoolState>]

```

In this case there is complete port to port connectivity so the <ConnectivityMatrix> is not required. In addition since there are sufficient ports to handle all wavelength signals the <RBPoolState> element is not needed.

Hence the attention will be focused on the <ResourcePool> sub-TLV:

```

<ResourcePool> ::=
<ResourceBlockInfo>[<ResourceBlockAccessibility>...][<ResourceWaveCon
straints>...]

```

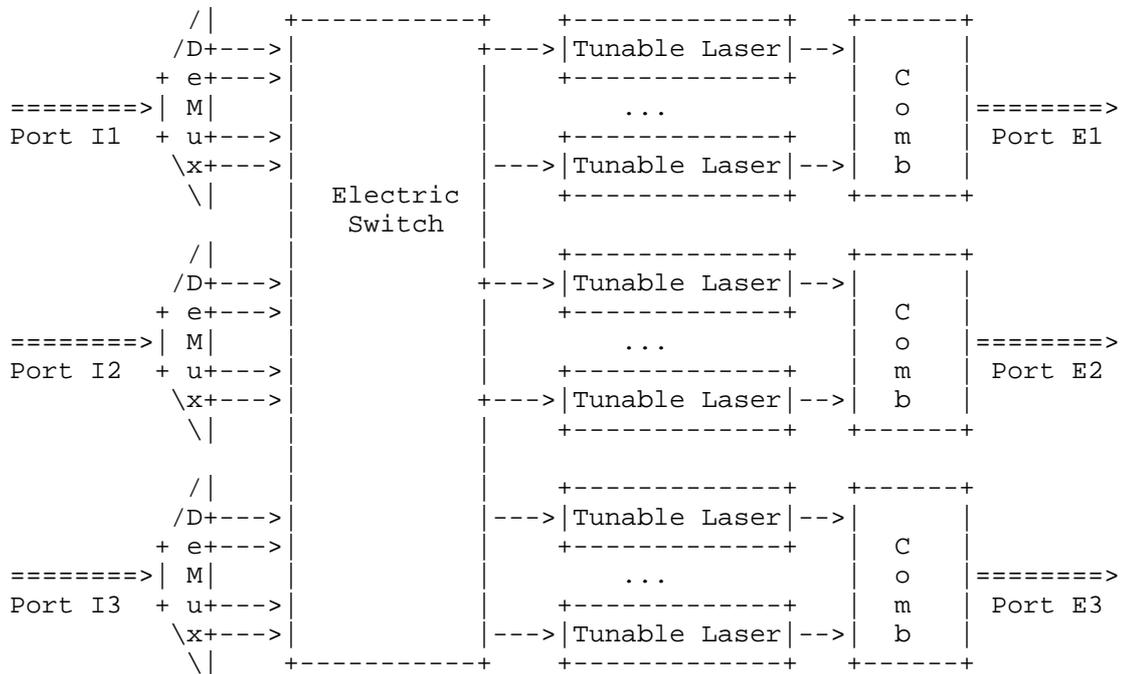
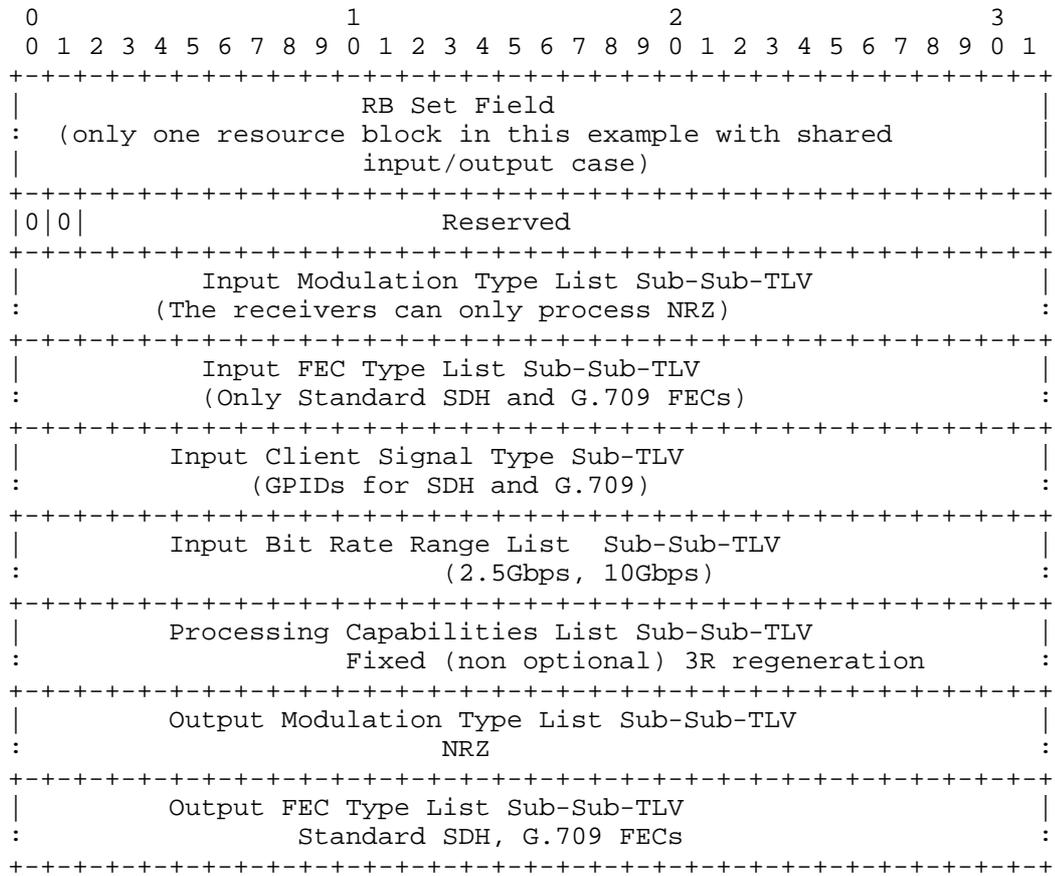


Figure 2 An optical switch built around an electronic switching fabric.

The resource block information will tell us about the processing constraints of the receivers, transmitters and the electronic switch. The resource availability information, although very simple, tells us that all signals must traverse the electronic fabric (fixed connectivity). The resource wavelength constraints are not needed since there are no special wavelength constraints for the resources that would not appear as port/wavelength constraints.

<ResourceBlockInfo>:



Since there is fixed connectivity to resource blocks (the electronic switch) the <ResourceBlockAccessibility> is:

```

0                               1                               2                               3
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| Connectivity=1|Reserved          |
+-----+-----+-----+-----+-----+-----+-----+-----+
|                               Ingress Link Set Field A #1      |
:                               (All ingress links connect to resource) :
+-----+-----+-----+-----+-----+-----+-----+-----+
|                               RB Set Field A #1                |
:                               (trivial set only one resource block) :
+-----+-----+-----+-----+-----+-----+-----+-----+
|                               Egress Link Set Field B #1       |
:                               (All egress links connect to resource) :
+-----+-----+-----+-----+-----+-----+-----+-----+

```

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.
- [RFC2578] McCloghrie, K., Perkins, D., and J. Schoenwaelder, "Structure of Management Information Version 2 (SMIv2)", STD 58, RFC 2578, April 1999.
- [RFC3471] Berger, L., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Functional Description", RFC 3471, January 2003.
- [RFC4328] Papadimitriou, D., Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Extensions for G.709 Optical Transport Networks Control", RFC 4328, January 2006.
- [G.694.1] ITU-T Recommendation G.694.1, "Spectral grids for WDM applications: DWDM frequency grid", June, 2002.

8.2. Informative References

- [G.694.1] ITU-T Recommendation G.694.1, Spectral grids for WDM applications: DWDM frequency grid, June 2002.
- [G.694.2] ITU-T Recommendation G.694.2, Spectral grids for WDM applications: CWDM wavelength grid, December 2003.
- [Gen-Encode] G. Bernstein, Y. Lee, D. Li, W. Imajuku, "General Network Element Constraint Encoding for GMPLS Controlled Networks", work in progress: draft-ietf-ccamp-general-constraint-encode.
- [Otani] T. Otani, H. Guo, K. Miyazaki, D. Caviglia, "Generalized Labels for G.694 Lambda-Switching Capable Label Switching Routers", work in progress: draft-ietf-ccamp-gmpls-g-694-lambda-labels.
- [WSON-Frame] Y. Lee, G. Bernstein, W. Imajuku, "Framework for GMPLS and PCE Control of Wavelength Switched Optical Networks", work in progress: draft-ietf-ccamp-wavelength-switched-framework.

[WSO-Info] G. Bernstein, Y. Lee, D. Li, W. Imajuku, "Routing and Wavelength Assignment Information Model for Wavelength Switched Optical Networks", work in progress: draft-ietf-ccamp-rwa-info.

9. Contributors

Diego Caviglia
Ericsson
Via A. Negrone 1/A 16153
Genoa Italy

Phone: +39 010 600 3736
Email: diego.caviglia@(marconi.com, ericsson.com)

Anders Gavler
Acreo AB
Electrum 236
SE - 164 40 Kista Sweden

Email: Anders.Gavler@acreo.se

Jonas Martensson
Acreo AB
Electrum 236
SE - 164 40 Kista, Sweden

Email: Jonas.Martensson@acreo.se

Itaru Nishioka
NEC Corp.
1753 Simonumabe, Nakahara-ku, Kawasaki, Kanagawa 211-8666
Japan

Phone: +81 44 396 3287
Email: i-nishioka@cb.jp.nec.com

Authors' Addresses

Greg M. Bernstein (ed.)
Grotto Networking
Fremont California, USA

Phone: (510) 573-2237
Email: gregb@grotto-networking.com

Young Lee (ed.)
Huawei Technologies
1700 Alma Drive, Suite 100
Plano, TX 75075
USA

Phone: (972) 509-5599 (x2240)
Email: ylee@huawei.com

Dan Li
Huawei Technologies Co., Ltd.
F3-5-B R&D Center, Huawei Base,
Bantian, Longgang District
Shenzhen 518129 P.R.China

Phone: +86-755-28973237
Email: danli@huawei.com

Wataru Imajuku
NTT Network Innovation Labs
1-1 Hikari-no-oka, Yokosuka, Kanagawa
Japan

Phone: +81-(46) 859-4315
Email: imajuku.wataru@lab.ntt.co.jp

Jianrui Han
Huawei Technologies Co., Ltd.
F3-5-B R&D Center, Huawei Base,
Bantian, Longgang District
Shenzhen 518129 P.R.China

Phone: +86-755-28972916
Email: hanjianrui@huawei.com

Intellectual Property Statement

The IETF Trust takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in any IETF Document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights.

Copies of Intellectual Property disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a 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 any standard or specification contained in an IETF Document. Please address the information to the IETF at ietf-ipr@ietf.org.

Disclaimer of Validity

All IETF Documents and the information contained therein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY, THE IETF TRUST AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION THEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Acknowledgment

Funding for the RFC Editor function is currently provided by the Internet Society.

Network Working Group

Y. Lee
Huawei
G. Bernstein
Grotto Networking
D. Li
Huawei
G. Martinelli
Cisco

Internet Draft

Intended status: Informational

April 12, 2011

Expires: October 2011

A Framework for the Control of Wavelength Switched Optical Networks
(WSO) with Impairments
draft-ietf-ccamp-wson-impairments-06.txt

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at <http://www.ietf.org/ietf/lid-abstracts.txt>

The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>

This Internet-Draft will expire on October 12, 2011.

Copyright Notice

Copyright (c) 2011 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Abstract

As an optical signal progresses along its path it may be altered by the various physical processes in the optical fibers and devices it encounters. When such alterations result in signal degradation, these processes are usually referred to as "impairments". These physical characteristics may be important constraints to consider when using a GMPLS control plane to support path setup and maintenance in wavelength switched optical networks.

This document provides a framework for applying GMPLS protocols and the PCE architecture to support Impairment Aware Routing and Wavelength Assignment (IA-RWA) in wavelength switched optical networks.

Table of Contents

1. Introduction.....	3
2. Terminology.....	4
3. Applicability.....	5
4. Impairment Aware Optical Path Computation.....	6
4.1. Optical Network Requirements and Constraints.....	7
4.1.1. Impairment Aware Computation Scenarios.....	7
4.1.2. Impairment Computation and Information Sharing Constraints.....	8
4.1.3. Impairment Estimation Process.....	10
4.2. IA-RWA Computation and Control Plane Architectures.....	11
4.2.1. Combined Routing, WA, and IV.....	13
4.2.2. Separate Routing, WA, or IV.....	13
4.2.3. Distributed WA and/or IV.....	14
4.3. Mapping Network Requirements to Architectures.....	15
5. Protocol Implications.....	17
5.1. Information Model for Impairments.....	17
5.2. Routing.....	18
5.3. Signaling.....	19
5.4. PCE.....	19
5.4.1. Combined IV & RWA.....	19

5.4.2. IV-Candidates + RWA.....	20
5.4.3. Approximate IA-RWA + Separate Detailed IV.....	21
6. Security Considerations.....	23
7. IANA Considerations.....	23
8. References.....	24
8.1. Normative References.....	24
8.2. Informative References.....	25
9. Acknowledgments.....	26

1. Introduction

Wavelength Switched Optical Networks (WSONs) are constructed from subsystems that may include Wavelength Division Multiplexed (WDM) links, tunable transmitters and receivers, Reconfigurable Optical Add/Drop Multiplexers (ROADM), wavelength converters, and electro-optical network elements. A WSON is a wavelength division multiplexed (WDM)-based optical network in which switching is performed selectively based on the center wavelength of an optical signal.

As an optical signal progresses along its path it may be altered by the various physical processes in the optical fibers and devices it encounters. When such alterations result in signal degradation, these processes are usually referred to as "impairments". Optical impairments accumulate along the path (without 3R regeneration) traversed by the signal. They are influenced by the type of fiber used, the types and placement of various optical devices and the presence of other optical signals that may share a fiber segment along the signal's path. The degradation of the optical signals due to impairments can result in unacceptable bit error rates or even a complete failure to demodulate and/or detect the received signal.

In order to provision an optical connection (an optical path) through a WSON certain path continuity, resource availability and impairments constraints must be met to determine viable and optimal paths through the network. The determination of paths is known as Impairment Aware Routing and Wavelength Assignment (IA-RWA).

Generalized Multi-Protocol Label Switching (GMPLS) [RFC3945] includes a set of control plane protocols that can be used to operate data networks ranging from packet switch capable networks, through those networks that use time division multiplexing, and WDM. [RFC4054] gives an overview of some critical optical impairments and their routing (path selection) implications for GMPLS. The Path Computation Element (PCE) architecture [RFC4655] defines functional components that can be used to compute and suggest appropriate paths in connection-oriented traffic-engineered networks.

This document provides a framework for applying GMPLS protocols and the PCE architecture to the control and operation of IA-RWA for WSONs. To aid in this process this document also provides an overview of the subsystems and processes that comprise WSONs, and describes IA-RWA so that the information requirements can be identified to explain how the information can be modeled for use by GMPLS and PCE systems. This work will facilitate the development of protocol solution models and protocol extensions within the GMPLS and PCE protocol families.

2. Terminology

Add/Drop Multiplexers (ADM): An optical device used in WDM networks composed of one or more line side ports and typically many tributary ports.

CWDM: Coarse Wavelength Division Multiplexing.

DWDM: Dense Wavelength Division Multiplexing.

FOADM: Fixed Optical Add/Drop Multiplexer.

GMPLS: Generalized Multi-Protocol Label Switching.

IA-RWA: Impairment Aware Routing and Wavelength Assignment

Line side: In WDM system line side ports and links typically can carry the full multiplex of wavelength signals, as compared to tributary (add or drop ports) that typically carry a few (typically one) wavelength signals.

OXC: Optical cross connect. An optical switching element in which a signal on any input port can reach any output port.

PCC: Path Computation Client. Any client application requesting a path computation to be performed by the Path Computation Element.

PCE: Path Computation Element. An entity (component, application, or network node) that is capable of computing a network path or route based on a network graph and applying computational constraints.

PCEP: PCE Communication Protocol. The communication protocol between a Path Computation Client and Path Computation Element.

ROADM: Reconfigurable Optical Add/Drop Multiplexer. A wavelength selective switching element featuring input and output line side ports as well as add/drop tributary ports.

RWA: Routing and Wavelength Assignment.

Transparent Network: A wavelength switched optical network that does not contain regenerators or wavelength converters.

Translucent Network: A wavelength switched optical network that is predominantly transparent but may also contain limited numbers of regenerators and/or wavelength converters.

Tributary: A link or port on a WDM system that can carry significantly less than the full multiplex of wavelength signals found on the line side links/ports. Typical tributary ports are the add and drop ports on an ADM and these support only a single wavelength channel.

Wavelength Conversion/Converters: The process of converting information bearing optical signal centered at a given wavelength to one with "equivalent" content centered at a different wavelength. Wavelength conversion can be implemented via an optical-electronic-optical (OEO) process or via a strictly optical process.

WDM: Wavelength Division Multiplexing.

Wavelength Switched Optical Networks (WSONs): WDM based optical networks in which switching is performed selectively based on the center wavelength of an optical signal.

3. Applicability

There are deployment scenarios for WSON networks where not all possible paths will yield suitable signal quality. There are multiple reasons behind this choice; here below is a non-exhaustive list of examples:

- o WSON is evolving using multi-degree optical cross connects in a way that network topologies are changing from rings (and interconnected rings) to general mesh. Adding network equipment such as amplifiers or regenerators, to make all paths feasible, leads to an over-provisioned network. Indeed, even with over provisioning, the network could still have some infeasible paths.
- o Within a given network, the optical physical interface may change over the network life, e.g., the optical interfaces might be upgraded to higher bit-rates. Such changes could result in paths being unsuitable for the optical signal. Moreover, the optical physical interfaces are typically provisioned at various stages of the network's life span as needed by traffic demands.

- o There are cases where a network is upgraded by adding new optical cross connects to increase network flexibility. In such cases existing paths will have their feasibility modified while new paths will need to have their feasibility assessed.
- o With the recent bit rate increases from 10G to 40G and 100G over a single wavelength, WSON networks will likely be operated with a mix of wavelengths at different bit rates. This operational scenario will impose impairment constraints due to different physical behavior of different bit rates and associated modulation formats.

Not having an impairment aware control plane for such networks will require a more complex network design phase that takes into account evolving network status in term of equipments and traffic at the beginning stage. This could result in over-engineering the DWDM network with additional regenerators and optical amplifiers. In addition, network operations such as path establishment, will require significant pre-design via non-control plane processes resulting in significantly slower network provisioning.

4. Impairment Aware Optical Path Computation

The basic criteria for path selection is whether one can successfully transmit the signal from a transmitter to a receiver within a prescribed error tolerance, usually specified as a maximum permissible bit error ratio (BER). This generally depends on the nature of the signal transmitted between the sender and receiver and the nature of the communications channel between the sender and receiver. The optical path utilized (along with the wavelength) determines the communications channel.

The optical impairments incurred by the signal along the fiber and at each optical network element along the path determine whether the BER performance or any other measure of signal quality can be met for a signal on a particular end-to-end path.

Impairment-aware path calculation also needs to take into account when regeneration is used along the path. [WSON-Frame] provides background on the concept of optical translucent networks which contains transparent elements and electro-optical elements such as OEO regenerations. In such networks a generic light path can go through a number of regeneration points.

Regeneration points could happen for two reasons:

- (i) wavelength conversion to assist RWA to avoid wavelength blocking. This is the impairment free case covered by [WSON-Frame].

(ii) the optical signal without regeneration would be too degraded to meet end to end BER requirements. This is the case when RWA takes into consideration impairment estimation covered by this document.

In the latter case an optical path can be seen as a set of transparent segments. The optical impairments calculation needs to be reset at each regeneration point so each transparent segment will have its own impairment evaluation.

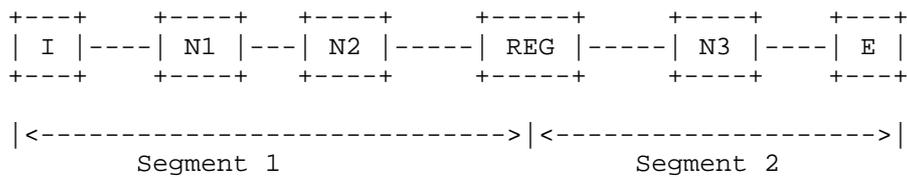


Figure 1 Optical path as a set of transparent segments

For example, Figure 1 represents an optical path from node I to node E with a regeneration point REG in between. It is feasible from an impairment validation perspective if both segments (I, N1, N2, REG) and (REG, N3, E) are feasible.

4.1. Optical Network Requirements and Constraints

This section examines the various optical network requirements and constraints that an impairment aware optical control plane may have to operate under. These requirements and constraints motivate the IA-RWA architectural alternatives to be presented in the following section. Different optical networks contexts can be broken into two main criteria: (a) the accuracy required in the estimation of impairment effects, and (b) the constraints on the impairment estimation computation and/or sharing of impairment information.

4.1.1. Impairment Aware Computation Scenarios

A. No concern for impairments or Wavelength Continuity Constraints

This situation is covered by existing GMPLS with local wavelength (label) assignment.

B. No concern for impairments but Wavelength Continuity Constraints

This situation is applicable to networks designed such that every possible path is valid for the signal types permitted on the network. In this case impairments are only taken into account during network design and after that, for example during optical path computation,

they can be ignored. This is the case discussed in [WSON-Frame] where impairments may be ignored by the control plane and only optical parameters related to signal compatibility are considered.

C. Approximated Impairment Estimation

This situation is applicable to networks in which impairment effects need to be considered but there is sufficient margin such that they can be estimated via approximation techniques such as link budgets and dispersion [G.680],[G.sup39]. The viability of optical paths for a particular class of signals can be estimated using well defined approximation techniques [G.680], [G.sup39]. This is the generally known as linear case where only linear effects are taken into account. Note that adding or removing an optical signal on the path should not render any of the existing signals in the network as non-viable. For example, one form of non-viability is the occurrence of transients in existing links of sufficient magnitude to impact the BER of existing signals.

Much work at ITU-T has gone into developing impairment models at this and more detailed levels. Impairment characterization of network elements may be used to calculate which paths are conformant with a specified BER for a particular signal type. In such a case, the impairment aware (IA) path computation can be combined with the RWA process to permit more optimal IA-RWA computations. Note that the IA path computation may also take place in a separate entity, i.e., a PCE.

D. Detailed Impairment Computation

This situation is applicable to networks in which impairment effects must be more accurately computed. For these networks, a full computation and evaluation of the impact to any existing paths needs to be performed prior to the addition of a new path. Currently no impairment models are available from ITU-T and this scenario is outside the scope of this document.

4.1.2. Impairment Computation and Information Sharing Constraints

In GMPLS, information used for path computation is standardized for distribution amongst the elements participating in the control plane and any appropriately equipped PCE can perform path computation. For optical systems this may not be possible. This is typically due to only portions of an optical system being subject to standardization. In ITU-T recommendations [G.698.1] and [G.698.2] which specify single channel interfaces to multi-channel DWDM systems only the single

channel interfaces (transmit and receive) are specified while the multi-channel links are not standardized. These DWDM links are referred to as "black links" since their details are not generally available. Note however the overall impact of a black link at the single channel interface points is limited by [G.698.1] and [G.698.2].

Typically a vendor might use proprietary impairment models for DWDM spans and to estimate the validity of optical paths. For example, models of optical nonlinearities are not currently standardized. Vendors may also choose not to publish impairment details for links or a set of network elements in order not to divulge their optical system designs.

In general, the impairment estimation/validation of an optical path for optical networks with "black links" (path) could not be performed by a general purpose impairment aware (IA) computation entity since it would not have access to or understand the "black link" impairment parameters. However, impairment estimation (optical path validation) could be performed by a vendor specific impairment aware computation entity. Such a vendor specific IA computation, could utilize standardized impairment information imported from other network elements in these proprietary computations.

In the following the term "black links" will be used to describe these computation and information sharing constraints in optical networks. From the control plane perspective the following options are considered:

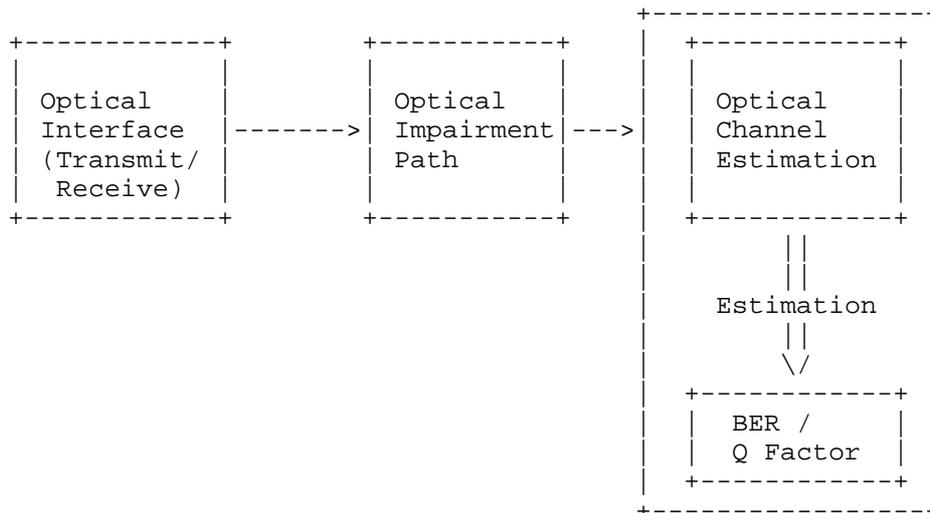
1. The authority in control of the "black links" can furnish a list of all viable paths between all viable node pairs to a computational entity. This information would be particularly useful as an input to RWA optimization to be performed by another computation entity. The difficulty here is for larger networks such a list of paths along with any wavelength constraints could get unmanageably large.
2. The authority in control of the "black links" could provide a PCE like entity that would furnish a list of viable paths/wavelengths between two requested nodes. This is useful as an input to RWA optimizations and can reduce the scaling issue previously mentioned. Such a PCE like entity would not need to perform a full RWA computation, i.e., it would not need to take into account current wavelength availability on links. Such an approach may require PCEP extensions for both the request and response information.

3. The authority in control of the "black links" can provide a PCE that performs full IA-RWA services. The difficulty is this requires the one authority to also become the sole source of all RWA optimization algorithms and such.

In all the above cases it would be the responsibility of the authority in control of the "black links" to import the shared impairment information from the other NEs via the control plane or other means as necessary.

4.1.3. Impairment Estimation Process

The Impairment Estimation Process can be modeled through the following functional blocks. These blocks are independent of any Control Plane architecture, that is, they can be implemented by the same or by different control plane functions as detailed in following sections.



Starting from functional block on the left the Optical Interface represents where the optical signal is transmitted or received and defines the properties at the end points path. Even the no-impairment case like scenario B in section 4.1.1 needs to consider a minimum set of interface characteristics. In such case only a few parameters used to assess the signal compatibility will be taken into account (see [WSON-Frame]). For the impairment-aware case these parameters may be sufficient or not depending on the accepted level of approximation (scenarios C and D). This functional block highlights the need to

consider a set of interface parameters during an Impairment Validation Process.

The block "Optical Impairment Path" represents all kinds of impairments affecting a wavelength as it traverses the networks through links and nodes. In the case where the control plane has no IV this block will not be present. Otherwise, this function must be implemented in some way via the control plane. Options for this will be given in the next section on architectural alternatives. This block implementation (e.g. through routing, signaling or PCE) may influence the way the control plane distributes impairment information within the network.

The last block implements the decision function for path feasibility. Depending on the IA level of approximation this function can be more or less complex. For example in case of no IA only the signal class compatibility will be verified. In addition to feasible/not-feasible result, it may be worthwhile for decision functions to consider the case in which paths can be likely-to-be-feasible within some degree of confidence. The optical impairments are usually not fixed values as they may vary within ranges of values according to the approach taken in the physical modeling (worst-case, statistical or based on typical values). For example, the utilization of the worst-case value for each parameter within impairment validation process may lead to marking some paths as not-feasible while they are very likely to be feasible in reality.

4.2. IA-RWA Computation and Control Plane Architectures

From a control plane point of view optical impairments are additional constraints to the impairment-free RWA process described in [WSON-Frame]. In impairment aware routing and wavelength assignment (IA-RWA), there are conceptually three general classes of processes to be considered: Routing (R), Wavelength Assignment (WA), and Impairment Validation (estimation) (IV).

Impairment validation may come in many forms, and maybe invoked at different levels of detail in the IA-RWA process. From a process point of view the following three forms of impairment validation will be considered:

- o IV-Candidates

In this case an Impairment Validation (IV) process furnishes a set of paths between two nodes along with any wavelength restrictions such that the paths are valid with respect to optical impairments. These

paths and wavelengths may not be actually available in the network due to its current usage state. This set of paths could be returned in response to a request for a set of at most K valid paths between two specified nodes. Note that such a process never directly discloses optical impairment information. Note that that this case includes any paths between source and destination that may have been "pre-validated".

In this case the control plane simply makes use of candidate paths but does not know any optical impairment information. Another option is when the path validity is assessed within the control plane. The following cases highlight this situation.

- o IV-Approximate Verification

Here approximation methods are used to estimate the impairments experienced by a signal. Impairments are typically approximated by linear and/or statistical characteristics of individual or combined components and fibers along the signal path.

- o IV-Detailed Verification

In this case an IV process is given a particular path and wavelength through an optical network and is asked to verify whether the overall quality objectives for the signal over this path can be met. Note that such a process never directly discloses optical impairment information.

The next two cases refer to the way an impairment validation computation can be performed.

- o IV-Centralized

In this case impairments to a path are computed at a single entity. The information concerning impairments, however, may still be gathered from network elements. Depending how information is gathered this may put additional requirements on routing protocols. This will be detailed in later sections.

- o IV-Distributed

In the distributed IV process, approximate degradation measures such as OSNR, dispersion, DGD, etc. are accumulated along the path via a signaling like protocol. Each node on the path may already perform some part of the impairment computation (i.e. distributed). When the accumulated measures reach the destination node a decision on the

impairment validity of the path can be made. Note that such a process would entail revealing an individual network element's impairment information but it does not generally require distributing optical parameters to the entire network.

The Control Plane must not preclude the possibility to operate one or all the above cases concurrently in the same network. For example there could be cases where a certain number of paths are already pre-validated (IV-Candidates) so the control plane may setup one of those path without requesting any impairment validation procedure. On the same network however the control plane may compute a path outside the set of IV-Candidates for which an impairment evaluation can be necessary.

The following subsections present three major classes of IA-RWA path computation architectures and reviews some of their respective advantages and disadvantages.

4.2.1. Combined Routing, WA, and IV

From the point of view of optimality, reasonably good IA-RWA solutions can be achieved if the path computation entity (PCE) can conceptually/algorithmically combine the processes of routing, wavelength assignment and impairment validation.

Such a combination can take place if the PCE is given: (a) the impairment-free WSON network information as discussed in [WSON-Frame] and (b) impairment information to validate potential paths.

4.2.2. Separate Routing, WA, or IV

Separating the processes of routing, WA and/or IV can reduce the need for sharing of different types of information used in path computation. This was discussed for routing separate from WA in [WSON-Frame]. In addition, as was discussed some impairment information may not be shared and this may lead to the need to separate IV from RWA. In addition, if IV needs to be done at a high level of precision it may be advantageous to offload this computation to a specialized server.

The following conceptual architectures belong in this general category:

- o R+WA+IV -- separate routing, wavelength assignment, and impairment validation.

- o R + (WA & IV) -- routing separate from a combined wavelength assignment and impairment validation process. Note that impairment validation is typically wavelength dependent hence combining WA with IV can lead to efficiencies.
- o (RWA)+IV - combined routing and wavelength assignment with a separate impairment validation process.

Note that the IV process may come before or after the RWA processes. If RWA comes first then IV is just rendering a yes/no decision on the selected path and wavelength. If IV comes first it would need to furnish a list of possible (valid with respect to impairments) routes and wavelengths to the RWA processes.

4.2.3. Distributed WA and/or IV

In the non-impairment RWA situation [WSON-Frame] it was shown that a distributed wavelength assignment (WA) process carried out via signaling can eliminate the need to distribute wavelength availability information via an interior gateway protocol (IGP). A similar approach can allow for the distributed computation of impairment effects and avoid the need to distribute impairment characteristics of network elements and links via routing protocols or by other means. An example of such an approach is given in [Martinelli] and utilizes enhancements to RSVP signaling to carry accumulated impairment related information. So the following conceptual options belong to this category:

- o RWA + D(IV) - Combined routing and wavelength assignment and distributed impairment validation.
- o R + D(WA & IV) -- routing separate from a distributed wavelength assignment and impairment validation process.

Distributed impairment validation for a prescribed network path requires that the effects of impairments be calculated by approximate models with cumulative quality measures such as those given in [G.680]. For such a system to be interoperable the exact encoding of the techniques from [G.680] would need to be agreed upon.

If distributed WA is being done at the same time as distributed IV then it is necessary to accumulate impairment related information for all wavelengths that could be used. This is somewhat winnowed down as potential wavelengths are discovered to be in use, but could be a significant burden for lightly loaded high channel count networks.

4.3. Mapping Network Requirements to Architectures

Figure 2 shows process flows for three main architectural alternatives to IA-RWA when approximate impairment validation suffices. Figure 3 shows process flows for two main architectural alternatives when detailed impairment verification is required.

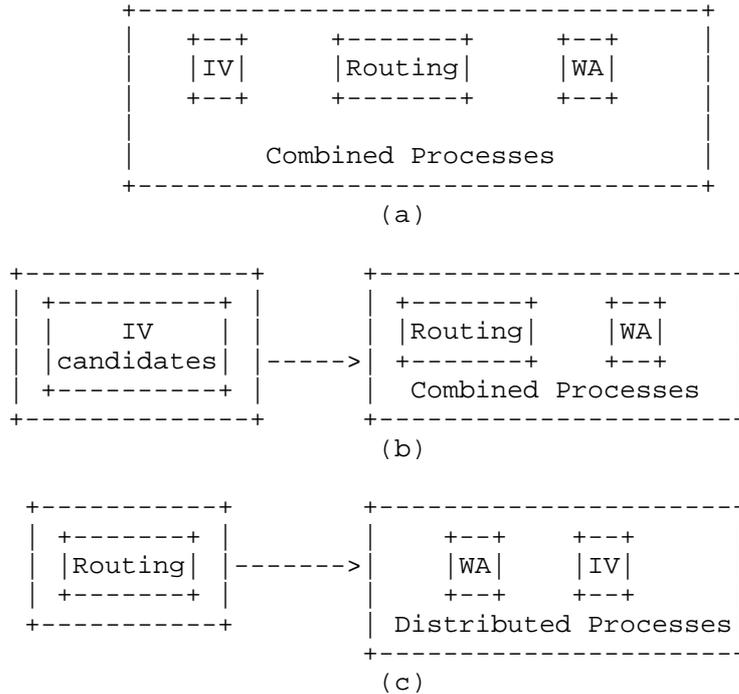


Figure 2 Process flows for the three main approximate impairment architectural alternatives.

The advantages, requirements and suitability of these options are as follows:

o Combined IV & RWA process

This alternative combines RWA and IV within a single computation entity enabling highest potential optimality and efficiency in IA-RWA. This alternative requires that the computational entity knows impairment information as well as non-impairment RWA information. This alternative can be used with "black links", but would then need to be provided by the authority controlling the "black links".

o IV-Candidates + RWA process

This alternative allows separation of impairment information into two computational entities while still maintaining a high degree of potential optimality and efficiency in IA-RWA. The candidates IV process needs to know impairment information from all optical network elements, while the RWA process needs to know non-impairment RWA information from the network elements. This alternative can be used with "black links", but the authority in control of the "black links" would need to provide the functionality of the IV-candidates process. Note that this is still very useful since the algorithmic areas of IV and RWA are very different and prone to specialization.

o Routing + Distributed WA and IV

In this alternative a signaling protocol is extended and leveraged in the wavelength assignment and impairment validation processes. Although this doesn't enable as high a potential degree of optimality of optimality as (a) or (b), it does not require distribution of either link wavelength usage or link/node impairment information. Note that this is most likely not suitable for "black links".

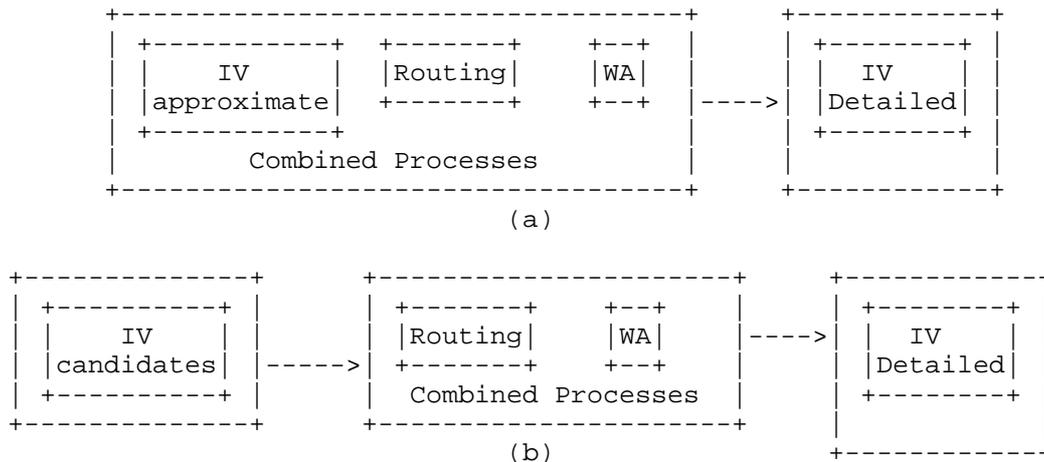


Figure 3 Process flows for the two main detailed impairment validation architectural options.

The advantages, requirements and suitability of these detailed validation options are as follows:

o Combined approximate IV & RWA + Detailed-IV

This alternative combines RWA and approximate IV within a single computation entity enabling highest potential optimality and

efficiency in IA-RWA; then has a separate entity performing detailed impairment validation. In the case of "black links" the authority controlling the "black links" would need to provide all functionality.

- o Candidates-IV + RWA + Detailed-IV

This alternative allows separation of approximate impairment information into a computational entity while still maintaining a high degree of potential optimality and efficiency in IA-RWA; then a separate computation entity performs detailed impairment validation. Note that detailed impairment estimation is not standardized.

5. Protocol Implications

The previous IA-RWA architectural alternatives and process flows make differing demands on a GMPLS/PCE based control plane. This section discusses the use of (a) an impairment information model, (b) PCE as computational entity assuming the various process roles and consequences for PCEP, (c) any needed extensions to signaling, and (d) extensions to routing. The impacts to the control plane for IA-RWA are summarized in Figure 4.

IA-RWA Option	PCE	Sig	Info Model	Routing
Combined IV & RWA	Yes	No	Yes	Yes
IV-Candidates + RWA	Yes	No	Yes	Yes
Routing + Distributed IV, RWA	No	Yes	Yes	No

Figure 4 IA-RWA architectural options and control plane impacts.

5.1. Information Model for Impairments

As previously discussed all IA-RWA scenarios to a greater or lesser extent rely on a common impairment information model. A number of ITU-T recommendations cover detailed as well as approximate impairment characteristics of fibers and a variety of devices and subsystems. A well integrated impairment model for optical network

elements is given in [G.680] and is used to form the basis for an optical impairment model in a companion document [Imp-Info].

It should be noted that the current version of [G.680] is limited to the networks composed of a single WDM line system vendor combined with OADMs and/or PXC's from potentially multiple other vendors, this is known as situation 1 and is shown in Figure 1-1 of [G.680]. It is planned in the future that [G.680] will include networks incorporating line systems from multiple vendors as well as OADMs and/or PXC's from potentially multiple other vendors, this is known as situation 2 and is shown in Figure 1-2 of [G.680].

The case of distributed impairment validation actually requires a bit more than an impairment information model. In particular, it needs a common impairment "computation" model. In the distributed IV case one needs to standardize the accumulated impairment measures that will be conveyed and updated at each node. Section 9 of [G.680] provides guidance in this area with specific formulas given for OSNR, residual dispersion, polarization mode dispersion/polarization dependent loss, effects of channel uniformity, etc... However, specifics of what intermediate results are kept and in what form would need to be standardized.

5.2. Routing

Different approaches to path/wavelength impairment validation gives rise to different demands placed on GMPLS routing protocols. In the case where approximate impairment information is used to validate paths GMPLS routing may be used to distribute the impairment characteristics of the network elements and links based on the impairment information model previously discussed.

Depending on the computational alternative the routing protocol may need to advertise information necessary to impairment validation process. This can potentially cause scalability issues due to the high amount of data that need to be advertised. Such issue can be addressed separating data that need to be advertised rarely and data that need to be advertised more frequently or adopting other form of awareness solutions described in previous sections (e.g. centralized and/or external IV entity).

In term of approximated scenario (see Section 4.1.1.) the model defined by [G.680] will apply and routing protocol will need to gather information required for such computation.

In the case of distributed-IV no new demands would be placed on the routing protocol.

5.3. Signaling

The largest impacts on signaling occur in the cases where distributed impairment validation is performed. In this case, it is necessary to accumulate impairment information as previously discussed. In addition, since the characteristics of the signal itself, such as modulation type, can play a major role in the tolerance of impairments, this type of information will need to be implicitly or explicitly signaled so that an impairment validation decision can be made at the destination node.

It remains for further study if it may be beneficial to include additional information to a connection request such as desired egress signal quality (defined in some appropriate sense) in non-distributed IV scenarios.

5.4. PCE

In section 4.3. a number of computation architectural alternatives were given that could be used to meet the various requirements and constraints of section 4.1. Here the focus is how these alternatives could be implemented via either a single PCE or a set of two or more cooperating PCEs, and the impacts on the PCEP protocol.

5.4.1. Combined IV & RWA

In this situation, shown in Figure 2(a), a single PCE performs all the computations needed for IA-RWA.

- o TE Database Requirements: WSON Topology and switching capabilities, WSON WDM link wavelength utilization, and WSON impairment information
- o PCC to PCE Request Information: Signal characteristics/type, required quality, source node, destination node
- o PCE to PCC Reply Information: If the computations completed successfully then the PCE returns the path and its assigned wavelength. If the computations could not complete successfully it would be potentially useful to know the reason why. At a very crude level it is of interest to know if this was due to lack of wavelength availability or impairment considerations or a bit of both. The information to be conveyed is for further study.

5.4.2. IV-Candidates + RWA

In this situation, as shown in Figure 2(b), two separate processes are involved in the IA-RWA computation. This requires two cooperating path computation entities: one for the Candidates-IV process and another for the RWA process. In addition, the overall process needs to be coordinated. This could be done with yet another PCE or this functionality can be added to one of previously defined entities. This later option requires the RWA entity to also act as the overall process coordinator. The roles, responsibilities and information requirements for these two entities when instantiated as PCEs are given below.

RWA and Coordinator PCE (RWA-Coord-PCE):

Responsible for interacting with PCC and for utilizing Candidates-PCE as needed during RWA computations. In particular it needs to know to use the Candidates-PCE to obtain potential set of routes and wavelengths.

- o TE Database Requirements: WSON Topology and switching capabilities and WSON WDM link wavelength utilization (no impairment information).
- o PCC to RWA-PCE request: same as in the combined case.
- o RWA-PCE to PCC reply: same as in the combined case.
- o RWA-PCE to IV-Candidates-PCE request: The RWA-PCE asks for a set of at most K routes along with acceptable wavelengths between nodes specified in the original PCC request.
- o IV-Candidates-PCE reply to RWA-PCE: The Candidates-PCE returns a set of at most K routes along with acceptable wavelengths between nodes specified in the RWA-PCE request.

IV-Candidates-PCE:

The IV-Candidates PCE is responsible for impairment aware path computation. It needs not take into account current link wavelength utilization, but this is not prohibited. The Candidates-PCE is only required to interact with the RWA-PCE as indicated above and not the initiating PCC. (Note: RWA-Coord PCE is also a PCC with respect to the IV-Candidate)

- o TE Database Requirements: WSON Topology and switching capabilities and WSON impairment information (no information link wavelength utilization required).

Figure 5 shows a sequence diagram for the interactions between the PCC, RWA-Coord PCE and IV-Candidates PCE.

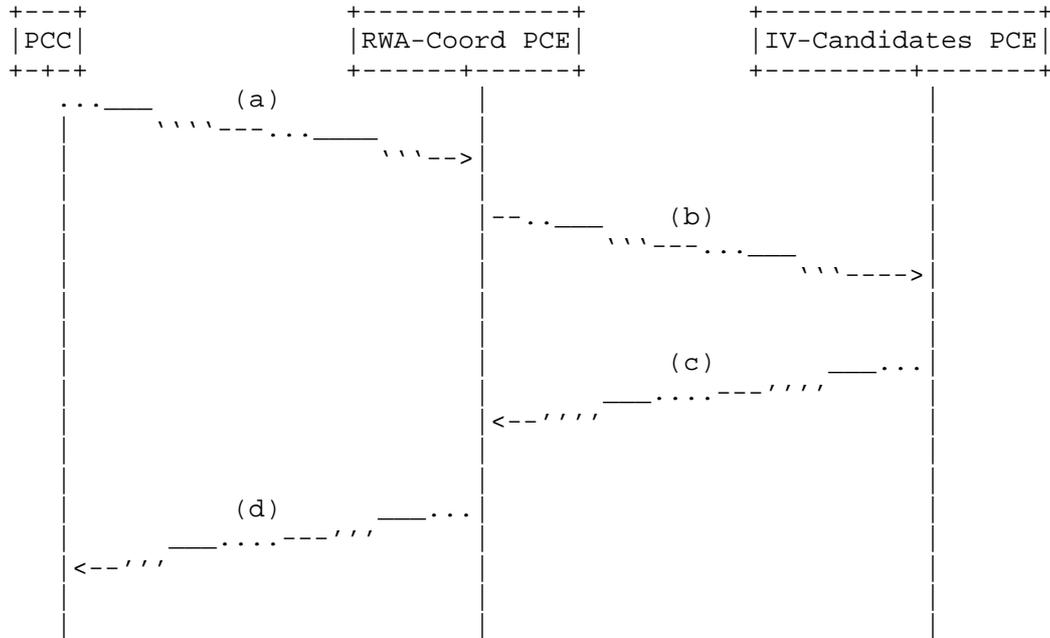


Figure 5 Sequence diagram for the interactions between PCC, RWA-Coordinating-PCE and the IV-Candidates-PCE.

In step (a) the PCC requests a path meeting specified quality constraints between two nodes (A and Z) for a given signal represented either by a specific type or a general class with associated parameters. In step (b) the RWA-Coordinating-PCE requests up to K candidate paths between nodes A and Z and associated acceptable wavelengths. In step (c) The IV-Candidates PCE returns this list to the RWA-Coordinating PCE which then uses this set of paths and wavelengths as input (e.g. a constraint) to its RWA computation. In step (d) the RWA-Coordinating PCE returns the overall IA-RWA computation results to the PCC.

5.4.3. Approximate IA-RWA + Separate Detailed IV

Previously Figure 3 showed two cases where a separate detailed impairment validation process could be utilized. It is possible to place the detailed validation process into a separate PCE. Assuming that a different PCE assumes a coordinating role and interacts with

the PCC it is possible to keep the interactions with this separate IV-Detailed-PCE very simple.

IV-Detailed-PCE:

- o TE Database Requirements: The IV-Detailed-PCE will need optical impairment information, WSON topology, and possibly WDM link wavelength usage information. This document puts no restrictions on the type of information that may be used in these computations.
- o Coordinating-PCE to IV-Detailed-PCE request: The coordinating-PCE will furnish signal characteristics, quality requirements, path and wavelength to the IV-Detailed-PCE.
- o IV-Detailed-PCE to Coordinating-PCE reply: The reply is essential an yes/no decision as to whether the requirements could actually be met. In the case where the impairment validation fails it would be helpful to convey information related to cause or quantify the failure, e.g., so a judgment can be made whether to try a different signal or adjust signal parameters.

Figure 6 shows a sequence diagram for the interactions for the process shown in Figure 3(b). This involves interactions between the PCC, RWA-PCE (acting as coordinator), IV-Candidates-PCE and the IV-Detailed-PCE.

In step (a) the PCC requests a path meeting specified quality constraints between two nodes (A and Z) for a given signal represented either by a specific type or a general class with associated parameters. In step (b) the RWA-Coordinating-PCE requests up to K candidate paths between nodes A and Z and associated acceptable wavelengths. In step (c) The IV-Candidates-PCE returns this list to the RWA-Coordinating PCE which then uses this set of paths and wavelengths as input (e.g. a constraint) to its RWA computation. In step (d) the RWA-Coordinating-PCE request a detailed verification of the path and wavelength that it has computed. In step (e) the IV-Detailed-PCE returns the results of the validation to the RWA-Coordinating-PCE. Finally in step (f) IA-RWA-Coordinating PCE returns the final results (either a path and wavelength or cause for the failure to compute a path and wavelength) to the PCC.

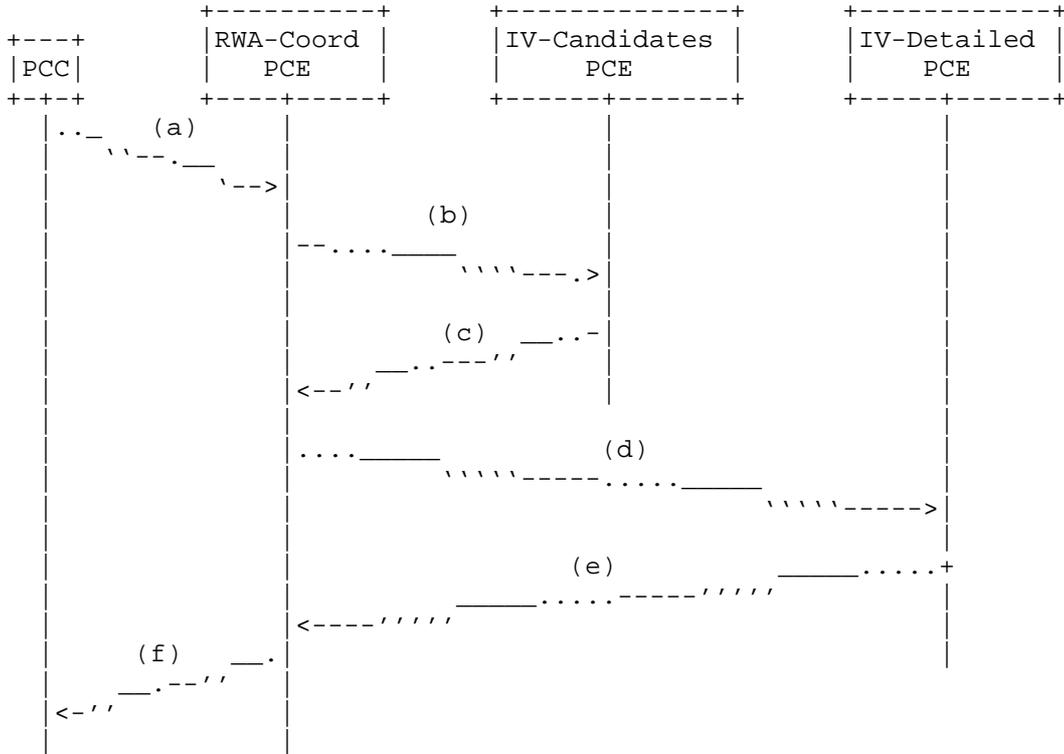


Figure 6 Sequence diagram for the interactions between PCC, RWA-Coordinating-PCE, IV-Candidates-PCE and IV-Detailed-PCE.

6. Security Considerations

This document discusses a number of control plane architectures that incorporate knowledge of impairments in optical networks. If such architecture is put into use within a network it will by its nature contain details of the physical characteristics of an optical network. Such information would need to be protected from intentional or unintentional disclosure.

7. IANA Considerations

This draft does not currently require any consideration from IANA.

8. References

8.1. Normative References

- [G.650.1] ITU-T Recommendation G.650.1, Definitions and test methods for linear, deterministic attributes of single-mode fibre and cable, June 2004.
- [G.650.2] ITU-T Recommendation G.650.2, Definitions and test methods for statistical and non-linear related attributes of single-mode fibre and cable, July 2007.
- [G.650.3] ITU-T Recommendation G.650.3
- [G.652] ITU-T Recommendation G.652, Characteristics of a single-mode optical fibre and cable, June 2005.
- [G.653] ITU-T Recommendation G.653, Characteristics of a dispersion-shifted single-mode optical fibre and cable, December 2006.
- [G.654] ITU-T Recommendation G.654, Characteristics of a cut-off shifted single-mode optical fibre and cable, December 2006.
- [G.655] ITU-T Recommendation G.655, Characteristics of a non-zero dispersion-shifted single-mode optical fibre and cable, March 2006.
- [G.656] ITU-T Recommendation G.656, Characteristics of a fibre and cable with non-zero dispersion for wideband optical transport, December 2006.
- [G.661] ITU-T Recommendation G.661, Definition and test methods for the relevant generic parameters of optical amplifier devices and subsystems, March 2006.
- [G.662] ITU-T Recommendation G.662, Generic characteristics of optical amplifier devices and subsystems, July 2005.
- [G.671] ITU-T Recommendation G.671, Transmission characteristics of optical components and subsystems, January 2005.
- [G.680] ITU-T Recommendation G.680, Physical transfer functions of optical network elements, July 2007.
- [G.691] ITU-T Recommendation G.691, Optical interfaces for multichannel systems with optical amplifiers, November 1998.

- [G.692] ITU-T Recommendation G.692, Optical interfaces for single channel STM-64 and other SDH systems with optical amplifiers, March 2006.
- [G.872] ITU-T Recommendation G.872, Architecture of optical transport networks, November 2001.
- [G.957] ITU-T Recommendation G.957, Optical interfaces for equipments and systems relating to the synchronous digital hierarchy, March 2006.
- [G.959.1] ITU-T Recommendation G.959.1, Optical Transport Network Physical Layer Interfaces, March 2006.
- [G.694.1] ITU-T Recommendation G.694.1, Spectral grids for WDM applications: DWDM frequency grid, June 2002.
- [G.694.2] ITU-T Recommendation G.694.2, Spectral grids for WDM applications: CWDM wavelength grid, December 2003.
- [G.698.1] ITU-T Recommendation G.698.1, Multichannel DWDM applications with Single-Channel optical interface, December 2006.
- [G.698.2] ITU-T Recommendation G.698.2, Amplified multichannel DWDM applications with Single-Channel optical interface, July 2007.
- [G.Sup39] ITU-T Series G Supplement 39, Optical system design and engineering considerations, February 2006.
- [RFC3945] Mannie, E., Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Architecture", RFC 3945, October 2004.
- [RFC4054] Strand, J., Ed., and A. Chiu, Ed., "Impairments and Other Constraints on Optical Layer Routing", RFC 4054, May 2005.
- [RFC4655] Farrel, A., Vasseur, J.-P., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, August 2006.

8.2. Informative References

- [WSON-Frame] Y. Lee, G. Bernstein, W. Imajuku, "Framework for GMPLS and PCE Control of Wavelength Switched Optical Networks", work in progress: draft-ietf-ccamp-wavelength-switched-framework.

[Imp-Info] G. Bernstein, Y. Lee, D. Li, "A Framework for the Control and Measurement of Wavelength Switched Optical Networks (WSON) with Impairments", work in progress: draft-bernstein-wson-impairment-info.

[Martinelli] G. Martinelli (ed.) and A. Zanardi (ed.), "GMPLS Signaling Extensions for Optical Impairment Aware Lightpath Setup", Work in Progress: draft-martinelli-ccamp-optical-imp-signaling.

9. Acknowledgments

This document was prepared using 2-Word-v2.0.template.dot.

Copyright (c) 2011 IETF Trust and the persons identified as authors of the code. All rights reserved.

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

- o Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
- o Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
- o Neither the name of Internet Society, IETF or IETF Trust, nor the names of specific contributors, may be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

Authors' Addresses

Young Lee (ed.)
Huawei Technologies
1700 Alma Drive, Suite 100
Plano, TX 75075
USA

Phone: (972) 509-5599 (x2240)
Email: ylee@huawei.com

Greg M. Bernstein (ed.)
Grotto Networking
Fremont California, USA

Phone: (510) 573-2237
Email: gregb@grotto-networking.com

Dan Li
Huawei Technologies Co., Ltd.
F3-5-B R&D Center, Huawei Base,
Bantian, Longgang District
Shenzhen 518129 P.R.China

Phone: +86-755-28973237
Email: danli@huawei.com

Giovanni Martinelli
Cisco
Via Philips 12
20052 Monza, Italy

Phone: +39 039 2092044
Email: giomarti@cisco.com

Contributor's Addresses

Ming Chen
Huawei Technologies Co., Ltd.
F3-5-B R&D Center, Huawei Base,
Bantian, Longgang District
Shenzhen 518129 P.R.China

Phone: +86-755-28973237

Email: mchen@huawei.com

Rebecca Han
Huawei Technologies Co., Ltd.
F3-5-B R&D Center, Huawei Base,
Bantian, Longgang District
Shenzhen 518129 P.R.China

Phone: +86-755-28973237
Email: hanjianrui@huawei.com

Gabriele Galimberti
Cisco
Via Philips 12,
20052 Monza, Italy

Phone: +39 039 2091462
Email: ggalimbe@cisco.com

Alberto Tanzi
Cisco
Via Philips 12,
20052 Monza, Italy

Phone: +39 039 2091469
Email: altanzi@cisco.com

David Bianchi
Cisco
Via Philips 12,
20052 Monza, Italy

Email: davbianc@cisco.com

Moustafa Kattan
Cisco
Dubai 500321
United Arab Emirates

Email: mkattan@cisco.com

Dirk Schroetter
Cisco

Email: dschroet@cisco.com

Daniele Ceccarelli
Ericsson

Via A. Negrone 1/A
Genova - Sestri Ponente
Italy

Email: daniele.ceccarelli@ericsson.com

Elisa Bellagamba
Ericsson
Farogatan 6,
Kista 164 40
Sweeden

Email: elisa.bellagamba@ericsson.com

Diego Caviglia
Ericsson
Via A. negrone 1/A
Genova - Sestri Ponente
Italy

Email: diego.caviglia@ericsson.com

Network Working Group
Internet Draft
Intended status: Standards Track
Expires: September 2011

Y. Lee
Huawei
G. Bernstein
Grotto Networking

March 8, 2011

OSPF Enhancement for Signal and Network Element Compatibility for
Wavelength Switched Optical Networks

draft-ietf-ccamp-wson-signal-compatibility-ospf-04.txt

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at <http://www.ietf.org/ietf/lid-abstracts.txt>

The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>

This Internet-Draft will expire on August 8, 2011.

Copyright Notice

Copyright (c) 2011 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Abstract

This document provides GMPLS OSPF routing enhancements to support signal compatibility constraints associated with WSON network elements. These routing enhancements are required in common optical or hybrid electro-optical networks where not all of the optical signals in the network are compatible with all network elements participating in the network.

This compatibility constraint model is applicable to common optical or hybrid electro optical systems such as OEO switches, regenerators, and wavelength converters since such systems can be limited to processing only certain types of WSON signals.

Conventions used in this document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC-2119 [RFC2119].

Table of Contents

- 1. Introduction.....3
 - 1.1. Revision History.....3
- 2. The Optical Node Property TLV.....4
 - 2.1. Sub-TLV Details.....4
 - 2.1.1. Resource Block Information.....4
 - 2.1.2. Resource Block Accessibility.....5
 - 2.1.3. Resource Block Wavelength Constraints.....5
 - 2.1.4. Resource Block Pool State.....6
- 3. Security Considerations.....6
- 4. IANA Considerations.....6
- 5. References.....8
 - 5.1. Normative References.....8
- 6. Contributors.....9

Authors' Addresses.....9
Intellectual Property Statement.....10
Disclaimer of Validity.....10

1. Introduction

The documents [WSON-Frame, WSON-Info, RWA-Encode] explain how to extend the wavelength switched optical network (WSON) control plane to allow both multiple WSON signal types and common hybrid electro optical systems as well hybrid systems containing optical switching and electro-optical resources. In WSON, not all of the optical signals in the network are compatible with all network elements participating in the network. Therefore, signal compatibility is an important constraint in path computation in a WSON.

This document provides GMPLS OSPF routing enhancements to support signal compatibility constraints associated with general WSON network elements. These routing enhancements are required in common optical or hybrid electro-optical networks where not all of the optical signals in the network are compatible with all network elements participating in the network.

This compatibility constraint model is applicable to common optical or hybrid electro optical systems such as OEO switches, regenerators, and wavelength converters since such systems can be limited to processing only certain types of WSON signals.

1.1. Revision History

From 00 to 01: The details of the encodings for compatibility moved from this document to [RWA_Encode].

From 01 to 02: Editorial changes.

From 02 to 03: Add a new Top Level Node TLV, Optical Node Property TLV to carry WSON specific node information.

From 03 to 04: Add a new sub-TLV, Block Shared Access Wavelength Availability TLV to be consistent with [RWA-Encode] and editorial changes.

2. The Optical Node Property TLV

[RFC 3630] defines OSPF TE LSA using an opaque LSA. This document adds a new top level TLV for use in the OSPF TE LSA: the Optical Node Property TLV.

The Optical Node Property TLV contains all WSON-specific node properties and signal compatibility constraints. The detailed encodings of these properties are defined in [RWA-Encode].

The following sub-TLVs of the Optical Node Property TLV are defined:

Value	Length	Sub-TLV Type
TBA	variable	Resource Block Information
TBA	variable	Resource Pool Accessibility
TBA	variable	Resource Block Wavelength Constraints
TBA	variable	Resource Pool State
TBA	variable	Block Shared Access Wavelength Availability

The detail encodings of these sub-TLVs are found in [RWA-Encode] as indicated in the table below.

Sub-TLV Type	Section[RWA-Encode]
Resource Block Information	4.1
Resource Pool Accessibility	3.1
Resource Block Wavelength Constraints	3.2
Resource Pool State	3.3
Block Shared Access Wavelength Availability	3.4

2.1. Sub-TLV Details

Among the sub-TLVs defined above, the Resource Pool State sub-TLV is dynamic in nature while the rest are static. As such, it will be separated out from the rest and make use of multiple TE LSA instances per source, per [RFC3630] multiple instance capability.

2.1.1. Resource Block Information

Resource Block Information sub-TLVs are used to convey relatively static information about individual resource blocks including the resource block properties and the number of resources in a block.

There are seven nested sub-TLVs defined in the Resource Block Information sub-TLV.

Value	Length	Sub-TLV Type
TBA	variable	Input Modulation Format List
TBA	variable	Input FEC Type List

TBA	variable	Input Bit Range List
TBA	variable	Input Client Signal List
TBA	variable	Processing Capability List
TBA	variable	Output Modulation Format List
TBA	variable	Output FEC Type List

The detail encodings of these sub-TLVs are found in [RWA-Encode] as indicated in the table below.

Name	Section[RWA-Encode]
Input Modulation Format List	4.2
Input FEC Type List	4.3
Input Bit Range List	4.4
Input Client Signal List	4.5
Processing Capability List	4.6
Output Modulation Format List	4.7
Output FEC Type List	4.8

2.1.2. Resource Pool Accessibility

This sub-TLV describes the structure of the resource pool in relation to the switching device. In particular it indicates the ability of an ingress port to reach a resource block and of a resource block to reach a particular egress port.

2.1.3. Resource Block Wavelength Constraints

Resources, such as wavelength converters, etc., may have a limited input or output wavelength ranges. Additionally, due to the structure of the optical system not all wavelengths can necessarily reach or leave all the resources. Resource Block Wavelength Constraints sub-TLV describe these properties.

2.1.4. Resource Pool State

This sub-TLV describes the usage state of a resource that can be encoded as either a list of 16 bit integer values or a bit map indicating whether a single resource is available or in use. This information can be relatively dynamic, i.e., can change when a connection is established or torn down.

2.1.5. Block Shared Access Wavelength Availability

Resources blocks may be accessed via a shared fiber. If this is the case then wavelength availability on these shared fibers is needed to understand resource availability.

3. Security Considerations

This document does not introduce any further security issues other than those discussed in [RFC 3630], [RFC 4203].

4. IANA Considerations

This document introduces a new Top Level Node TLV (Optical Node Property TLV) under the OSPF TE LSA defined in [RFC 3630].

Value TLV Type

TBA Optical Node Property

IANA is to allocate a new TLV Type and its Value for this Top Level Node TLV.

This document also introduces the following sub-TLVs associated with the Optical Node Property TLV as defined in Section 2.1 as follows:

Value	Length	Sub-TLV Type
TBA	variable	Resource Block Information
TBA	variable	Resource Pool Accessibility
TBA	variable	Resource Block Wavelength Constraints
TBA	variable	Resource Pool State
TBA	variable	Block Shared Access Wavelength Availability

IANA is to allocate new sub-TLV Types and their Values for these sub-TLVs defined under the Optical Node Property TLV.

There are seven nested sub-TLVs defined in the Resource Block Information sub-TLV as follows:

Value	Length	Sub-TLV Type
-------	--------	--------------

TBA	variable	Input Modulation Format List
TBA	variable	Input FEC Type List
TBA	variable	Input Bit Range List
TBA	variable	Input Client Signal List
TBA	variable	Processing Capability List
TBA	variable	Output Modulation Format List
TBA	variable	Output FEC Type List

IANA is to allocate new Sub-TLV Types and their Values for these Sub-TLVs defined under the Resource Block Information Sub-TLV.

5. References

5.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC3471] Berger, L., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Functional Description", RFC 3471, January 2003.
- [RFC3630] Katz, D., Kompella, K., and Yeung, D., "Traffic Engineering (TE) Extensions to OSPF Version 2", RFC 3630, September 2003.
- [G.694.1] ITU-T Recommendation G.694.1, "Spectral grids for WDM applications: DWDM frequency grid", June, 2002.
- [RFC4202] Kompella, K., Ed., and Y. Rekhter, Ed., "Routing Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4202, October 2005
- [RFC4203] Kompella, K., Ed., and Y. Rekhter, Ed., "OSPF Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4203, October 2005.
- [RFC4328] Papadimitriou, D., Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Extensions for G.709 Optical Transport Networks Control", RFC 4328, January 2006.
- [RFC5307] Kompella, K., Ed., and Y. Rekhter, Ed., "IS-IS Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 5307, October 2008.
- [OSPF-Node] R. Aggarwal and K. Kompella, "Advertising a Router's Local Addresses in OSPF TE Extensions", draft-ietf-ospf-te-node-addr, work in progress.
- [Lambda-Labels] T. Otani, H. Guo, K. Miyazaki, D. Caviglia, "Generalized Labels for G.694 Lambda-Switching Capable Label Switching Routers", draft-ietf-ccamp-gmpls-g-694-lambda-labels, work in progress.

[WSON-Frame] Y. Lee, G. Bernstein, W. Imajuku, "Framework for GMPLS and PCE Control of Wavelength Switched Optical Networks", draft-ietf-ccamp-rwa-WSON-Framework, work in progress.

[WSON-Info] Y. Lee, G. Bernstein, D. Li, W. Imajuku, "Routing and Wavelength Assignment Information Model for Wavelength Switched Optical Networks", draft-ietf-ccamp-rwa-info work in progress.

[RWA-Encode] G. Bernstein, Y. Lee, D. Li, W. Imajuku, "Routing and Wavelength Assignment Information Encoding for Wavelength Switched Optical Networks", draft-ietf-ccamp-rwa-wson-encode, work in progress.

6. Authors and Contributors

Authors' Addresses

Young Lee (ed.)
Huawei Technologies
1700 Alma Drive, Suite 100
Plano, TX 75075
USA

Phone: (972) 509-5599 (x2240)
Email: ylee@huawei.com

Greg M. Bernstein (ed.)
Grotto Networking
Fremont California, USA

Phone: (510) 573-2237
Email: gregb@grotto-networking.com

Intellectual Property Statement

The IETF Trust takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in any IETF Document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights.

Copies of Intellectual Property disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a 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 any standard or specification contained in an IETF Document. Please address the information to the IETF at ietf-ipr@ietf.org.

Disclaimer of Validity

All IETF Documents and the information contained therein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY, THE IETF TRUST AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION THEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Acknowledgment

Funding for the RFC Editor function is currently provided by the Internet Society.

Network Working Group
Internet Draft
Intended status: Standards Track

Expires: September 2011

G. Bernstein
Grotto Networking
Sugang Xu
NICT
Y.Lee
Huawei
G. Martinelli
Cisco
Hiroaki Harai
NICT

March 12, 2011

Signaling Extensions for Wavelength Switched Optical Networks
draft-ietf-ccamp-wson-signaling-01.txt

Abstract

This memo provides extensions to Generalized Multi-Protocol Label Switching (GMPLS) signaling for control of wavelength switched optical networks (WSON). Such extensions are necessary in WSONs under a number of conditions including: (a) when optional processing, such as regeneration, must be configured to occur at specific nodes along a path, (b) where equipment must be configured to accept an optical signal with specific attributes, or (c) where equipment must be configured to output an optical signal with specific attributes. In addition this memo provides mechanisms to support distributed wavelength assignment with bidirectional LSPs, and choice in distributed wavelength assignment algorithms. These extensions build on previous work for the control of lambda and G.709 based networks.

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at
<http://www.ietf.org/ietf/lid-abstracts.txt>

The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>

This Internet-Draft will expire on September 12, 2011.

Copyright Notice

Copyright (c) 2011 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Conventions used in this document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

Table of Contents

1. Introduction.....	3
2. Terminology.....	3
3. Requirements for WSON Signaling.....	4
3.1. WSON Signal Characterization.....	4
3.2. Per LSP Network Element Processing Configuration.....	5
3.3. Bi-Directional Distributed Wavelength Assignment.....	5
3.4. Distributed Wavelength Assignment Support.....	7
3.5. Out of Scope.....	7
4. WSON Signal Traffic Parameters, Attributes and Processing.....	7
4.1. Traffic Parameters for Optical Tributary Signals.....	7
4.2. Signal Attributes and Processing.....	8
4.2.1. Modulation Type sub-TLV.....	8
4.2.2. FEC Type sub-TLV.....	10
4.2.3. Regeneration Processing TLV.....	13
5. Bidirectional Lightpath Setup.....	14
5.1. Possible Solutions for Bidirectional Lightpath.....	14

5.2. Bidirectional Lightpath Signaling Procedure.....	15
5.3. Backward Compatibility Considerations.....	16
6. RWA Related.....	16
6.1. Wavelength Assignment Method Selection.....	16
7. Security Considerations.....	17
8. IANA Considerations.....	18
9. Acknowledgments.....	18
10. References.....	19
10.1. Normative References.....	19
10.2. Informative References.....	19
Author's Addresses.....	21
Intellectual Property Statement.....	22
Disclaimer of Validity.....	23

1. Introduction

This memo provides extensions to Generalized Multi-Protocol Label Switching (GMPLS) signaling for control of wavelength switched optical networks (WSON). Fundamental extensions are given to permit simultaneous bi-directional wavelength assignment while more advanced extensions are given to support the networks described in [WSON-Frame] which feature connections requiring configuration of input, output, and general signal processing capabilities at a node along a LSP

These extensions build on previous work for the control of lambda and G.709 based networks.

2. Terminology

CWDM: Coarse Wavelength Division Multiplexing.

DWDM: Dense Wavelength Division Multiplexing.

FOADM: Fixed Optical Add/Drop Multiplexer.

ROADM: Reconfigurable Optical Add/Drop Multiplexer. A reduced port count wavelength selective switching element featuring ingress and egress line side ports as well as add/drop side ports.

RWA: Routing and Wavelength Assignment.

Wavelength Conversion/Converters: The process of converting an information bearing optical signal centered at a given wavelength to one with "equivalent" content centered at a different wavelength. Wavelength conversion can be implemented via an optical-electronic-optical (OEO) process or via a strictly optical process.

WDM: Wavelength Division Multiplexing.

Wavelength Switched Optical Networks (WSON): WDM based optical networks in which switching is performed selectively based on the center wavelength of an optical signal.

AWG: Arrayed Waveguide Grating.

OXC: Optical Cross Connect.

Optical Transmitter: A device that has both a laser tuned on certain wavelength and electronic components, which converts electronic signals into optical signals.

Optical Responder: A device that has both optical and electronic components. It detects optical signals and converts optical signals into electronic signals.

Optical Transponder: A device that has both an optical transmitter and an optical responder.

Optical End Node: The end of a wavelength (optical lambdas) lightpath in the data plane. It may be equipped with some optical/electronic devices such as wavelength multiplexers/demultiplexer (e.g. AWG), optical transponder, etc., which are employed to transmit/terminate the optical signals for data transmission.

3. Requirements for WSON Signaling

The following requirements for GMPLS based WSON signaling are in addition to the functionality already provided by existing GMPLS signaling mechanisms.

3.1. WSON Signal Characterization

WSON signaling MUST convey sufficient information characterizing the signal to allow systems along the path to determine compatibility and perform any required local configuration. Examples of such systems include intermediate nodes (ROADMs, OXCs, Wavelength converters, Regenerators, OEO Switches, etc...), links (WDM systems) and end systems (detectors, demodulators, etc...). The details of any local configuration processes are out of the scope of this document.

From [WSON-Frame] we have the following list of WSON signal characteristic information:

List 1. WSON Signal Characteristics

1. Optical tributary signal class (modulation format).
2. FEC: whether forward error correction is used in the digital stream and what type of error correcting code is used
3. Center frequency (wavelength)
4. Bit rate
5. G-PID: General Protocol Identifier for the information format

The first three items on this list can change as a WSON signal traverses a network with regenerators, OEO switches, or wavelength converters. An ability to control wavelength conversion already exists in GMPLS signaling along with the ability to share client signal type information (G-PID). In addition, bit rate is a standard GMPLS signaling traffic parameter. It is referred to as Bandwidth Encoding in [RFC3471]. This leaves two new parameters: modulation format and FEC type, needed to fully characterize the optical signal.

3.2. Per LSP Network Element Processing Configuration

In addition to configuring a network element (NE) along an LSP to input or output a signal with specific attributes, we may need to signal the NE to perform specific processing, such as 3R regeneration, on the signal at a particular NE. In [WSON-Frame] we discussed three types of processing not currently covered by GMPLS:

- (A) Regeneration (possibly different types)
- (B) Fault and Performance Monitoring
- (C) Attribute Conversion

The extensions here MUST provide for the configuration of these types of processing at nodes along an LSP.

3.3. Bi-Directional Distributed Wavelength Assignment

WSON signaling MAY support distributed wavelength assignment consistent with the wavelength continuity constraint for bi-directional connections. The following cases MAY be separately supported:

- (a) Where the same wavelength is used for both upstream and downstream directions
- (b) Where different wavelengths can be used for both upstream and downstream directions.

The need for the same wavelength on both directions mainly comes from the color constraint on some edges' hardware. In fact, the edges can be classified into two types, i.e. without and with the wavelength-port mapping re-configurability.

Without the mapping re-configurability at edges, the edge nodes must use the same wavelength in both directions. For example, (1) transponders are only connected to fixed AWGs (i.e. multiplexer/de-multiplexer) ports directly, or (2) transponders are connected to the add/drop ports of ROADM and each port is mapped to a fixed dedicated wavelength.

On the other hand, with mapping re-configurability at edges, the edge nodes can use different wavelengths in different directions. For example, in edge nodes, transponders are connected to add/drop ports of colorless ROADM. Thus, the wavelength-port remapping problem can be solved locally by appropriately configuring the colorless ROADM. If the colorless ROADM consists of OXC and AWGs, the OXC is configured appropriately.

The edges of data-plane in WSON can be constructed in different types based on cost and flexibility concerns. Without re-configurability we should consider the constraint of the same wavelength usage on both directions, but have lower costs. While, with wavelength-port mapping re-configurability we can relax the constraint, but have higher costs.

These two types of edges will co-exist in WSON mesh, till all the edges are unified by the same type. The existence of the first type edges presents a requirement of the same wavelength usage on both directions, which must be supported.

Moreover, if some carriers prefer easy management of lightpath usage, say use the same wavelength on both directions to reduce the burden on lightpath management, the same wavelength usage would be beneficial.

In cases of equipment failure, etc., fast provisioning used in quick recovery is critical to protect Carriers/Users against system loss. This requires efficient signaling which supports distributed

wavelength assignment, in particular when the centralized wavelength assignment capability is not available.

3.4. Distributed Wavelength Assignment Support

WSON signaling MAY support the selection of a specific distributed wavelength assignment method.

This method is beneficial in cases of equipment failure, etc., where fast provisioning used in quick recovery is critical to protect carriers/users against system loss. This requires efficient signaling which supports distributed wavelength assignment, in particular when the centralized wavelength assignment capability is not available.

As discussed in the [WSON-Frame] different computational approaches for wavelength assignment are available. One method is the use of distributed wavelength assignment. This feature would allow the specification of a particular approach when more than one is implemented in the systems along the path.

3.5. Out of Scope

This draft does not address signaling information related to optical impairments.

4. WSON Signal Traffic Parameters, Attributes and Processing

As discussed in [WSON-Frame] single channel optical signals used in WSONs are called "optical tributary signals" and come in a number of classes characterized by modulation format and bit rate. Although WSONs are fairly transparent to the signals they carry, to ensure compatibility amongst various networks devices and end systems it can be important to include key lightpath characteristics as traffic parameters in signaling [WSON-Frame].

4.1. Traffic Parameters for Optical Tributary Signals

In [RFC3471] we see that the G-PID (client signal type) and bit rate (byte rate) of the signals are defined as parameters and in [RFC3473] they are conveyed Generalized Label Request object and the RSVP SENDER_TSPEC/FLOWSPEC objects respectively.

4.2. Signal Attributes and Processing

Section 3.2. gave the requirements for signaling to indicate to a particular NE along an LSP what type of processing to perform on an optical signal or how to configure that NE to accept or transmit an optical signal with particular attributes.

One way of accomplishing this is via a new EXPLICIT_ROUTE subobject. Reference [RFC3209] defines the EXPLICIT_ROUTE object (ERO) and a number of subobjects, while reference [RFC5420] defines general mechanisms for dealing with additional LSP attributes. Although reference [RFC5420] defines a RECORD_ROUTE object (RRO) attributes subobject, it does not define an ERO subobject for LSP attributes.

Regardless of the exact coding for the ERO subobject conveying the input, output, or processing instructions. This new "processing" subobject would follow a subobject containing the IP address, or the interface identifier [RFC3477], associated with the link on which it is to be used along with any label subobjects [RFC3473].

The contents of this new "processing" subobject would be a list of TLVs that could include:

- o Modulation Type TLV (input and/or output)
- o FEC Type TLV (input and/or output)
- o Processing Instruction TLV

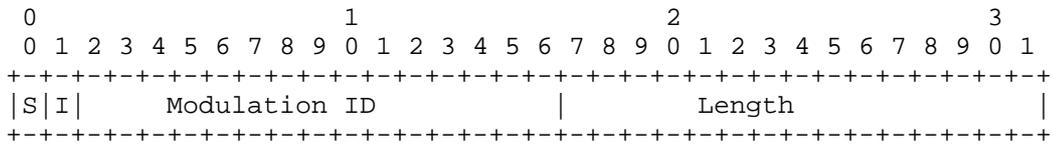
Currently the only processing instruction TLV currently defined is for regeneration. The [WSON-Info] and [WSON-Encoding] provides the details for these specifics sub-TLVs.

Possible encodings and values for these TLV are given in below.

4.2.1. Modulation Type sub-TLV

The encoding for modulation type sub-TLV is defined in [WSON-Encode] Section 4.2.1.

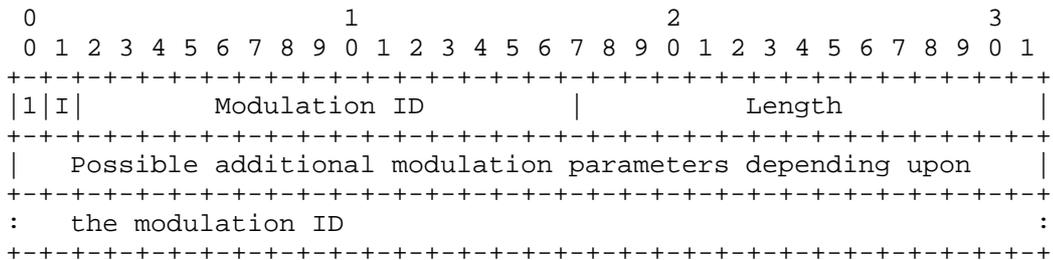
It may come in two different formats: a standard modulation field or a vendor specific modulation field. Both start with the same 32 bit header shown below.



Where S bit set to 1 indicates a standardized modulation format and S bit set to 0 indicates a vendor specific modulation format. The length is the length in bytes of the entire modulation type field.

Where I bit set to 1 indicates an input modulation format and where I bit set to 0 indicates an output modulation format. Note that the source modulation type is implied when I bit is set to 0 and that the sink modulation type is implied when I bit is set to 1. For signaling purposes only the output form (I=0) is needed.

The format for the standardized type is given by:



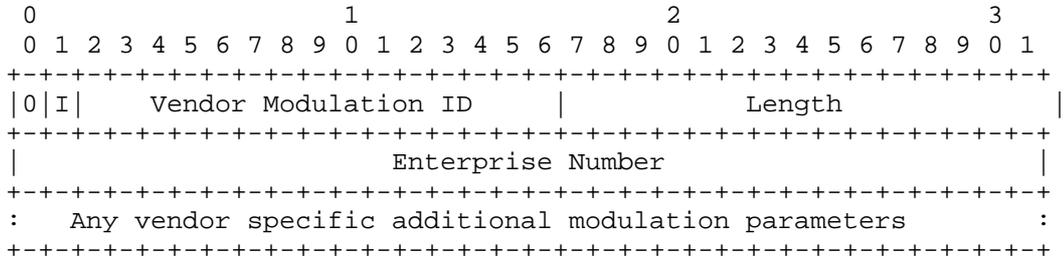
Modulation ID

Takes on the following currently defined values:

- 0 Reserved
- 1 optical tributary signal class NRZ 1.25G
- 2 optical tributary signal class NRZ 2.5G
- 3 optical tributary signal class NRZ 10G
- 4 optical tributary signal class NRZ 40G
- 5 optical tributary signal class RZ 40G

Note that future modulation types may require additional parameters in their characterization.

The format for vendor specific modulation is given by:



Vendor Modulation ID

This is a vendor assigned identifier for the modulation type.

Enterprise Number

A unique identifier of an organization encoded as a 32-bit integer. Enterprise Numbers are assigned by IANA and managed through an IANA registry [RFC2578].

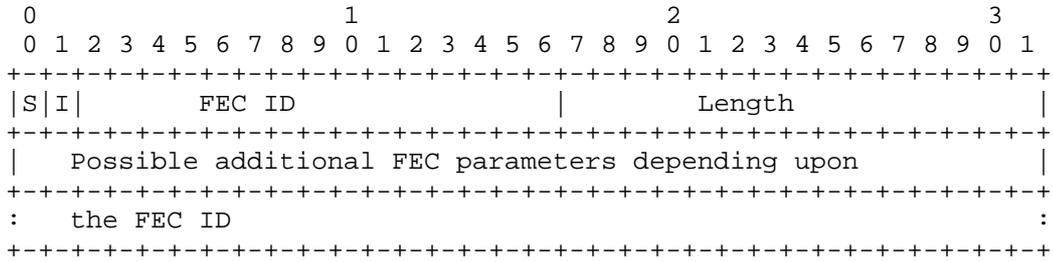
Vendor Specific Additional parameters

There can be potentially additional parameters characterizing the vendor specific modulation.

4.2.2. FEC Type sub-TLV

The encoding for FEC Type TLV is defined in [WSON-Encode] Section 4.3.1.

It indicates the FEC type output at particular node along the LSP. The FEC type sub-TLV comes in two different types: a standard FEC field or a vendor specific FEC field. Both start with the same 32 bit header shown below.

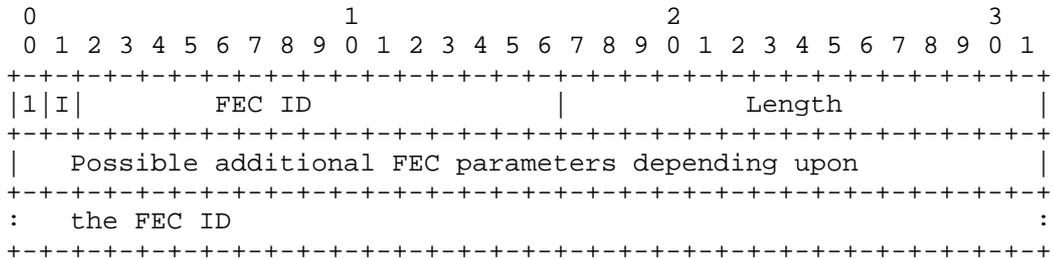


Where S bit set to 1 indicates a standardized FEC format and S bit set to 0 indicates a vendor specific FEC format. The length is the length in bytes of the entire FEC type field.

Where the length is the length in bytes of the entire FEC type field.

Where I bit set to 1 indicates an input FEC format and where I bit set to 0 indicates an output FEC format. Note that the source FEC type is implied when I bit is set to 0 and that the sink FEC type is implied when I bit is set to 1. Only the output form (I=0) is used in signaling.

The format for standard FEC field is given by:



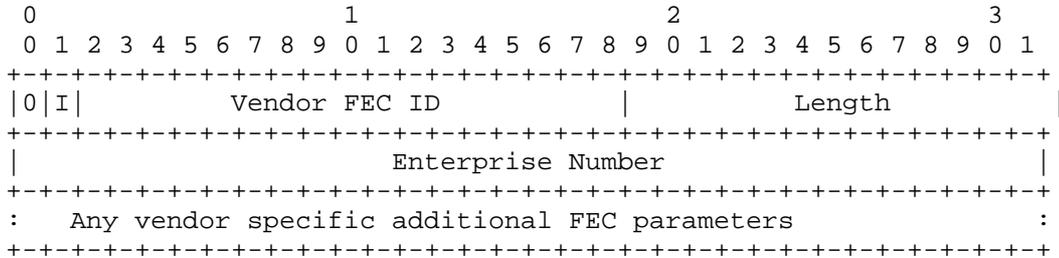
Takes on the following currently defined values for the standard FEC ID:

- 0 Reserved
- 1 G.709 RS FEC
- 2 G.709V compliant Ultra FEC

- 3 G.975.1 Concatenated FEC (RS(255,239)/CSOC(n0/k0=7/6,J=8))
- 4 G.975.1 Concatenated FEC (BCH(3860,3824)/BCH(2040,1930))
- 5 G.975.1 Concatenated FEC (RS(1023,1007)/BCH(2407,1952))
- 6 G.975.1 Concatenated FEC (RS(1901,1855)/Extended Hamming Product Code (512,502)X(510,500))
- 7 G.975.1 LDPC Code
- 8 G.975.1 Concatenated FEC (Two orthogonally concatenated BCH codes)
- 9 G.975.1 RS(2720,2550)
- 10 G.975.1 Concatenated FEC (Two interleaved extended BCH (1020,988) codes)

Where RS stands for Reed-Solomon and BCH for Bose-Chaudhuri-Hocquengham.

The format for vendor-specific FEC field is given by:



Vendor FEC ID

This is a vendor assigned identifier for the FEC type.

Enterprise Number

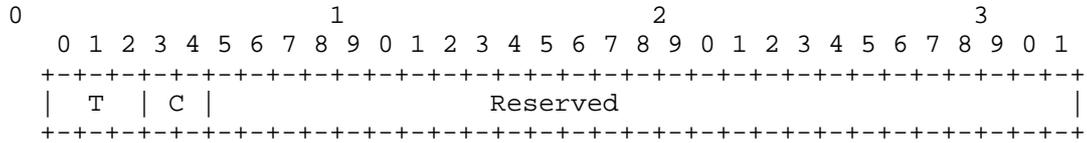
A unique identifier of an organization encoded as a 32-bit integer. Enterprise Numbers are assigned by IANA and managed through an IANA registry [RFC2578].

Vendor Specific Additional FEC parameters

There can be potentially additional parameters characterizing the vendor specific FEC.

4.2.3. Regeneration Processing TLV

The Regeneration Processing TLV is used to indicate that this particular node is to perform the specified type of regeneration processing on the signal.



Where T bit indicates the type of regenerator:

- T=0: Reserved
- T=1: 1R Regenerator
- T=2: 2R Regenerator
- T=3: 3R Regenerator

Where C bit indicates the capability of regenerator:

- C=0: Reserved
- C=1: Fixed Regeneration Point
- C=2: Selective Regeneration Pools

Note that the use of the C field is optional in signaling.

5. Bidirectional Lightpath Setup

With the wavelength continuity constraint in CI-incapable [RFC3471] WSONs, where the nodes in the networks cannot support wavelength conversion, the same wavelength on each link along a unidirectional lightpath should be reserved. In addition to the wavelength continuity constraint, requirement 3.2 gives us another constraint on wavelength usage in data plane, in particular, it requires the same wavelength to be used in both directions. [WSON-Frame] in section 6.1 reports on the implication to GMPLS signaling related to both bi-directionality and Distributed Wavelengths Assignment.

5.1. Possible Solutions for Bidirectional Lightpath

A first classification is using a unique bidirectional LSP (as defined by [RFC3471]) two unidirectional LSPs as per [RFC2205] approach, so possible options are the following:

- o Bidirectional LSP
 1. Current [RFC3471], [RFC3473] co-routed approach. The label distribution is based on Label_Set and Upstream_Label objects. In case of specific constraints such as the same wavelengths in both directions, it may require several signaling attempts using information from the Acceptable_Label_Set received from path error messages.
 2. Using a specific LSP_ATTRIBUTE or a newly defined Upstream_Label_Set object. This mechanism seems to be more efficient (i.e. one signaling attempt) in case of distributed wavelength assignment and same wavelength in both directions.
- o Two Unidirectional LSPs. This solution has been always available as per [RFC3209] however recent work introduces the association concept [RFC4872] and [ASSOC-Info]. Recent transport evolutions [ASSOC-ext] provide a way to associate two unidirectional LSPs as a bidirectional LSP. In line with this, a small extension can make this approach work for the WSON case.

5.2. Bidirectional Lightpath Signaling Procedure

[TO BE UPDATED ACCORDING TO THE BIDIRECTIONAL METHOD CHOSEN FOR WSON either new objects or assoc]

Considering the system configuration mentioned above, it is needed to add a new function into RSVP-TE to support bidirectional lightpath with same wavelength on both directions.

The lightpath setup procedure is described below:

1. Ingress node adds the new type lightpath indication in an LSP_ATTRIBUTES object. It is propagated in the Path message in the same way as that of a Label Set object for downstream;
2. On reception of a Path message containing both the new type lightpath indication in an LSP_ATTRIBUTES object and Label Set object, the receiver of message along the path checks the local LSP database to see if the Label Set TLVs are acceptable on both directions jointly. If there are acceptable wavelengths, then copy the values of them into new Label Set TLVs, and forward the Path message to the downstream node. Otherwise the Path message will be terminated, and a PathErr message with a "Routing problem/Label Set" indication will be generated;
3. On reception of a Path message containing both such a new type lightpath indication in an LSP_ATTRIBUTES object and an Upstream Label object, the receiver MUST terminate the Path message using a PathErr message with Error Code "Unknown Attributes TLV" and Error Value set to the value of the new type lightpath TLV type code;
4. On reception of a Path message containing both the new type lightpath indication in an LSP_ATTRIBUTES object and Label Set object, the egress node verifies whether the Label Set TLVs are acceptable, if one or more wavelengths are available on both directions, then any one available wavelength could be selected. A Resv message is generated and propagated to upstream node;
5. When a Resv message is received at an intermediate node, if it is a new type lightpath, the intermediate node allocates the label to interfaces on both directions and update internal database for this bidirectional same wavelength lightpath, then configures the local ROADM or OXC on both directions.

Except the procedure related to Label Set object, the other processes will be left untouched.

5.3. Backward Compatibility Considerations

Due to the introduction of new processing on Label Set object, it is required that each node in the lightpath is able to recognize the new type lightpath indication Flag carried by an LSP_ATTRIBUTES object, and deal with the new Label Set operation correctly. It is noted that this new extension is not backward compatible.

According to the descriptions in [RFC5420], an LSR that does not recognize a TLV type code carried in this object MUST reject the Path message using a PathErr message with Error Code "Unknown Attributes TLV" and Error Value set to the value of the Attributes Flags TLV type code.

An LSR that does not recognize a bit set in the Attributes Flags TLV MUST reject the Path message using a PathErr message with Error Code "Unknown Attributes Bit" and Error Value set to the bit number of the new type lightpath Flag in the Attributes Flags. The reader is referred to the detailed backward compatibility considerations expressed in [RFC5420].

6. RWA Related

6.1. Wavelength Assignment Method Selection

Routing + Distributed wavelength assignment (R+DWA) is one of the options defined by the [WSOON-Frame]. The output from the routing function will be a path but the wavelength will be selected on a hop-by-hop basis.

Under this hypothesis the node initiating the signaling process needs to declare its own wavelength availability (through a label_set object). Each intermediate node may delete some labels due to connectivity constraints or its own assignment policy. At the end, the destination node has to make the final decision on the wavelength assignment among the ones received through the signaling process.

As discussed in [HZang00] a number of different wavelength assignment algorithms maybe employed. In addition as discussed in [WSOON-Frame] the wavelength assignment can be either for a unidirectional lightpath or for a bidirectional lightpath constrained to use the same lambda in both directions.

A simple TLV could be used to indication wavelength assignment directionality and wavelength assignment method. This would be placed in an LSP_REQUIRED_ATTRIBUTES object per [RFC5420]. The use of a TLV in the LSP required attributes object was pointed out in [Xu].

[TO DO: The directionality stuff needs to be reconciled with the earlier material]

Unique Wavelength: 0 same wavelength in both directions, 1 may use different wavelengths [TBD: shall we use only 1 bit]

Wavelength Assignment Method: 0 unspecified (any), 1 First-Fit, 2 Random, 3 Least-Loaded (multi-fiber). Others TBD.

0										1										2										3									
0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9
Unique WL										WA Method										Reserved																			

7. Security Considerations

This document has no requirement for a change to the security models within GMPLS and associated protocols. That is the OSPF-TE, RSVP-TE, and PCEP security models could be operated unchanged.

However satisfying the requirements for RWA using the existing protocols may significantly affect the loading of those protocols. This makes the operation of the network more vulnerable to denial of service attacks. Therefore additional care maybe required to ensure that the protocols are secure in the WSON environment.

Furthermore the additional information distributed in order to address the RWA problem represents a disclosure of network capabilities that an operator may wish to keep private. Consideration should be given to securing this information.

8. IANA Considerations

TBD. Once finalized in our approach we will need identifiers for such things and modulation types, modulation parameters, wavelength assignment methods, etc...

9. Acknowledgments

Anyone who provide comments and helpful inputs

10. References

10.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC2578] McCloghrie, K., Perkins, D., and J. Schoenwaelder, "Structure of Management Information Version 2 (SMIV2)", STD 58, RFC 2578, April 1999.
- [RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, December 2001.
- [RFC3471] Berger, L., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Functional Description", RFC 3471, January 2003.
- [RFC3473] Berger, L., Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions", RFC 3473, January 2003.
- [RFC3477] Kompella, K. and Y. Rekhter, "Signalling Unnumbered Links in Resource ReSerVation Protocol - Traffic Engineering (RSVP-TE)", RFC 3477, January 2003.
- [RFC5420] Farrel, A., Ed., Papadimitriou, D., Vasseur, J.-P., and A. Ayyangar, " Encoding of Attributes for MPLS LSP Establishment Using Resource Reservation Protocol Traffic Engineering (RSVP-TE)", RFC 5420, February 2006.

10.2. Informative References

- [WSON-CompOSPF] Y. Lee, G. Bernstein, "OSPF Enhancement for Signal and Network Element Compatibility for Wavelength Switched Optical Networks", work in progress: draft-lee-ccamp-wson-signal-compatibility-OSPF.
- [WSON-Frame] Y. Lee, G. Bernstein, W. Imajuku, "Framework for GMPLS and PCE Control of Wavelength Switched Optical Networks", work in progress: draft-bernstein-ccamp-wavelength-switched-03.txt, February 2008.

- [HZang00] H. Zang, J. Jue and B. Mukherjee, "A review of routing and wavelength assignment approaches for wavelength-routed optical WDM networks", *Optical Networks Magazine*, January 2000.
- [Xu] S. Xu, H. Harai, and D. King, "Extensions to GMPLS RSVP-TE for Bidirectional Lightpath the Same Wavelength", work in progress: draft-xu-rsvp-te-bidir-wave-01, November 2007.
- [Winzer06] Peter J. Winzer and Rene-Jean Essiambre, "Advanced Optical Modulation Formats", *Proceedings of the IEEE*, vol. 94, no. 5, pp. 952-985, May 2006.
- [G.959.1] ITU-T Recommendation G.959.1, *Optical Transport Network Physical Layer Interfaces*, March 2006.
- [G.694.1] ITU-T Recommendation G.694.1, *Spectral grids for WDM applications: DWDM frequency grid*, June 2002.
- [G.694.2] ITU-T Recommendation G.694.2, *Spectral grids for WDM applications: CWDM wavelength grid*, December 2003.
- [G.Sup43] ITU-T Series G Supplement 43, *Transport of IEEE 10G base-R in optical transport networks (OTN)*, November 2006.
- [RFC4427] Mannie, E., Ed., and D. Papadimitriou, Ed., "Recovery (Protection and Restoration) Terminology for Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4427, March 2006.
- [RFC4872] Lang, J., Rekhter, Y., and Papadimitriou, D., "RSVP-TE Extensions in Support of End-to-End Generalized Multi-Protocol Label Switching (GMPLS) Recovery", RFC 4872,
- [ASSOC-Info] Berger, L., Faucheur, F., and A. Narayanan, "Usage of The RSVP Association Object", draft-ietf-ccamp-assoc-info-00 (work in progress), October 2010.
- [ASSOC-Ext] Zhang, F., Jing, R., "RSVP-TE Extension to Establish Associated Bidirectional LSP", draft-zhang-mpls-tp-rsvp-te-ext-associated-lsp-03 (work in progress), February 2011.

Author's Addresses

Greg M. Bernstein (editor)
Grotto Networking
Fremont California, USA

Phone: (510) 573-2237
Email: gregb@grotto-networking.com

Nicola Andriolli
Scuola Superiore Sant'Anna, Pisa, Italy
Email: nick@sssup.it

Alessio Giorgetti
Scuola Superiore Sant'Anna, Pisa, Italy
Email: a.giorgetti@sssup.it

Lin Guo
Key Laboratory of Optical Communication and Lightwave Technologies
Ministry of Education
P.O. Box 128, Beijing University of Posts and Telecommunications,
P.R.China
Email: guolintom@gmail.com

Hiroaki Harai
National Institute of Information and Communications Technology
4-2-1 Nukui-Kitamachi, Koganei,
Tokyo, 184-8795 Japan

Phone: +81 42-327-5418
Email: harai@nict.go.jp

Yuefeng Ji
Key Laboratory of Optical Communication and Lightwave Technologies
Ministry of Education
P.O. Box 128, Beijing University of Posts and Telecommunications,
P.R.China
Email: jyf@bupt.edu.cn

Daniel King
Old Dog Consulting

Email: daniel@olddog.co.uk

Young Lee (editor)
Huawei Technologies

1700 Alma Drive, Suite 100
Plano, TX 75075
USA

Phone: (972) 509-5599 (x2240)
Email: ylee@huawei.com

Sugang Xu
National Institute of Information and Communications Technology
4-2-1 Nukui-Kitamachi, Koganei,
Tokyo, 184-8795 Japan

Phone: +81 42-327-6927
Email: xsg@nict.go.jp

Giovanni Martinelli
Cisco
Via Philips 12
20052 Monza, IT

Phone: +39 039-209-2044
Email: giomarti@cisco.com

Intellectual Property Statement

The IETF Trust takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in any IETF Document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights.

Copies of Intellectual Property disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a 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 any standard or specification contained in an IETF Document. Please address the information to the IETF at ietf-ipr@ietf.org.

Disclaimer of Validity

All IETF Documents and the information contained therein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY, THE IETF TRUST AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION THEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Acknowledgment

Funding for the RFC Editor function is currently provided by the Internet Society.

Internet Engineering Task Force
Internet-Draft
Intended status: Informational
Expires: April 4, 2011

M. Kattan, Ed.
G. Martinelli
D. Bianchi
Cisco
N. Ibrahim
MOT/OGERO
March 2011

WSON Wavelength Property Information
draft-kattan-wson-property-01

Abstract

Wavelength Switched Optical Network will extend GMPLS protocols to to manage wavelength across DWDM optical networks. In many situations the control plane needs to know additional information regarding wavelengths. The current proposal identify a way to carry some property information along with wavelength information. Control plane can leverage the knowledge of such properties during its operations.

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on April 4, 2011.

Copyright Notice

Copyright (c) 2010 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents

carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

- 1. Introduction 3
 - 1.1. Requirements Language 3
- 2. Scenarios 3
- 3. Lambda Properties Definitions 5
- 4. Lambda Properties Encoding 6
 - 4.1. OSPF Extensions 6
 - 4.2. RSVP Extensions 7
- 5. Acknowledgements 8
- 6. IANA Considerations 8
- 7. Security Considerations 8
- 8. References 8
 - 8.1. Normative References 8
 - 8.2. Informative References 8
- Authors' Addresses 9

1. Introduction

One the current Generalized MPLS (GMPLS) evolutions is toward the Wavelength Switched Optical Networks (WSON) as described in [I-D.ietf-ccamp-rwa-wson-framework]. A related work is defined within [I-D.ietf-ccamp-gmpls-g-694-lambda-labels] defining the GMPLS label in a format suitable for Lambda Switched Capable (LSC equipments).

Today's WSON networks are implemented through DWDM technologies and they treat all light paths as equal regardless of the type of data, bandwidth and mission criticality of the traffic it is carrying.

This draft suggests the introduction of some properties like prioritizing light paths for scenarios such as restoration, fiber congestion and resource contention. This could be achieved in assigning properties information to each light path. Following sections will describe some scenarios where such information will be useful. How those information are assigned is out of the scope of this draft.

1.1. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

2. Scenarios

The following list identify several scenarios occurring in operating WSON networks where some wavelength information will help. Note that these scenarios are triggered by the availability of new reconfigurable equipments allowing new level of flexibility within DWDM networks.

Example of this hardware would be multi-degree Reconfigurable Optical Add Drop Multiplexers or ROADMs to support mesh DWDM networks. Fiber 1 is an example of a meshed DWDM network where multiple light paths are being set up to and from node C.

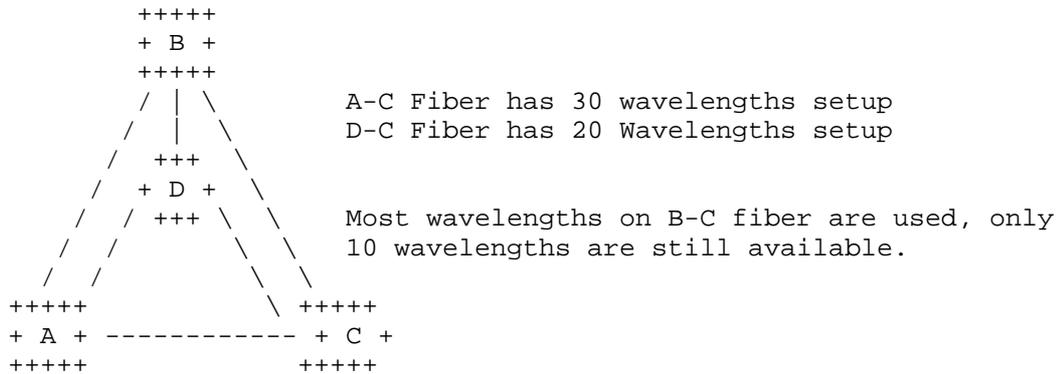


Figure 1

(a) Prioritize then Restore

With the reference to Figure 2 we can consider a dual fiber cut on the path A-C and D-C. A lambda prioritization might be used to ensure high priority light paths be served first. This will ensure both a faster restoration time compared to other channels as well as the ability of high priority light paths to grab first (before other lower priority light paths) the available resources on the working fiber.

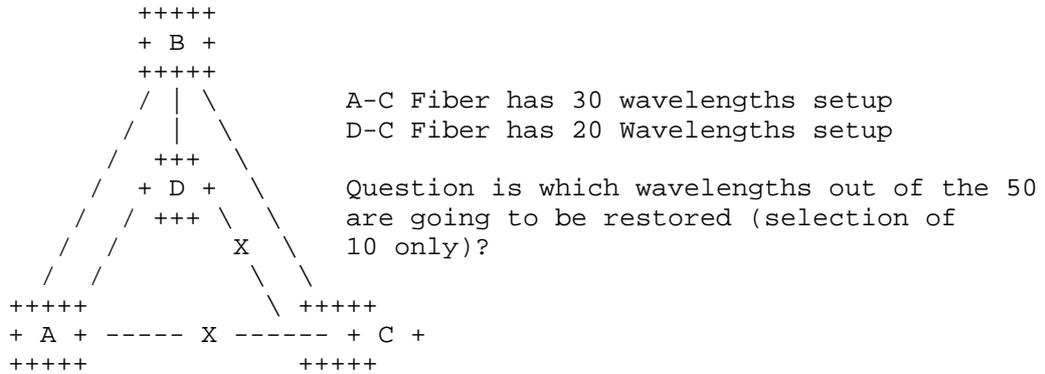


Figure 2

(b) Revertive Operation

In this scenario, a fiber is being restored and hence having a high priority light paths restored first might or might not be desirable. Setting a revertive or not revertive option would be useful in this scenario. Moreover, in the event of multiple fiber cuts with only one fiber restored as an example, prioritizing light paths will ensure higher priority traffic will get the best service as well as up time once the WSON restoration mechanism kicks in. Other possibilities include defining some others lambda properties like a "no not restore bit" or "Wait time to restore" to allow the control plane operates according to different restoration strategies.

(c) Network Optimization

Similar to revertive operation, prioritizing light paths will also be useful in network optimization. High priority traffic will always get the option to ride on the best available fiber path. Also high priority light path could be provided with the option to get the best performance OI parameters to chose from.

(d) Service Level Agreement support

This could be useful for DWDM service providers where light paths are tagged with different parameters so that to create a desirable and configurable level of SLA. This SLA could be derived from bandwidth (100G, 40G and 10G), traffic type (TDM vs IP/Eth or FC payload) or just a network management defined requirement.

(e) Resource Contention

In the event of one or multiple fiber cuts, we could be faced with a situation whereby the number of light paths to be restored is larger than the available light path resources on the working fiber (see Figure 2 above). Having light paths prioritization together with a wait-time-to-restore will ensure that the high priority traffic will be served first and hence will be able to grab the available resources first.

3. Lambda Properties Definitions

This section provide a list of wavelengths properties that worths to include in a control plane.

Priority. This information will allow a preferred treatment to a

wavelength with higher priority.

Do Not Restore. If this information will not restore try to restore the wavelength after a failure.

Lambda-Timer. This timer can be used as either hold-off-timer or wait-time-to-restore to control how the wavelength is managed during a protection and/or restoration actions.

4. Lambda Properties Encoding

The lambda priority will be encoded over three bits. There are different encoding possibility depending on the protocol used to distribute this information over the control plane.

It worth noting that GMPLS extension in [RFC4202] and [RFC4203] already define LSP priority bandwidth within Interface Switching Capability Descriptor sub-TLV. This concept however does not suffice for WSON LSP for the scenario represented above.

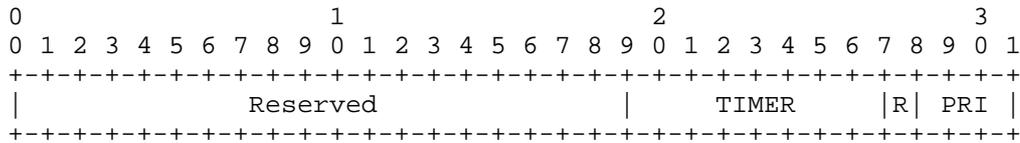


Figure 3

The 3 bits PRI field represent the lambda priority encoding. Zero means no priority, Seven means maximum priority

R is the "Do not restore bit". If set the wavelength will be exclude from any kind restoration

Timer is a timer to delay restoration/protection actions on the wavelengths. 8 bits with a granularity of 1 second will allow up to 255 seconds of delay on restoration.

4.1. OSPF Extensions

In order to improve the WSON path computation it make sense to add such information through the chosen IGP. Current WSON proposal are available for OSPF-TE extentions.

Document [I-D.ietf-ccamp-rwa-wson-encode] report the information on how to encode Dynamic Link Information through the label set specification.

Efficient encoding through a Link Attributes shall be identified. An initial proposal may looks like the label set attribute as explained in the following picture. The wavelength property encoding will be a sub-TLV (type TBD) of the link TLV. The set of

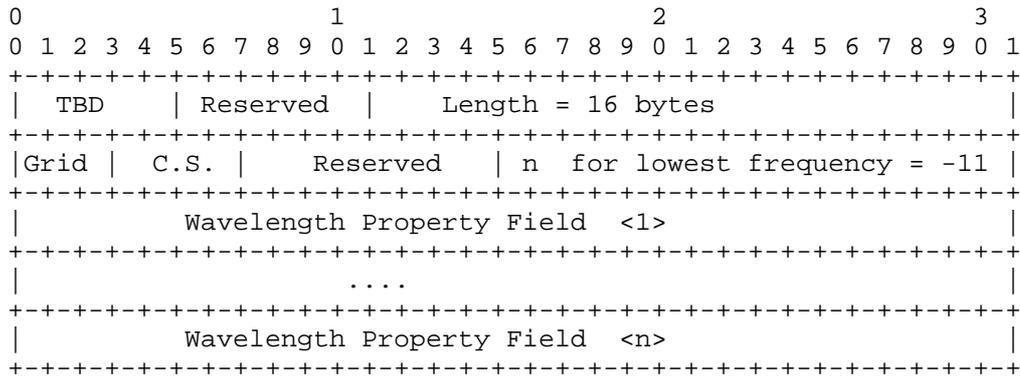


Figure 4

Where:

TBD: is the sub-TLV type (to be defined)

The Grid provide the current WSON wavelength encoding in use and must match with the label set defined in [I-D.ietf-ccamp-general-constraint-encode].

A list of Wavelength property field, defined n Figure 4 in an order they match with the last label set advertised.

4.2. RSVP Extensions

WSON signalling extentions are reported through [draft-bernstein-ccamp-wson-signaling-07]. In addition to this a new LSP_ATTRIBUTES as defined in [RFC5420] will be required to carry the lambda priority information.

A new LSP_ATTRIBUTE shall include the Wavelength Property Field as defined in Figure 4

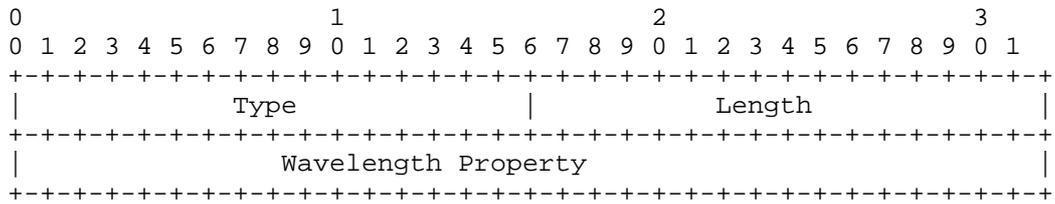


Figure 5

In this case the only one wavelenght property object will be required.

5. Acknowledgements

6. IANA Considerations

This memo includes no request to IANA.

All drafts are required to have an IANA considerations section (see the update of RFC 2434 [I-D.narten-iana-considerations-rfc2434bis] for a guide). If the draft does not require IANA to do anything, the section contains an explicit statement that this is the case (as above). If there are no requirements for IANA, the section will be removed during conversion into an RFC by the RFC Editor.

7. Security Considerations

All drafts are required to have a security considerations section. See RFC 3552 [RFC3552] for a guide.

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.

8.2. Informative References

[I-D.ietf-ccamp-general-constraint-encode]
 Bernstein, G., "General Network Element Constraint Encoding for GMPLS Controlled Networks",

draft-ietf-ccamp-general-constraint-encode-03 (work in progress), October 2010.

[I-D.ietf-ccamp-gmpls-g-694-lambda-labels]
Otani, T., Rabbat, R., Shiba, S., Guo, H., Miyazaki, K., Caviglia, D., Li, D., and T. Tsuritani, "Generalized Labels for Lambda-Switching Capable Label Switching Routers", draft-ietf-ccamp-gmpls-g-694-lambda-labels-07 (work in progress), April 2010.

[I-D.ietf-ccamp-rwa-wson-encode]
Bernstein, G., "Routing and Wavelength Assignment Information Encoding for Wavelength Switched Optical Networks", draft-ietf-ccamp-rwa-wson-encode-05 (work in progress), July 2010.

[I-D.ietf-ccamp-rwa-wson-framework]
Bernstein, G., Lee, Y., and W. Imajuku, "Framework for GMPLS and PCE Control of Wavelength Switched Optical Networks (WSON)", draft-ietf-ccamp-rwa-wson-framework-07 (work in progress), October 2010.

[I-D.narten-iana-considerations-rfc2434bis]
Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", draft-narten-iana-considerations-rfc2434bis-09 (work in progress), March 2008.

[RFC2629] Rose, M., "Writing I-Ds and RFCs using XML", RFC 2629, June 1999.

[RFC3552] Rescorla, E. and B. Korver, "Guidelines for Writing RFC Text on Security Considerations", BCP 72, RFC 3552, July 2003.

Authors' Addresses

Moustafa Kattan (editor)
Cisco
DUBAI, 500321
UNITED ARAB EMIRATES

Phone: +14085275101
Email: mkattan@cisco.com

Giovanni Martinelli
Cisco
Italy

Phone: +39 039 209 2044
Email: giomarti@cisco.com

David Bianchi
Cisco
Italy

Phone: +39 039 2091505
Email: davbianc@cisco.com

Nazih Ibrahim
MOT/OGERO
Lebabon

Phone: +9613327863
Email: nibrahim@ogero.com

CCAMP Working Group
Internet-Draft
Intended status: Proposed Standard
Expires: September 15, 2011

Khuzema Pithewan
Rajan Rao
Ashok Kunjidhapatham
Biao Lu
Mohit Misra
Infinera
Lyndon Ong
Ciena
March 14, 2011

Signaling Extensions for Generalized MPLS (GMPLS) Control of
G.709 Optical Transport Networks
draft-khuzema-ccamp-gmpls-signaling-g709-01.txt

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at
<http://www.ietf.org/lid-abstracts.html>

The list of Internet-Draft Shadow Directories can be accessed at
<http://www.ietf.org/shadow.html>

Copyright and License Notice

Copyright (c) 2011 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of

the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Abstract

As OTN network capabilities continue to evolve, there is an increased need to support GMPLS control for the same. [RFC4328] introduced GMPLS signaling extensions for supporting early version of G.709 [G.709-v1]. The basic routing considerations from signaling perspective is also specified in [RFC4328].

The recent revision of ITU-T Recommendation G.709 [G.709-v3] and [GSUP.43] have introduced new ODU containers (both fixed and flexible) and additional ODU multiplexing capabilities, enabling support for optimal service aggregation.

This document extends [RFC4328] to provide GMPLS signaling support for the new OTN capabilities defined in [G.709-v3] and [GSUP.43]. The signaling extensions described in this document caters to ODU layer switching only. Optical Channel Layer switching considerations in [RFC4328] are not modified in this document.

Table of Contents

1. Introduction	4
2. Conventions used in this document	5
3. Overview of GMPLS Signaling Extensions required for the Evolving OTN	5
4. Extensions to G.709 Traffic Parameters	6
4.1. Usage of Bit_Rate and Tolerance for ODUflex Service	7
5. New Generalized Label Format	8
5.1 Multi-stage Label	8
5.2. Label format for NVC or Multiplier > 1	10
6. Usage of Multi-stage Label	10
7. Label Distribution Rules	12
8. Interoperability Considerations	13
9. Examples	14
10. Security Considerations	16
11. IANA Considerations	16
12. References	16
12.1. Normative References	16
12.2. Informative References	17
13. Acknowledgements	17
Author's Addresses	17
Appendix A: Abbreviations & Terminology	18
Appendix B : RFC4328 and G.709v3	20

1. Introduction

Generalized Multi-Protocol Label Switching (GMPLS) [RFC3945] extends MPLS from supporting Packet Switching Capable (PSC) interfaces and switching to include support of four new classes of interfaces and switching: Layer-2 Switching (L2SC), Time-Division Multiplex (TDM), Lambda Switch (LSC), and Fiber-Switch (FSC) Capable. A functional description of the extensions to MPLS signaling that are needed to support these new classes of interfaces and switching is provided in [RFC3471].

ITU-T Recommendations G.709 and G.872 provide specifications for OTN interface and network architecture respectively. As OTN network capabilities continue to evolve; there is an increased need to support GMPLS control for the same.

GMPLS signaling extensions to support [G.709-v1] OTN interfaces are specified in [RFC4328]. Further extensions are required to support the new capabilities introduced since [G.709-v1]. Following are the features added in OTN since the first version [G.709-v1].

(a) OTU Containers:

Pre-existing Containers: OTU1, OTU2 and OTU3

New Containers introduced in [G.709-v3]: OTU2e and OTU4

New Containers introduced in [GSUP.43]: OTU1e, OTU3e1 and OTU3e2

(b) Fixed ODU Containers:

Pre-existing Containers: ODU1, ODU2 and ODU3

New Containers introduced in [G.709-v3]: ODU0, ODU2e and ODU4

New Containers introduced in [GSUP.43]: ODU1e, ODU3e1 and ODU3e2

(c) Flexible ODU Containers:

ODUflex for CBR and GFP-F mapped services. ODUflex uses 'n' number of OPU Tributary Slots where 'n' is different from the number of OPU Tributary Slots used by the Fixed ODU Containers.

(d) Tributary Slot Granularity:

OPU2 and OPU3 support two Tributary Slot Granularities: (i) 1.25Gbps and (ii) 2.5Gbps.

(e) ODU Multiplexing Hierarchy:

Multi-stage multiplexing of LO-ODUs into HO-ODU is supported. Also, multiplexing could be heterogeneous (meaning LO-ODUs of different rates can be multiplexed into the same HO-ODU).

OTN networks support switching at two layers: (i) ODU Layer - TDM Switching and (ii) OCH Layer - Lambda (LSC) Switching. The nodes on the network may support one or both the switching types. When

multiple switching types are supported MLN based routing [RFC5339] is assumed.

This document extends [RFC4328] to provide GMPLS signaling support for the new OTN capabilities defined in [G.709-v3] and [GSUP.43]. This complies with the requirements outlined in the framework document [G.709-FRAME]. The signaling extensions described in this document caters to ODU layer switching only. Optical Channel Layer switching considerations in [RFC4328] are not modified in this document.

Following are the extensions described in this document:

(i) G.709 Traffic Parameters defined in [RFC4328] is extended to include Bit Rate (in bytes/second) and Tolerance (in ppm) fields for supporting ODUflex service.

(ii) New Generalized Label Format is introduced to provide compact encoding of Tributary Slot information and support multi-stage ODU multiplexing.

2. Conventions used in this document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document is to be interpreted as described in RFC-2119 [RFC2119].

In addition, the reader is assumed to be familiar with the terminology used in ITU-T [G.709-v3], [G.872] and [GSUP.43], as well as [RFC4201] and [RFC4203].

3. Overview of GMPLS Signaling Extensions required for the Evolving OTN

The GMPLS signaling extensions introduced in [RFC4328] cover OTN switching requirement pertaining to [G.709-v1]. The signaling objects defined in [RFC4328] need to be further extended to cover the new capabilities added to OTN since the first version of G.709 [G.709-v1]. A brief overview of the extensions required are captured below:

(a) Support for the new ODU containers

The new ODU containers added since [G.709-v1] are listed in the section-1. SignalType attribute defined in [RFC4328] need to be extended to cover the new signal types. This is captured in [OSPF-EXTN-FOR-OTN].

(b) Support for ODUflex

Unlike the other ODUj signal types, ODUflex requires an user specified bit-rate (together with a Tolerance value) to be mapped to 'n' TSs of an higher-order container. Even within the same Tributary Slot Granularity, the Tributary Slot size varies among the ODU container of different rate. This results in ODUflex service of certain bit-rate and tolerance requiring different number of TSs on different higher order ODU containers. The present way of specifying bandwidth requirement (via NMC field in G.709 Traffic Parameters) will not work for ODUflex. G.709 Traffic Parameters object need to be extended to include Bit-Rate (in bytes/sec) and Tolerance (in ppm) fields as well.

(c) Support for ODU multiplexing hierarchy

The G.709 Traffic Parameter and Generalized Label Format defined in [RFC4328] supports single stage multiplexing only. A new Generalized Label Format need to be introduced to support specification of multi-stage label.

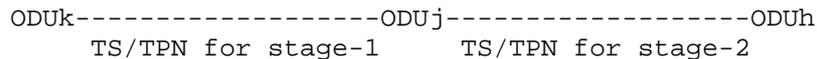


Figure-1: Multi-stage Label

(d) Support for different OPU Tributary Slot Granularities

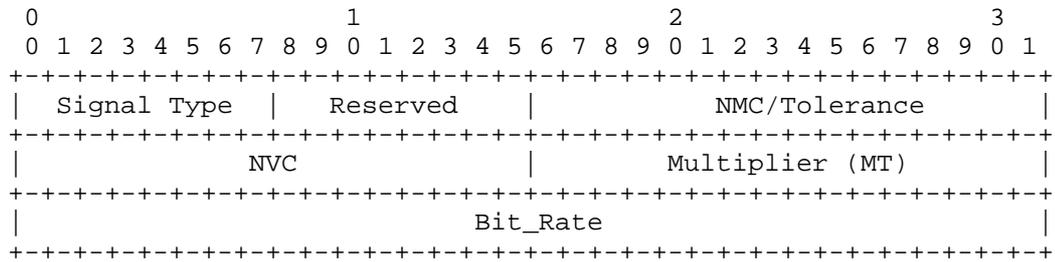
The G.709 Traffic Parameters and Generalized Label Format defined in [RFC4328] supports 2.5Gbps Tributary Slot Granularity only. With [G.709-v3], two types of tributary slots are supported - viz., 1.25Gbps and 2.5Gbps. The Generalized Label Format need to be equipped with Tributary Slot Type indicator to facilitate interpretation of the encoded TS information.

(e) Exchange of Tributary Port Number

A Tributary Port Number (TPN) in MSI field of OPU-OH is used to correlate the TSs used for mapping a LO-ODU on a HO-ODU. This needs to be exchanged along with the Label such that each neighbor on a span knows the TPN value to expect for a given ODUj mapping. This applies to each stage associated with a multi-stage label. The Generalized Label Format needs to be extended to include TPN value for each stage of multiplexing.

4. Extensions to G.709 Traffic Parameters

G.709 Traffic Parameters defined in [RFC4328] is extended to include additional fields in support of ODUflex service as explained in the previous section. The modified object format is captured below:



Signal Type

As explained in the previous section, Signal Type attribute needs to be extended to cover the new ODU containers defined in more recent G.709 specification [G.709-v3].

Value	Type
4	ODU4 (100Gbps)
5	ODU0 (1.25Gbps)
10	ODUflex
11	ODU1e (10Gbps Ethernet [GSUP.43])
12	ODU2e (10Gbps Ethernet)
13	ODU3e1 (40Gbps Ethernet [GSUP.43])
14	ODU3e2 (40Gbps Ethernet [GSUP.43])
15-255	Reserved (for future)

NMC/Tolerance

This field is redefined from the original definition in [RFC4328]. NMC field defined in [RFC4328] can not be fixed value for an end-to-end circuit involving dissimilar OTN link types. For example, ODU2e requires 9 TS on ODU3 and 8 TS on ODU4. Usage of NMC field is deprecated and should be used only with [RFC4328] generalized label format for backwards compatibility reasons.

For the new generalized label format as defined in this document this field is interpreted as Tolerance. The unit of tolerance is ppm and is encoded as unsigned integer. For signal types other than ODUflex, Tolerance field should be coded as 0.

Bit_Rate

Bit_Rate is used when signal Type is ODUflex. For all the other signal types, this field should be coded as zero.

4.1. Usage of Bit_Rate and Tolerance for ODUflex Service

Bit_Rate and Tolerance are used together to compute number of Tributary slots required for ODUFlex(CBR) traffic on a given higher order ODU container. The computation of Number of Tributary Slot (n) is as follows.

$$n = \frac{\text{Ceiling of Bit_Rate} * (1 + \text{Tolerance})}{\text{ODTUK.ts nominal bit rate} * (1 - \text{HO OPUk bit rate tolerance})}$$

5. New Generalized Label Format

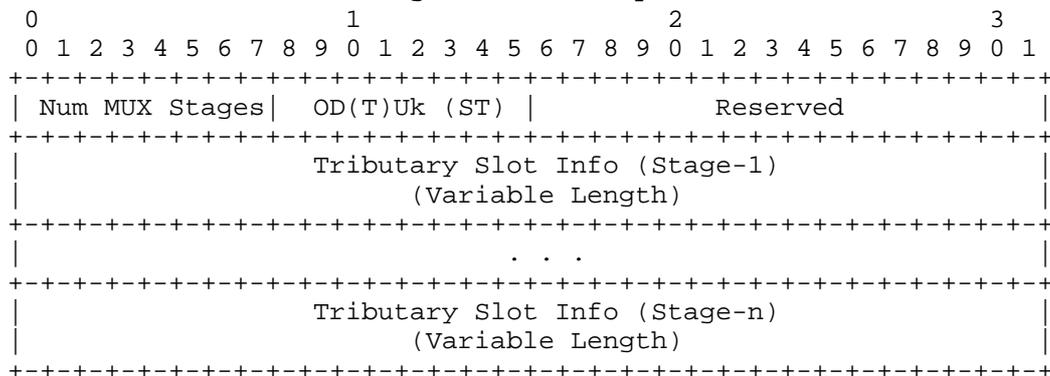
As explained in section 3, the Generalized Label format defined in [RFC4328] can not accommodate the new features added in [G.709v3]. Further the label format as defined in [RFC4328] is not scalable for large number of Tributary Slots (at 1.25G granularity) associated with bigger containers such as ODU3 and ODU4.

The Generalized Label for G.709 may contain one or more multi-stage Label.

5.1 Multi-stage Label

A multi-stage label includes TS and TPN information for all the stages of a multi-stage multiplexing hierarchy.

The format of a multi-stage label is explained below.



Num MUX Stages

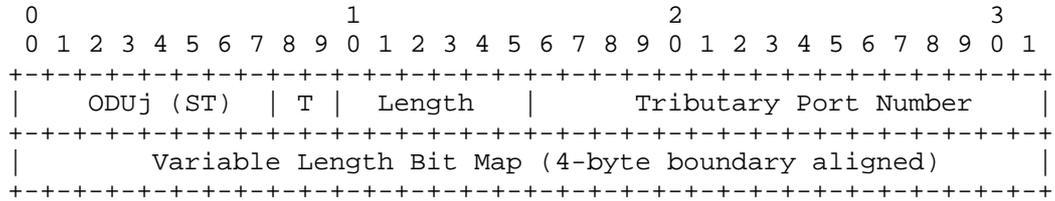
This field indicates the number of multiplexing stages specified by the label.

OD(T)Uk

This field encodes the signal type of HO OD(T)Uk container.

Tributary Slot Info

Tributary Slot Information for a single stage is encoded as follows.



ODUj

This field indicates the signal type of a LO-ODU being multiplexed into its immediate HO-ODU.

T

This is a 2 bit field, which defines the granularity of tributary slots for this multiplexing stage. It can take following values

T field	TS Granularity type
0	1.25Gbps
1	2.5Gbps
2-3	Reserved (for future use)

Length

This field indicates the number of valid Bits in the of Bit Map excluding the filler bits.

Tributary Port Number(TPN)

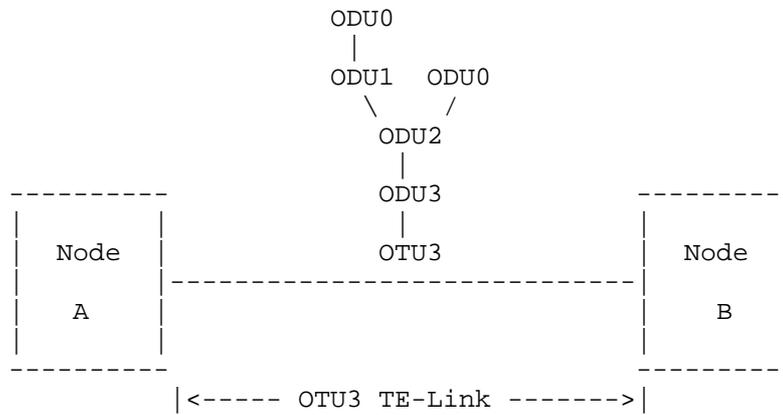
This field is encoded with TPN value assigned for a ODTUjk or ODTUk.ts on a OPUk. TPN assignment could be fixed or flexible.

For fixed TPN assignment scheme, TPN value need not be specified. In this case, TPN value should be coded as 0xFFFFFFFF.

For flexible TPN assignment scheme, TPN value should contain the assigned logical value. Not all the bits of TPN are used. Only a subset of bits are used depending on the ODTU type.

Bit Map

Multi-stage Label helps in implicit creation of ODU3 and ODU2 layers as part of ODU1 LSP setup and thus eliminates the need for the creation of the FA LSPs/TELinks.



Label Format:

Stage-1: ODU3<-ODU2/TPN/Trib Slots

Stage-2: ODU2<-ODU1/TPN/Trib Slots

Figure-2: Multi-stage Label on OTUk Link

Example-2:

Assume on an ODU3 FA LSP/TE-Link (B-C-D), signaling of ODU1 LSP requires termination of ODU2. Multi-stage Label helps in implicit creation of ODU2 layer as part of ODU1 LSP setup (A-B-D-E).

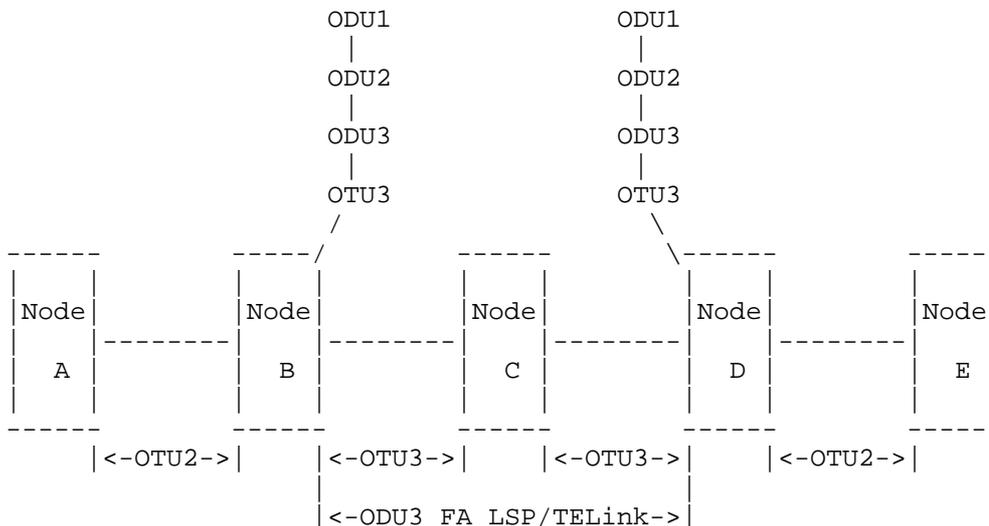


Figure-3: Multi-stage Label on ODUk Link

Note: Multi-stage Label is NOT intended to facilitate the creation of FA-LSP or Hierarchical LSP. It is basically used to eliminate the need for FA-LSP in some obvious scenarios.

7. Label Distribution Rules

This document does not change the existing label distribution procedures defined in [RFC4328] except that the new ODU label should be processed as follows.

A. Sending Side

When Generalized Label Request is received on given node for setting up an ODU LSP from its upstream neighbor, it reserves the bandwidth required for the ODU Layer being switched and also the terminating HO-ODUs layers involved. It sends upstream label and suggested label (if applicable) to the downstream node and downstream label via PATH Message and downstream label to the upstream node via RESV Message.

Note that Label can also be explicitly specified by source node.

The encoding of Generalized Label is as follows:

Case-1: ODUk mapping into OTUk

Number of MUX stages = 0

Tributary Slot information is not included.

Case-2: ODUj mux into ODUk

Number of MUX Stages = 1.

Stage-1: Length = <number of TSs on ODUk>.

TPN = <specified as per Section 5>

TS BitMap = <TSs reserved for ODUj are set to 1>

Case-3 ODUh mux into ODUj into ODUk

Number of MUX Stages = 2.

Stage-1: Length = <number of TSs on ODUk>.

TPN = <specified as per Section 5>

TS BitMap = <TSs reserved for ODUj are set to 1>

Stage-2: Length = <number of TSs on ODUj>.

TPN = <specified as per Section 5>

TS BitMap = <TSs reserved for ODUh are set to 1>

B. Receiving Side

The decoding of the Generalized Label is as follows:

Case-1: ODUk mapping into OTUk

For ODUk to OTUk mapping, the Tributary Slot Information is not expected.

Case-2: ODUj mux into ODUk

For ODUj to ODUk multiplexing, one MUX stage Label is expected. The node extracts the Bit Map field in Tributary Slot Info using the Length field. The position of Bit in the Bitmap interpreted as the Tributary Slot Number. The value stored in the bit indicates if it is reserved for the ODUj.

Case-3: ODUh mux into ODUj into ODUk

For ODUh mux into ODUj into ODUk, two MUX stage Label is expected. Each stage is further decoded as explained in case-2 above.

8. Interoperability Considerations

The neighbor nodes on a TE-Link span should exchange the signaling stack versions (via some link discovery mechanism) in order to determine the Generalized Label Format to use.

In the following example, Switch B and C are running the newer version of signaling stack (that support the new G.709 Traffic Parameters and Generalized Label Format) while Switch A is running the older version.

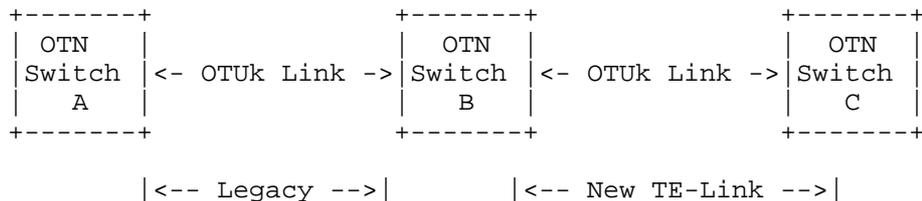


Figure-4: OTUk TE-Link

Link A-B: G.709-v1 version (2001) based OTUk link
TSG: 2.5G;
Label format: as per RFC 4328

Link B-C: G.709-v3 version based OTUk link (12/09)
TSG: 1.25G;
Label format: new label format proposed in this draft.

For an ODU2 connection going from A-C,
On link A-B : NMC is set to 4 & [RFC4328] label format is used.

On link B-C : NMC is not used & new label format is used.

9. Examples

Example-1 : ODUj LSP over OTUk Links

Consider the network topology shown in the Figure-5 below:

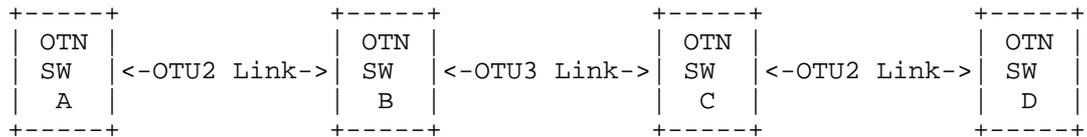


Figure-5: OTN Signaling Example

Assumptions:

(1) ODU2 links between OTN-Switches A & B and C & D support 1.25Gbps TS Granularity.

(2) ODU3 link between OTN-Switches B & C supports TS Granularity of 2.5Gbps only. Hence, ODU0 switching on this link is possible only through ODU3-ODU2-ODU0 or ODU3-ODU1-ODU0 multiplexing hierarchies.

G.709 Traffic Parameters and Generalized Label for ODU0 LSP from node A to D is captured below:

A. G.709 Traffic Parameters

```

Signal Type = ODU0
NMC/Tolerance = 0    // NMC is not used.
NVC = 0
Multiplier (MT) = 1
Bit_Rate = 0
    
```

B. Generalized Label Format:

	A to B	B to C	C to D
# of Stages	1	2	1
Stage-1	ODU2<--ODU0 TSG = 1.25G #TSs = 8 TPN = <1..8> BMap = 4bytes	ODU3<--ODU2 TSG = 2.5G #TSs = 16 TPN = <1..4> BMap = 4bytes	ODU2<--ODU0 TSG = 1.25G #TSs = 8 TPN = <1..8> BMap = 4bytes
Stage-2	N/A	ODU2<--ODU0 TSG = 1.25G #TSs = 8 TPN = <1..8> BMap = 4bytes	N/A

Example 2: ODUj LSP over ODUk FA-LSP/TE-Link

Refer to Figure-3 in section 6. The G.709 Traffic Parameters and Generalized Label for ODU1 LSP from Node A to E is captured below:

A. G.709 Traffic Parameters:

Signal Type = ODU1
 NMC/Tolerance = 0 // NMC is not used.
 NVC = 0
 Multiplier (MT) = 1
 Bit_Rate = 0

B. Generalized Label Format:

	A to B	B to D	D to E
# of Stages	1	2	1
Stage-1	ODU2<--ODU1 TSG = 1.25G #TSs = 8 TPN = <1..4> BMap = 4bytes	ODU3<--ODU2 TSG = 2.5G #TSs = 16 TPN = <1..4> BMap = 4bytes	ODU2<--ODU1 TSG = 1.25G #TSs = 8 TPN = <1..4> BMap = 4bytes
Stage-2	N/A	ODU2<--ODU1 TSG = 1.25G #TSs = 8 TPN = <1..4> BMap = 4bytes	N/A

10. Security Considerations

There are no additional security implications to Signaling protocol due to the extensions captured in this document.

11. IANA Considerations

TBD

12. References

12.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels".
- [RFC3630] Katz, D., Kompella, K., and D. Yeung, "Traffic Engineering (TE) Extensions to OSPF Version 2", RFC 3630
- [RFC3471] Berger, L., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Functional Description", RFC 3471, January 2003.
- [RFC4201] Kompella, K., Rekhter, Y., and L. Berger, "Link Bundling in MPLS Traffic Engineering (TE)"
- [RFC4203] Kompella, K. and Y. Rekhter, "OSPF Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)"
- [RFC4204] Lang, J., Ed., "Link Management Protocol (LMP)", RFC 4204, October 2005.
- [RFC4328] Papadimitriou, D., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Extensions for G.709 Optical Transport Networks Control", RFC 4328, January 2006.
- [RFC5339] Le Roux, JL. and D. Papadimitriou, "Evaluation of Existing GMPLS Protocols against Multi-Layer and Multi-Region Networks (MLN/MRN)", RFC 5339, September 2008.
- [VCAT-LCAS] G. Bernstein (ed.), D. Caviglia, R. Rabbat and H. van Helvoort, "Operating Virtual Concatenation (VCAT) and the Link Capacity Adjustment Scheme (LCAS) with Generalized Multi-Protocol Label Switching (GMPLS)", draft-bernstein-ccamp-gmpls-vcat-lcas-11.txt, March 09, 2011

[G.709-v3] ITU-T, "Interfaces for the Optical Transport Network (OTN)", G.709 Recommendation, December 2009.

12.2. Informative References

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

[G.709-v1] ITU-T, "Interface for the Optical Transport Network (OTN)," G.709 Recommendation (and Amendment 1), February 2001 (October 2001).

[G.872] ITU-T, "Architecture of optical transport networks", November 2001 (11 2001).

[G.709-FRAME] F. Zhang, D. Li, H. Li, S. Belotti, "Framework for GMPLS and PCE Control of G.709 Optical Transport Networks", draft-zhang-ccamp-gmpls-g709-framework-02, work in progress.

[WSON-FRAME] Y. Lee, G. Bernstein, W. Imajuku, "Framework for GMPLS and PCE Control of Wavelength Switched Optical Networks (WSON)", draft-ietf-ccamp-rwa-wson-framework, work in progress.

[OSPF-EXTN-FOR-OTN] S. Bardalai, R. Rao, A. Kunjidhapatham, K. Pithewan, "OSPF TE Extensions for GMPLS Control of G.709 Optical Transport Networks", draft-ashok-ccamp-gmpls-ospf-g709-02, work in progress.

13. Acknowledgements

Authors would like to thank Lou Berger, Steve Balls and Radhakrishna Valiveti for review comments and suggestions.

Author's Addresses

Khuzema Pithewan
Infinera Corporation
169, Java Drive
Sunnyvale, CA-94089, USA
Email: kpithewan@infinera.com

Mohit Misra
Infinera Corporation
169, Java Drive

Sunnyvale, CA-94089, USA
Email: mmisra@infinera.com

Rajan Rao
Infinera Corporation
169, Java Drive
Sunnyvale, CA-94089, USA
Email: rrao@infinera.com

Ashok Kunjidhpatham
Infinera Corporation
169, Java Drive
Sunnyvale, CA-94089, USA
Email: akunjidhpatham@infinera.com

Biao Lu
Infinera Corporation
169, Java Drive
Sunnyvale, CA-94089, USA
Email: blu@infinera.com

Lyndon Ong
Ciena
10480 Ridgeview Court
Cupertino, CA 95014, USA
EMail: lyong@ciena.com

Appendix A: Abbreviations & Terminology

A.1 Abbreviations:

CBR	Constant Bit Rate
GFP	Generic Framing Procedure
HO-ODU	Higher Order ODU
LSC	Lambda Switch Capable
LSP	Label Switched Path
LO-ODU	Lower Order ODU
ISCD	Interface Switch Capability Descriptor
OCC	Optical Channel Carrier
OCG	Optical Carrier Group
OCh	Optical Channel (with full functionality)
OChr	Optical Channel (with reduced functionality)
ODTUG	Optical Data Tributary Unit Group
ODU	Optical Channel Data Unit
OMS	Optical Multiplex Section
OMU	Optical Multiplex Unit
OPS	Optical Physical Section

OPU	Optical Channel Payload Unit
OSC	Optical Supervisory Channel
OTH	Optical Transport Hierarchy
OTM	Optical Transport Module
OTN	Optical Transport Network
OTS	Optical Transmission Section
OTU	Optical Channel Transport Unit
OTUkV	Functionally Standardized OTUk
SCSI	Switch Capability Specific Information
TDM	Time Division Multiplex
TPN	Tributary Port Number
TS	Tributary Slot or Time Slot

A.2 Terminology

1. ODUk and ODUj

ODUk refers to the ODU container that is directly mapped to an OTU container. ODUj refers to the lower order ODU container that is mapped to an higher order ODU container via multiplexing.

2. LO-ODU and HO-ODU

LO-ODU refers to the ODU client layer of lower rate that is mapped to an ODU server layer of higher rate via multiplexing. HO-ODU refers to the ODU server layer of higher rate that supports mapping of one or more ODU client layers of lower rate.

In multi-stage multiplexing case, a given ODU layer can be a client for one stage (interpreted as LO-ODU) and at the same time server for another stage (interpreted as HO-ODU). In this case, the notion of LO-ODU and HO-ODU needs to be interpreted in a recursive manner.

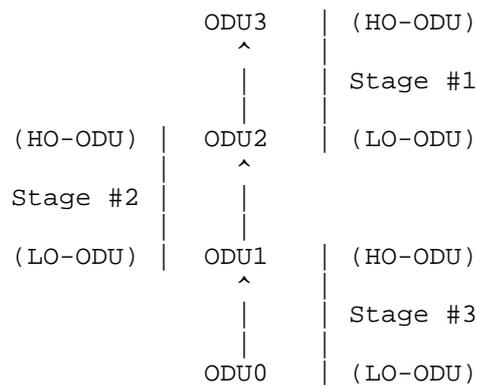


Figure-6 : LO-ODU and HO-ODU

3. Single Stage Multiplexing

When ODU multiplexing hierarchy involves only two levels (ODUk and ODUj), it is referred as single stage multiplexing.

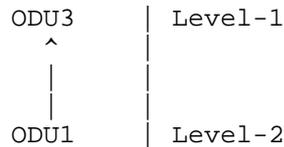


Figure-7: Single Stage Multiplexing

4. Multi Stage Multiplexing

When ODU multiplexing hierarchy involves more than two levels, it is referred as multi-stage multiplexing. Two adjoining levels form a multiplexing stage.

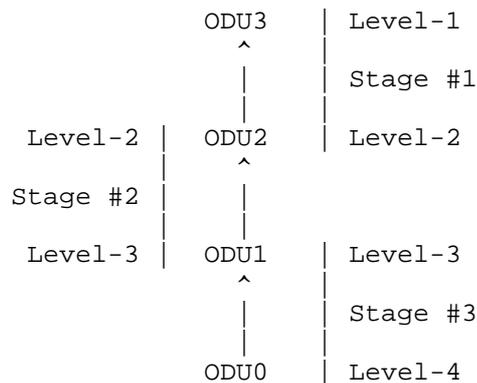


Figure-8 : Multi Stage Multiplexing

Appendix B : RFC4328 and G.709v3

B.1 G.709 Traffic Parameters

The G.709 Traffic Parameters defined in [RFC4328] does not work well for the new features introduced in [G.709-v3]. The basic draw-backs are:

- (a) NMC attribute defined in G.709 Traffic Parameters does not apply end-to-end especially when links with different TSG are involved in the path of a LSP.

(b) ODUflex needs absolute nominal rate and tolerance to be specified.

B.2 Label Format

The Label format defined in [RFC4328] is not scalable/extensible to cover the new ODU rates defined in [G.709-v3]. Some of the limitations are captured below:

(a) The bit-fields defined to represent TSs for specific ODU rates are not future proof. The reserved bits are not sufficient to cover the future ODU types.

(b) The label format assumes 2.5G Tributary Slot Granularity. It needs to be redefined for 1.25G Tributary Slot Granularity.

(c) One Tributary Slot information is coded in 4 bytes. ODU3 and ODU4 requires 32 and 80 TSs respectively. This would dramatically increase the label size and thus impact the scalability.

Internet Engineering Task Force
Internet-Draft
Intended status: Informational
Expires: September 2, 2011

R. Kunze, Ed.
Deutsche Telekom AG
March 1, 2011

A framework for Black Link Management and Control
draft-kunze-black-link-management-framework-00

Abstract

This document provides a framework that describes a solution space for the control and management of optical interfaces according to the Black Link approach as specified by ITU-T [ITU.G698.2] and beyond (friendly wavelength). In particular, it examines topological elements and related network management measures.

Optical Routing and Wavelength assignment based on WSON is out of scope. This document concentrates on the management of optical interfaces. The application of a dynamic control plane, e.g. for auto-discovery or for the distribution of interface parameters, is complementary. Anyway, this work is not in conflict with WSON but leverages and supports related work already done for management plane and control plane.

Furthermore, support for Fast Fault Detection, to e.g. trigger Protection Switching is provided by the WDM interface capability of the client interface (e.g. ITU-T G.709) is out of scope for this work. Additionally the wavelength ordering process and the process how to determine the demand for a new wavelength from A to Z is out of scope.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on September 2, 2011.

Copyright Notice

Copyright (c) 2011 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	3
1.1. Requirements Language	4
2. Terminology and Definitions	4
3. DWDM Black Link Management Solution Space	5
3.1. Description of Client Network Layer - WDM connection	6
3.1.1. Traditional WDM deployments	6
3.1.2. Black Link Deployments	7
4. Black Link Operation scenarios	9
4.1. Bringing into service	9
4.2. Configuration Management	9
4.3. In service (performance management)	9
4.4. Fault Clearance	9
5. Black Link Control and Management Solutions	9
5.1. BL Separate Operation and Management Approaches	10
5.1.1. Direct connection to the management system	11
5.1.2. Indirect connection to the WDM management system	13
5.2. Control Plane Considerations	14
5.2.1. Black Link deployment with common control plane	14
5.2.2. Black Link deployment with an separate control plane	15
6. Acknowledgements	15
7. IANA Considerations	15
8. Security Considerations	15
9. Contributors	15
10. References	16
10.1. Normative References	16
10.2. Informative References	16

1. Introduction

The usage of the Black Link approach in carrier long haul and aggregation networks adds a further option for operators to facilitate their networks. The integration of optical coloured interfaces into routers and other types of clients could lead to a lot of benefits regarding an efficient and optimized data transport for higher layer services.

Carriers deploy their networks today as a combination of transport and packet infrastructure. This ensures high available and flexible data transport. Both network technologies are managed usually by different operational units using different management concepts. This is the status quo in many carrier networks today. In the case of a black link or friendly wavelength deployment, where the coloured interface moves into the client (e.g. router), it is necessary to establish a management connection between the client providing the coloured interface and the corresponding EMS (Element Management System) of the transport network to ensure that the coloured interface parameters can be managed in the same way as traditional deployments allow this.

The objective of this document is to provide a framework that describes the solution space for the control and management of WDM Black Links as specified by ITU-T [ITU.G698.2] and beyond (friendly wavelength). In particular, it examines topological elements and related network management measures.

Optical Routing and Wavelength assignment based on WSON is out of scope. This document concentrates on the management of optical interfaces. The application of a dynamic control plane, e.g. for auto-discovery or distribute interface parameters, is complementary. Anyway, this work is not in conflict with WSON but leverages and supports related work already done for management plane and control plane.

Furthermore, support for Fast Fault Detection, to e.g. trigger Protection Switching is provided by the WDM interface capability of the client interface (e.g. ITU-T G.709) is out of scope for this work. Additionally the wavelength ordering process and the process how to determine the demand for a new wavelength from A to Z is out of scope.

Note that Control and Management Plane are two separate entities that are handling the same information in different ways. This document covers management as well as control plane considerations in different cases of BL (Black Link) and friendly wavelength operation.

1.1. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

2. Terminology and Definitions

Black Link: The Black Link [ITU.G698.2] allows supporting an optical transmitter/receiver pair of one or different vendors to inject a DWDM channel and run it over an optical network composed of amplifiers, filters, add-drop multiplexers from a different vendor. Therefore the standard defines the ingress and egress parameters at the interface Ss and Rs.

Coloured Interface: The term coloured interface defines the single channel optical interface that is used to bridge long distances and is directly connected with a DWDM system. Coloured interfaces operate on a fix wavelength or within a wavelength band (tunability). Coloured interface is a more generic term and it is a superset of the Black Link.

Friendly Wavelength: A friendly wavelength is a wavelength that is generated or originated by an optical interface that is not part of the WDM system but completely managed and known by the WDM system.

Alien Wavelength: Alien Wavelength: An alien wavelength is a wavelength that is generated or originated by an optical interface that is not part of the WDM system and not managed and known by the WDM system.

Forward error correction (FEC): FEC is an important way of improving the performance of high-capacity long haul optical transmission systems. Employing FEC in optical transmission systems yields system designs that can accept relatively large BER (much more than 10⁻¹²) in the optical transmission line (before decoding).

Intra-domain Interface (IaDI): The intra-domain interface (line site of the optical system) is a physical interface within an optical administrative or vendor domain and is implemented as:

- a. standardized single channel interface specified according to G.698.2 (standardized optical interface AND OTUk according G.709 or
- b. proprietary single channel interface proprietary optical interface OR functionally specified OTUkV according G.709, i.e. proprietary FEC.

Inter-Domain Interface(IrDI): The inter-domain interface is a physical interface that represents the boundary between two administrative domains as well as the boundary between Client and optical domain.

Management Plane: Management Plane: The management plane supports FCAPS (Fault, Configuration, Accounting, Performance and Security Management) capabilities for carrier networks.

Control Plane: The control plane supports signalling, path computation, routing, path setup and restoration.

Client Network Layer: The client network layer is the layer above (on top) the WDM layer, from the perspective of the WDM layer.

Transponder: A Transponder is a network element that performs O/E/O (Optical /Electrical/Optical)conversion. In this document is referred only transponders with 3R (rather than 2R or 1R regeneration) as defined in [ITU.G.872]

3. DWDM Black Link Management Solution Space

Basically the Black Link and Friendly Wavelength management deals with aspects needed for setup, tear down and maintenance of wavelengths, which are demanded by a client network layer (the layer above WDM). The following types of WDM networks are considered for BL and FW management:

- a. Passive WDM
- b. Legacy point to point WDM systems
- c. Legacy WDM systems with OADMs
- d. Transparent optical networks supporting specific IPoDWDM functions, interfaces or protocols

Table 1 provides a list of tasks, which are related to BL management, It is indicated which domain (optical or client) is responsible for a task. The relevance of a task for each type of WDM network is also indicated.

Task	Domain	a	b	c	d
determination of centre frequency	client	R	R	R	R
configuration of centre frequency at colored IF	optical	NR	NR	R	R
path computation of wavelength	optical	NR	NR	R	R
routing of wavelength	optical	NR	NR	R	R
wavelength setup across optical network	client	?	?	R	R
detection of wavelength fault	optical	R	R	R	R
fault isolation, identification of root failure	optical	NR	R	R	R
repair actions within optical network	optical	R	R	R	R
protection switching of wavelength	optical	NR	NR	R	R
restoration of wavelength	optical	NR	NR	R	R

Table 1: List of tasks related to BL management

Furthermore the following deployment cases will be considered:

- a. Exclusive Black Link deployment
- b. Black Link deployment in combination with grey client network interfaces

Case b) is motivated by the usage of legacy equipment using the traditional connection as described in Figure 1 combined with the BL approach.

3.1. Description of Client Network Layer - WDM connection

3.1.1. Traditional WDM deployments

The ordinary connection of a client layer network towards a WDM system is based today on client interfaces (grey) bridging short or intermediate distances between client and WDM system. The Optical Signal incoming into the WDM system must be converted (OEO conversion) to corresponding WDM wavelength grid and the power level that is applicable for the WDM transmission path. This conversion is done by a component termed as transponder (see Figure 1).

After that OEO conversion the signal complies with the parameters that are specified for a certain WDM link.

Figure 1 shows the traditional Client - WDM interconnection using transponders for wavelength conversion. IrDI and IaDI as defined in Section 2 specifying the different demarcation areas related to external and internal connections

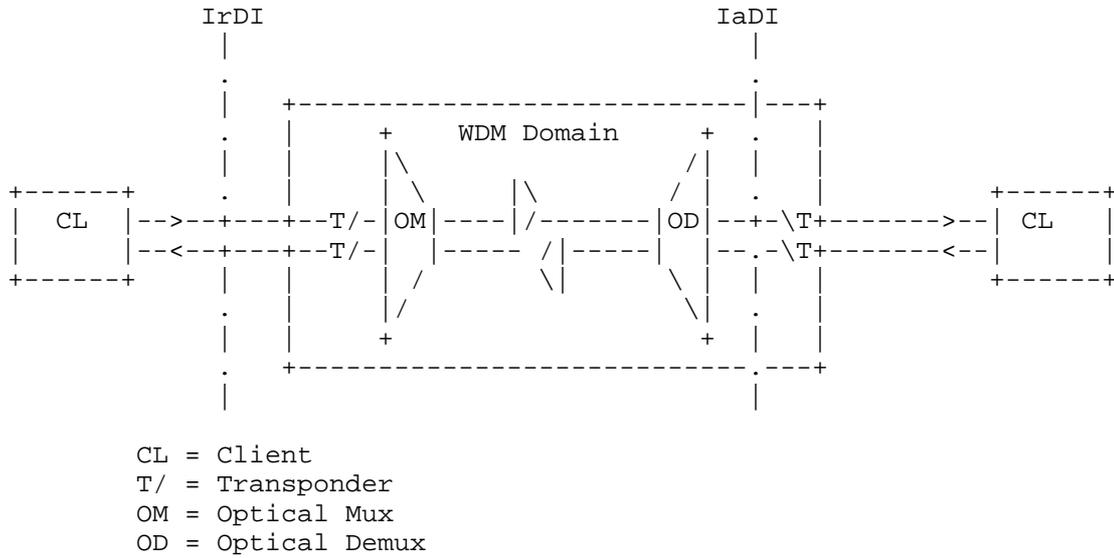


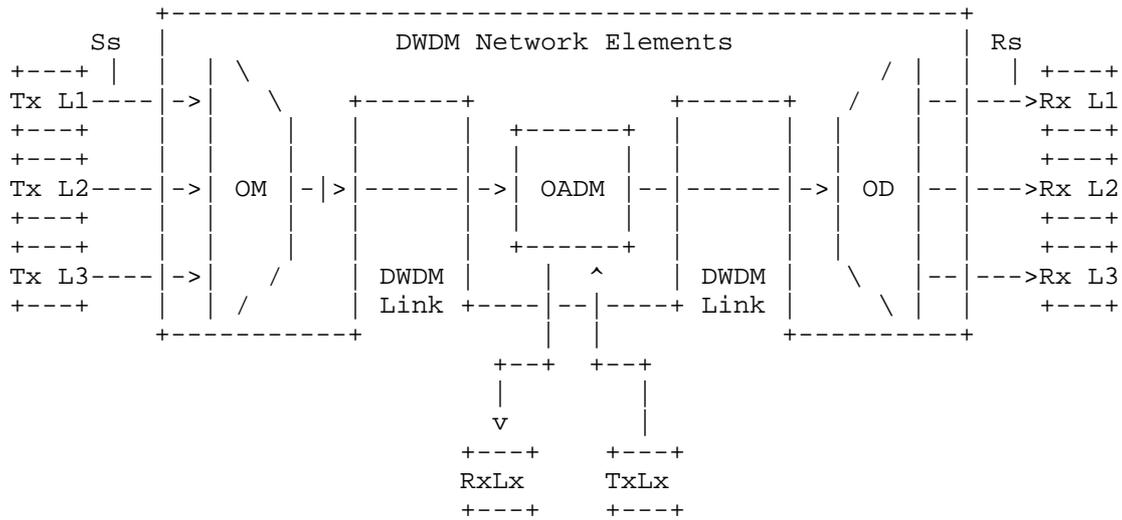
Figure 1: Inter and Intra-Domain Interface Identification

The management and control of WDM and client layer is done by different control and management solutions. Different operational units are responsible for client and WDM layer.

3.1.2. Black Link Deployments

In case of a black link deployment Figure 2 the DWDM transceiver is located directly at the client and the grey interfaces will be saved. In that case a solution must be found to manage that coloured interface in the same way as in the traditional case. This requirement must be fulfilled especially in the cases where legacy equipment and Black Link/Friendly Wavelength interfaces will be used in parallel or together and the operational situation is unchanged.

Figure 2 shows a set of reference points, for the linear "black-link" approach, for single-channel connection (Ss and Rs) between transmitters (Tx) and receivers (Rx). Here the WDM network elements include an OM and an OD (which are used as a pair with the opposing element), one or more optical amplifiers and may also include one or more OADMs.



Ss = reference point at the DWDM network element tributary output
 Rs = reference point at the DWDM network element tributary input
 Lx = Lambda x
 OM = Optical Mux
 OD = Optical Demux
 OADM = Optical Add Drop Mux

from Fig. 5.1/G.698.2

Figure 2: Linear Black Link

Independent from the WDM networks that are considered the black link must perform as well in mixed setups with both legacy and Black Link/Friendly Wavelength equipment.

4. Black Link Operation scenarios

A Comparison of the black link with the traditional operation scenarios provides an insight of similarities and distinctions in operation and management. The following four use cases provide an overview about operation and maintenance processes.

4.1. Bringing into service

tbd.

4.2. Configuration Management

tbd.

4.3. In service (performance management)

tbd.

4.4. Fault Clearance

tbd.

5. Black Link Control and Management Solutions

Operation and management of WDM systems is traditionally seen as a homogenous group of tasks that could be carried out best when a single management system or an umbrella management system is used. Each WDM vendor provides a management system that also administrates the wavelengths.

This old operational approach was predicted on a high amount/rate of connection oriented traffic in carrier networks. This behaviour has been changed completely. Today IP is the dominating traffic in the network and from the operating perspective it is more beneficial to use a common management and operation approach. Due to a long history of operational separation it must be possible to manage and operate Black Link deployments with the traditional approach too.

Therefore from the operational point of view in a pure Black Link or in a mixed setup with legacy equipment (transponders) there are two approaches to manage and operate the network.

1. Separate operation and management of client and Transport network
 - a. Direct link to the management system (e.g. EMS, OSS)
 - b. Indirect link to the management system; using a protocol

between the peer node and the directly connected WDM system node to exchange management information

2. Common operation and management of IP and Transport network

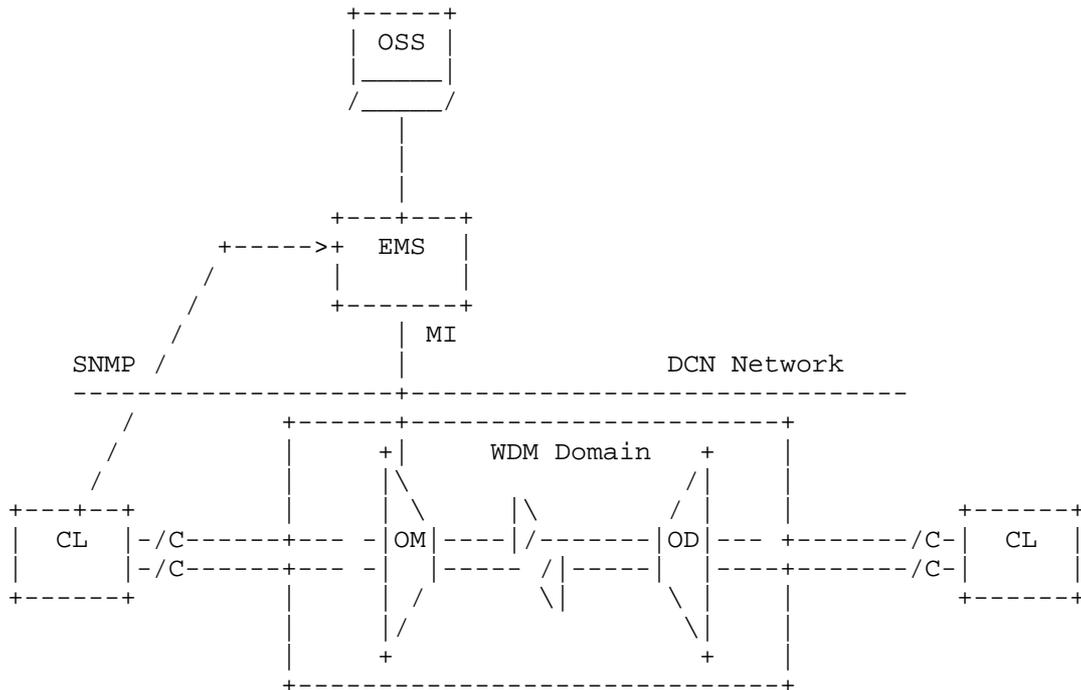
The first option keeps the status quo in large carrier networks as mentioned above. In that case it must be ensured that the full FCAPS Management (Fault, Configuration, Accounting, Performance and Security) capabilities are supported. This means from the management staff point of view nothing changes. The transceiver/receiver optical interface will be part of the optical management domain and will be managed from the transport management staff.

The second option should be favoured if the underlying WDM transport network is mainly used to interconnect IP nodes and the service creation and restoration will be done on higher layers (e.g. IP/MPLS). Then it is more beneficial have a higher level of integration and a common management will be more efficient.

5.1. BL Separate Operation and Management Approaches

5.1.1. Direct connection to the management system

As depicted in Figure 3 one possibility to manage the optical interface within the client is a direct connection to the management system of the optical domain. This ensures manageability as usual.



- CL = Client
- /C = Coloured Interface
- OM = Optical Mux
- OD = Optical Demux
- EMS = Element Management System
- MI= Management Interface

Figure 3: Connecting BL on Transport Management

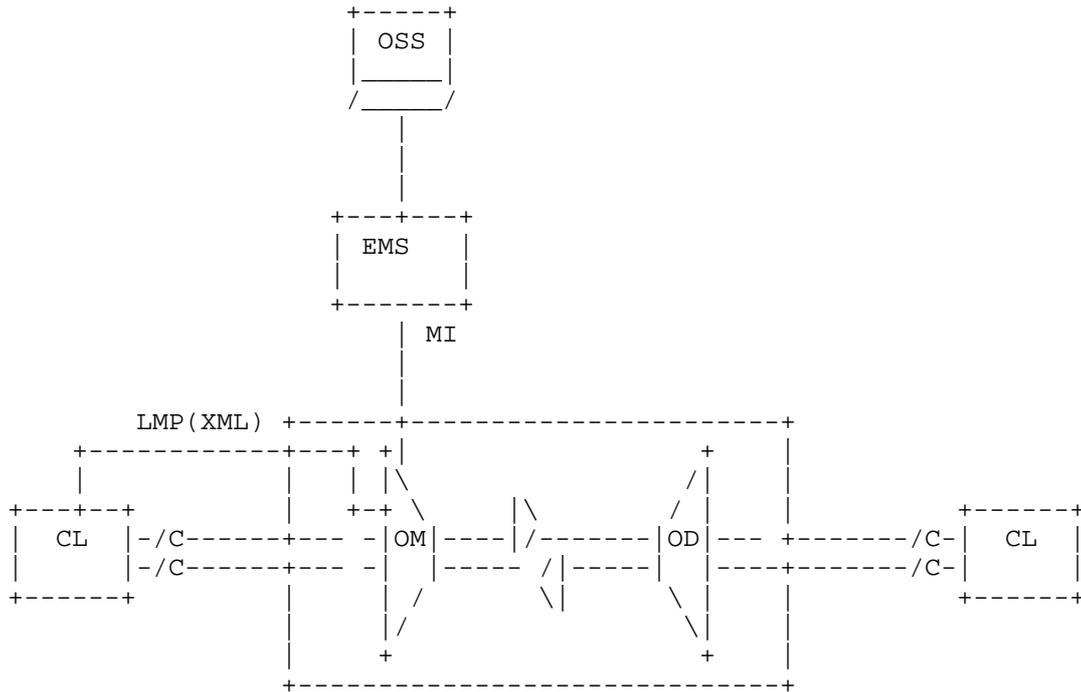
The exchange of management information between client and management system assumes that some form of a direct link exists between the client node and the WDM management system (e.g. EMS). This may be an Ethernet Link or a DCN connection.

It must be ensured that the optical interface can be managed in a standardized way to enable interoperable solutions between different optical interface vendors and vendors of the optical network management software. RFC 3591 [RFC3591] defines manage objects for the optical interface type but does not cover the scenarios described by this framework document. Therefore an extension to this MIB for the optical interface has been drafted in [Black-Link-MIB]. In that case SNMP is used to exchange data between client and management system of the WDM domain.

Note that a software update of the interface components of the client does not lead obligatory to an update of the software of the EMS and vice versa.

5.1.2. Indirect connection to the WDM management system

The alternative as shown in Figure 4 can be used in cases where a more automated relationship between transport node and router is aspired. In that case a combination of rudimentary control plane features and manual management will be used. It is a first step into a more control plane oriented operation model.



CL = Client
 /C = Coloured Interface
 OM = Optical Mux
 OD = Optical Demux
 EMS= Element Management System
 MI= Management Interface

Figure 4: Direct connection between peer node and first optical network node

For information exchange between client and the direct connected node of the optical transport network LMP as specified in RFC 4209

[RFC4209] can (should) be used. This extension of LMP may be used between a peer node and an adjacent optical network node as depicted in Figure 4.

Recently LMP based on RFC 4209 does not support the transmission of configuration data (information). This functionality has to be added to the existing extensions of the protocol. The use LMP-WDM assumes that some form of a control channel exists between the client node and the WDM equipment. This may be a dedicated lambda, an Ethernet Link, or a DCN. It is proposed to use an out of band signalling over a separate link or DCN to ensure a high availability.

5.2. Control Plane Considerations

Basically it is not mandatory necessary to run a control plane in Black Link or friendly wavelength scenarios at least not in simple black link case where clients will be connected point to point using a simple WDM infrastructure (multiplexer and amplifier). As a first step it is possible to configure the entire link using the standard management system and a direct connection of the router or client to the EMS of the transport network. Configuration information will be exchanged using SNMP (see sections Section 5.1.1).

Looking at the control plane the following two scenarios may be considered:

- a. A common control plane for transport and client network; this implies a single operation unit responsible for both client and transport network management.
- b. A separate control plane for client and optical network without any interaction

As mentioned in chapter Section 5.1.2 some control plane features like LMP in an enhanced version could be used.

In such simple scenario it is imaginable to use only LMP to exchange information between the nodes of the optical domain. LMP must be run between the both end-points of the link and between the edge node and the first optical network node.

5.2.1. Black Link deployment with common control plane

tbd.

5.2.2. Black Link deployment with an separate control plane

tbd.

6. Acknowledgements

The author would like to thank Ulrich Drafts for the very good teamwork during the last years and the initial thoughts related to the packet optical integration. Furthermore the author would like to thank all people involved within Deutsche Telekom for the support and fruitful discussions.

7. IANA Considerations

This memo includes no request to IANA.

8. Security Considerations

This document has no requirement for a change to the security models within GMPLS, associated protocols and management interfaces. That is the LMP security models could be operated unchanged.

9. Contributors

Arnold Mattheus
Deutsche Telekom
Darmstadt
Germany
email arnold.Mattheus@telekom.de

Manuel Paul
Deutsche Telekom
Berlin
Germany
email Manuel.Paul@telekom.de

Josef Roese
Deutsche Telekom
Darmstadt
Germany
email j.roese@telekom.de

Frank Luennemann
Deutsche Telekom
Muenster
Germany
email Frank.Luennemann@telekom.de

10. References

10.1. Normative References

- [ITU.G.872] International Telecommunications Union, "Architecture of optical transport networks", ITU-T Recommendation G.872, November 2001.
- [ITU.G698.2] International Telecommunications Union, "Amplified multichannel dense wavelength division multiplexing applications with single channel optical interfaces", ITU-T Recommendation G.698.2, November 2009.
- [ITU.G709] International Telecommunications Union, "Interface for the Optical Transport Network (OTN)", ITU-T Recommendation G.709, March 2003.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC3591] Lam, H-K., Stewart, M., and A. Huynh, "Definitions of Managed Objects for the Optical Interface Type", RFC 3591, September 2003.
- [RFC4209] Fredette, A. and J. Lang, "Link Management Protocol (LMP) for Dense Wavelength Division Multiplexing (DWDM) Optical Line Systems", RFC 4209, October 2005.

10.2. Informative References

- [Black-Link-MIB] Internet Engineering Task Force, "A SNMP MIB to manage the optical parameters characteristic of a DWDM Black-Link", draft-galimbe-kunze-black-link-mib-00 draft-galimbe-kunze-black-link-mib-00, March 2011.

Author's Address

Ruediger Kunze (editor)
Deutsche Telekom AG
Berlin, 10589
DE

Phone: +49 30 3497 3152
EMail: ruediger.kunze@telekom.de

INTERNET-DRAFT
Intended Status: Proposed Standard
Expires: September 15, 2011

A. Malis, ed.
Verizon Communications
A. Lindem, ed.
Ericsson
March 14, 2011

Updates to ASON Routing for OSPFv2 Protocols (RFC 5787bis)
draft-malis-ccamp-rfc5787bis-02.txt

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at
<http://www.ietf.org/lid-abstracts.html>

The list of Internet-Draft Shadow Directories can be accessed at
<http://www.ietf.org/shadow.html>

Copyright and License Notice

Copyright (c) 2011 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Abstract

The ITU-T has defined an architecture and requirements for operating an Automatically Switched Optical Network (ASON).

The Generalized Multiprotocol Label Switching (GMPLS) protocol suite is designed to provide a control plane for a range of network technologies including optical networks such as time division multiplexing (TDM) networks including SONET/SDH and Optical Transport Networks (OTNs), and lambda switching optical networks.

The requirements for GMPLS routing to satisfy the requirements of ASON routing, and an evaluation of existing GMPLS routing protocols are provided in other documents. This document defines extensions to the OSPFv2 Link State Routing Protocol to meet the requirements for routing in an ASON.

Note that this work is scoped to the requirements and evaluation expressed in RFC 4258 and RFC 4652 and the ITU-T Recommendations current when those documents were written. Future extensions or revisions of this work may be necessary if the ITU-T Recommendations are revised or if new requirements are introduced into a revision of RFC 4258.

Table of Contents

1.	Introduction	4
1.1.	Conventions Used in This Document	5
2.	Routing Areas, OSPF Areas, and Protocol Instances	5
3.	Terminology and Identification	6
4.	Reachability	6
5.	Link Attribute	7
5.1.	Local Adaptation	7
5.2.	Bandwidth Accounting	8
6.	Routing Information Scope	8
6.1.	Link Advertisement (Local and Remote TE Router ID Sub-TLV)	9
6.2.	Reachability Advertisement (Local TE Router ID sub-TLV)	10
7.	Routing Information Dissemination	11
7.1.	Import/Export Rules	11
7.2.	Loop Prevention	11
7.2.1.	Inter-RA Export Upward/Downward Sub-TLVs	12
7.2.2.	Inter-RA Export Upward/Downward Sub-TLV Processing	13
8.	OSPFv2 Scalability	13
9.	Security Considerations	14
10.	IANA Considerations	14
10.1.	Sub-TLVs of the Link TLV	14

- 10.2. Sub-TLVs of the Node Attribute TLV 15
- 10.3. Sub-TLVs of the Router Address TLV 15
- 11. Management Considerations 16
 - 11.1. Routing Area (RA) Isolation 16
 - 11.2 Routing Area (RA) Topology/Configuration Changes 16
- 12. References 17
 - 12.1. Normative References 17
 - 12.2. Informative References 17
- 13. Acknowledgements 18
- Appendix A. ASON Terminology 19
- Appendix B. ASON Routing Terminology 20
- Authors' Addresses 21

1. Introduction

The Generalized Multiprotocol Label Switching (GMPLS) [RFC3945] protocol suite is designed to provide a control plane for a range of network technologies including optical networks such as time division multiplexing (TDM) networks including SONET/SDH and Optical Transport Networks (OTNs), and lambda switching optical networks.

The ITU-T defines the architecture of the Automatically Switched Optical Network (ASON) in [G.8080].

[RFC4258] describes the routing requirements for the GMPLS suite of routing protocols to support the capabilities and functionality of ASON control planes identified in [G.7715] and in [G.7715.1].

[RFC4652] evaluates the IETF Link State routing protocols against the requirements identified in [RFC4258]. Section 7.1 of [RFC4652] summarizes the capabilities to be provided by OSPFv2 [RFC2328] in support of ASON routing. This document describes the OSPFv2 specifics for ASON routing.

Multi-layer transport networks are constructed from multiple networks of different technologies operating in a client-server relationship. The ASON routing model includes the definition of routing levels that provide scaling and confidentiality benefits. In multi-level routing, domains called routing areas (RAs) are arranged in a hierarchical relationship. Note that as described in [RFC4652], there is no implied relationship between multi-layer transport networks and multi-level routing. The multi-level routing mechanisms described in this document work for both single-layer and multi-layer networks.

Implementations may support a hierarchical routing topology (multi-level) for multiple transport network layers and/or a hierarchical routing topology for a single transport network layer.

This document describes the processing of the generic (technology-independent) link attributes that are defined in [RFC3630], [RFC4202], and [RFC4203] and that are extended in this document. As described in Section 5.2, technology-specific traffic engineering attributes and their processing may be defined in other documents that complement this document.

Note that this work is scoped to the requirements and evaluation expressed in [RFC4258] and [RFC4652] and the ITU-T Recommendations current when those documents were written. Future extensions of revisions of this work may be necessary if the ITU-T Recommendations are revised or if new requirements are introduced into a revision of

[RFC4258].

1.1. Conventions Used in This Document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

The reader is assumed to be familiar with the terminology and requirements developed in [RFC4258] and the evaluation outcomes described in [RFC4652].

General ASON terminology is provided in Appendix A. ASON routing terminology is described in Appendix B.

2. Routing Areas, OSPF Areas, and Protocol Instances

An ASON routing area (RA) represents a partition of the data plane, and its identifier is used within the control plane as the representation of this partition.

RAs are hierarchically contained: a higher-level (parent) RA contains lower-level (child) RAs that in turn MAY also contain RAs, etc. Thus, RAs contain RAs that recursively define successive hierarchical RA levels. Routing information may be exchanged between levels of the RA hierarchy, i.e., Level N+1 and N, where Level N represents the RAs contained by Level N+1. The links connecting RAs may be viewed as external links (inter-RA links), and the links representing connectivity within an RA may be viewed as internal links (intra-RA links). The external links to an RA at one level of the hierarchy may be internal links in the parent RA. Intra-RA links of a child RA MAY be hidden from the parent RA's view. [RFC4258]

An ASON RA can be mapped to an OSPF area, but the hierarchy of ASON RA levels does not map to the hierarchy of OSPF areas. Instead, successive hierarchical levels of RAs MUST be represented by separate instances of the protocol. Thus, inter-level routing information exchange (as described in Section 7) involves the export and import of routing information between protocol instances.

An ASON RA may therefore be identified by the combination of its OSPF instance identifier and its OSPF area identifier. With proper and careful network-wide configuration, this can be achieved using just the OSPF area identifier, and this process is RECOMMENDED in this document. These concepts are discussed in Section 7.

A key ASON requirement is the support of multiple transport planes or layers. Each transport node has associated topology (links and

reachability) which is used for ASON routing.

3. Terminology and Identification

This section describes the mapping of key ASON entities to OSPF entities. Appendix A contains a complete glossary of ASON routing terminology.

There are three categories of identifiers used for ASON routing (G7715.1): transport plane names, control plane identifiers for components, and SCN addresses. This section discusses the mapping between ASON routing identifiers and corresponding identifiers defined for GMPLS routing, and how these support the physical (or logical) separation of transport plane entities and control plane components. GMPLS supports this separation of identifiers and planes.

In the context of OSPF Traffic Engineering (TE), an ASON transport node corresponds to a unique OSPF TE node. An OSPF TE node is uniquely identified by the TE Router Address TLV [RFC3630]. In this document, this TE Router Address is referred to as the TE Router ID, which is in the ASON transport plane name space. The TE Router ID should not be confused with the OSPF Router ID which uniquely identifies an OSPF router within an OSPF routing domain [RFC2328] and is in a name space for control plane components.

Note: The Router Address top-level TLV definition, processing, and usage are unchanged from [RFC3630]. This TLV specifies a stable OSPF TE node IP address, i.e., the IP address is always reachable when there is IP connectivity to the associated OSPF TE node.

ASON defines a Routing Controller (RC) as an entity that handles (abstract) information needed for routing and the routing information exchange with peering RCs by operating on the Routing Database (RDB). ASON defines a Protocol Controller (PC) as an entity that handles protocol-specific message exchanges according to the reference point over which the information is exchanged (e.g., E-NNI, I-NNI), and internal exchanges with the Routing Controller (RC) [RFC4258]. In this document, an OSPF router advertising ASON TE topology information will perform both the functions of the RC and PC. Each OSPF router is uniquely identified by its OSPF Router ID [RFC2328].

4. Reachability

Reachability in ASON refers to the set of endpoints reachable in the transport plane by a node or the reachable endpoints of a level N. Reachable entities are identified in the transport plane name space

(ASON SNPP name space). In order to advertise blocks of reachable address prefixes, a summarization mechanism is introduced that is based on the techniques described in [RFC5786]. For ASON reachability advertisement, blocks of reachable address prefixes are advertised together with the associated data plane node. The data plane node is identified in the control plane by its TE Router ID, as discussed in section 6.

In order to support ASON reachability advertisement, the Node Attribute TLV defined in [RFC5786] is used to advertise the combination of a TE Router ID and its set of associated reachable address prefixes. The Node Attribute TLV can contain the following sub-TLVs:

- TE Router ID sub-TLV: Length: 4; Defined in Section 6.2
- Node IPv4 Local Address sub-TLV: Length: variable; [RFC5786]
- Node IPv6 Local Address sub-TLV: Length: variable; [RFC5786]

A router may support multiple transport nodes as discussed in section 6, and, as a result, may be required to advertise reachability (ASON TRIs) separately for each transport node. As a consequence, it MUST be possible for the router to originate more than one TE LSA containing the Node Attribute TLV when used for ASON reachability advertisement.

Hence, the Node Attribute TLV [RFC5786] advertisement rules must be relaxed for ASON. A Node Attribute TLV MAY appear in more than one TE LSA originated by the RC when the RC is advertising reachability information for a different transport node identified by the Local TE Router Sub-TLV (refer to section 6.1).

5. Link Attribute

With the exception of local adaptation (described below), the mapping of link attributes and characteristics to OSPF TE Link TLV Sub-TLVs is unchanged [RFC4652]. OSPF TE Link TLV Sub-TLVs are described in [RFC3630] and [RFC4203]. Advertisement of this information SHOULD be supported on a per-layer basis, i.e., one TE LSA per unique switching capability and bandwidth granularity combination.

5.1. Local Adaptation

Local adaptation is defined as a TE link attribute (i.e., sub-TLV) that describes the cross/inter-layer relationships.

The Interface Switching Capability Descriptor (ISCD) TE Attribute [RFC4202] identifies the ability of the TE link to support cross-connection to another link within the same layer. When advertising

link adaptation, it also identifies the ability to use a locally terminated connection that belongs to one layer as a data link for another layer (adaptation capability). However, the information associated with the ability to terminate connections within that layer (referred to as the termination capability) is advertised with the adaptation capability.

For instance, a link between two optical cross-connects will contain at least one ISCD attribute describing the Lambda Switching Capable (LSC) switching capability. Conversely, a link between an optical cross-connect and an IP/MPLS Label Switching Router (LSR) will contain at least two ISCD attributes, one for the description of the LSC termination capability and one for the Packet Switching Capable (PSC) adaptation capability.

In OSPFv2, the Interface Switching Capability Descriptor (ISCD) is a sub-TLV (type 15) of the top-level Link TLV (type 2) [RFC4203]. The adaptation and termination capabilities are advertised using two separate ISCD sub-TLVs within the same top-level Link TLV.

An interface MAY have more than one ISCD sub-TLV, [RFC4202] and [RFC4203]. Hence, the corresponding advertisements should not result in any compatibility issues.

5.2. Bandwidth Accounting

GMPLS routing defines an Interface Switching Capability Descriptor (ISCD) that provides, among other things, the available (maximum/minimum) bandwidth per priority available for Label Switched Path (LSPs). One or more ISCD sub-TLVs can be associated with an interface, [RFC4202] and [RFC4203]. This information, combined with the Unreserved Bandwidth Link TLV sub-TLV [RFC3630], provides the basis for bandwidth accounting.

In the ASON context, additional information may be included when the representation and information in the other advertised fields are not sufficient for a specific technology, e.g., SDH. The definition of technology-specific information elements is beyond the scope of this document. Some technologies will not require additional information beyond what is already defined in [RFC3630], [RFC4202], and [RFC4203].

6. Routing Information Scope

For ASON routing, the control plane component routing adjacency topology (i.e., the associated Protocol Controller (PC) connectivity) and the transport topology are NOT assumed to be congruent [RFC4258]. Hence, a single OSPF router (i.e., the PC) MUST be able to advertise

on behalf of multiple transport layer nodes. The OSPF routers are identified by OSPF Router ID and the transport nodes are identified by TE Router ID.

The Router Address TLV [RFC3630] is used to advertise the TE Router ID associated with the advertising Routing Controller. TE Router IDs for additional transport nodes are advertised through specification of the Local TE Router Identifier in the Local and Remote TE Router TE sub-TLV and the Local TE Router Identifier sub-TLV described in the sections below. These Local TE Router Identifiers are typically used as the local endpoints for TE Label Switched Paths (LSPs) terminating on the associated transport node.

It MAY be feasible for multiple OSPF Routers to advertise TE information for the same transport node. However, this is not considered a required use case and is not discussed further.

6.1. Link Advertisement (Local and Remote TE Router ID Sub-TLV)

An OSPF router advertising on behalf of multiple transport nodes will require additional information to distinguish the link endpoints amongst the subsumed transport nodes. In order to unambiguously specify the transport topology, the local and remote transport nodes MUST be identified by TE router ID.

For this purpose, a new sub-TLV of the OSPFv2 TE LSA top-level Link TLV is introduced that defines the Local and Remote TE Router ID.

The Type field of the Local and Remote TE Router ID sub-TLV is assigned a value TBD. The Length field takes the value 8. The Value field of this sub-TLV contains 4 octets of the Local TE Router Identifier followed by 4 octets of the Remote TE Router Identifier. The value of the Local and Remote TE Router Identifier SHOULD NOT be set to 0.

The format of the Local and Remote TE Router ID sub-TLV is:

```

0                               1                               2                               3
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-----+-----+-----+-----+-----+-----+-----+-----+
|                                     |                                     |
|                                     Type                                     | Length (8)
|                                     |                                     |
|                                     Local TE Router Identifier           |
|                                     |                                     |
|                                     Remote TE Router Identifier           |
|                                     |                                     |
+-----+-----+-----+-----+-----+-----+-----+-----+

```

This sub-TLV MUST be included as a sub-TLV of the top-level Link TLV

if the OSPF router is advertising on behalf of one or more transport nodes having TE Router IDs different from the TE Router ID advertised in the Router Address TLV. Therefore, it MUST be included if the OSPF router is advertising on behalf of multiple transport nodes.

Note: The Link ID sub-TLV identifies the other end of the link (i.e., Router ID of the neighbor for point-to-point links) [RFC3630]. When the Local and Remote TE Router ID Sub-TLV is present, it MUST be used to identify local and remote transport node endpoints for the link and the Link-ID sub-TLV MUST be ignored. The Local and Remote ID sub-TLV, if specified, MUST only be specified once.

6.2. Reachability Advertisement (Local TE Router ID sub-TLV)

When an OSPF router is advertising on behalf of multiple transport nodes, the routing protocol MUST be able to associate the advertised reachability information with the correct transport node.

For this purpose, a new sub-TLV of the OSPFv2 TE LSA top-level Node Attribute TLV is introduced. This TLV associates the local prefixes (see above) to a given transport node identified by TE Router ID.

The Type field of the Local TE Router ID sub-TLV is assigned a value TBD. The Length field takes the value 4. The Value field of this sub-TLV contains the Local TE Router Identifier [RFC3630] encoded over 4 octets.

The format of the Local TE Router ID sub-TLV is:

```

0                               1                               2                               3
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-----+-----+-----+-----+-----+-----+-----+-----+
|                                     |                                     |
|                                     Type                                     | Length (4)
|                                     |                                     |
|                                     Local TE Router Identifier           |
+-----+-----+-----+-----+-----+-----+-----+-----+

```

This sub-TLV MUST be included as a sub-TLV of the top-level Node Attribute TLV if the OSPF router is advertising on behalf of one or more transport nodes having TE Router IDs different from the TE Router ID advertised in the Router Address TLV. Therefore, it MUST be included if the OSPF router is advertising on behalf of multiple transport nodes.

7. Routing Information Dissemination

An ASON routing area (RA) represents a partition of the data plane, and its identifier is used within the control plane as the representation of this partition. An RA may contain smaller RAs inter-connected by links. ASON RA levels do not map directly to OSPF areas. Rather, hierarchical levels of RAs are represented by separate OSPF protocol instances.

Routing controllers (RCs) supporting multiple RAs disseminate information downward and upward in this ASON hierarchy. The vertical routing information dissemination mechanisms described in this section do not introduce or imply hierarchical OSPF areas. RCs supporting RAs at multiple levels are structured as separate OSPF instances with routing information exchange between levels described by import and export rules between these instances. The functionality described herein does not pertain to OSPF areas or OSPF Area Border Router (ABR) functionality.

7.1 Import/Export Rules

RCs supporting RAs disseminate information upward and downward in the hierarchy by importing/exporting routing information as TE LSAs. TE LSAs are area-scoped opaque LSAs with opaque type 1 [RFC3630]. The information that MAY be exchanged between adjacent levels includes the Router Address, Link, and Node Attribute top-level TLVs.

The imported/exported routing information content MAY be transformed, e.g., filtered or aggregated, as long as the resulting routing information is consistent. In particular, when more than one RC is bound to adjacent levels and both are allowed to import/export routing information, it is expected that these transformations are performed in a consistent manner. Definition of these policy-based mechanisms is outside the scope of this document.

In practice, and in order to avoid scalability and processing overhead, routing information imported/exported downward/upward in the hierarchy is expected to include reachability information (see Section 4) and, upon strict policy control, link topology information.

7.2 Loop Prevention

When more than one RC is bound to an adjacent level of the ASON hierarchy, and is configured to export routing information upward or downward, a specific mechanism is required to avoid looping of routing information. Looping is the re-advertisement of routing information into an RA that had previously advertised that routing

information upward or downward into an upper or lower level RA in the ASON hierarchy. For example, without loop prevention mechanisms, this could happen when the RC advertising routing information downward in the hierarchy is not the same one that advertises routing information upward in the hierarchy.

7.2.1 Inter-RA Export Upward/Downward Sub-TLVs

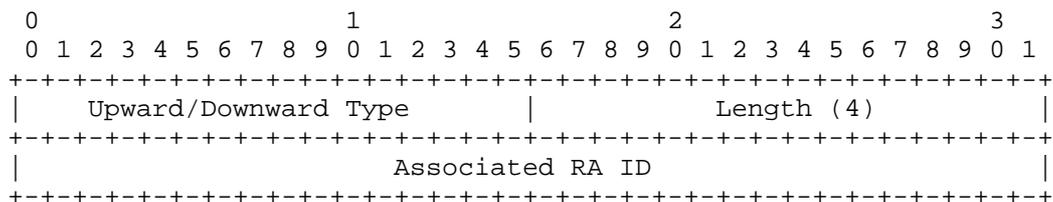
The Inter-RA Export Sub-TLVs can be used to prevent the re-advertisement of OSPF TE routing information into an RA which previously advertised that information. The type value TBD will indicate that the associated routing information has been exported downward. The type value TBD will indicate that the associated routing information has been exported upward. While it is not required for routing information exported downward, both Sub-TLVs will include the Routing Area (RA) ID from the which the routing information was exported. This RA is not necessarily the RA originating the routing information but RA from which the information was immediately exported.

These additional Sub-TLVs MAY be included in TE LSAs that include any of the following top-level TLVs:

- Router Address top-level TLV
- Link top-level TLV
- Node Attribute top-level TLV

The Type field of the Inter-RA Export Upward and Inter-RA Export Downward sub-TLVs are respectively assigned the values TBD1 and TBD2. The Length of the Associated RA ID TLV is 4 octets. The Value field in these sub-TLVs contains the associated RA ID. The RA ID value must be a unique identifier for the RA within the ASON routing domain.

The format of the Inter-RA Export Upward and Inter-RA Export Downward Sub-TLVs is graphically depicted below:



7.2.2 Inter-RA Export Upward/Downward Sub-TLV Processing

TE LSAs MAY be imported or exported downward or upward in the ASON routing hierarchy. The direction and advertising RA ID are advertised in an Inter-RA Export Upward/Downward Sub-TLV. They MUST be retained and advertised in the receiving RA with the associated routing information.

When exporting routing information upward in the ASON routing hierarchy, any information received from a level above, i.e., tagged with an Inter-RA Export Downward Sub-TLV, MUST NOT be exported upward. Since an RA at level N is contained by a single RA at level N+1, this is the only checking that is necessary and the associated RA ID is used solely for informational purposes.

When exporting routing information downward in the ASON routing hierarchy, any information received from a level below, i.e., tagged with an Inter-RA Export Upward Sub-TLV MUST NOT be exported downward if the target RA ID matches the RA ID associated with the routing information. This additional checking is required for routing information exported downward since a single RA at level N+1 may contain multiple RAs at level N in the ASON routing hierarchy. In order words, routing information MUST NOT be exported downward into the RA from which it was received.

8. OSPFv2 Scalability

The extensions described herein are only applicable to ASON routing domains and it is not expected that the attendant reachability (see Section 4) and link information will ever be mixed with global or local IP routing information. If there were ever a requirement for a given RC to participate in both domains, separate OSPFv2 instances would be utilized. However, in a multi-level ASON hierarchy, the potential volume of information could be quite large and the recommendations in this section SHOULD be followed by RCs implementing this specification.

- Routing information exchange upward/downward in the hierarchy between adjacent RAs SHOULD, by default, be limited to reachability information. In addition, several transformations such as prefix aggregation are RECOMMENDED to reduce the amount of information imported/exported by a given RC when such transformations will not impact consistency.
- Routing information exchange upward/downward in the ASON hierarchy involving TE attributes MUST be under strict policy control. Pacing and min/max thresholds for triggered updates are strongly RECOMMENDED.

- The number of routing levels MUST be maintained under strict policy control.

9. Security Considerations

This document specifies the contents and processing of OSPFv2 TE LSAs [RFC3630] and [RFC4202]. The TE LSA extensions defined in this document are not used for SPF computation, and have no direct effect on IP routing. Additionally, ASON routing domains are delimited by the usual administrative domain boundaries.

Any mechanisms used for securing the exchange of normal OSPF LSAs can be applied equally to all TE LSAs used in the ASON context. Authentication of OSPFv2 LSA exchanges (such as OSPF cryptographic authentication [RFC2328] and [RFC5709]) can be used to secure against passive attacks and provide significant protection against active attacks. [RFC5709] defines a mechanism for authenticating OSPFv2 packets by making use of the HMAC algorithm in conjunction with the SHA family of cryptographic hash functions.

If a stronger authentication were believed to be required, then the use of a full digital signature [RFC2154] would be an approach that should be seriously considered. Use of full digital signatures would enable precise authentication of the OSPF router originating each OSPF link-state advertisement, and thereby provide much stronger integrity protection for the OSPF routing domain.

10. IANA Considerations

This document is classified as Standards Track. It defines new sub-TLVs for inclusion in OSPF TE LSAs. According to the assignment policies for the registries of code points for these sub-TLVs, values must be assigned by IANA [RFC3630].

The following subsections summarize the required sub-TLVs.

10.1. Sub-TLVs of the Link TLV

This document defines the following sub-TLVs of the Link TLV advertised in the OSPF TE LSA:

- Local and Remote TE Router ID sub-TLV
- Associated RA ID sub-TLV
- Inter-RA Export Upward sub-TLV
- Inter-RA Export Downward sub-TLV

Codepoints for these Sub-TLVs should be allocated from the "Types for sub-TLVs of TE Link TLV (Value 2)" registry standards action range (0

- 32767) [RFC3630].

Note that the same values for the Associated RA ID sub-TLV, Inter-RA Export Upward sub-TLV, and Inter-RA Export Downward Sub-TLV MUST be used when they appear in the Link TLV, Node Attribute TLV, and Router Address TLV.

10.2. Sub-TLVs of the Node Attribute TLV

This document defines the following sub-TLVs of the Node Attribute TLV advertised in the OSPF TE LSA:

- Local TE Router ID sub-TLV
- Associated RA ID sub-TLV
- Inter-RA Export Upward sub-TLV
- Inter-RA Export Downward sub-TLV

Codepoints for these Sub-TLVs should be assigned from the "Types for sub-TLVs of TE Node Attribute TLV (Value 5)" registry standards action range (0 - 32767) [RFC5786].

Note that the same values for the Associated RA ID sub-TLV, Inter-RA Export Upward sub-TLV, and Inter-RA Export Downward Sub-TLV MUST be used when they appear in the Link TLV, Node Attribute TLV, and Router Address TLV.

10.3. Sub-TLVs of the Router Address TLV

The Router Address TLV is advertised in the OSPF TE LSA [RFC3630]. Since this TLV currently has no Sub-TLVs defined, a "Types for sub-TLVs of Router Address TLV (Value 1)" registry must be defined.

The registry guidelines for the assignment of types for sub-TLVs of the Router Address TLV are as follows:

- o Types in the range 0-32767 are to be assigned via Standards Action.
- o Types in the range 32768-32777 are for experimental use; these will not be registered with IANA, and MUST NOT be mentioned by RFCs.
- o Types in the range 32778-65535 are not to be assigned at this time. Before any assignments can be made in this range, there MUST be a Standards Track RFC that specifies IANA Considerations that covers the range being assigned.

This document defines the following sub-TLVs for inclusion in the

Router Address TLV:

- Associated RA ID sub-TLV
- Inter-RA Export Upward sub-TLV
- Inter-RA Export Downward sub-TLV

Codepoints for these Sub-TLVs should be allocated from the "Types for sub-TLVs of Router Address TLV (Value 1)" registry standards action range (0 - 32767).

Note that the same values for the Associated RA ID sub-TLV, Inter-RA Export Upward sub-TLV, and Inter-RA Export Downward Sub-TLV MUST be used when they appear in the Link TLV, Node Attribute TLV, and Router Address TLV.

11. Management Considerations

11.1. Routing Area (RA) Isolation

If the RA Identifier is mapped to the OSPF Area ID as recommended in section 2.0, OSPF [RFC2328] implicitly provides isolation. On any intra-RA link, packets will only be accepted if the area-id in the OSPF packet header matches the area ID for the OSPF interface on which the packet was received. Hence, RCs will only establish adjacencies and exchange reachability information (see Section 4.0) with RCs in the same RC. Other mechanisms for RA isolation are beyond the scope of this document.

11.2 Routing Area (RA) Topology/Configuration Changes

The GMPLS Routing for ASON requirements [RFC4258] dictate that the routing protocol MUST support reconfiguration and SHOULD support architectural evolution. OSPF [RFC2328] includes support for the dynamic introduction or removal of ASON reachability information through the flooding and purging of OSPF opaque LSAs [RFC5250]. Also, when an RA is partitioned or an RC fails, stale LSAs SHOULD NOT be used unless the advertising RC is reachable. The configuration of OSPF RAs and the policies governing the redistribution of ASON reachability information between RAs are implementation issues outside of the OSPF routing protocol and beyond the scope of this document.

12. References

12.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC2328] Moy, J., "OSPF Version 2", STD 54, RFC 2328, April 1998.
- [RFC3630] Katz, D., Kompella, K., and D. Yeung, "Traffic Engineering (TE) Extensions to OSPF Version 2", RFC 3630, September 2003.
- [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.
- [RFC4203] Kompella, K., Ed., and Y. Rekhter, Ed., "OSPF Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4203, October 2005.
- [RFC5250] Berger, L., Bryskin, I., Zinin, A., and R. Coltun, "The OSPF Opaque LSA Option", RFC 5250, July 2008.
- [RFC5786] Aggarwal, R. and K. Kompella, "Advertising a Router's Local Addresses in OSPF TE Extensions", RFC 5786, March 2010.

12.2. Informative References

- [RFC2154] Murphy, S., Badger, M., and B. Wellington, "OSPF with Digital Signatures", RFC 2154, June 1997.
- [RFC4258] Brungard, D., Ed., "Requirements for Generalized Multi-Protocol Label Switching (GMPLS) Routing for the Automatically Switched Optical Network (ASON)", RFC 4258, November 2005.
- [RFC4652] Papadimitriou, D., Ed., Ong, L., Sadler, J., Shew, S., and D. Ward, "Evaluation of Existing Routing Protocols against Automatic Switched Optical Network (ASON) Routing Requirements", RFC 4652, October 2006.
- [RFC5709] Bhatia, M., Manral, V., Fanto, M., White, R., Barnes, M., Li, T., and R. Atkinson, "OSPFv2 HMAC-SHA

Cryptographic Authentication", RFC 5709, October 2009.

For information on the availability of ITU Documents, please see <http://www.itu.int>.

- [G.7715] ITU-T Rec. G.7715/Y.1306, "Architecture and Requirements for the Automatically Switched Optical Network (ASON)", June 2002.
- [G.7715.1] ITU-T Draft Rec. G.7715.1/Y.1706.1, "ASON Routing Architecture and Requirements for Link State Protocols", November 2003.
- [G.805] ITU-T Rec. G.805, "Generic functional architecture of transport networks)", March 2000.
- [G.8080] ITU-T Rec. G.8080/Y.1304, "Architecture for the Automatically Switched Optical Network (ASON)," 2006 (and Amendments 1 and 2).

13. Acknowledgements

The editors would like to thank Dimitri Papadimitriou for editing RFC 5787, from which this document is derived, and Lyndon Ong and Remi Theillaud for their useful comments and suggestions.

Appendix A. ASON Terminology

This document makes use of the following terms:

Administrative domain: (See Recommendation [G.805].) For the purposes of [G7715.1], an administrative domain represents the extent of resources that belong to a single player such as a network operator, a service provider, or an end-user. Administrative domains of different players do not overlap amongst themselves.

Control plane: performs the call control and connection control functions. Through signaling, the control plane sets up and releases connections, and may restore a connection in case of a failure.

(Control) Domain: represents a collection of (control) entities that are grouped for a particular purpose. The control plane is subdivided into domains matching administrative domains. Within an administrative domain, further subdivisions of the control plane are recursively applied. A routing control domain is an abstract entity that hides the details of the RC distribution.

External NNI (E-NNI): interfaces located between protocol controllers between control domains.

Internal NNI (I-NNI): interfaces located between protocol controllers within control domains.

Link: (See Recommendation G.805.) A "topological component" that describes a fixed relationship between a "subnetwork" or "access group" and another "subnetwork" or "access group". Links are not limited to being provided by a single server trail.

Management plane: performs management functions for the transport plane, the control plane, and the system as a whole. It also provides coordination between all the planes. The following management functional areas are performed in the management plane: performance, fault, configuration, accounting, and security management.

Management domain: (See Recommendation G.805.) A management domain defines a collection of managed objects that are grouped to meet organizational requirements according to geography, technology, policy, or other structure, and for a number of functional areas such as configuration, security, (FCAPS), for the purpose of providing control in a consistent manner. Management domains can be disjoint, contained, or overlapping. As such, the resources

within an administrative domain can be distributed into several possible overlapping management domains. The same resource can therefore belong to several management domains simultaneously, but a management domain shall not cross the border of an administrative domain.

Subnetwork Point (SNP): The SNP is a control plane abstraction that represents an actual or potential transport plane resource. SNPs (in different subnetwork partitions) may represent the same transport resource. A one-to-one correspondence should not be assumed.

Subnetwork Point Pool (SNPP): A set of SNPs that are grouped together for the purposes of routing.

Termination Connection Point (TCP): A TCP represents the output of a Trail Termination function or the input to a Trail Termination Sink function.

Transport plane: provides bidirectional or unidirectional transfer of user information, from one location to another. It can also provide transfer of some control and network management information. The transport plane is layered; it is equivalent to the Transport Network defined in Recommendation G.805.

User Network Interface (UNI): interfaces are located between protocol controllers between a user and a control domain. Note: There is no routing function associated with a UNI reference point.

Appendix B. ASON Routing Terminology

This document makes use of the following terms:

Routing Area (RA): an RA represents a partition of the data plane, and its identifier is used within the control plane as the representation of this partition. Per [G.8080], an RA is defined by a set of sub-networks, the links that interconnect them, and the interfaces representing the ends of the links exiting that RA. An RA may contain smaller RAs inter-connected by links. The limit of subdivision results in an RA that contains two sub-networks interconnected by a single link.

Routing Database (RDB): a repository for the local topology, network topology, reachability, and other routing information that is updated as part of the routing information exchange and may additionally contain information that is configured. The RDB may contain routing information for more than one routing area (RA).

Routing Components: ASON routing architecture functions. These functions can be classified as protocol independent (Link Resource Manager or LRM, Routing Controller or RC) or protocol specific (Protocol Controller or PC).

Routing Controller (RC): handles (abstract) information needed for routing and the routing information exchange with peering RCs by operating on the RDB. The RC has access to a view of the RDB. The RC is protocol independent.

Note: Since the RDB may contain routing information pertaining to multiple RAs (and possibly to multiple layer networks), the RCs accessing the RDB may share the routing information.

Link Resource Manager (LRM): supplies all the relevant component and TE link information to the RC. It informs the RC about any state changes of the link resources it controls.

Protocol Controller (PC): handles protocol-specific message exchanges according to the reference point over which the information is exchanged (e.g., E-NNI, I-NNI), and internal exchanges with the RC. The PC function is protocol dependent.

Authors' Addresses

Andrew G. Malis
Verizon Communications
117 West St.
Waltham MA 02451 USA

EMail: andrew.g.malis@verizon.com

Acee Lindem
Ericsson
102 Carric Bend Court
Cary, NC 27519

EMail: acee.lindem@ericsson.com

IETF
Internet Draft

Ping Pan
Mohana Srinivas
Rajan Rao
Biao Lu
(Infinera)
Sam Aldrin
(Huawei)

Expires: September 14, 2011

March 14, 2011

Supporting Shared Mesh Protection in MPLS-TP Networks

draft-pan-shared-mesh-protection-01.txt

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79. This document may not be modified, and derivative works of it may not be created, and it may not be published except as an Internet-Draft.

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79. This document may not be modified, and derivative works of it may not be created, except to publish it as an RFC and to translate it into languages other than English.

This document may contain material from IETF Documents or IETF Contributions published or made publicly available before November 10, 2008. The person(s) controlling the copyright in some of this material may not have granted the IETF Trust the right to allow modifications of such material outside the IETF Standards Process. Without obtaining an adequate license from the person(s) controlling the copyright in such materials, this document may not be modified outside the IETF Standards Process, and derivative works of it may not be created outside the IETF Standards Process, except to format it for publication as an RFC or to translate it into languages other than English.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that

other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at <http://www.ietf.org/ietf/lid-abstracts.txt>

The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>

This Internet-Draft will expire on September 14, 2011.

Copyright Notice

Copyright (c) 2011 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document.

Abstract

Shared mesh protection is a common protection and recovery mechanism in transport networks, where multiple paths can share the same set of network resources for protection purposes.

In the context of MPLS-TP, it has been explicitly requested as a part of the overall solution (Req. 67, 68 and 69 in RFC5654 [1]).

It's important to note that each MPLS-TP LSP may be associated with transport network resources. In event of network failure, it may require explicit activation on the protecting paths before switching user traffic over.

In this memo, we define a lightweight signaling mechanism for protecting path activation in shared mesh protection-enabled MPLS-TP networks.

Table of Contents

1. Introduction.....	3
2. Background.....	4
3. Problem Definition.....	5
4. Protection Switching.....	6
5. Activation Operation Overview.....	7
6. Protocol Definition.....	9
6.1. Activation Messages.....	9
6.2. Message Encapsulation.....	10
6.3. Reliable Messaging.....	12
6.4. Message Scoping.....	12
7. Processing Rules.....	13
7.1. Enable a protecting path.....	13
7.2. Disable a protecting path.....	13
7.3. Get protecting path status.....	14
7.4. Acknowledgement with STATUS.....	14
7.5. Preemption.....	14
8. Security Consideration.....	15
9. IANA Considerations.....	15
10. Normative References.....	15
11. Acknowledgments.....	15

1. Introduction

Shared mesh protection is a common protection and recovery mechanism in transport networks, where multiple paths can share the same set of network resources for protection purposes.

In the context of MPLS-TP, it has been explicitly requested as a part of the overall solution (Req. 67, 68 and 69 in RFC5654 [1]). Its operation has been further outlined in Section 4.7.6 of MPLS-TP Survivability Framework [2].

It's important to note that each MPLS-TP LSP may be associated with transport network resources. In event of network failure, it may require explicit activation on the protecting paths before switching user traffic over.

In this memo, we define a lightweight signaling mechanism for protecting path activation in shared mesh protection-enabled MPLS-TP networks.

Here are the key design goals:

1. Fast: The protocol is to activate the previously configured protecting paths in a timely fashion, with minimal transport and processing overhead. The goal is to support 50msec end-to-end traffic switch-over in large transport networks.
2. Reliable message delivery: Activation and deactivation operation have serious impact on user traffic. This requires the protocol to adapt a low-overhead reliable messaging mechanism.
3. Modular: Depending on deployment scenarios, the signaling may need to support functions such as preemption, resource re-allocation and bi-directional activation in a modular fashion.

Here are some of the conventions used in this document. The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC-2119 [RFC2119].

2. Background

Transport network protection can be typically categorized into three types:

Cold Standby: In this type of protection, the nodes will only negotiate and establish backup path after the detection of network failure.

Hot Standby: The protecting paths are established prior to network failure. This is also known as "make-before-break". Upon the detection of network failure, the edge nodes will switch data traffic into pre-established backup path immediately.

Warm Standby: The nodes will negotiate and reserve protecting path prior to network failure. However, data forwarding path will not be programmed. Upon the detection of network failure, the nodes will send explicit messages to relevant nodes to "wake up" the protecting path.

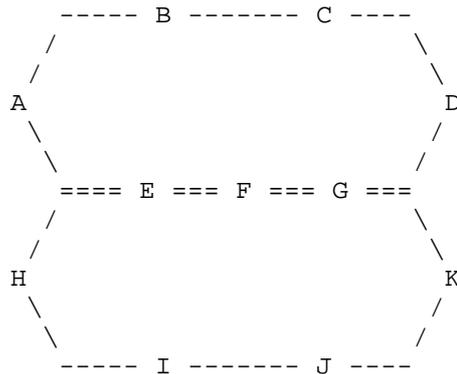
The activation signaling defined in this memo is to support warm standby in the context of MPLS-TP.

Further, the activation procedure may be triggered using the failure notification methods defined in MPLS-TP OAM specifications.

3. Problem Definition

In this section, we describe the operation of shared mesh protection in the context of MPLS-TP networks, and outline some of the relevant definitions.

We refer to the figure below for illustration:



Working paths: $X = \{A, B, C, D\}$, $Y = \{H, I, J, K\}$

Protecting paths: $X' = \{A, E, F, G, D\}$, $Y' = \{H, E, F, G, K\}$

The links between E, F and G are shared by both protecting paths. All paths are established via MPLS-TP control plane prior to network failure.

All paths are assumed to be bi-directional. An edge node is denoted as a headend or tailend for a particular path in accordance to the path setup direction.

Initially, the operators setup both working and protecting paths. During setup, the operators specify the network resources for each path.

The working path X and Y will configure the appropriate resources on the intermediate nodes, however, the protecting paths, X' and Y', will reserve the resources on the nodes, but won't occupy them.

Depending on network planning requirements (such as SRLG), X' and Y' may share the same set of resources on node E, F and G. The resource

assignment is a part of the control-plane CAC operation taking place on each node.

At some time, link B-C is cut. Node A will detect the outage, and initiate activation messages to bring up the protecting path X'. The intermediate nodes, E, F and G will program the switch fabric and configure the appropriate resources. Upon the completion of the activation, A will switch the user traffic to X'.

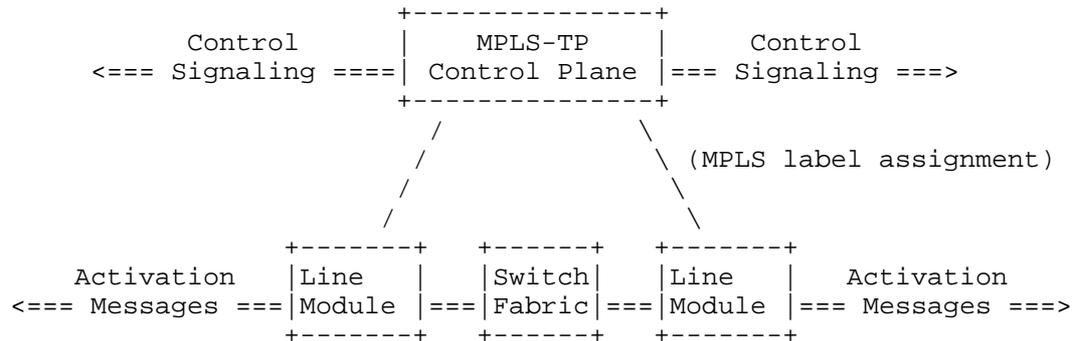
The operation may have extra caveat:

1. Preemption: Protecting paths X' and Y' may share the same resources on node E, F or G due to resource constraints. Y' has higher priority than that of X'. In the previous example, X' is up and running. When there is a link outage on I-J, H can activate its protecting path Y'. On E, F or G, Y' can take over the resources from X' for its own traffic. The behavior is acceptable with the condition that A should be notified about the preemption action.
2. Over-subscription (1:N): A unit of network resource may be reserved by one or multiple protecting paths. In the example, the network resources on E-F and F-G are shared by two protecting paths, X' and Y'. In deployment, the over-subscription ratio is an important factor on network resource utilization.

4. Protection Switching

The entire activation and switch-over operation need to be within the range of milliseconds to meet customer's expectation [1]. This section illustrates how this may be achieved on MPLS-TP-enabled transport switches. Note that this is for illustration of protection switching operation, not mandating the implementation itself.

The diagram below illustrates the operation.



Typical MPLS-TP user flows (or, LSP's) are bi-directional, and setup as co-routed or associated tunnels, with a MPLS label for each of the upstream and downstream traffic. On this particular type of transport switch, the control-plane can download the labels to the line modules. Subsequently, the line module will maintain a label lookup table on all working and protecting paths.

Upon the detection of network failure, the headend nodes will transmit activation messages along the MPLS LSP's. When receiving the messages, the line modules can locate the associated protecting path from the label lookup table, and perform activation procedure by programming the switching fabric directly. Upon its success, the line module will swap the label, and forward the activation messages to the next hop.

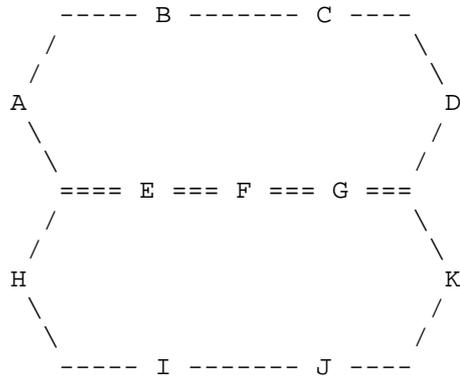
In summary, the activation procedure involves efficient path lookup and switch fabric re-programming.

To achieve the tight end-to-end switch-over budget, it's possible to implement the entire activation procedure with hardware-assistance (such as in FPGA or ASIC).

The activation messages are encapsulated with a MPLS-TP Generic Associated Channel Header (GACH) [3]. Detailed message encoding is explained in Section 6.

5. Activation Operation Overview

In this section, we describe the activation procedure using the same figure shown before:



Working paths: $X = \{A, B, C, D\}$, $Y = \{H, I, J, K\}$

Protecting paths: $X' = \{A, E, F, G, D\}$, $Y' = \{H, E, F, G, K\}$

Upon the detection of working path failure, the edge nodes, A, D, H and K may trigger the activation messages to activate the protecting paths, and redirect user traffic immediately after.

We assume that there is a consistent definition of priority levels among the paths throughout the network. At activation time, each node may rely on the priority levels to potentially preempt other paths.

When the nodes detect path preemption on a particular node, they should inform all relevant nodes to free the resources.

To optimize traffic protection and resource management, each headend should periodically poll the protecting paths about resource availability. The intermediate nodes have the option to inform the current resource utilization.

Note that, upon the detection of a working path failure, both headend and tailend may initiate the activation simultaneously (known as bi-directional activation). This may expedite the activation time. However, both headend and tailend nodes need to coordinate the order of protecting paths for activation, since there may be multiple protecting paths for each working path (i.e., 1:N protection). For clarity, we will describe the operation from headend in the memo. The tailend operation will be available in the subsequent revisions.

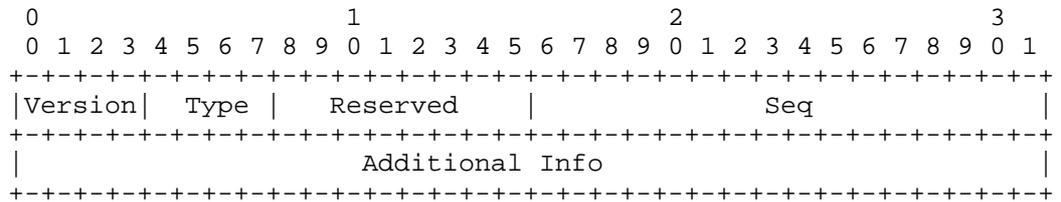
6. Protocol Definition

6.1. Activation Messages

The activation requires the following messages:

- o ENABLE: this is initiated by the headend nodes to activate a protecting path
- o DISABLE: this is initiated by the headend nodes to disable a protecting path and free the associated network resources
- o GET: this is initiated by the headend to gather resource availability information on a particular protecting path
- o NOTIFY: this is initiated by the intermediate nodes and terminate on the headend nodes to report preemption or protection failure conditions
- o STATUS: this is the acknowledgement message for ENABLE, DISABLE, GET, and NOTIFY messages, and contains the relevant status information

Each activation message has the following format:

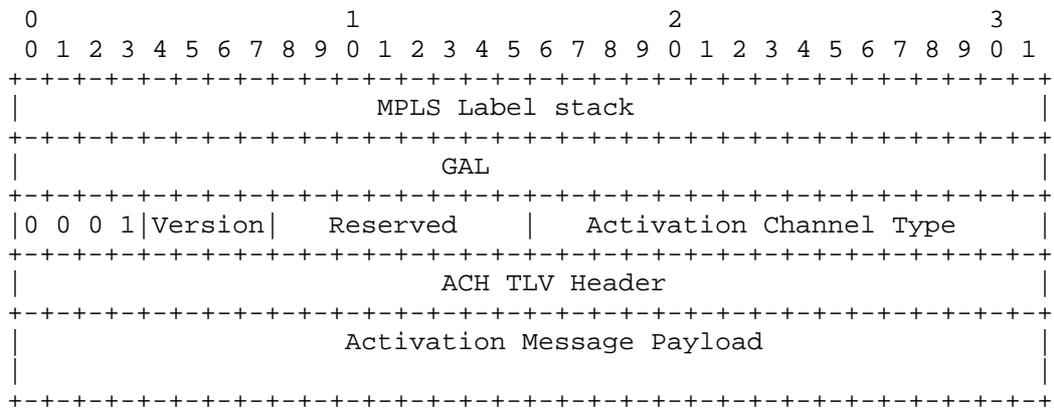


- o Version: 1
- o Type:
 - o ENABLE 1
 - o DISABLE 2
 - o GET 3
 - o STATUS 4

- o NOTIFY 5
- o Reserved: This field is reserved for future use
- o Seq: This uniquely identifies a particular message. This field is defined to support reliable message delivery
- o Additional Info: the message-specific data

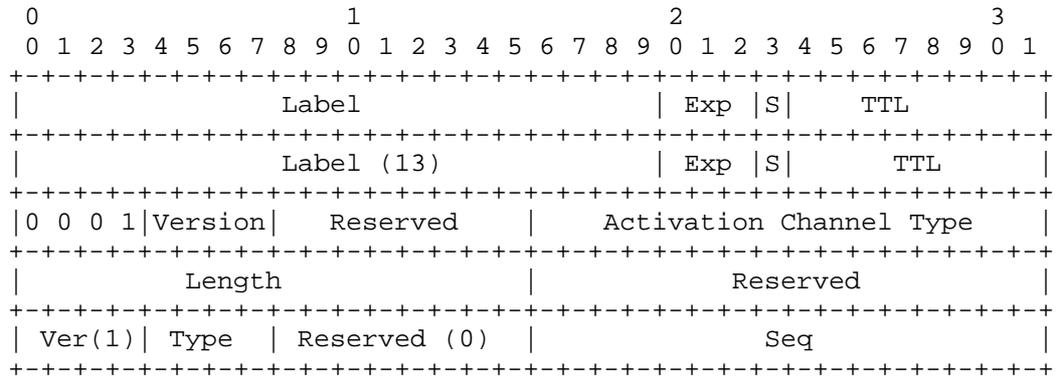
6.2. Message Encapsulation

Activation messages use MPLS labels to identify the paths. Further, the messages are encapsulated in GAL/GACH:

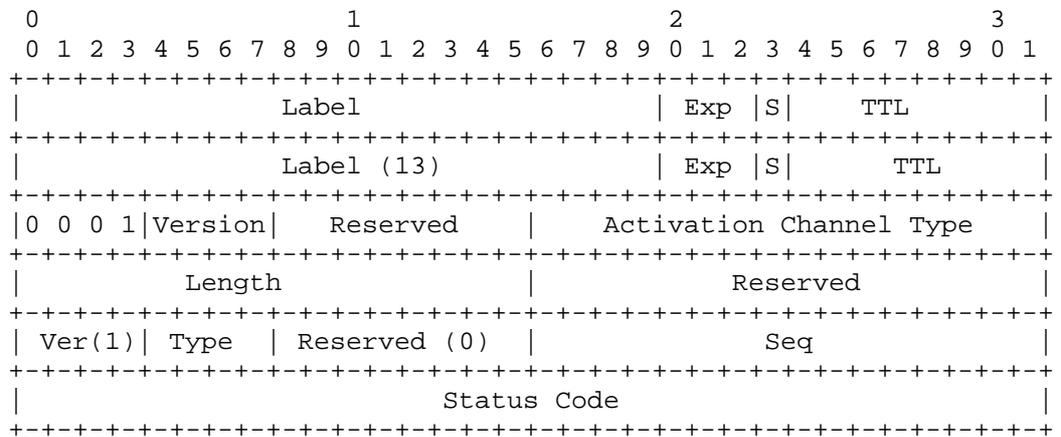


- o GAL is described in [3]
- o Activation Channel Type is the GACH channel number assigned to the protocol. This uniquely identifies the activation messages.
- o ACH TLV Header contains the message length, and is described in [3]

Specifically, ENABLE, DISABLE and GET messages have the following message format:



Both STATUS and NOTIFY messages have the following message format:



Currently, the status code used for acknowledging and preemption notification has the following definition:

- o 1xx: OK
 - . 101: end-to-end ack
- o 2xx: message processing errors
 - . 201: no such path
- o 3xx: processing issues:

- . 301: no more resource for the path
- . 302: preempted by another path
- . 303: system failure
- o 4xx: informative data:
 - . 401: shared resource has been taken by other paths

Further, for preemption notification, we may consider of using the existing MPLS-TP OAM messaging. More details will be available in the future revisions.

6.3. Reliable Messaging

The activation procedure adapts a simple two-way handshake reliable messaging.

Each node maintains a sequence number generator. Each new sending message will have a new sequence number. After sending a message, the node will wait for a response with the same sequence number.

Specifically, upon the generation of ENABLE, DISABLE, GET and NOTIFY messages, the message sender expects to receive a STATUS in reply with same sequence number.

If a sender is not getting the reply (STATUS) within a time interval, it will retransmit the same message with a new sequence number, and starts to wait again. After multiple retries (by default, 3), the sender will declare activation failure, and alarm the operators for further service.

6.4. Message Scoping

Activation signaling uses MPLS label TTL to control how far the message would traverse. Here are the processing rules on each intermediate node:

- o On receive, if the message has label TTL = 0, the node must drop the packet without further processing
- o The receiving node must always decrement the label TTL value by one. If TTL = 0 after the decrement, the node must process the message. Otherwise, the node must forward the message without further processing (unless, of course, the node is headend or tailend)

- o On transmission, the node will adjust the TTL value. For hop-by-hop messages, TTL = 1. Otherwise, TTL = 0xFF, by default.

7. Processing Rules

7.1. Enable a protecting path

Upon the detection of network failure on a working path, the headend node identifies the corresponding MPLS-TP label and initiates the protection switching by sending an ENABLE message.

ENABLE messages always use MPLS label TTL = 1 to force hop-by-hop process. Upon reception, a next-hop node will locate the corresponding path and activate the path.

If the Enable message is received on an intermediate node, due to label TTL expiry, the message is processed and then propagated to the next hop of the MPLS TP LSP, by setting the MPLS TP label TTL = 1. The intermediate node may NOT respond back to the headend node with STATUS message.

The headend node will declare the success of the activation only when it gets a positive reply from the tailend node. This requires that the tailend nodes must reply STATUS messages to the headend nodes in all cases.

If the headend node is not receiving the acknowledgement within a time interval, it will retransmit another ENABLE message with a different Seq number.

If the headend node is not receiving a positive reply within a longer time interval, it will declare activation failure.

If an intermediate node cannot activate a protecting path, it will reply an NOTIFY message to report failure. When the headend node receives a NOTIFY message for failure, it must initiate DISABLE messages to clean up networks resources on all the relevant nodes on the path.

7.2. Disable a protecting path

The headend removes the network resources on a path by sending DISABLE messages.

In the message, the MPLS label represents the path to be de-activated. The MPLS TTL is one to force hop-by-hop processing.

Upon reception, a node will de-activate the path, by freeing the resources from the data-plane.

As a part of the clean-up procedure, each DISABLE message must traverse through and be processed on all the nodes of the corresponding path. When the DISABLE message reaches to the tailend node, the tailend is required to reply with a STATUS message to the headend.

The de-activation process is complete when the headend receives the corresponding STATUS message from the tailend.

7.3. Get protecting path status

The operators have the option to trigger GET messages from the headend to check on the protecting path periodically or on-demand. The process procedure on each node is very similar to that of ENABLE messages on the intermediate nodes, except the GET messages should not trigger any network resource re-programming.

Upon reception, the node will check the availability of resources.

If the resource is no longer available, the node will reply a NOTIFY with error conditions.

7.4. Acknowledgement with STATUS

The STATUS message is the acknowledgement packet to all messages, and may be generated by any node in the network.

Each STATUS message must use the same sequence number as the corresponding message (ENABLE, DISABLE, GET and NOTIFY).

When replying to headend, the tailend nodes must originate STATUS messages with a large MPLS TTL value (0xff, by default).

7.5. Preemption

The preemption operation typically takes place when processing an ENABLE message.

If the activating network resources have been used by another path and carrying user traffic, the node needs to compare the priority levels.

If the existing path has higher priority, the node needs to reject the ENABLE message by sending a STATUS message to the corresponding headend to inform the unavailability of network resources.

If the new path has higher priority, the node will reallocate the resource to the new path, and send an NOTIFY message to old path's headend node to inform about the preemption.

8. Security Consideration

The protection activation takes place in a controlled networking environment. Nevertheless, it is expected that the edge nodes will encapsulate and transport external traffic into separated tunnels, and the intermediate nodes will never have to process them.

9. IANA Considerations

Activation messages are encapsulated in MPLS-TP with a specific GACH channel type that needs to be assigned by IANA.

10. Normative References

- [1] RFC 5654: Requirements of an MPLS Transport Profile, B. Niven-Jenkins, D. Brungard, M. Betts, N. Sprecher, S. Ueno, September 2009
- [2] IETF draft, Multiprotocol Label Switching Transport Profile Survivability Framework (draft-ietf-mpls-tp-survive-fw-06.txt), N. Sprecher, A. Farrel, June 2010
- [3] RFC5586 - Vigoureux,, M., Bocci, M., Swallow, G., Aggarwal, R., and D. Ward, "MPLS Generic Associated Channel", May 2009.
- [4] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [5] Crocker, D. and Overell, P.(Editors), "Augmented BNF for Syntax Specifications: ABNF", RFC 2234, Internet Mail Consortium and Demon Internet Ltd., November 1997.

11. Acknowledgments

Authors like to thank Eric Osborne, Lou Berger, Nabil Bitar and Deborah Brungard for detailed feedback on the earlier work, and the guidance and recommendation for this proposal.

We also thank Maneesh Jain, Mohit Misra, Yalin Wang, Ted Sprague, Ann Gui and Tony Jorgenson for discussion on network operation, feasibility and implementation methodology.

Authors' Addresses

Ping Pan
Email: ppan@infinera.com

Sri Mohana Satya Srinivas Singamsetty
Email: ssingamsetty@infinera.com

Rajan Rao
Email: rrao@infinera.com

Biao Lu
Email: blu@infinera.com

Sam Aldrin
Email: sam.aldrin@huawei.com

Internet Engineering Task Force
Internet-Draft
Intended status: Standards Track
Expires: January 12, 2012

P. Peloso, Ed.
Alcatel-Lucent
G. Martinelli
Cisco
J. Meuric
France Telecom
C. Margaria
Nokia Siemens Networks
July 11, 2011

OSPF-TE Extensions for WSON-specific Network Element Constraints
draft-peloso-ccamp-wson-ospf-oeo-04

Abstract

The original content of this internet draft was to propose some extensions to OSPF encoding in the context of Wavelength Switched Optical Networks, especially for internal constraints of optical network elements. General description can be found in the framework document.

This update of the document still aims at specifying the detailed structure of OSPF LSAs for WSONs. Nevertheless, the proposed LSA layout slightly differs from the current content of the information model and encodings drafts. As a result, the following sections highlight the differences between both approaches and summarize why the authors think these CCAMP's drafts would benefit from an update according to the proposed description.

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on January 12, 2012.

Copyright Notice

Copyright (c) 2011 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1.	Introduction	4
1.1.	Requirements Language	4
2.	Information Model	4
2.1.	Summary of Information Model Changes	5
2.2.	Node Information (WSON specific)	6
2.2.1.	Label Restrictions	6
2.2.2.	Resource Pools, Resource Blocks and Resource Description Containers	7
2.2.3.	Resource Pool Accessibility	11
2.2.4.	Resource Pool ID	12
2.2.5.	Resource Block State	12
2.2.6.	Resource Description	13
2.2.7.	Resource Pool Wavelength Constraints	14
2.2.8.	Shared Access Available Wavelengths	15
3.	Encoding	15
3.1.	Node related generic encodings	15
3.2.	Node related WSON specific encodings	16
3.2.1.	Label Restrictions	16
3.2.2.	Id Set Field	16
3.2.3.	Resource Pool Accessibility	17
3.2.4.	Resource Block State	18
3.2.5.	Resource Description	18
3.2.6.	Resource Pool Wavelength Constraints	21
3.2.7.	Shared Access Available Wavelengths	22
3.2.8.	Resource Pool	22
3.2.9.	Resource Description Container	23
3.3.	Link related encodings	23
4.	OSPF-TE Extensions	24
4.1.	Introduction	24
4.2.	Link top level TLV	25
4.3.	Node Attribute top level TLV	26
4.4.	Resource Pool top level TLV	26
4.5.	Resource Description Container top level TLV	26
4.5.1.	Resource Description sub-TLV	27
5.	Acknowledgements	27
6.	Contributors	27
7.	IANA Considerations	27
8.	Security Considerations	28
9.	References	28
9.1.	Normative References	28
9.2.	Informative References	29
Appendix A.	Solution(s) Evaluation	29
A.1.	RBNFs Comparison	30
A.2.	Depiction of the considered cases for evaluation	32
A.3.	Comparing evaluation of the solutions	34
Authors' Addresses	35

1. Introduction

The original content of this internet draft was to propose some extensions to OSPF encoding in the context of Wavelength Switched Optical Networks, especially for internal constraints of optical network elements. General description can be found in the framework document [RFC6163].

This update of the document still aims at specifying the detailed structure of OSPF LSAs for WSONs. Nevertheless, the proposed LSA layout slightly differs from the current content of the information model [I-D.ietf-ccamp-rwa-info] and encodings [I-D.ietf-ccamp-rwa-wson-encode] drafts. As a result, the following sections highlight the differences between both approaches and summarize why the authors think these CCAMP's drafts would benefit from an update according to the proposed description.

More specifically, the sections below follow the scope of current documents, namely information model, encodings and OSPF-TE extensions. Building the latter allowed to identify some improvements which are described in the two former parts. In both, the line has been drawn between the optical information that can be specified by using generic protocol extensions and the one requiring some WSON-specific objects, as agreed by the working group.

1.1. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

2. Information Model

This section provides a model of information needed by the routing and wavelength assignment (RWA) process in wavelength switched optical networks (WSONs). The purpose of the information described in this model is to facilitate constrained optical path computation in WSONs. This model takes into account compatibility constraints between WSON signal attributes and network elements but does not include constraints due to optical impairments.

The reduced Backus-Naur form (RBNF) syntax of [RFC5511] is used to aid in defining the RWA information model.

The text in the following reports every WSON information model modification compared to [I-D.ietf-ccamp-rwa-info]. Whenever a RBNF term is used without explicit definition we assume the same format

and semantic of the original information model.

An initial sub section here below reports a summary of changes introduced by this document.

2.1. Summary of Information Model Changes

In this document, most of the concepts and definitions from [I-D.ietf-ccamp-rwa-info] remain the same. For instance, a "Resource Block" is still "a group of devices with same features and same connectivity constraints".

Compared to the aforementioned document, the following main changes should be noticed:

1. The Resource Pool entity is introduced into the model, allowing the definition of several resource entities per node, which can be advertised independantly. A "Resource Pool" is defined as a group of resource blocks with same connectivity constraints. Several Resource Pools can be defined to associate them with different properties. The goal is to decrease the size of OSPF advertisement upon LSP changes (setup or tear down).
2. The connectivity matrix, defining the node capabilities on interconnection of external links, is used in order to describe connectivity constraints between node-external links and the resource pools. Two advantages can be stressed. First, it gathers all the static information into a node LSA, which OSPF-TE is not required to advertise upon LSP updates. Then it limits the number of connectivity representations introduced by [draft-ietf-rwa-info] (which proposes similar TLVs in different LSAs).
3. The scope of Resource Block Information is reduced, and focuses only on resource/device description. The described device are then efficiently instantiated by referring to these defined types. This allows to separate the physical resource characteristics from the way they are arranged in the node, thus having the description completely independent from the node design.

As a result, this method allows to share resource description for all the identical blocks of a node, thus decreasing the total size of information. Furthermore, as this information is very static and common to several resource blocks, it can be advertised and refreshed independently to any other information.

2.2. Node Information (WSON specific)

As presented in [RFC6163] a WSON node may contain electro-optical subsystems such as regenerators, wavelength converters or entire switching subsystems. The model present here can be used in characterizing the accessibility and availability of limited resources such as regenerators or wavelength converters as well as WSON signal attribute constraints of electro-optical subsystems. As such this information element is fairly specific to WSON technologies.

2.2.1. Label Restrictions

This section is a preamble presenting the Label Restriction entity, which is referred many times later in this document.

Wavelength constraint are used in different part of the information model, either as static constraints (in the resource pool as `RPWvlConstraints`, and the resource block `IngressWaveConstraint` and `EgressWaveConstraint`) or representing dynamic properties of a given element (`SharedAccessWvls` in resource pool). In the GMPLS context Wavelengths are physical instance of Labels.

The wavelength constraints used in this document, although having different semantic, refer to the same notion of list of wavelength. Those constraints apply in addition to either the incoming part of a device (or group of device), the outgoing part or both if the constraint is the same, which is for instance not unusual for static wavelength constraint.

To support this concept, this section defines a field:

`LABEL_RESTRICTIONS`

that carry a label set information and for which direction this label restriction is valid. The directions considered is upstream, downstream or both. The label set information is the one defined in [I-D.ietf-ccamp-rwa-info] as `AvailableLabel`.

This encoding is reused in different TLV or sub-TLV for different semantic but do not require to define a TLV per direction.

DELTA:

- Define a generic information for label restrictions
- Reuse generic label set and provide a compact representation

2.2.2. Resource Pools, Resource Blocks and Resource Description Containers

As presented in [RFC6163], a WSON node may include regenerators or wavelength converters arranged in shared pools, and can include OEO based WDM switches as well. There are plenty approaches used in the design of WDM switches containing regenerator or converters. However, from the point of view of path computation the following need to be known:

1. The nodes that support regeneration or wavelength conversion.
2. The accessibility and availability of OEO devices to convert from a given ingress wavelength on a particular ingress port to a desired egress wavelength on a particular egress port, which are summarized under the accessibility constraints.
3. Limitations on the types of signals that can be converted and the conversions that can be performed, namely the processing capabilities.

For modeling purposes and encoding efficiency regenerators or wavelength converters with identical limitations and/or processing and accessibility constraints are grouped into "blocks". Such blocks can consist of a single resource, though grouping resources into blocks leads to more efficient encodings. Then, these resource blocks are gathered once more into resource pool, for which the blocks share the same accessibility constraints. OEO devices sharing accessibility constraints are likely to being multiplexed on a given piece of equipment (like an Optical Amplifier, a splitter, a Wavelength Selective Switch port, a length of fiber...).

Definitions:

- Resource Block: A group of resources sharing both the same processing properties and the same accessibility constraints. Each Resource Block can contain a different number of resources, but all the resources constituting the block are identical devices.
- Resource Pool: A group of resources sharing the same accessibility constraints, hence a Resource Pool becomes a group of Resource Blocks sharing the same accessibility constraints. Each Resource Pool can contain a different number of blocks, each

of different size, as long as all the devices in the pool are subject to the same accessibility constraints regarding the way these are linked to ingress and egress links of the WSON node containing the pool.

The following picture represents the model of WSON nodes with the help of Resource Blocks and Resource Pools entities.

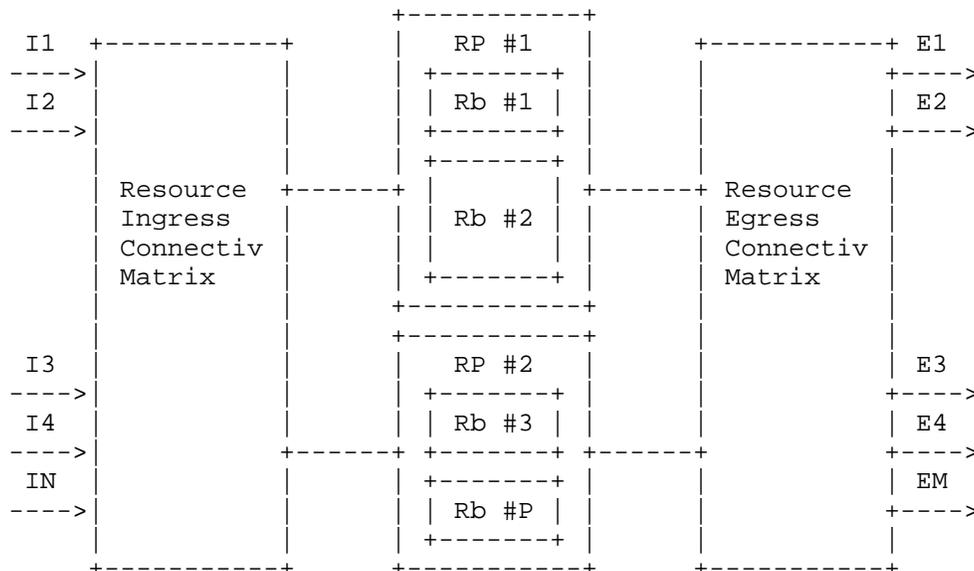


Figure 1

This figure shows a Resource Ingress Connectivity Matrix and another one of the egress, the model from [I-D.ietf-ccamp-rwa-info] gathers both these connectivity matrix inside a Resource Pool Accessibility item, which would lead to the following definition of a Resource Pool.

The following picture represents an abstracted model of the preceding node, that corresponds to the information model chosen in this document.

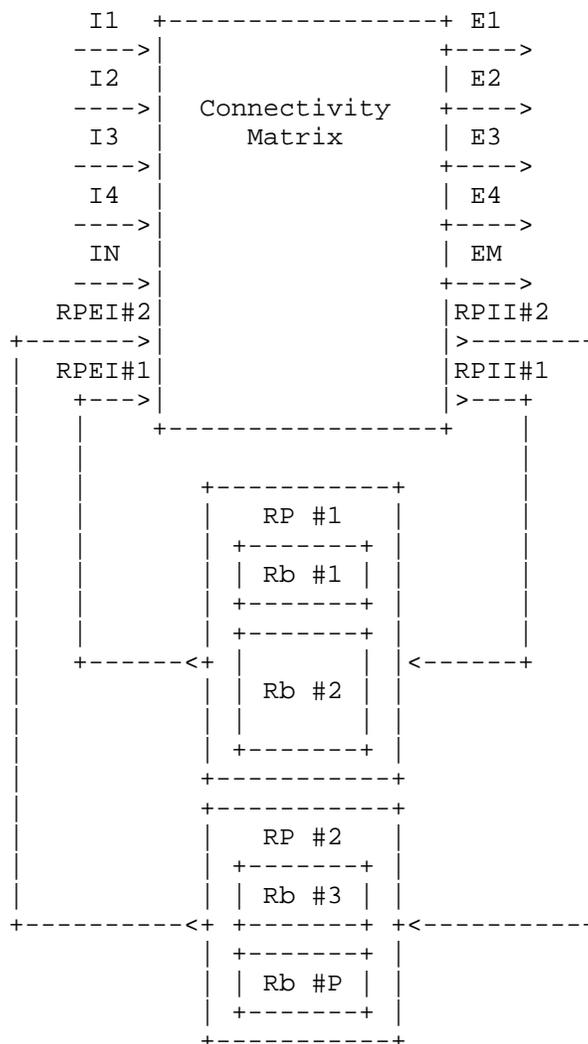


Figure 2

Since by definitions Resource Pools identify wavelength accessibility to regeneration resources, Section 2.2.3 details how to deal with accessibility constraints. This lead to the following definition of a Resource Pool.

```
<ResourcePool> ::= <ResourcePoolID> ([<SharedAccessWvls>]...)
                    ([<ResourceBlockState>]...)
                    ([<ResourcePoolWvlConstraints>]...)
```

- ResourcePoolID a unique (within node scope) number used to identify the pool,
- SharedAccessWvls represents the dynamic spectral availability coming from the usage of wavelengths by activated resources inside the pool,
- ResourceBlockStates are used to provide the dynamic availability of resources inside the pool.
- ResourcePoolWvlConstraints may be used to define the structural (static) spectral constraints of accessibility of the pool.

A WSON node having some OEO resource might have from 1 to P resource pools. The ResourcePool is created as an entity that will fit in a dedicated TLV (as sub-TLV) so the case of multiple Resource Pools will be handled by fitting one or more Resource Pool entity in each advertisement. The unique identifier ResourcePoolID allows to distinguish among all available pools.

As this document means to have one Resource Pool entity per physical pool of resources inside the node, inside a given node there is no reason for its pools not to share type of resources, hence their modeled representations refer to identical Resource Descriptions entities. In order to avoid unnecessary information flooding, this document gathers all these Resource Descriptions inside a dedicated entity, that is named Resource Description Container.

```
<ResourceDescriptionContainer> ::= <ResourceDescription>...
```

The Resource Description Container is a list of Resource Descriptions which, in turn, defines the features (i.e. physical characteristics) of each type of resources held inside the pools (of a given node).

DELTA:

- Introduced definition of Resource Pool.
- Introduced definition of Resource Pool ID.
- Introduced definition of Resource Description Container.
- Changed accordingly Figure 1 and 2 from [I-D.ietf-ccamp-rwa-info].

- Changed the RBNF from [I-D.ietf-ccamp-rwa-info].
- Changed the Resource Block Info into Resource Description (small semantic change, due to minor internal changes).
- Adapted some pieces of models which were related to Resource Block, to the Resource Pool level, namely: RPWvlConstraints

2.2.3. Resource Pool Accessibility

Every device inside a Resource Pool shares the same accessibility constraints, hence the accessibility is a property related to the pool. In order to depict the accessibility of a given pool, two pieces of information need to be described:

- Which ingress links of the node can be connected to the entry of the Resource Pool,
- Which egress links of the node can be connected to the exit of the Resource Pool.

Following remarks can be made concerning these accessibility information:

- These information share the same nature as the one of the Connectivity Matrix,
- These information are relatively static, changing only when the switching fabric of the node is changing (either failure or upgrade),

Hence, the accessibility information of every Resource Pool are embedded together inside the node own's Connectivity Matrix. The solution used to do that consists in using both Local Link Identifiers and Resource Pool Identifiers inside the Link Sets of the Connectivity Matrix. To keep unchanged the definition of the Link Set, 32 bits unnumbered IDs for the Resource Pool are needed (see Section 2.2.4). Thanks to this in the context of a node, the Connectivity Matrix is then providing associations between:

- On one side a set composed of a mix of: (1) ingress link(s) and (2) exit(s) of resource pool(s),
- On the other side a set composed of a mix of: (1) egress link(s) and (2) entry(ies) of resource pool(s).

Then the RBNF for the Connectivity Matrix becomes,

```
<ConnectivityMatrix> ::= <MatrixID> <ConnectType>
    (<IngressSetOfMixedLink&Pool> <EgressSetOfMixedLink&PoolSet>)...
```

The Resource Pool Accessibility information are optional, if not defined, Resource Pool is meant to have no accessibility constraints: from every node ingress port it's possible to reach the pool and the pool egress can reach every egress port of the node.

DELTA:

This section could be compared to the Resource Block Accessibility constraint, and this is a major change that is proposed here.

2.2.4. Resource Pool ID

In order to encode directly resource pools accessibility, inside the node's connectivity matrix, each Resource Pool needs to be identified alike an internal link with one ID on each side (ingress and egress), and then requires a Resource Pool ID. For each Resource Pool, WSON node assigns one identifier to each side of the pool. This identifier is a non-zero 32-bit number that is unique within the scope of the WSON node that assigns it, hence the Resource Pool ID is composed by a couple of unique numbers.

Consider a (resource) pool inside WSON node A. WSON node A chooses two distincts identifiers for the pool (one for the ingress side and one for the egress side). Considering these identifiers being unique inside the scope of the WSON node A, implies that: no other (resource) pool inside WSON node A may be assigned the value corresponding to any of these two identifiers, neither any (unnumbered) link between WSON node A and any other node may be assigned a link local identifier (from the WSON node A perspective) value corresponding to any of these two identifiers.

Support for resource pools in routing includes carrying information about the identifiers of these pools. Specifically, when an LSR advertises a resource pool, the advertisement carries both the ingress and the egress identifiers of the link.

```
<RPoolID> ::= <RESOURCE_INGRESS_ID> <RESOURCE_EGRESS_ID>
```

2.2.5. Resource Block State

The Resource Block State keep track of the current usage of a resource block within a resource pool.

The state indicate for the resource the number of available resources

and optionally the total number (or maximum number) of resources. decoupling ResourceDescription from the ResourceBlock configuration and allowing a better aggregation of the ResourceDescription. The state available in info model is the following:

Resource Block State definition

```
<ResourceBlockState> ::= <ResourceBlockID> [<CountMaxResources>]
    <CountAvailableResources>
```

DELTA:

This definition of the Resource Block State allow to separate the total number of resources from the resource description (differing in this from [I-D.ietf-ccamp-rwa-info]). This enable a sharing of the resource description between all the pools, while the other solution requires that each pool holds the same number of devices to share the same ResourceBlockDescription (see Section 2.2.6).

2.2.6. Resource Description

The resource block information contains the pieces of information needed to fully identify the resource block static and dynamic information. The static information consist of the characteristics that do not depend on the LSPs using the resource block. In particular the wavelength constraints are the one of the OEO and are independent of the LSPs. the static information is described by a ResourceDescription, which can be valid for several resource blocks, then referenced by their ResourceBlockID.

The ResourceBlockID identifies a resource block, it is a node wide stable and unique identifier (inside the node context). The ResourceBlockID is defined in the ResourceBlockState TLV held in the Resource Pool TLV and used in the Resource Description TLV.

```
<ResourceDescription> := <ResourceBlockID>... <InputConstraints>
    <ProcessingCapabilities> <OutputConstraints>
```

with,

```
<InputConstraints> ::= [<IngressWaveConstraint>] [<modulation-list>]
    [<fec-list>] [<rate-range-list>] [<client-signal-list>]
```

```
<ProcessingCapabilities> ::= <RegenerationCapabilities>
    [<FaultPerfMon>] [<VendorSpecific>]
```

```
<OutputConstraints> ::= [<EgressWaveConstraint>] [<modulation-list>]
    [<fec-list>]
```

IngressWaveConstraint and EgressWaveConstraint are described in Section 2.2.7. The modulation-list and fec-list represent the list of modulation formats and FEC encoding available within the resource block. This information MAY be present in the advertisement, the absence of this information means that potentially all Modulation and FEC are accepted and possible cranchback may occur.

DELTA:

- Split between static (can be in a separate LSA or in the resource pool) and dynamic information.
- The maximum number of resource is in the state to allow better summarization of the resourceDescription
- The static information is describing the properties, the ResourceDescription is more explicit than resourceInfo in this context
- Changed the RBNF from [I-D.ietf-ccamp-rwa-info], make use of generic label restriction for the wavelength restrictions.

2.2.7. Resource Pool Wavelength Constraints

This field defines any constraint at wavelength level within a resource pool, and is meaningful only when a subset of wavelengths could be configurable within the Pool. This information is static since it depends on specific physical resources within the pools and changes only if there is a node reconfiguration (OEO pools added or removed from an optical node, change in the mux or demuxing devices). As there is an ingress side and an egress side of a pool, this item needs to modelize the wavelength usage on each side.

This field takes the format of a Label_Restrictions Section 2.2.1. At most two instances of this item can be needed: one for each sides (incoming / outgoing) of the pool.

The field is optional, when this field is not present it means there are no specific wavelength constraints imposed by pool. As an example this field is equivalent to the Maximum Bandwidth field defined within [RFC3630]. As the Maximum Bandwidth represents the true link capacity, the RESOURCE_POOL_WAVELENGTH_CONSTRAINTS represent the set of wavelengths that can possibly be configured on

the pool.

Note that the usable set of wavelengths could be limited by other constraints: e.g. currently in-use wavelength (see Section 2.2.8) or due to OEO device constraint on compliant wavelengths (see Wavelength Constraints in Section 2.2.6).

DELTA:

Only wavelength constrain. While physical constraints are grouped in another set.

2.2.8. Shared Access Available Wavelengths

The SHARED_ACCESS_AVAILABLE_WAVELENGTHS represents wavelength usage in a Resource Pool hence it is related with the Resource Pool dynamic state.

If a wavelength is in use within a pool, the same wavelength cannot be reused in the same pool however the pool will be available for a different wavelength depending on free resource blocks (Resource Pool definition as in Section 2.2.2). As there is an ingress side and egress side of a pool, this item needs to modelize the wavelength usage on each side. Hence, this representation automatically considers the case of wavelength conversion happening inside the pool.

This field takes the format of a Label_Restrictions Section 2.2.1. At most two instances of this item can be needed: one for each sides (incoming / outgoing) of the pool.

N.B.: Hence, SHARED_ACCESS_AVAILABLE_WAVELENGTHS has the same format as RESOURCE_POOL_WAVELENGTH_CONSTRAINTS defined in Section 2.2.7.

DELTA:

Only wavelength constraint. While physical constraints are grouped in another set.

3. Encoding

3.1. Node related generic encodings

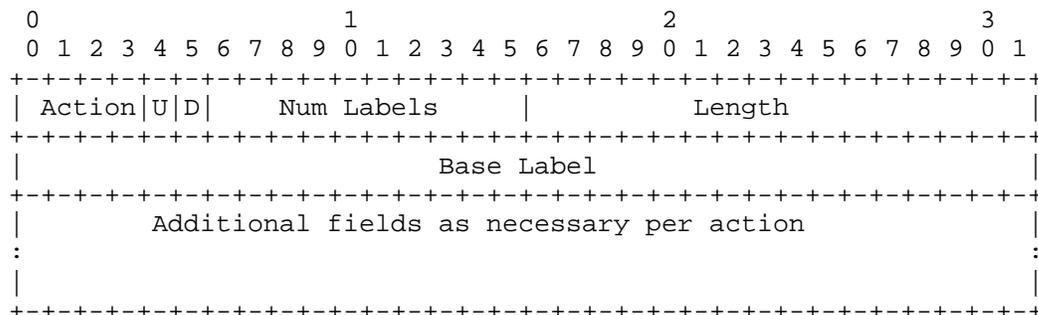
In this section we propose modification to [I-D.ietf-ccamp-general-constraint-encode].

3.2. Node related WSON specific encodings

This section refer to [I-D.ietf-ccamp-rwa-wson-encode]

3.2.1. Label Restrictions

Relatively to section 2.2 of [I-D.ietf-ccamp-general-constraint-encode] the LABEL_SET field is here slightly modified in order to define a Label Restrictions field.



Although it make sense only using the actions 0-Inclusive List, 2-Inclusive Range or 4-Bitmap. The U bit indicate a label set restriction valid at the upstream direction/incoming side of a resource pool/resource block. The D bit indicate a label set restriction valid at the downstream/outgoing side of a resource pool/resource block. At least one of U or D bit MUST be set, both U and D bit MAY be set.

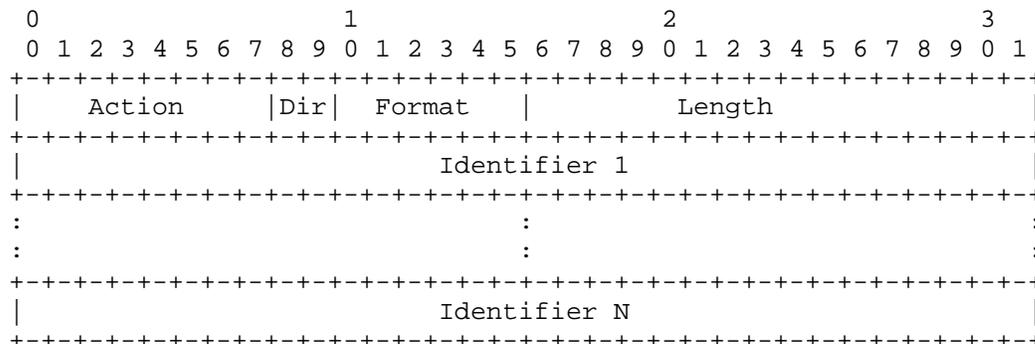
 DELTA:

The Num Labels field become 10 bits and this leave room for 1024 labels represented by this encoding. This encoding will be reused in specific TLVs, in case more than 1024 labels are needed multiple fields within TLVs can be used.

3.2.2. Id Set Field

With the introduction of resource description describing properties for a group of resource block we need to efficiently represent a set of IDs. To do so we introduce an IDSet field which has the same encoding as the LinkSet field defined in [I-D.ietf-ccamp-general-constraint-encode] but with a more generic description.

ID Set Field



The Action, Dir have the same encoding as in [I-D.ietf-ccamp-general-constraint-encode]. The Format field indicates the format and length of the Identifier:

- 0 -- 32 Bit unnumbered identifier
- 1 -- IPv4 identifier
- 2 -- IPv6 identifier

This field is used later to define a set of resource blocks (e.g. to list the resource blocks sharing the same resource description).

3.2.3. Resource Pool Accessibility

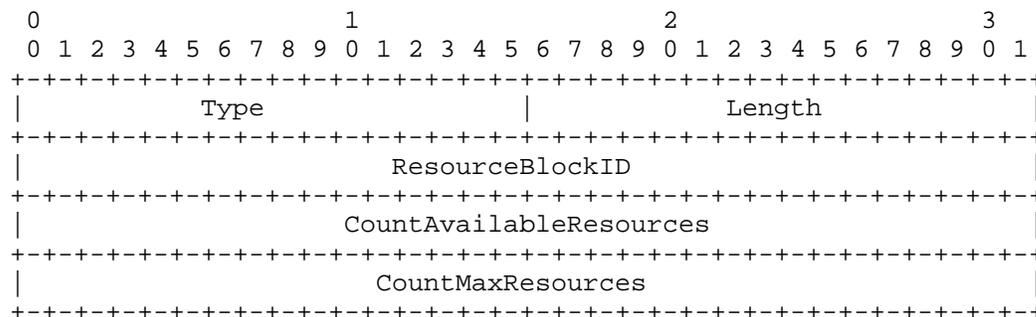
The Resource Pool Accessibility needs no encoding of its own. As explained in the Section 2.2.3 this piece of information is merged inside the Connectivity Matrix object which is actually not impacted by this solution.

Nota: The Link Sets held inside the Connectivity Matrix are composed of LINK_LOCAL_IDENTIFIERS (32 bits identifiers), and the solution to describe the Resource Pool Accessibility consists in using either RESOURCE_INGRESS_ID or RESOURCE_EGRESS_ID (also 32 bits identifiers) which are by definition different from the LINK_LOCAL_IDENTIFIERS (see Section 2.2.4).

DELTA: A major change here as the content of this field are moved inside Connectivity Matrix.

3.2.4. Resource Block State

This TLV indicate the state of a resource block as defined in Section 2.2.5. It defines the ResourceBlockId, and provides the number of free resources and maximum in this resource block. The ResourceBlockID field is a 32 bit node-wide identifier,



The information of the maximum number of resource is optional, this is encoded with a value of 0 in the CountMaxResource field, or with a Length value set to 8 instead of 12.

DELTA:

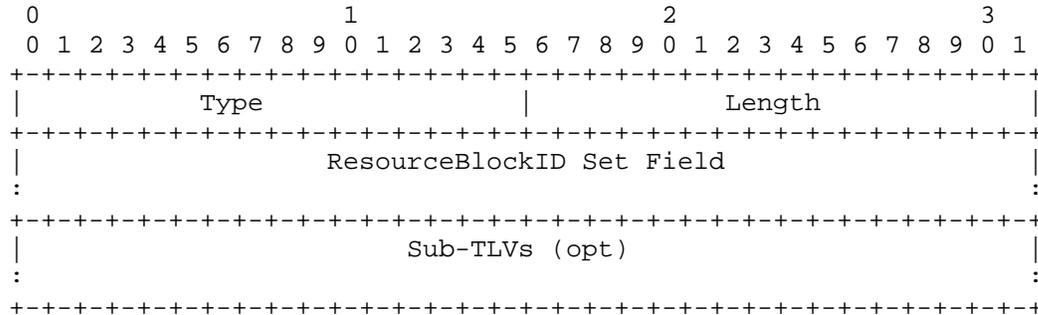
This is an adaptation of the resource pool status that fits the new definition of resource description.

3.2.5. Resource Description

Resource Description sub-TLVs represent the information described in Section 2.2.6.

The resource description TLV encoding follow the definition from Section 2.2.6 with a list of sub-sub TLV.

Resource Description TLV



The ResourceBlockID Set Field is encoded using the IDSet field encoding using the ResourceBlockID as identifier with format 0.

The Sub-Sub TLVs are defined as follow, the order does not matter. Each of the Sub-Sub-TLV defined in this document MAY be repeated more than once, on receipt all Sub-Sub-TLV MUST be taken into account. The resulting information is the union of all the element of the Sub-Sub-TLVs (all Sub-Sub-TLVs of this document describe lists). For example an implementation may choose to indicate that in total 4 label can be used as 4 Label constraint Sub-Sub-tlv, each of them with 1 label.

Info model	Type	Encoding
IngressWaveConstraint	Label Constraints	Label restriction, see Section 3.2.1.
Input modulation-list	Modulation List	A list of Modulation Format Fields, described in [I-D.ietf-ccamp-rwa-wson-encode] section 4.2.1.
Input fec-list	FEC List	A list of FEC type, described in [I-D.ietf-ccamp-rwa-wson-encode] section 4.3.1.
Input rate-range-list	Rate Range	A list of rate range field, described in [I-D.ietf-ccamp-rwa-wson-encode] section 4.4.1.
Input client-signal-list	Client Signal List	A list of GPids, described in [I-D.ietf-ccamp-rwa-wson-encode] section 4.5.

ProcessingCapabilities	Processing Capabilities	A list of Processing Capabilities Fields, except processing cap "Number of Resources", described in [I-D.ietf-ccamp-rwa-wson-encode] section 4.6.1.
EgressWaveConstraint	Label Constraints	Label restriction, see Section 3.2.1.
Output modulation-list	Modulation List	see Input modulation-list
Output fec-list	FEC List	see Input fec-list

Resource description Sub-Sub-TLVs and relation to info model

The Label Constraints Sub-Sub-TLV is used for IngressWaveConstraint and EgressWaveConstraint as the Label Restriction field carries the U and D bit to allow to distinguish a label restriction valid for incoming, outgoing or both.

The Modulation List Sub-Sub-TLV is similarly used for the input and output modulation list. The Sub-Sub-TLV contains a list of Modulation format field, which indicate if they are valid for the input (I bit set to 1) or for the output (I bit cleared). The list of Modulation format field MUST contain at least one ingress FEC modulation format. If no Egress modulation format is present in the list it is implied that no modulation format conversion is impossible, the egress modulation list is the same as the ingress modulation list and modulation format is not performed.

The FEC list Sub-Sub-TLV is also representing both Input and Output FEC list. The Sub-Sub-TLV is defined as a list of FEC Fields, conceptually being Sub-Sub-Sub-TLVs indicating via the I bit if they are valid for ingress or egress. At least one ingress FEC MUST be present in the list, if no egress modulation format is present in the list it is implied that the egress FEC list is the same as the ingress FEC list. In such case FEC format conversion MAY be performed.

The Processing Capabilities Sub-Sub-TLV is the same as in [I-D.ietf-ccamp-rwa-wson-encode] section 4.6.1. except for the maximum number of resource which is represented in the ResourceBlockState. The FEC and Modulation format conversion capabilities are expressed via the Modulation and FEC list by not including any egress modulation/fec in the respective lists.

Bit-Rate Range and Client Signal lists are unchanged from [I-D.ietf-ccamp-rwa-wson-encode]

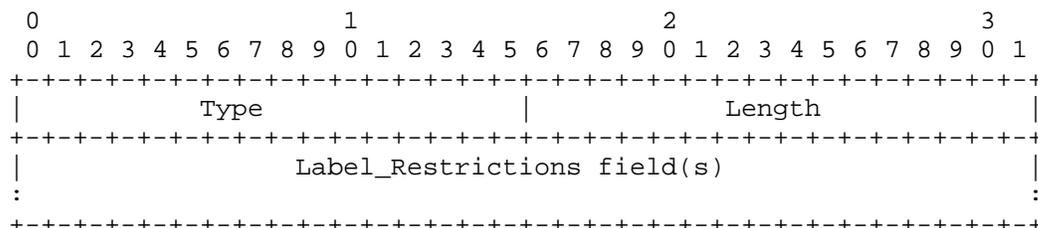
 DELTA:

- use a common TLV for the label restriction
- use a common TLV for the FEC list
- use a common TLV for the Modulation format list
- re-use indirectly (via ID Set) the general encoding LinkSet for RBlockId set
- More explicit statement on FEC and Modulation format conversion capabilities

3.2.6. Resource Pool Wavelength Constraints

This TLV is used to describe static wavelength constraint, it follows the encoding of Label_Restrictions field Section 3.2.1

RESOURCE_POOL_WAVELENGTH_CONSTRAINTS TLV



The Label_Restrictions field might be repeated several times depending on the U and D bit flags. In case multiple fields with the same U and D bits set to 1, the final resulting constrain will be the intersection of all Label_Restrictions. If multiple TLVs are present the resulting constraint is the intersection of all the TLV.

Example below:

- No RESOURCE_POOL_WAVELENGTH_CONSTRAINTS TLV meaning that these type of constraints are not described.
- A TLV present with one Label_Restrictions field with both the U or D bits MUST be set to 1. Which means the same constrains apply to both sides of the pool.

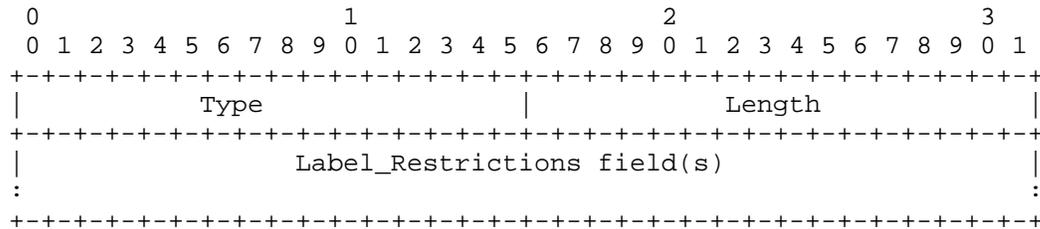
- A TLV present with three Label_Restrictions field presents, one field with U=1 so applicable upstream. The two other fields with D=1 so applicable downstream

DELTA: Small delta, just using the add-on bits to provide a direction/side semantic.

3.2.7. Shared Access Available Wavelengths

This TLV is used to describe dynamic wavelength availability, it follows the encoding of Label_Restrictions field. Section 3.2.1

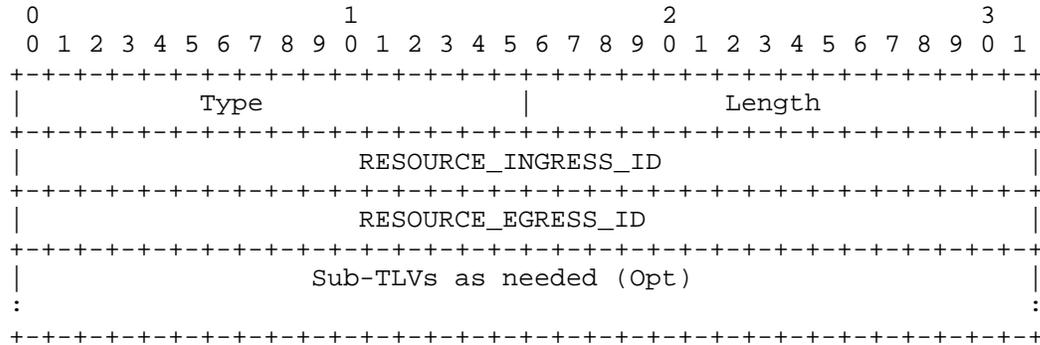
SHARED_ACCESS_AVAILABLE_WAVELENGTH TLV



The same rules and usage defined in Section 3.2.6 apply here.

3.2.8. Resource Pool

The RESOURCE_POOL TLV contains the preceding TLVs.



List of possible Sub-TLVs:

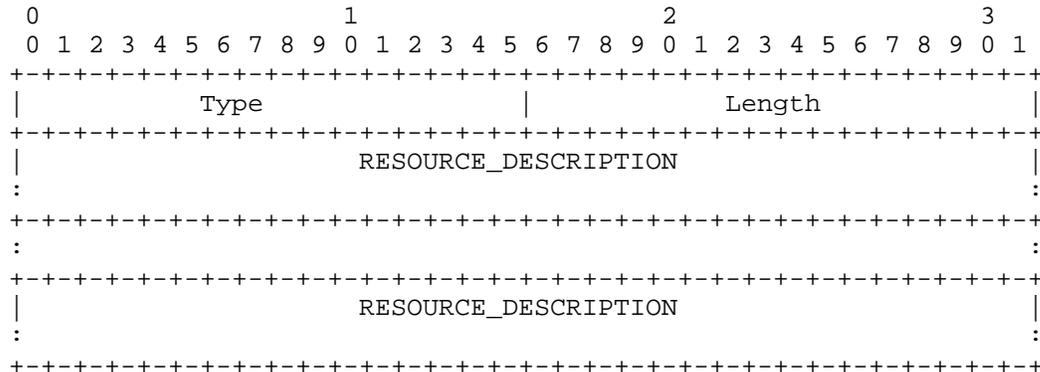
Name	Static/Dynamic
Resource Block State	Dynamic
Shared Access Available Wavelength	Dynamic
Resource Pool Wavelength Constraints	Static

DELTA:

Similar to Resource Pool inside [I-D.ietf-ccamp-rwa-wson-encode] with a different internal layout that allows for multiple instances.

3.2.9. Resource Description Container

The RESOURCE_DESCRIPTION_CONTAINER is a list of RESOURCE_DESCRIPTION. This one MAY be used to extract the static content of the previous TLV, in order to hold all this content inside this purposely defined static TLV. Then each one can be in separatly flooded entities (e.g. in separated LSAs see Section 4.1.



DELTA:

New item.

3.3. Link related encodings

This section does not differ from the equivalent in [I-D.ietf-ccamp-general-constraint-encode]

4. OSPF-TE Extensions

This section handles OSPF-TE extensions.

It starts with introducing the top view of the extensions provided by this draft. Then a sub-section dedicated for each top level TLV details the extensions relevant for this top level TLV.

4.1. Introduction

This introduction provides the layout of the preceding information model (Section 2) and encodings (Section 3) into top-level TLVs of opaque LSAs.

[RFC3630] introduces Link top level TLV (type 2). This document extends its content with the encodings depicted in Section 3.3. These extensions offer the capability to advertise restrictions on the list of available labels.

N.B.: This capability is specifically useful when these labels have a network wide semantic like suggested in a WSON context.

[RFC5786] introduces Node Attribute top level TLV (type 5). This document extends its content with the encodings depicted in Section 3.1. These extensions offer the capability to advertise restrictions on the switching capabilities of the node.

N.B.: This TLV is unique for a given node and contains static information only, hence no more than one LSA per node is expected to host such a TLV.

This document introduces a new top level TLV named RESOURCE_POOL (type value to be defined), which encodings are depicted in Section 3.2. RESOURCE_POOL TLV offers the capability to advertise one or multiple pools of OEO devices held in a given node. This object can carry resource descriptions, the available resources inside the pool(s) and the availability of wavelengths to reach the pool (refer to pool definition inside Section 2.2.2).

N.B.: A LSA can contain more than one RESOURCE_POOL top level TLV (allowing one LSA to advertise the description of all the pools of the originating node). Alternatively, a node can originate more than one LSA containing each RESOURCE_POOL top level TLVs (allowing each LSA to advertise an individual pool). In that case all the RESOURCE_POOL originated by the same node MUST have different RESOURCE_POOL_ID. As most of the information contained inside a RESOURCE_POOL are dynamic, an implementer may well choose to define one LSA per pool of resources in order to reduce the quantity of information flooded upon change in resource usage.

This document introduces another new top level TLV named

RESOURCE_DESCRIPTION_CONTAINER (type value to be defined), which encoding is depicted in Section 3.2.9.

RESOURCE_DESCRIPTION_CONTAINER TLV contains a list of RESOURCE_DESCRIPTION valid in the scope of the originating node. A given node MUST NOT originate more than one LSA containing RESOURCE_DESCRIPTION_CONTAINER TLV. An LSA containing a RESOURCE_DESCRIPTION_CONTAINER TLV MUST NOT contain any additional top level TLV.

N.B.: This TLV is designed to be unique in the scope of the originating node and to gather all the resource descriptions relevant in this scope.

Summarizing Table

Top-TLV	Type	Name	Instances	Static/Dynamic
2	Link		1 per fiber	Mix
5	Node Attribute		1 per Node	Static
TBD	Resource Pool		1 per Pool	Dynamic
TBD	Resource Desc Cont		1 per Node	Static

DELTA:

- Renamed the Node Optical Property tlv into Resource Pool TLV
- Allow multiple instance of Resource Pool TLV
- Introduced an optional new TLV named Resource Description

4.2. Link top level TLV

This section refer to [I-D.ietf-ccamp-gmpls-general-constraints-ospf-te].

The following new sub-TLVs are added to the Link top level TLV (type 2).

Sub-TLV Type	Length	Name
TBD	variable	Port Label Restrictions
TBD	variable	Available Wavelengths
TBD	variable	Shared Backup Wavelengths

In Link TLV, all the sub-TLV listed above are optional.

4.3. Node Attribute top level TLV

This section refer to [I-D.ietf-ccamp-gmpls-general-constraints-ospf-te].

The following new sub-TLVs are added to the Node Attribute top level TLV (type 5).

Sub-TLV Type	Length	Name
TBD	variable	Connectivity Matrix
TBD	variable	Port Label Restrictions
TBD	variable	Shared Risk Node Group

In Node Attribute, all the sub-TLV listed above are optional. None of them contain sub-TLV.

4.4. Resource Pool top level TLV

This section refer to [I-D.ietf-ccamp-wson-signal-compatibility-ospf]

The following sub-TLVs are created for the Resource Pool top level TLV.

Sub-TLV Type	Length	Name
TBD	variable	Resource Block State
TBD	variable	Shared Access Available Wavelength
TBD	variable	Resource Pool Wavelength Constraints

In Resource Pool, all the sub-TLV listed above are optional.

4.5. Resource Description Container top level TLV

This section refer to [I-D.ietf-ccamp-wson-signal-compatibility-ospf]

The following sub-TLVs are created for the Resource Description Container top level TLV.

Sub-TLV Type	Length	Name
TBD	variable	Resource Description

4.5.1. Resource Description sub-TLV

The following sub-TLVs are created for the Resource Pool top level TLV.

Sub-TLV Type	Length	Name
TBD	variable	Modulation List
TBD	variable	FEC List
TBD	variable	Rate Range List
TBD	variable	Client Signal List
TBD	variable	Processing Capabilities
TBD	variable	Label Constraints

In Resource Description, all the sub-TLV listed above are optional.

5. Acknowledgements

This template was derived from an initial version written by Pekka Savola and contributed by him to the xml2rfc project.

This document shares common material with the documents quoted, which seems fair as the target of this version is to highlight differences.

The editors wish to thank Ramon Casellas for his constructive comments.

6. Contributors

Daniele Ceccarelli
Ericsson
Via A. Negrone 1/A
Genova - Sestri Ponente
Italy

Email: daniele.ceccarelli@ericsson.com

7. IANA Considerations

This memo requires many requests to IANA, which will be completed in a latter version.

8. Security Considerations

All drafts are required to have a security considerations section. See RFC 3552 [RFC3552] for a guide.

9. References

9.1. Normative References

- [I-D.ietf-ccamp-general-constraint-encode]
Bernstein, G., "General Network Element Constraint Encoding for GMPLS Controlled Networks", draft-ietf-ccamp-general-constraint-encode-04 (work in progress), December 2010.
- [I-D.ietf-ccamp-gmpls-general-constraints-ospf-te]
Zhang, F., Lee, Y., Han, J., Bernstein, G., Xu, Y., Zhang, G., Li, D., Chen, M., and Y. Ye, "OSPF-TE Extensions for General Network Element Constraints", draft-ietf-ccamp-gmpls-general-constraints-ospf-te-00 (work in progress), March 2011.
- [I-D.ietf-ccamp-rwa-wson-encode]
Bernstein, G., Lee, Y., Li, D., Imajuku, W., and J. Han, "Routing and Wavelength Assignment Information Encoding for Wavelength Switched Optical Networks", draft-ietf-ccamp-rwa-wson-encode-11 (work in progress), March 2011.
- [I-D.ietf-ccamp-wson-signal-compatibility-ospf]
Lee, Y. and G. Bernstein, "OSPF Enhancement for Signal and Network Element Compatibility for Wavelength Switched Optical Networks", draft-ietf-ccamp-wson-signal-compatibility-ospf-04 (work in progress), March 2011.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC2629] Rose, M., "Writing I-Ds and RFCs using XML", RFC 2629, June 1999.
- [RFC3552] Rescorla, E. and B. Korver, "Guidelines for Writing RFC Text on Security Considerations", BCP 72, RFC 3552, July 2003.
- [RFC3630] Katz, D., Kompella, K., and D. Yeung, "Traffic Engineering

(TE) Extensions to OSPF Version 2", RFC 3630, September 2003.

- [RFC4202] Kompella, K. and Y. Rekhter, "Routing Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4202, October 2005.
- [RFC5511] Farrel, A., "Routing Backus-Naur Form (RBNF): A Syntax Used to Form Encoding Rules in Various Routing Protocol Specifications", RFC 5511, April 2009.
- [RFC5786] Aggarwal, R. and K. Kompella, "Advertising a Router's Local Addresses in OSPF Traffic Engineering (TE) Extensions", RFC 5786, March 2010.

9.2. Informative References

- [I-D.ietf-ccamp-rwa-info] Bernstein, G., Lee, Y., Li, D., and W. Imajuku, "Routing and Wavelength Assignment Information Model for Wavelength Switched Optical Networks", draft-ietf-ccamp-rwa-info-11 (work in progress), March 2011.
- [I-D.narten-iana-considerations-rfc2434bis] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", draft-narten-iana-considerations-rfc2434bis-09 (work in progress), March 2008.
- [RFC6163] Lee, Y., Bernstein, G., and W. Imajuku, "Framework for GMPLS and Path Computation Element (PCE) Control of Wavelength Switched Optical Networks (WSONs)", RFC 6163, April 2011.

Appendix A. Solution(s) Evaluation

Within this section we try evaluate the amount of information that needs to be exchanged through routing advertisements. For this evaluation we consider some minimum optical node reference design to make a OEO extension future proof.

This sections starts with summarizing the LSAs needed to depict a node with both the solution depicted by this document and the solution depicted by [I-D.ietf-ccamp-rwa-info]. Afterwards, the hypothesis concerning the node features that will serve as a basis for the solution evaluation will be presented, before the actual results of the solutions evaluations (both the one of this document

and the one of [I-D.ietf-ccamp-rwa-info]).

A.1. RBNFs Comparison

In this section we try compare the how TLVs are composed according two this draft proposal versus existing WSON solutions. The goal here is to provide the all reference for and easy understanding where two solutions are different. Numbers will be provided in the next section.

The evaluation will be done on the Resource Pool top-level TLV since the Node address and Link TLV are considered equivalent.

WSON Drafts. According to [I-D.ietf-ccamp-wson-signal-compatibility-ospf] in section 2 defines the Optical Node Property TLV which collect the WSON specific information. This TLV is composed of the following:

```
<ResourcePool> ::= [<ResourceBlockInformation>]...
  [<ResourceBlockAccessibility>]... [<ResourceBlockWvlConstraint>]...
  [<ResourceBlockPoolState>...] [<SharedAccessWvls>...]
```

- a) Resource Block Information. Defined as : ([<ResourceSet>] <InputConstraints> <ProcessingCapabilities> <OutputConstraints>). A resource block information defines here the number of devices inside the block.
- b) Resource Block Accessibility. Defined as (<PoolIngressMatrix> <PoolEgressMatrix>) which is expanded in tuples like (<INGRESS_LINK_SET><ResourceSet>)* and (<EGRESS_LINK_SET><ResourceSet>)*. Note that INGRESS/EGRESS_LINK_SET is a name defined here for the link set field used in the [I-D.ietf-ccamp-rwa-info] document.
- c) Resource Block Wavelength Constraints. Defined as <IngressWaveConstraints><EgressWaveConstraints>. This is expanded in <ResourceSet>INPUT_WAVELENGTH_SET OUTPUT_WAVELENGTH_SET, for the static constraints of resource blocks.
- d) Shared Access Wavelengths. Defined as <IngressWaveConstraints><EgressWaveConstraints>. This is expanded in <ResourceSet>INPUT_WAVELENGTH_SET OUTPUT_WAVELENGTH_SET, for the shared fibers between blocks.

- e) Resource Block Pool State. <ResourceSet> <USAGE_STATE_BITMAP>

In current proposal there are two types of TLV.

First the Resource Pool TLV (with an instance per pool) is composed of the following information:

```
<ResourcePool> ::= <ResourcePoolID> [<ResourceDescription>]...
  [<ResourcePoolWvlConstraints>]... [<SharedAccessWvls>]...
  [<ResourceBlockState>]...
```

- a) Resource Description. Which is defined as: (<RBlockID>...) <InputConstraints> <ProcessingCapabilities> <OutputConstraints>. This is equivalent to the item a) above without the number of devices inside the resource block, which allow this definition to be usable by any block. The number of available resource of a given type inside the pool being specified by the Resource Block State below. When a Resource Description Container TLV is defined by a Node, the Resource Pool TLV of this same node SHOULD NOT contain any Resource Description sub-TLV.
- b) Resource Block State. Where RBlockState is defined as <RBlockID> [<NumResources>] <NumberOfAvailableResources>. This field efficiently report how many of a given resource type is available inside the pool or not.
- c) Shared Access Available Wavelength. This is composed of a Label Restriction field and SHOULD used to depict the dynamic constraints of the pool.
- d) Resource Pool Wavelength Constraints. This is composed of a Label Restriction field and MAY be used to depict the static constraints of the pool.

Second the Resource Descriptor Container TLV (with a single instance per node) is used to gather all the Resource Descriptions of a given node, as these are static information composed of the following information:

```
<ResourceDescriptionContainer> ::= <ResourceDescription>...
```

- a) Resource Description. Which is defined as: (<RBlockID>...) <InputConstraints> <ProcessingCapabilities> <OutputConstraints>. This is equivalent to the item a) above.

A.2. Depiction of the considered cases for evaluation

For the sake of the comparison we have considered the following parameters and values characterizing the optical node design:

- o Node Degree Connectivity: 4, 8 and 16.
- o WDM capacity: 100 wavelengths.
- o Switching capacity. Defines the total node switching capability and is calculated as Node Degree Connectivity x 100 wavelengths.
- o Regeneration Capability. We assume a value of 5% of the total switching capacity.
- o Add/Drop Capability. We assume a typical value of 25% of the switching capacity. So in the average up to 30 wavelengths per incoming fiber can be added/dropped within the optical node.
- o Resource pool setup and capabilities. A physical resource pool contains a mix of Add/Drop and Regeneration capabilities. This has the effect of increasing the number of resource pool advertized. Resource pool can be fully flexible (connected to any port), partial (only to some port) or Fixed (can only be connected to one direction). This parameter influences the complexity of the connectivity matrix.
- o Number of Regenerator types. For a given node the number of OEO capabilities is limited, it is typically decided by the type of electrical equipment and optical modules (emitting laser and optical receiver).
- o Blocking Ratio. The Spatial/Spectral blocking ratio indicates how much port-based/wavelength based blocking a node is experiencing.

For example considering the typical design it results in the following static layout:

- o 3 OEO pools each having 3 Resource Block inside.
- o Connectivity Matrix: $(8+30+30)$ 64x64 if considering one connectivity matrix. Ingress=64x3, Egress=3x64 (considering the OEO access with a multiple-wavelength link).

The following types of nodes and node designs were considered in this evaluation:

Node Types and designs

Node Type	Nodal Degree	Pool Type	Blocking
Small(S), Flexible	4	Partial	None
Small(S), Fixed(port)	4	Fixed	Port
Small(S), Fixed(label)	4	Partial	Lambda
Middle(M), Flexible	8	Flexible	None
Large(L), Flexible	16	Flexible	None

For the small nodes, 5 different type of regenerators are considered, for the Middle and Large ones 10 different type of regenerators are considered. Based on those designs we derived the following important figures:

- o Number of resourcePool : depends on the pool type and connectivity, which depend on the port blocking and number of Add/Drop and Regenerator capacity.
- o Number of resourceBlock. There is two numbers to be considered here : the number of resourceBlock for a given resource pool (this document) and total number of resourceBlock ([I-D.ietf-ccamp-rwa-info]). In this document the number of resource block within a resource pool is, worst case, the number of possible regenerator types, whereas in [I-D.ietf-ccamp-rwa-info] the number of resource block depends on the number of OEO types and on the connectivity.
- o Number of connectivity matrix/number of pairs/link per pairs. The number of sub-matrix increase depending on the port blocking ratio, the number of pair in one connectivity matrix depends on the wavelength restrictions. Those two criteria do not depend on which information model is considered. The number of link per set is increased by the number of resource pool in this draft.

Those numbers for each node are shown in the following table:

Details of information elements per node

Node Type	# Pools	Resource	Blocks	Matrix/Pair/Links
S, Flexible	6	5 (30)		1/1/10 (1/1/1)
S, Fixed(port)	12	5 (60)		4/4/4 (4/4/1)
S, Fixed(label)	6	5 (30)		4/1/10 (4/1/1)
M, Flexible	3	10 (30)		1/1/11 (1/1/1)
L, Flexible	5	10 (50)		1/1/21 (1/1/1)

Nota: Values for [I-D.ietf-ccamp-rwa-wson-encode] are between brackets.

For further reading easiness the above table could be further expanded as the following one:

Details of information elements per node

Node Type	#Pools	#Device Type	#Blocks	#ResProp TLV	Matrix/Pair/Links
S, Flexible	6	5 (30)	30	5 (25)	1/1/10 (1/1/1)
S, Fixed(port)	12	5 (60)	60	5 (45)	4/4/4 (4/4/1)
S,Fixed(label)	6	5 (30)	30	5 (25)	4/1/10 (4/1/1)
M, Flexible	3	10 (30)	30	10 (35)	1/1/11 (1/1/1)
L, Flexible	5	10 (50)	50	10 (40)	1/1/21 (1/1/1)

Nota: Values for [I-D.ietf-ccamp-rwa-wson-encode] are between brackets.

A.3. Comparing evaluation of the solutions

Based on those key information model elements both the tables "LSA size" indicate the size of the LSAs in this document and in [I-D.ietf-ccamp-rwa-wson-encode]. Number of flooded LSAs of a given type are indicated between brackets (when bigger than 1).

Solution of this document - Average size (and number) of LSAs per node type (unit: bytes)

Node Type	Node Attr	LSA Resource	Pool	LSA Resource	Desc	LSA
S, Flexible	117		120 (6)			524
S, Fixed(port)	692		120 (12)			644
S, Fixed(label)	620		120 (6)			524
M, Flexible	127		120 (3)			904
L, Flexible	209		120 (5)			984

Solution of [I-D.ietf-ccamp-rwa-wson-encode] - Average size (and number) of LSAs per node type (unit: bytes)

Node Type	Node Attr	LSA Optical	Node LSA
S, Flexible	49		2801
S, Fixed(port)	340		2980
S, Fixed(label)	132		4118
M, Flexible	52		2980
L, Flexible	54		2809

The Resource Description Container LSA contains several resource description TLVs. This LSA is smaller than the corresponding in [I-D.ietf-ccamp-rwa-wson-encode] mainly because the resource description do not depend on the port/lambda connectivity and number of device per block, thus allowing a better sharing of the information depicting the oeo capabilities.

The following summarizing table indicates the size of the sum of all LSA and the average size per update. In this document all the dynamic part is in the resource pool, allowing a more efficient updating behavior. The evaluation for [I-D.ietf-ccamp-rwa-wson-encode] are best case/worst case; the best case being an update of the RBState TLV and SharedAccessPool TLV only, which requires a multi-instance implementation of OSPF.

Summarizing Table (unit:bytes)

Node Type	Total LSA size	Total number of LSA	Avg size of an update
S, Flexible	1361 (2850)	8 (2)	120 (616/2801)
S, Fixed(port)	2776 (5411)	14 (2)	120 (1192/2980)
S, Fixed(label)	1864 (2941)	8 (2)	120 (616/4118)
M, Flexible	1391 (3032)	5 (2)	120 (448/2980)
L, Flexible	1793 (4172)	7 (2)	120 (720/2809)

Nota: Values for [I-D.ietf-ccamp-rwa-wson-encode] are between brackets

The node design considered are typical case, a worst case can be a node with high nodal degree, with lots of port and wavelength constraints. With considering a nodal degree of 8, resulting in 28 resource pool and 140 resource blocks, the total size is 9816 (11820) with 30 (2) LSAs.

Authors' Addresses

Pierre Peloso (editor)
Alcatel-Lucent
R.te de Villejust
Nozay, 91620
France

Phone: +33 130 702 662
Email: pierre.peloso@alcatel-lucent.com

Giovanni Martinelli
Cisco
Monza, 20900
Italy

Phone: +39 039 209 2044
Email: giomarti@cisco.com

Julien Meuric
France Telecom
2, av. Pierre Marzin
Lannion, 22307
France

Phone: +33 296 052 828
Email: julien.meuric@orange-ftgroup.com

Cyril Margaria
Nokia Siemens Networks
St-Martin str. 76
Munchen, 81541
Germany

Phone: +49-89-5159-16934
Email: cyril.margaria@nsn.com

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: September 15, 2011

X. Fu
M. Betts
Q. Wang
ZTE
D. McDysan
A. Malis
Verizon
March 14, 2011

GMPLS extensions to communicate latency as a traffic engineering
performance metric
draft-wang-ccamp-latency-te-metric-03

Abstract

Latency is such requirement that must be achieved according to the Service Level Agreement (SLA) between customers and service providers. Network Performance Objective (NPO) defined in ITU-T Y.1540 and Y.1541 is used for describing the meaning and numerical values performance parameters traversing multiple packet networks. The definitions of the packet network performance parameters are often also used as the basis of SLAs service providers, but possibly with different numerical values. A SLA is a part of a service contract where the level of service is formally defined between service providers and customers. For example, the service level includes platinum, golden, silver and bronze. Different service level may associate with different protection/restoration requirement. Latency can also be associated with different service level. The user may select a private line provider based on the ability to meet a latency SLA.

The key driver for latency is stock/commodity trading applications that use data base mirroring. A few milli seconds can impact a transaction. Financial or trading companies are very focused on end-to-end private pipe line latency optimizations that improve things 2-3 ms. Latency and latency SLA is one of the key parameters that these "high value" customers use to select a private pipe line provider. Other key applications like video gaming, conferencing and storage area networks require stringent latency and bandwidth.

This document describes the requirements and mechanisms to communicate latency as a traffic engineering performance metric in today's network which is consisting of potentially multiple layers of packet transport network and optical transport network in order to meet the latency SLA between service provider and his customers. This document also extends RSVP-TE and IGP to support these requirement. These extensions are intended to advertise and convey

the latency information of nodes and links as traffic engineering performance metric.

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on September 15, 2011.

Copyright Notice

Copyright (c) 2011 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1.	Introduction	4
1.1.	Conventions Used in This Document	5
2.	Requirements Identification and Solution Consideration	6
2.1.	Requirements Identification	6
2.2.	Solution Consideration	7
3.	Control Plane Solution	9
3.1.	Latency Advertisement	10
3.1.1.	Routing Extensions	10
3.1.1.1.	OSPF-TE Extension	10
3.1.1.2.	IS-IS-TE Extension	11
3.1.1.3.	Routing Extensions for Bundle Link/Composite Link	11
3.2.	Latency SLA Parameters Conveying	11
3.2.1.	Signaling Extensions	11
3.2.1.1.	Latency SLA Parameters ERO subobject	12
3.2.1.2.	Signaling Procedure	14
3.3.	Latency Accumulation and Verification	15
3.3.1.	Signaling Extensions	15
3.3.1.1.	Latency Accumulation Object	15
3.3.1.2.	Required Latency Object	17
3.3.1.3.	Signaling Procedures	17
4.	Security Considerations	19
5.	IANA Considerations	19
6.	References	19
6.1.	Normative References	19
6.2.	Informative References	20
	Authors' Addresses	20

1. Introduction

In a network, latency, a synonym for delay, is an expression of how much time it takes for a packet/frame of data to get from one designated point to another. In some usages, latency is measured by sending a packet/frame that is returned to the sender and the round-trip time is considered the latency of bidirectional co-routed or associated LSP. One way time is considered as the latency of unidirectional LSP. The one way latency may not be half of the round-trip latency in the case that the transmit and receive directions of the path are of unequal lengths.

Latency on a connection has two sources: Node latency which is caused by the node as a result of process time in each node and: Link latency as a result of packet/frame transit time between two neighbouring nodes or a FA-LSP/Composit Link [CL-REQ]. Latency variation is a parameter that is used to indicate the variation range of the latency value. These values should be made available to the control plane and management plane prior to path computation. This allows path computation to select a path that will meet the latency SLA.

In many cases, latency is a sensitive topic. For example, two stock exchanges (e.g., one in Chicago and another in New York) need to communicate with each other. A few ms can result in large impact on service. Some customers would pay for the latency performance. SLA contract which includes the requirement of latency is signed between service providers and customers. Service provider should assure that the network path latency MUST be limited to a value lower than the upper limit. In the future, latency optimization will be needed by more and more customers. For example, some customers pay for a private pipe line with latency constraint (e.g., less than 10 ms) which connects to Data Center. If this "provisioned" latency of this private pipe line couldn't meet the SLA, service provider may transfer customer's service to other Data Centers. Service provider may have many layers of pre-defined restoration for this transfer, but they have to duplicate restoration resources at significant cost. So service provider needs some mechanisms to avoid the duplicate restoration and reduce the network cost.

Measurement mechanism for link latency has been defined in many technologies. For example, the measurement mechanism for link latency has been provided in ITU-T [G.8021] and [Y.1731] for Ethernet. The link transit latency between two Ethernet equipments can be measured by using this mechanism. Similarly, overhead byte and measurement mechanism of latency has been provided in OTN (i.e., ITU-T [G.709]). In order to measure the link latency between two OTN nodes, PM&TCM which include Path Latency Measurement field and flag

used to indicate the beginning of measurement of latency is added to the overhead of ODUk. Node latency can also be recorded at each node by recording the process time between the beginning and the end. The measurement mechanism of links and nodes is out scope of this document.

Current operation and maintenance mode of latency measurement is high in cost and low in efficiency. The latency can only be measured after the connection has been established, if the measurement indicates that the latency SLA is not met then another path is computed, set up and measured. This "trial and error" process is very inefficient. To avoid this problem a means of making an accurate prediction of latency before a path is establish is required.

This document describes the requirements and mechanisms to communicate latency as a traffic engineering performance metric in today's network which is consisting of potentially multiple layers of packet transport network and optical transport network in order to meet the latency SLA between service provider and his customers. This document extends IGP to advertise and convey the latency attributes and latency variation as traffic engineering performance metric. Thus path computation entity can have a good knowledge of the latency traffic engineering database.

This document extends RSVP-TE protocol to accumulate (e.g., sum) latency information of links and nodes along one LSP across multi-domain (e.g., Inter-AS, Inter-Area or Multi-Layer) so that an latency verification can be made at source node. One-way and round-trip latency collection along the LSP by signaling protocol can be supported. So the end points of this LSP can verify whether the total amount of latency could meet the latency agreement between operator and his user.

When RSVP-TE signaling is used, the source can determine if the latency requirement is met much more rapidly than performing the actual end-to-end latency measurement.

The required latency could be signaled by RSVP-TE (i.e., Path and Resv message). Intermediate nodes could reject the request (Path or Resv message) if the accumulated latency is not achievable. this is essential in multiple AS use cases, but may not be needed in a single IGP level/area if the IGP is extended to convey latency information.

1.1. Conventions Used in This Document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this

document are to be interpreted as described in [RFC2119].

2. Requirements Identification and Solution Consideration

2.1. Requirements Identification

End-to-end service optimization based on latency is a key requirement for service provider. This type of function will be adopted by their "premium" service customers. They would like to pay for this "premium" service. After these premium services are deployed, they will also expand to their own customers. Following key requirements associated with latency is identified.

- o Communication latency of links and nodes including latency and latency variation as a traffic engineering performance metric is a very important requirement.
- o End-to-end service optimization based on latency constraint is a key requirement for service provider. Latency on a route level will help carriers' customers to make his provider selection decision.
 - * Path computation entity MUST have the capability to compute one end-to-end path with latency constraint. For example, it MUST have the capability to compute a route with x amount bandwidth and less than y ms of latency limit based on the latency traffic engineering database.
 - * It should also support combined routing constraints with pre-defined priorities, e.g., SRLG diversity, latency and cost.
- o One end-to-end LSP may be across some Composite Links [CL-REQ]. Even if the transport technology (e.g., OTN) implementing the component links is identical, the latency characteristics of the component links may differ. In order to assign the LSP to one of component links with different latency characteristics, following related requirements are from [CL-REQ].
 - * The solution SHALL provide a means to indicate that a traffic flow shall select a component link with the minimum latency value.
 - * The solution SHALL provide a means to indicate that a traffic flow shall select a component link with a maximum acceptable latency value as specified by protocol.

- * The solution SHALL provide a means to indicate that a traffic flow shall select a component link with a maximum acceptable latency variation value as specified by protocol.
- o RSPV-TE should support the accumulation (e.g., sum) of latency information of links and nodes along one LSP across multi-domain (e.g., Inter-AS, Inter-Area or Multi-Layer) so that an latency validation decision can be made at the source node. One-way and round-trip latency collection along the LSP by signaling protocol and latency verification at the end of LSP should be supported.

2.2. Solution Consideration

- o The latency performance metric MUST be advertised into path computation entity by IGP (etc., OSPF-TE or IS-IS-TE) to perform route computation and network planning based on latency SLA target.
 - * Data plane is responsible for measuring the latency (e.g., latency and latency variation). Latency measurement can be provided by different technologies. This information will be provided to the Control Plane. In order to monitor the performance, pro-active latency measurement is required. Generally, every 15 minutes or 24 hours, the value of latency and latency variation should be collected. Similarly, on demand latency measurement is required due to the goal of maintenance. This can be done every fixed time interval (e.g., 5 minutes or 1 hour). The method used to measure the latency of links and nodes is out scope of this document.
 - * Control plane is responsible for advertising and collecting the latency value of links and nodes by IGP (i.e., OSPF-TE/IS-IS-TE). Latency characteristics of these links and nodes may change dynamically. In order to control IGP messaging and avoid being unstable when the latency and latency variation value changes, a threshold and a limit on rate of change MUST be configured to control plane.
- o When the Composite Links [CL-REQ] is advertised into IGP, there are following solution consideration.
 - * The latency of composite link may be the range (e.g., at least minimum and maximum) latency value of all component links. The latency of composite link may also be the maximum latency value of all component links. In these cases, only partial information is transmitted in the IGP. So the path computation entity has insufficient information to determine whether a particular path can support its delay requirements. This leads

to signaling crankback.

- * The IGP may be extended to advertise latency of each component link within one Composite Link.
- o In order to assign the LSP to one of component links with different latency characteristics, RSVP-TE message MUST convey latency SLA parameter to the end points of Composite Links where it can select one of component links or trigger the creation of lower layer connection which MUST meet latency SLA parameter.
- * The RSVP-TE message needs to carry a indication of request minimum latency, maximum acceptable latency value and maximum acceptable delay variation value for the component link selection or creation. The composite link will take these parameters into account when assigning traffic of LSP to a component link.
- o One end-to-end LSP (e.g., in IP/MPLS or MPLS-TP network) may traverse a FA-LSP of server layer (e.g., OTN rings). The boundary nodes of the FA-LSP SHOULD be aware of the latency information of this FA-LSP (e.g., latency and latency variation).
- * If the FA-LSP is able to form a routing adjacency and/or as a TE link in the client network, the latency value of the FA-LSP can be as an input to a transformation that results in a FA traffic engineering metric and advertised into the client layer routing instances. Note that this metric will include the latency of the links and nodes that the trail traverses.
- * If the latency information of the FA-LSP changes (e.g., due to a maintenance action or failure in OTN rings), the boundary node of the FA-LSP will receive the TE link information advertisement including the latency value which is already changed and if it is over than the threshold and a limit on rate of change, then it will compute the total latency value of the FA-LSP again. If the total latency value of FA-LSP changes, the client layer MUST also be notified about the latest value of FA. The client layer can then decide if it will accept the increased latency or request a new path that meets the latency requirement.
- * When one end-to-end LSP traverse a server layer, there will be some latency constraint requirement for the segment route in server layer. So RSVP-TE message needs to carry a indication of request minimum latency, maximum acceptable latency value and maximum acceptable delay variation value for the FA selection or FA-LSP creation. The boundary nodes of FA-LSP

will take these parameters into account for FA selection or FA-LSP creation.

- o Restoration, protection and equipment variations can impact "provisioned" latency (e.g., latency increase). The change of one end-to-end LSP latency performance MUST be known by source and/or sink node. So it can inform the higher layer network of a latency change. The latency change of links and nodes will affect one end-to-end LSP's total amount of latency. Applications can fail beyond an application-specific threshold. Some remedy mechanism could be used.
- * Some customers may insist on having the ability to re-route if the latency SLA is not being met. If a "provisioned" end-to-end LSP latency could not meet the latency agreement (e.g., latency or latency variation) between operator and his user, then re-routing could be triggered based on the local policy. Pre-defined or dynamic re-routing could be triggered to handle this case. The latency performance of pre-defined or dynamic re-routing LSP MUST meet the latency SLA parameter. In the case of predefined re-routing, the large amounts of redundant capacity may have a significant negative impact on the overall network cost. Dynamic re-routing also has to face the risk of resource limitation. So the choice of mechanism MUST be based on SLA or policy. In the case where the latency SLA cannot be met after a re-route is attempted, control plane should report an alarm to management plane. It could also try restoration for several times which could be configured.
- * As a result of the change of links and nodes latency in the LSP, current LSP may be frequently switched to a new LSP with a appropriate latency value. In order to avoid this, the solution SHOULD indicate the switchover of the LSP according to maximum acceptable change latency value.

3. Control Plane Solution

In order to meet the requirements which have been identified in section 3, this document defines following four phases.

- o The first phase is the advertisement of the latency information by routing protocol (i.e., OSPF-TE/IS-IS-TE), including latency of nodes and links, a FA-LSP or Composite Link [CL-REQ] between two neighbour and latency variation, so path computation entity can be aware of the latency of nodes and links.

- o In the second phase, path computation entity is responsible for end-to-end path computation with latency constraint (e.g., less than 10 ms) combining other routing constraint parameters (e.g., SRLG, cost and bandwidth). How does the path computation entity compute the latency variation of one end-to-end connection can be referred to ITU-T Y.1540.
- o The third phase is to convey the latency SLA parameters for the selection or creation of component link or FA/FA-LSP. One end-to-end LSP may be across some Composite Links or server layers, so it can convey latency SLA parameters by RSVP-TE message.
- o The last phase is the latency collection and verification. This stage could be optional. It could accumulate (e.g., sum) latency information along the LSP across multi-domain (e.g., Inter-AS, Inter-Area or Multi-Layer) by RSVP-TE signaling message to verify the total latency at the end of path.

3.1. Latency Advertisement

A node in the packet transport network or optical transport network can detect the latency value of link which connects to it. Also the node latency can be recorded at every node. Then latency values of TE links, Composite Links [CL-REQ] or FAs, latency values of nodes and latency variation are notified to the IGP. If any latency values change and over than the threshold and a limit on rate of change, then the change MUST be notified to the IGP again. As a result, path computation entity can have every node and link latency values and latency variation in its view of the network, and it can compute one end-to-end path with latency constraint. It needs to extend IGP protocol (i.e., OSPF-TE/IS-IS-TE).

3.1.1. Routing Extensions

Following is the extensions to OSPF-TE/IS-IS-TE to support the advertisement of the node latency value, link latency and latency variation.

3.1.1.1. OSPF-TE Extension

OSPF-TE routing protocol can be used to carry latency performance metric by adding a sub-TLV to the TE link defined in [RFC4203]. As defined in [RFC3630] and [RFC4203], the top-level TLV can take one of two values (1) Router address or (2) Link. Latency sub-TLV of link is added behind the top-level TLV. It includes estimated latency and latency variation value.

This link attribute may also take into account the latency of a

network element (node), i.e., the latency between the incoming port and the outgoing port of a network element. If the link attribute were to include node latency AND link latency, then when the latency calculation is done for paths traversing links on the same node then the node latency can be subtracted out. Following is the link Latency sub-TLV format.

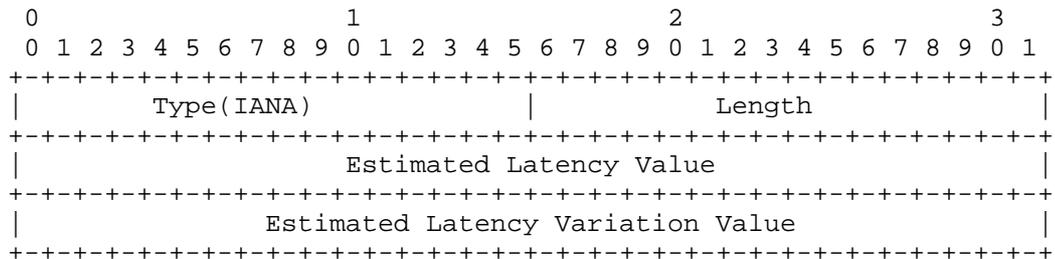


Figure 1: Format of the Latency sub-TLV

- o Estimated Latency Value: a value indicates the latency of link or node.
- o Estimated Latency Variation Value: a value indicates the variation range of the estimated latency value.

3.1.1.2. IS-IS-TE Extension

TBD

3.1.1.3. Routing Extensions for Bundle Link/Composite Link

[Editor Notes:Some discussion have been raised in RTGWG Mailing list.]

Some people are discussing having an IGP adjacency (and metric) for a composite link but a separate advertisement that contains parameters, such as those listed here.

3.2. Latency SLA Parameters Conveying

3.2.1. Signaling Extensions

This document defines extensions to and describes the use of RSVP-TE [RFC3209], [RFC3471], [RFC3473] to explicitly convey the latency SLA parameter for the selection or creation of component link or FA/FA-LSP. Specifically, in this document, Latency SLA Parameters TLV are defined and added into ERO as a subobject.

3.2.1.1. Latency SLA Parameters ERO subobject

A new OPTIONAL subobject of the EXPLICIT_ROUTE Object (ERO) is used to specify the latency SLA parameters including a indication of request minimum latency, request maximum acceptable latency value and request maximum acceptable latency variation value. It can be used for the following scenarios.

- o One end-to-end LSP may traverse a server layer FA-LSP. This subobject of ERO can indicate that FA selection or FA-LSP creation shall be based on this latency constraint. The boundary nodes of multi-layer will take these parameters into account for FA selection or FA-LSP creation.
- o One end-to-end LSP may be across some Composite Links [CL-REQ]. This subobject of ERO can indicate that a traffic flow shall select a component link with some latency constraint values as specified in this subobject.

This Latency SLA Parameters ERO subobject has the following format. It follows a subobject containing the IP address, or the link identifier [RFC3477], associated with the TE link on which it is to be used.

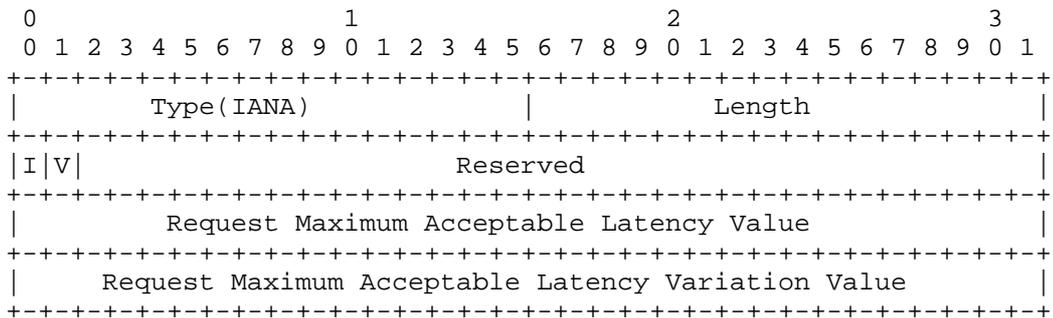


Figure 2: Format of Latency SLA Parameters TLV

- o I bit: a one bit field indicates whether a traffic flow shall select a component link with the minimum latency value or not. It can also indicate whether one end-to-end LSP shall select a FA or trigger a FA-LSP creation with the minimum latency value or not when it traverse a server layer.
- o V bit: a one bit field indicates whether a traffic flow shall select a component link with the minimum latency variation value or not. It can also indicate whether one end-to-end LSP shall select a FA or trigger a FA-LSP creation with the minimum latency

variation value or not when it traverse a server layer.

- o Request Maximum Acceptable Latency Value: a value indicates that a traffic flow shall select a component link with a maximum acceptable latency value. It can also indicate one end-to-end LSP shall select a FA or trigger a FA-LSP creation with a maximum acceptable latency value when it traverse a server layer.
- o Request Maximum Acceptable Latency Variation Value: a value indicates that a traffic flow shall select a component link with a maximum acceptable latency variation value. It can also indicate one end-to-end LSP shall select a FA or trigger a FA-LSP creation with a maximum acceptable latency variation value when it traverse a server layer.

Following is an example about how to use these parameters. Assume there are following component links within one composite link.

- o Component link1: latency = 5ms, latency variation = 15 us
- o Component link2: latency = 10ms, latency variation = 6 us
- o Component link3: latency = 20ms, latency variation = 3 us
- o Component link4: latency = 30ms, latency variation = 1 us

Assume there are following request information.

- o Request minimum latency = FALSE
- o Request minimum latency variation= FALSE
- o Maximum Acceptable Latency Value= 15 ms
- o Maximum Acceptable Latency Variation Value = 10us

Only Component link2 could be qualified.

- o Request minimum latency = FALSE
- o Request minimum latency variation= FALSE
- o Maximum Acceptable Latency Value= 35 ms
- o Maximum Acceptable Latency Variation Value = 10us

Component link2/3/4 could be qualified. Which component link is selected depends on local policy.

- o Request minimum latency = FALSE
- o Request minimum latency variation= TRUE
- o Maximum Acceptable Latency Value= 35 ms
- o Maximum Acceptable Latency Variation Value = 10us

Only Component link4 could be qualified.

- o Request minimum latency = TRUE
- o Request minimum latency variation= FALSE
- o Maximum Acceptable Latency Value= 35 ms
- o Maximum Acceptable Latency Variation Value = 10us

Only Component link2 could be qualified.

- Request minimum latency = TRUE
- Request minimum latency variation= TRUE
- Maximum Acceptable Latency Value= 35 ms
- Maximum Acceptable Latency Variation Value = 10us

In this case, there is no any qualified component links.

3.2.1.2. Signaling Procedure

When a intermediate node receives a PATH message containing ERO and finds that there is a Latency SLA Parameters ERO subobject immediately behind the IP address or link address sub-object related to itself, if the node determines that it's a region edge node of FA-LSP or an end point of a composite link [CL-REQ], then, this node extracts latency SLA parameters (i.e., request minimum, request maximum acceptable and request maximum acceptable latency variation value) from Latency SLA Parameters ERO subobject. This node used these latency parameters for FA selection, FA-LSP creation or component link selection. If the intermediate node couldn't support the latency SLA, it MUST generate a PathErr message with a "Latency

SLA unsupported" indication (TBD by INNA). If the intermediate node couldn't select a FA or component link, or create a FA-LSP which meet the latency constraint defined in Latency SLA Parameters ERO subobject, it must generate a PathErr message with a "Latency SLA parameters couldn't be met" indication (TBD by INNA).

3.3. Latency Accumulation and Verification

Latency accumulation and verification applies where the full path of an multi-domain (e.g., Inter-AS, Inter-Area or Multi-Layer) TE LSP can't be or is not determined at the ingress node of the TE LSP. This is most likely to arise owing to TE visibility limitations. If all domains support to communicate latency as a traffic engineering metric parameter, one end-to-end optimized path with delay constraint (e.g., less than 10 ms) which satisfies latency SLAs parameter could be computed by BRPC [RFC5441] in PCE. Otherwise, it could use the mechanism defined in this section to accumulat the latency of each links and nodes along the path which is across multi-domain. Latency accumulation and verification also applies where not all domains could support the communication latency as a traffic engineering metric parameter.

One domain may need to know that other domains support latency accumulation. It could be discovered in some automatic way. PCEs in different domains may play a role here. It is for further study.

3.3.1. Signaling Extensions

3.3.1.1. Latency Accumulation Object

An Latency Accumulation Object is defined in this document to support the accumulation and verification of the latency. This object which can be carried in a Path/Resv message may includes two sub-TLVs. Latency Accumulation Object has the following format.

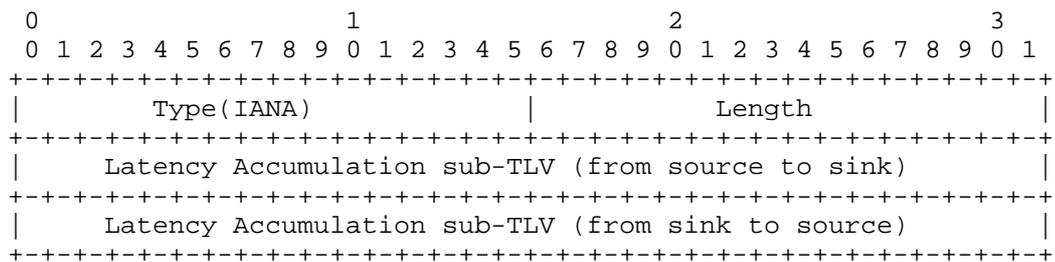


Figure 3: Format of Accumulated Latency Object

- o Latency Accumulation sub-TLV (from source to sink):It is used to accumulate the latency from source to sink along the unidirectional or bidirectional LSP. A Path message for unidirectional and bidirectional LSP must includes this sub-TLV. When sink node receives the Path message including this sub-TLV, it must copy this sub-TLV into Resv message. So the source node can receive the latency accumulated value (i.e., sum) from itself to sink node which can be used for latency verification.
- o Latency Accumulation sub-TLV (from sink to source):It is used to accumulate the latency from sink to source along the bidirectional LSP. A Resv message for the bidirectional LSP must includes this sub-TLV. So the source node can get the latency accumulated value (i.e., sum) of round-trip which can be used for latency verification.

3.3.1.1.1. Latency Accumulation sub-TLV

The Sub-TLV format is defined in the next picture.

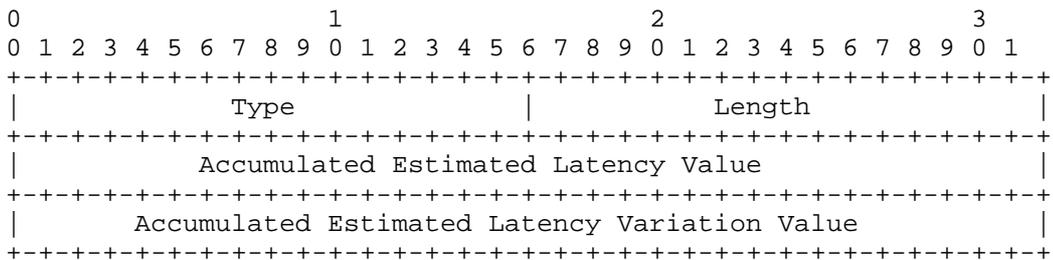


Figure 4: Format of Latency Accumulation sub-TLV

- o Type: sub-TLV type
 - * 0: It indicates the sub-TLV is for the latency accumulation from source to sink node along the LSP.
 - * 1: It indicates the sub-TLV is for the latency accumulation from sink to source node along the LSP.
- o Length: length of the sub-TLV value in bytes.
- o Accumulated Estimated Latency Value: a value indicates the sum of each links and nodes' latency along one direction of LSP.
- o Accumulated Estimated Latency Variation Value: a value indicates the sume of each links and nodes' latency variation along one direction of LSP. Since latecnry variation is accumulated non-

linearly. Latency variation accumulatoin should be in a lower priority.

3.3.1.2. Required Latency Object

A required latency could be signaled by RSVP-TE message (i.e., Path and Resv). This object is carried in the LSP_ATTRIBUTES object of Path/Resv message, object that is defined in [RFC5420]. Intermediate nodes could reject the request (Path or Resv message) if the accumulated latency exceeds require latency value in the Required Latency Object.

If the accumulated latency is not achievable, there is no necessary to accumulate the latency for remaining domain or nodes. In order to balance the load across network links more efficiently if the absolute minimum latency is not required, intermediate nodes could choose a cost-effective path if the requested latency could easily be met. Note that this would apply inter-AS if the IGP is extended to advertise latency.

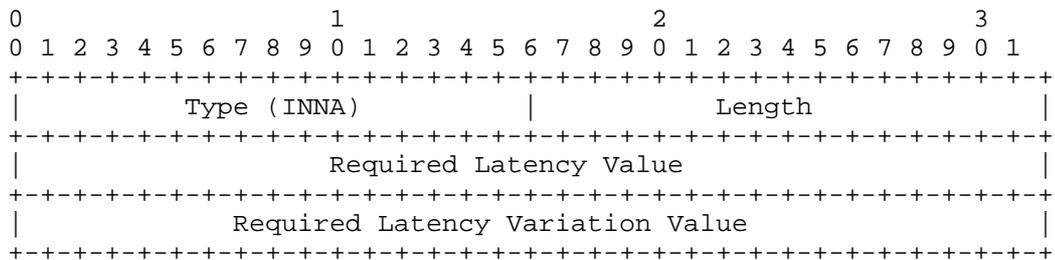


Figure 5: Required Latency Object

- o Required Latency Value: The accumulated estimated latency value should not exceed this value.
- o Required Latency Variation Value: The accumulated estimated latency variation value should not exceed this value.

3.3.1.3. Signaling Procedures

When the source node desires to accumulate (i.e., sum) the total latency of one end-to-end LSP, the "Latency Accumulating desired" flag (value TBD) should be set in the LSP_ATTRIBUTES object of Path/Resv message, object that is defined in [RFC5420]. If the source node makes the intermediate node have the capability to verify the accumulated latency, the "Latency Verifying desired" flag (value TBD) should be also set in the LSP_ATTRIBUTES object of Path/Resv message.

A source node initiates latency accumulation for a given LSP by adding Latency Accumulation object to the Path message. The Latency Accumulation object only includes one sub-TLV (sub-TLV type=0) where it is going to accumulate the latency value of each links and nodes along path from source to sink. If latency verifying is desired, the source node also adds the Required Latency Object to the Path message.

When the downstream node receives Path message and if the "Latency Accumulating desired" is set in the LSP_ATTRIBUTES, it accumulates the latency of link and node based on the accumulated latency value of the sub-TLV (sub-TLV type=0) in Latency Accumulation object before it sends Path message to downstream.

If the "Latency Verifying desired" is set in the LSP_ATTRIBUTES, downstream node will check whether the Accumulated Estimated Latency and Variation value exceeds the Required Latency and Variation value. If the accumulated latency is not achievable, there is no necessary to accumulate the latency for remaining domain or nodes. It MUST generate a error message with a "Accumulated Latency couldn't meet the required latency" indication (TBD by INNA).

If the intermediate node (e.g., entry node of one domain) couldn't support the latency accumulation function, it MUST generate a error message with a "Latency Accumulation unsupported" indication (TBD by INNA).

If the intermediate node (e.g., entry node of one domain) couldn't support the latency verify function, it MUST generate a error message with a "Latency Verify unsupported" indication (TBD by INNA).

When the sink node of LSP receives the Path message and the "Latency Accumulating desired" is set in the LSP_ATTRIBUTES, it copy the Accumulated Estimated Latency and Variation value in the Latency Accumulation sub-TLV (sub-TLV type=0) of Path message into the one of Resv message which will be forwarded hop by hop in the upstream direction until it arrives the source node. Then source node can get the latency sum value from source to sink for unidirectional and bidirectional LSP.

If the LSP is a bidirectional one and the "Latency Accumulating desired" is set in the LSP_ATTRIBUTES, it adds another Latency Accumulation sub-TLV (sub-TLV type=1) into the Latency Accumulation object of Resv message where latency of each links and nodes along path will be accumulated from sink to source into this sub-TLV.

If the LSP is a bidirectional one and the "Latency Verifying desired" is set in the LSP_ATTRIBUTES, it copy the Required Latency and

Variation value in the Required Latency Object of Path message into the one of Resv message.

When the upstream node receives Resv message and if the "Latency Accumulating desired" is set in the LSP_ATTRIBUTES, it accumulates the latency of link and node based on the latency value in sub-TLV (sub-TLV type=1) before it continues to send Resv message.

If the "Latency Verifying desired" is set in the LSP_ATTRIBUTES, it will check whether the latency sum of Accumulated Estimated Latency and Variation value in each Latency Accumulation sub-TLV exceeds the Required Latency and Variation value. If the accumulated latency is not achievable, there is no necessary to accumulate the latency for remaining domain or nodes. It MUST generate an error message with a "Accumulated Latency couldn't meet the required latency" indication (TBD by INNA).

After source node receives Resv message, it can get the total latency value of one way or round-trip from Latency Accumulation object. So it can confirm whether the latency value meets the latency SLA or not.

4. Security Considerations

TBD

5. IANA Considerations

TBD

6. References

6.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, December 2001.
- [RFC3473] Berger, L., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource Reservation Protocol-Traffic Engineering (RSVP-TE) Extensions", RFC 3473, January 2003.
- [RFC3477] Kompella, K. and Y. Rekhter, "Signalling Unnumbered Links

in Resource ReSerVation Protocol - Traffic Engineering (RSVP-TE)", RFC 3477, January 2003.

[RFC3630] Katz, D., Kompella, K., and D. Yeung, "Traffic Engineering (TE) Extensions to OSPF Version 2", RFC 3630, September 2003.

[RFC4203] Kompella, K. and Y. Rekhter, "OSPF Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4203, October 2005.

6.2. Informative References

[CL-REQ] C. Villamizar, "Requirements for MPLS Over a Composite Link", draft-ietf-rtgwg-cl-requirement-02 .

[G.709] ITU-T Recommendation G.709, "Interfaces for the Optical Transport Network (OTN)", December 2009.

Authors' Addresses

Xihua Fu
ZTE

Email: fu.xihua@zte.com.cn

Malcolm Betts
ZTE

Email: malcolm.betts@zte.com.cn

Qilei Wang
ZTE

Email: wang.qilei@zte.com.cn

Dave McDysan
Verizon

Email: dave.mcdysan@verizon.com

Andrew Malis
Verizon

Email: andrew.g.malis@verizon.com

Network Working Group
Internet Draft
Category: Standards Track

Fatai Zhang
Huawei
Guoying Zhang
CATR
Sergio Belotti
Alcatel-Lucent
D. Ceccarelli
Ericsson
March 11, 2011

Expires: September 11 2011

Generalized Multi-Protocol Label Switching (GMPLS) Signaling
Extensions for the evolving G.709 Optical Transport Networks Control

draft-zhang-ccamp-gmpls-evolving-g709-07.txt

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at <http://www.ietf.org/ietf/lid-abstracts.txt>.

The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>.

This Internet-Draft will expire on September 11, 2011.

Abstract

Recent progress in ITU-T Recommendation G.709 standardization has introduced new ODU containers (ODU0, ODU4, ODU2e and ODUFlex) and enhanced Optical Transport Networking (OTN) flexibility. Several

recent documents have proposed ways to modify GMPLS signaling protocols to support these new OTN features.

It is important that a single solution is developed for use in GMPLS signaling and routing protocols. This solution must support ODUk multiplexing capabilities, address all of the new features, be acceptable to all equipment vendors, and be extensible considering continued OTN evolution.

This document describes the extensions to the Generalized Multi-Protocol Label Switching (GMPLS) signaling to control the evolving Optical Transport Networks (OTN) addressing ODUk multiplexing and new features including ODU0, ODU4, ODU2e and ODUFlex.

Conventions used in this document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

Table of Contents

1. Introduction	3
2. Terminology	3
3. GMPLS Extensions for the Evolving G.709 - Overview	4
4. Extensions for Traffic Parameters for the Evolving G.709	5
4.1. Usage of ODUFlex(CBR) Traffic Parameter	6
4.2. Example of ODUFlex(CBR) Traffic Parameter	7
5. Generalized Label	8
5.1. New definition of ODUk Label	8
5.2. Examples	12
5.3. Label Distribution Procedure	13
5.3.1. Notification on Label Error	14
5.4. Supporting Virtual Concatenation and Multiplication	15
5.5. Supporting Multiplexing Hierarchy	15
5.6. Control Plane Backward Compatibility Considerations	16
6. Security Considerations	18
7. IANA Considerations	18
8. References	18
8.1. Normative References	18
8.2. Informative References	20
9. Authors' Addresses	20
Acknowledgment	22

1. Introduction

Generalized Multi-Protocol Label Switching (GMPLS) [RFC3945] extends MPLS to include Layer-2 Switching (L2SC), Time-Division Multiplex (e.g., SONET/SDH, PDH, and ODU), Wavelength (OCh, Lambdas) Switching, and Spatial Switching (e.g., incoming port or fiber to outgoing port or fiber). [RFC3471] presents a functional description of the extensions to Multi-Protocol Label Switching (MPLS) signaling required to support Generalized MPLS. RSVP-TE-specific formats and mechanisms and technology specific details are defined in [RFC3473].

With the evolution and deployment of G.709 technology, it is necessary that appropriate enhanced control technology support be provided for G.709. [RFC4328] describes the control technology details that are specific to foundation G.709 Optical Transport Networks (OTN), as specified in the ITU-T Recommendation G.709 [G709-V1], for ODUk deployments without multiplexing.

In addition to increasing need to support ODUk multiplexing, the evolution of OTN has introduced additional containers and new flexibility. For example, ODU0, ODU2e, ODU4 containers and ODUflex are developed in [G709-V3].

In addition, the following issues require consideration:

- Support for hitless adjustment of ODUflex, which is to be specified in ITU-T G.hao.
- Support for Tributary Port Number. The Tributary Port Number has to be negotiated on each link for flexible assignment of tributary ports to tributary slots in case of LO-ODU over HO-ODU (e.g., ODU2 into ODU3).

Therefore, it is clear that [RFC4328] has to be updated or superceded in order to support ODUk multiplexing, as well as other ODU enhancements introduced by evolution of OTN standards.

This document updates [RFC4328] extending the G.709 ODUk traffic parameters and also presents a new OTN label format which is very flexible and scalable.

2. Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

3. GMPLS Extensions for the Evolving G.709 - Overview

New features for the evolving OTN, for example, new ODU0, ODU2e, ODU4 and ODUflex containers are specified in [G709-V3]. The corresponding new signal types are summarized below:

- Optical Channel Transport Unit (OTUk):
 - . OTU4
- Optical Channel Data Unit (ODUk):
 - . ODU0
 - . ODU2e
 - . ODU4
 - . ODUflex

A new Tributary Slot (TS) granularity (i.e., 1.25 Gbps) is also described in [G709-V3]. Thus, there are now two TS granularities for the foundation OTN ODU1, ODU2 and ODU3 containers. The TS granularity at 2.5 Gbps is used on legacy interfaces while the new 1.25 Gbps will be used for the new interfaces.

In addition to the support of ODUk mapping into OTUk ($k = 1, 2, 3, 4$), the evolving OTN [G.709-V3] encompasses the multiplexing of ODUj ($j = 0, 1, 2, 2e, 3, flex$) into an ODUk ($k > j$), as described in Section 3.1.2 of [OTN-frwk].

Virtual Concatenation (VCAT) of OPUk (OPUk-Xv, $k = 1/2/3$, $X = 1...256$) are also supported by [OTN-V3]. Note that VCAT of OPU0 / OPU2e / OPU4 / OPUflex are not supported per [OTN-V3].

[RFC4328] describes GMPLS signaling extensions to support the control for G.709 Optical Transport Networks (OTN) [G709-V1]. However, [RFC4328] needs to be updated because it does not provide the means to signal all the new signal types and related mapping and multiplexing functionalities. Moreover, it supports only the deprecated auto-MSI mode which assumes that the Tributary Port Number is automatically assigned in the transmit direction and not checked in the receive direction.

This document extends the G.709 traffic parameters described in [RFC4328] and presents a new OTN label format which is very flexible and scalable. Additionally, procedures about Tributary Port Number assignment through control plane are also provided in this document.

4. Extensions for Traffic Parameters for the Evolving G.709

The traffic parameters for G.709 are defined as follows:



The Signal Type should be extended to cover the new Signal Type introduced by the evolving OTN. The new Signal Type is extended as follows:

Value	Type
0	Not significant
1	ODU1 (i.e., 2.5 Gbps)
2	ODU2 (i.e., 10 Gbps)
3	ODU3 (i.e., 40 Gbps)
4	ODU4 (i.e., 100 Gbps)
5	Reserved (for future use)
6	OCh at 2.5 Gbps
7	OCh at 10 Gbps
8	OCh at 40 Gbps
9	OCh at 100 Gbps
10~19	Reserved (for future use)
20	ODU0 (i.e., 1.25 Gbps)
21~30	Reserved (for future use)
31	ODU2e (i.e., 10Gbps for FC1200 and GE LAN)
32	ODUflex(GFP) (i.e., 1.25*N Gbps)
33	ODUflex(CBR) (i.e., 1.25*N Gbps)
34~255	Reserved (for future use)

In case of ODUflex(CBR), the Bit_Rate and Tolerance fields are used together to represent the actual bandwidth of ODUflex, where:

- The Bit_Rate field indicates the nominal bit rate of ODUflex(CBR) encoded as a 32-bit IEEE single-precision floating-point number (referring to [RFC4506] and [IEEE]).
- The Tolerance field indicates the bit rate tolerance (part per million, ppm) of the ODUflex(CBR) encoded as an unsigned integer, which is bounded in 0~100ppm.

For example, for an ODUflex(CBR) service with Bit_Rate = 2.5Gbps and Tolerance = 100ppm, the actual bandwidth of the ODUflex is:

$$2.5\text{Gbps} * (1 - 100\text{ppm}) \sim 2.5\text{Gbps} * (1 + 100\text{ppm})$$

In case of other ODUk signal types, the Bit_Rate and Tolerance fields are not necessary and MUST be filled with 0.

The usage of the NMC, NVC and Multiplier (MT) fields are the same as [RFC4328].

4.1. Usage of ODUflex(CBR) Traffic Parameter

In case of ODUflex(CBR), the information of Bit_Rate and Tolerance in the ODUflex traffic parameter is used to determine the total number of tributary slots N in the HO ODUk link to be reserved. Here:

N = Ceiling of

$$\frac{\text{ODUflex(CBR) nominal bit rate} * (1 + \text{ODUflex(CBR) bit rate tolerance})}{\text{ODTUK.ts nominal bit rate} * (1 - \text{HO OPUk bit rate tolerance})}$$

Therefore, a node receiving a Path message containing ODUflex(CBR) traffic parameter can allocate precise number of tributary slots and set up the cross-connection for the ODUflex service.

The table below shows the actual bandwidth of the tributary slot of ODUk (in Gbps), referring to [G709-V3].

ODUk	Minimum	Nominal	Maximum
ODU2	1.249 384 632	1.249 409 620	1.249 434 608
ODU3	1.254 678 635	1.254 703 729	1.254 728 823
ODU4	1.301 683 217	1.301 709 251	1.301 735 285

Note that:

Minimum bandwidth of ODUTk.ts =
 ODTUk.ts nominal bit rate * (1 - HO OPUk bit rate tolerance)

Maximum bandwidth of ODTUk.ts =
 ODTUk.ts nominal bit rate * (1 + HO OPUk bit rate tolerance)

Where: HO OPUk bit rate tolerance = 20ppm

For different ODUk, the bandwidths of the tributary slot are different, and so the total number of tributary slots to be reserved for the ODUflex(CBR) may not be the same on different HO ODUk links. This is why the traffic parameter should bring the actual bandwidth information other than the NMC field.

4.2. Example of ODUflex(CBR) Traffic Parameter

This section gives an example to illustrate the usage of ODUflex(CBR) traffic parameter.

As shown in Figure 1, assume there is an ODUflex(CBR) service requesting a bandwidth of (2.5Gbps, +/-100ppm) from node A to node C. In other words, the ODUflex traffic parameter indicates that Signal Type is 33 (ODUflex(CBR)), Bit_Rate is 2.5Gbps and Tolerance is 100ppm.

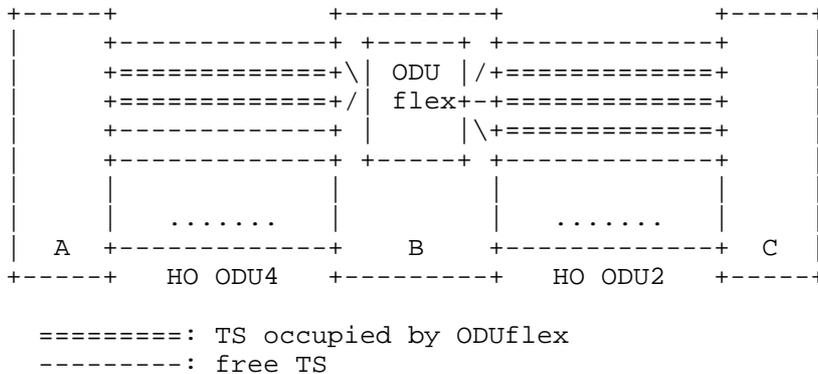


Figure 1 - Example of ODUflex(CBR) Traffic Parameter

- On the HO ODU4 link between node A and B:

The maximum bandwidth of the ODUflex equals $2.5\text{Gbps} * (1 + 100\text{ppm})$, and the minimum bandwidth of the tributary slot of ODU4 equals $1.301\ 683\ 217\text{Gbps}$, so the total number of tributary slots $N1$ to be reserved on this link is:

$$N1 = \text{ceiling} (2.5\text{Gbps} * (1 + 100\text{ppm}) / 1.301\ 683\ 217) = 2$$

- On the HO ODU2 link between node B and C:

The maximum bandwidth of the ODUflex equals $2.5\text{Gbps} * (1 + 100\text{ppm})$, and the minimum bandwidth of the tributary slot of ODU2 equals $1.249\ 384\ 632\text{Gbps}$, so the total number of tributary slots $N2$ to be reserved on this link is:

$$N2 = \text{ceiling} (2.5\text{Gbps} * (1 + 100\text{ppm}) / 1.249\ 384\ 632) = 3$$

5. Generalized Label

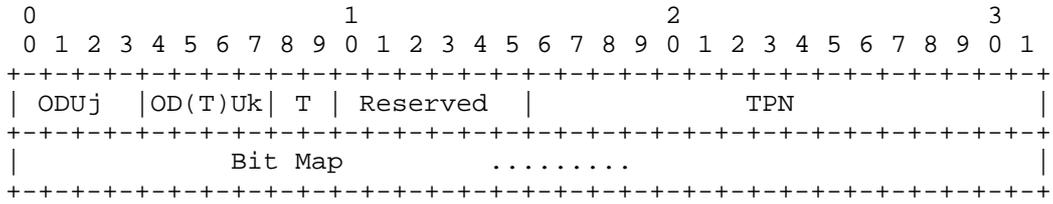
[RFC3471] has defined the Generalized Label which extends the traditional label by allowing the representation of not only labels which travel in-band with associated data packets, but also labels which identify time-slots, wavelengths, or space division multiplexed positions. The format of the corresponding RSVP-TE Generalized Label object is defined in the Section 2.3 of [RFC3473].

However, for different technologies, we usually need use specific label rather than the Generalized Label. For example, the label format described in [RFC4606] could be used for SDH/SONET, the label format in [RFC4328] for G.709.

In this document, a new ODUk label format is defined, the information model of which is described in Section 4.10 of [OTN-info].

5.1. New definition of ODUk Label

In order to be compatible with new types of ODU signal and new types of tributary slot, the following new ODUk label format is defined:



ODUj and OD(T)Uk (4 bits respectively): indicate that LO ODUj is multiplexed into HO ODUk(k>j), or LO ODUj is mapped into OTUk (j=k).

ODUj field	Signal type
-----	-----
0	LO ODU0
1	LO ODU1
2	LO ODU2
3	LO ODU3
4	LO ODU4
5	LO ODU2e
6	LO ODUflex
7-15	Reserved (for future use)

OD(T)Uk field	Signal type
-----	-----
0	Reserved (for future use)
1	HO ODU1 / OTU1
2	HO ODU2 / OTU2
3	HO ODU3 / OTU3
4	HO ODU4 / OTU4
5-15	Reserved (for future use)

T (2 bits): indicates the type of tributary slot of HO ODUk when LO ODUj is multiplexed into the HO ODUk (j<k). Currently, two types of tributary slot are defined in [G709-V3], the 1.25Gbps tributary slot and the 2.5Gbps tributary slot.

T field	TS type
-----	-----
0	1.25Gbps TS granularity
1	2.5Gbps TS granularity
2-3	Reserved (for future use)

In case of LO ODU_j mapped into OTU_k (j=k), this field is not necessary and should be ignored.

TPN (16 bits): indicates the Tributary Port Number (TPN) for the assigned Tributary Slot(s).

- In case of LO ODU_j multiplexed into HO ODU1/ODU2/ODU3, only the lower 6 bits of TPN field is significant and the other bits of TPN MUST be set to 0.
- In case of LO ODU_j multiplexed into HO ODU4, only the lower 7 bits of TPN field is significant and the other bits of TPN MUST be set to 0.
- In case of ODU_j mapped into OTU_k (j=k), the TPN is not needed and this field MUST be set to 0.

As per [G709-V3], The TPN is used to allow for correct demultiplexing in the data plane. When an LO ODU_j is multiplexed into HO ODU_k occupying one or more TSs, a new TPN value is configured at the two end of the HO ODU_k link and is put into the related MSI byte(s) in the OPU_k overhead at the (traffic) ingress end of the link, so that the other end of the link can learn which TS(s) is/are used by the LO ODU_j in the data plane.

According to [G709-V3], the rules of TPN assignment should be as the following tables:

Table 1 - TPN Assignment Rules (2.5Gbps TS granularity)

HO ODU _k	LO ODU _j	TPN	TPN Assignment Rules
ODU2	ODU1	1~4	Fixed, = TS# occupied by ODU1
ODU3	ODU1	1~16	Fixed, = TS# occupied by ODU1
	ODU2	1~4	Flexible, != other existing LO ODU2s' TPNs

Table 2 - TPN Assignment Rules (1.25Gbps TS granularity)

HO ODUk	LO ODUj	TPN	TPN Assignment Rules
ODU1	ODU0	1~2	Fixed, = TS# occupied by ODU0
ODU2	ODU1	1~4	Flexible, != other existing LO ODUs' TPNs
	ODU0 & ODUflex	1~8	Flexible, != other existing LO ODU0s and ODUflexes' TPNs
ODU3	ODU1	1~16	Flexible, != other existing LO ODUs' TPNs
	ODU2	1~4	Flexible, != other existing LO ODU2s' TPNs
	ODU0 & ODU2e & ODUflex	1~32	Flexible, != other existing LO ODU0s and ODU2es and ODUflexes' TPNs
ODU4	Any ODU	1~80	Flexible, != ANY other existing LO ODUs' TPNs

Note that in the case of "Flexible", the value of TPN is not relevant to the TS number as per [G709-V3].

Bit Map (variable): indicates which tributary slots in HO ODUk that the LO ODUj will be multiplexed into. The sequence of the Bit Map is consistent with the sequence of the tributary slots in HO ODUk. Each bit in the bit map represents the corresponding tributary slot in HO ODUk with a value of 1 or 0 indicating whether the tributary slot will be used by LO ODUj or not.

The size of the bit map equals to the total number of the tributary slots of HO ODUk, which is deduced by the ODU(T)k and T fields.

In case of an ODUk mapped into OTUk, it's no need to indicate which tributary slots will be used, so the size of Bit Map is 0.

Padded bits are added behind the Bit Map to make the whole label a multiple of four bytes if necessary. Padded bit MUST be set to 0 and MUST be ignored.

5.2. Examples

The following examples are given in order to illustrate the label format described in the previous sections of this document.

(1) ODUk into OTUk mapping:

In such conditions, the downstream node along an LSP returns a label indicating that the ODU1 (ODU2 or ODU3 or ODU4) is directly mapped into the corresponding OTU1 (OTU2 or OTU3 or OTU4). The following example label indicates an ODU1 mapped into OTU1.

```

0                1                2                3
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-----+-----+-----+-----+-----+-----+-----+-----+
|0 0 0 1|0 0 0 1|0 0| Reserved |                All 0s                |
+-----+-----+-----+-----+-----+-----+-----+-----+

```

(2) ODUj into ODUk multiplexing:

In such conditions, this label indicates that an ODUj is multiplexed into several tributary slots of OPUk and then mapped into OTUk. Some instances are shown as follow:

- ODU0 into ODU2 Multiplexing:

```

0                1                2                3
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-----+-----+-----+-----+-----+-----+-----+-----+
|0 0 0 0|0 0 1 0|0 0| Reserved |                TPN = 2                |
+-----+-----+-----+-----+-----+-----+-----+-----+
|0 1 0 0 0 0 0 0|                Padded Bits (0)                |
+-----+-----+-----+-----+-----+-----+-----+-----+

```

This above label indicates an ODU0 multiplexed into the second tributary slot of ODU2, wherein the type of the tributary slot is 1.25Gbps, and the TPN value is 2.

- ODU1 into ODU2 Multiplexing with 1.25Gbps TS granularity:

```

0                1                2                3
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-----+-----+-----+-----+-----+-----+-----+-----+
|0 0 0 1|0 0 1 0|0 0| Reserved |                TPN = 1                |
+-----+-----+-----+-----+-----+-----+-----+-----+
|0 1 0 1 0 0 0 0|                Padded Bits (0)                |
+-----+-----+-----+-----+-----+-----+-----+-----+

```

This above label indicates an ODU1 multiplexed into the 2nd and the 4th tributary slot of ODU2, wherein the type of the tributary slot is 1.25Gbps, and the TPN value is 1.

- ODU2 into ODU3 Multiplexing with 2.5Gbps TS granularity:

```

0                               1                               2                               3
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-----+-----+-----+-----+-----+-----+-----+-----+
|0 0 1 0|0 0 1 1|0 1| Reserved |                               TPN = 1 |
+-----+-----+-----+-----+-----+-----+-----+-----+
|0 1 1 0 1 0 1 0 0 0 0 0 0 0 0| Padded Bits (0) |
+-----+-----+-----+-----+-----+-----+-----+-----+

```

This above label indicates an ODU2 multiplexed into the 2nd, 3rd, 5th and 7th tributary slot of ODU3, wherein the type of the tributary slot is 2.5Gbps, and the TPN value is 1.

5.3. Label Distribution Procedure

This document does not change the existing label distribution procedures [RFC4328] for GMPLS except that the new ODUk label should be processed as follows.

When a node receives a generalized label request for setting up an ODUj LSP from its upstream neighbor node, the node should generate an ODU label according to the signal type of the requested LSP and the free resources (i.e., free tributary slots of ODUk) that will be reserved for the LSP, and send the label to its upstream neighbor node.

In case of ODUj to ODUk multiplexing, the node should firstly determine the size of the Bit Map field according to the signal type and the tributary slot type of ODUk, and then set the bits to 1 in the Bit Map field corresponding to the reserved tributary slots. The node should also assign a valid TPN, which does not collided with other TPN value used by existing LO ODU connections in the selected HO ODU link, and configure the expected multiplex structure identifier (ExMSI) using this TPN. Then, the assigned TPN is filled into the label.

In case of ODUk to OTUk mapping, the node only needs to fill the ODUj and the ODUk fields with corresponding values in the label. Other bits are reserved and MUST be set to 0.

When receiving an ODU label from its downstream neighbor node, the node should learn which ODU signal type is multiplexed or mapped into which ODU signal type by analyzing the ODUj and the ODUk fields.

In case of ODUj to ODUk multiplexing, the node should firstly determine the size of the Bit Map field according to the signal type and the tributary slot type of ODUk, and then obtain which tributary slots in ODUk are reserved by its downstream neighbor node according to the position of the bits that are set to 1 in the Bit Map field, so that the node can multiplex the ODUj into the reserved tributary slots of ODUk after the LSP is established. The node should also get the TPN value assigned by its downstream neighbor node from the label, and fill the TPN into the related MSI byte(s) in the OPUK overhead in the data plane, so that the downstream neighbor node can check whether the TPN received from the data plane is consistent with the ExMSI and determine whether there is any mismatch defect.

In case of ODUk to OTUk mapping, the size of Bit Map field is 0 and no additional procedure is needed.

Note that the procedures of other label related objects (e.g., Upstream Label, Label Set) are similar as described above.

Note also that the TPN in the label_ERO may not be assigned (i.e., TPN field = 0) if the TPN is requested to be assigned locally.

5.3.1. Notification on Label Error

When receiving an ODUk label from the neighbor node, the node should check the integrity of the label. An error message containing an "Unacceptable label value" indication ([RFC3209]) should be sent if one of the following cases occurs:

- The ODUj field does not match with the Traffic Parameters;
- The OD(T)Uk field does not match with the type of the selected link;
- The selected link only supports 2.5Gbps TS granularity while the T field in the label indicates the 1.25Gbps TS granularity;
- The label includes an invalid TPN value that breaks the TPN assignment rules;
- Not enough bits of Bit Map, or Bit Map with non-zero padding bits;

- The reserved resources (i.e., the number of "1" in the Bit Map field) do not match with the Traffic Parameters.

5.4. Supporting Virtual Concatenation and Multiplication

As per [VCAT], the VCGs can be created using Co-Signaled style or Multiple LSPs style.

In case of Co-Signaled style, the explicit ordered list of all labels reflects the order of VCG members, which is similar to [RFC4328]. In case of multiplexed virtually concatenated signals (NVC > 1), the first label indicates the components of the first virtually concatenated signal; the second label indicates the components of the second virtually concatenated signal; and so on. In case of multiplication of multiplexed virtually concatenated signals (MT > 1), the first label indicates the components of the first multiplexed virtually concatenated signal; the second label indicates components of the second multiplexed virtually concatenated signal; and so on.

In case of Multiple LSPs style, multiple control plane LSPs are created with a single VCG and the VCAT Call can be used to associate the control plane LSPs. The procedures are similar to section 6 of [VCAT].

5.5. Supporting Multiplexing Hierarchy

As described in [OTN-FRWK], one ODU_j connection can be nested into another ODU_k (j < k) connection, which forms the multiplexing hierarchy in the ODU layer. This is useful if there are some intermediate nodes in the network which only support ODU_k but not ODU_j switching.

For example, in Figure 2, assume that N3 is a legacy node which only supports [G709-V1] and does not support ODU0 switching. If an ODU0 connection between N1 and N5 is required, then we can create an ODU2 connection between N2 and N4 (or ODU1 / ODU3 connection, depending on policies and the capabilities of the two ends of the connection), and nest the ODU0 into the ODU2 connection. In this way, N3 only needs to perform ODU2 switching and does not need to be aware of the inner ODU0.

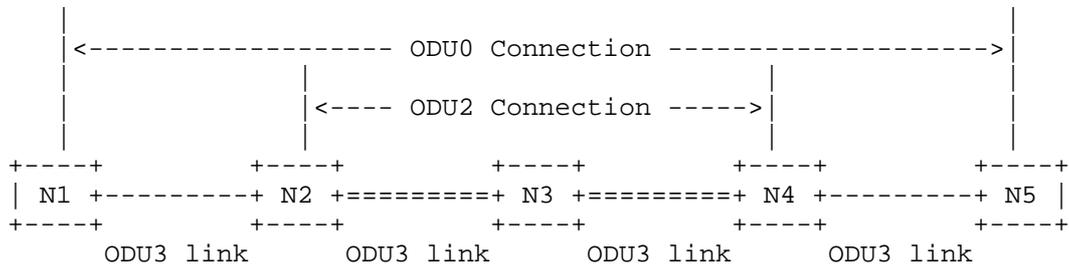


Figure 2 - Example of multiplexing hierarchy

The control plane signaling should support the provisioning of hierarchical multiplexing. Two methods are provided below (taking Figure 2 as example):

- The outer ODU2 connection is created in advance based on network planning, which is treated as a Forwarding Adjacency (FA). Then the inner ODU0 can be created using the resource of the ODU2 FA. In this case, the outer ODU2 and inner ODU0 connections are created separately, and the normal ODU connection creation procedure described in this document can be used.
- Using the multi-layer network signaling described in [RFC4206], [RFC6107] and [RFC6001] (including related modifications, if needed). That is, when the signaling message for ODU0 connection arrives at N2, a new RSVP session between N2 and N4 is triggered to create the ODU2 connection. This ODU2 connection is treated as an FA after it is created. And then the signaling procedure for the ODU0 connection can be continued using the resource of the ODU2 FA.

5.6. Control Plane Backward Compatibility Considerations

Since the [RFC4328] has been deployed in the network for the nodes that support [G709-V1] (herein we call them "legacy nodes"), backward compatibility SHOULD be taken into consideration when the new nodes (i.e., nodes that support [G709-V3]) and the legacy nodes are interworking.

For backward compatibility consideration, the new node SHOULD have the ability to generate and parse legacy labels.

- o For the legacy node, it always generates and sends legacy label to its upstream node, no matter the upstream node is new or legacy, as described in [RFC4328].
- o For the new node, it will generate and send legacy label if its upstream node is a legacy one, and generate and send new label if its upstream node is a new one.

One backwards compatibility example is shown in Figure 3:

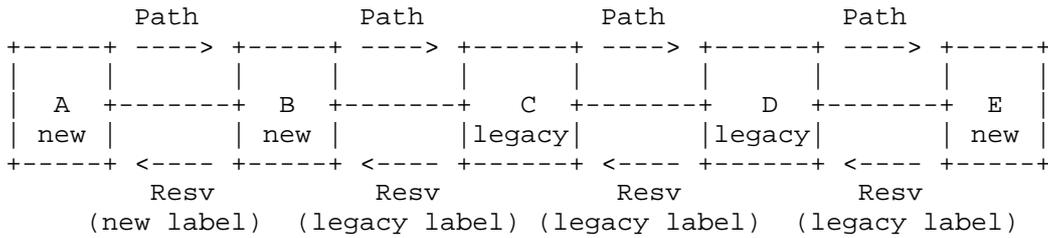


Figure 3 - Backwards compatibility example

As described above, for backward compatibility considerations, it is necessary for a new node to know whether the neighbor node is new or legacy.

One optional method is manual configuration. But it is recommended to use LMP to discover the capability of the neighbor node automatically, as described in [OTN-LMP].

When performing the HO ODU link capability negotiation:

- o If the neighbor node only support the 2.5Gbps TS and only support ODU1/ODU2/ODU3, the neighbor node should be treated as a legacy node.
- o If the neighbor node can support the 1.25Gbps TS, or can support other LO ODU types defined in [G709-V3]), the neighbor node should be treated as new node.
- o If the neighbor node returns a LinkSummaryNack message including an ERROR_CODE indicating nonsupport of HO ODU link capability negotiation, the neighbor node should be treated as a legacy node.

6. Security Considerations

This document introduces no new security considerations to the existing GMPLS signaling protocols. Referring to [RFC3473], further details of the specific security measures are provided. Additionally, [GMPLS-SEC] provides an overview of security vulnerabilities and protection mechanisms for the GMPLS control plane.

7. IANA Considerations

- G.709 SENDER_TSPEC and FLOWSPEC objects:

The traffic parameters, which are carried in the G.709 SENDER_TSPEC and FLOWSPEC objects, do not require any new object class and type based on [RFC4328]:

- o G.709 SENDER_TSPEC Object: Class = 12, C-Type = 5 [RFC4328]
- o G.709 FLOWSPEC Object: Class = 9, C-Type = 5 [RFC4328]

- Generalized Label Object:

The new defined ODU label (session 5) is a kind of generalized label. Therefore, the Class-Num and C-Type of the ODU label is the same as that of generalized label described in [RFC3473], i.e., Class-Num = 16, C-Type = 2.

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.
- [RFC4328] D. Papadimitriou, Ed. "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Extensions for G.709 Optical Transport Networks Control", RFC 4328, Jan 2006.
- [RFC3209] D. Awduche et al, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC3209, December 2001.

- [RFC3471] Berger, L., Editor, "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Functional Description", RFC 3471, January 2003.
- [RFC3473] L. Berger, Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions", RFC 3473, January 2003.
- [RFC3945] Mannie, E., "Generalized Multi-Protocol Label Switching (GMPLS) Architecture", RFC 3945, October 2004.
- [VCAT] G. Bernstein et al, "Operating Virtual Concatenation (VCAT) and the Link Capacity Adjustment Scheme (LCAS) with Generalized Multi-Protocol Label Switching (GMPLS)", draft-ietf-ccamp-gmpls-vcat-lcas-11.txt, March 9, 2011.
- [RFC4206] K. Kompella, Y. Rekhter, Ed., " Label Switched Paths (LSP) Hierarchy with Generalized Multi-Protocol Label Switching (GMPLS) Traffic Engineering (TE)", RFC 4206, October 2005.
- [RFC6107] K. Shiimoto, A. Farrel, "Procedures for Dynamically Signaled Hierarchical Label Switched Paths", RFC6107, February 2011.
- [RFC6001] Dimitri Papadimitriou et al, "Generalized Multi-Protocol Label Switching (GMPLS) Protocol Extensions for Multi-Layer and Multi-Region Networks (MLN/MRN)", RFC6001, February 21, 2010.
- [OTN-frwk] Fatai Zhang et al, "Framework for GMPLS and PCE Control of G.709 Optical Transport Networks", draft-ietf-ccamp-gmpls-g709-framework-02.txt, July 12, 2010.
- [OTN-info] S. Belotti et al, "Information model for G.709 Optical Transport Networks (OTN)", draft-bccg-ccamp-otn-g709-info-model-03.txt, Oct 18, 2010.
- [OTN-LMP] Fatai Zhang, Ed., "Link Management Protocol (LMP) extensions for G.709 Optical Transport Networks", draft-zhang-ccamp-gmpls-g.709-lmp-discovery-03.txt, May 13, 2010.
- [G709-V3] ITU-T, "Interfaces for the Optical Transport Network (OTN)", G.709/Y.1331, December 2009.

8.2. Informative References

- [G709-V1] ITU-T, "Interface for the Optical Transport Network (OTN)," G.709 Recommendation (and Amendment 1), February 2001 (November 2001).
- [G709-V2] ITU-T, "Interface for the Optical Transport Network (OTN)," G.709 Recommendation, March 2003.
- [G798-V2] ITU-T, "Characteristics of optical transport network hierarchy equipment functional blocks", G.798, December 2006.
- [G798-V3] ITU-T, "Characteristics of optical transport network hierarchy equipment functional blocks", G.798v3, consented June 2010.
- [RFC4506] M. Eisler, Ed., "XDR: External Data Representation Standard", RFC 4506, May 2006.
- [IEEE] "IEEE Standard for Binary Floating-Point Arithmetic", ANSI/IEEE Standard 754-1985, Institute of Electrical and Electronics Engineers, August 1985.
- [GMPLS-SEC] Fang, L., Ed., "Security Framework for MPLS and GMPLS Networks", Work in Progress, October 2009.

9. Authors' Addresses

Fatai Zhang
Huawei Technologies
F3-5-B R&D Center, Huawei Base
Bantian, Longgang District
Shenzhen 518129 P.R.China
Phone: +86-755-28972912
Email: zhangfatai@huawei.com

Guoying Zhang
China Academy of Telecommunication Research of MII
11 Yue Tan Nan Jie Beijing, P.R.China
Phone: +86-10-68094272
Email: zhangguoying@mail.ritt.com.cn

Sergio Belotti
Alcatel-Lucent
Optics CTO
Via Trento 30 20059 Vimercate (Milano) Italy
+39 039 6863033
Email: sergio.belotti@alcatel-lucent.it

Daniele Ceccarelli
Ericsson
Via A. Negrone 1/A
Genova - Sestri Ponente
Italy
Email: daniele.ceccarelli@ericsson.com

Yi Lin
Huawei Technologies
F3-5-B R&D Center, Huawei Base
Bantian, Longgang District
Shenzhen 518129 P.R.China
Phone: +86-755-28972914
Email: yi.lin@huawei.com

Yunbin Xu
China Academy of Telecommunication Research of MII
11 Yue Tan Nan Jie Beijing, P.R.China
Phone: +86-10-68094134
Email: xuyunbin@mail.ritt.com.cn

Pietro Grandi
Alcatel-Lucent
Optics CTO
Via Trento 30 20059 Vimercate (Milano) Italy
+39 039 6864930
Email: pietro_vittorio.grandi@alcatel-lucent.it

Diego Caviglia
Ericsson
Via A. Negrone 1/A
Genova - Sestri Ponente
Italy
Email: diego.caviglia@ericsson.com

Acknowledgment

This document was prepared using 2-Word-v2.0.template.dot.

Intellectual Property

The IETF Trust takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in any IETF Document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights.

Copies of Intellectual Property disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a 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 any standard or specification contained in an IETF Document. Please address the information to the IETF at ietf-ipr@ietf.org.

The definitive version of an IETF Document is that published by, or under the auspices of, the IETF. Versions of IETF Documents that are published by third parties, including those that are translated into other languages, should not be considered to be definitive versions of IETF Documents. The definitive version of these Legal Provisions is that published by, or under the auspices of, the IETF. Versions of these Legal Provisions that are published by third parties, including

those that are translated into other languages, should not be considered to be definitive versions of these Legal Provisions.

For the avoidance of doubt, each Contributor to the IETF Standards Process licenses each Contribution that he or she makes as part of the IETF Standards Process to the IETF Trust pursuant to the provisions of RFC 5378. No language to the contrary, or terms, conditions or rights that differ from or are inconsistent with the rights and licenses granted under RFC 5378, shall have any effect and shall be null and void, whether published or posted by such Contributor, or included with or in such Contribution.

Disclaimer of Validity

All IETF Documents and the information contained therein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY, THE IETF TRUST AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION THEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Copyright Notice

Copyright (c) 2010 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Network Working Group
Internet Draft
Category: Standards Track

Fatai Zhang
Dan Li
Huawei
F. Javier Jimenez Chico
O. Gonzalez de Dios
Telefonica Investigacion y Desarrollo
C. Margaria. C
Nokia Siemens Networks
March 11, 2011

Expires: September 11, 2011

GMPLS-based Hierarchy LSP creation
in Multi-Region and Multi-Layer Networks

draft-zhang-ccamp-gmpls-h-lsp-mln-03.txt

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at <http://www.ietf.org/ietf/lid-abstracts.txt>.

The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>.

This Internet-Draft will expire on September 11, 2011.

Abstract

This specification describes the hierarchy LSP creation models in the Multi-Region and Multi-Layer Networks (MRN/MLN), and provides the extensions to the existing protocol mechanisms described in [RFC4206], [RFC6107] and [RFC6001] to create the hierarchy LSP through multiple layer networks.

Conventions used in this document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

Table of Contents

1. Introduction	2
2. Terminology	3
3. Provisioning of FA-LSP in Server Layer Network	3
3.1. Selection of Switching Layers	3
3.2. Selection of Switching Granularity Levels	4
3.3. Selection of Adaptation Capabilities	6
4. Signaling Extension Requirements for Server Layer Selection ...	7
4.1. Model 1: Pre-provisioning of FA-LSP	8
4.2. Model 2: Signaling trigger server layer path computation .	9
4.3. Model 3: Full path computation at source node	9
5. ERO Sub-Object	10
5.1. Application of SERVER_LAYER_INFO sub-object	11
6. Security Considerations	12
7. IANA Considerations	12
8. Acknowledgments	12
9. References	12
10. Authors' Addresses	14

1. Introduction

Networks may comprise multiple layers which have different switching technologies or different switching granularity levels. The GMPLS technology is required to support control of such network.

[RFC5212] defines the concept of MRN/MLN and describes the framework and requirements of GMPLS controlled MRN/MLN. The GMPLS extension for MRN/MLN, including routing aspect and signaling aspect, is described in [RFC6001].

[RFC4206] and [RFC6107] describe how to set up a hierarchy LSP passing through multi-layer network and how to advertise the forwarding adjacency LSP (FA-LSP) created in the server layer network as a TE link via GMPLS signaling and routing protocols.

Based on these existing standards, this document further describes the provisioning of FA-LSP when the region nodes support multiple

It's possible that the edge node of a region is a hybrid node which has multiple ISCs in the server layer. In this case, selection of which server layer to create the FA-LSP is necessary.

Figure 2 shows an example multi-layer network, where node B and C are region edge nodes having three switching matrices which support, for instance, PSC, TDM and WDM switching, respectively. The three switching matrices are connected to each other by the internal links. Both the link between B and E and the link between E and C support TDM and WDM switching capabilities.

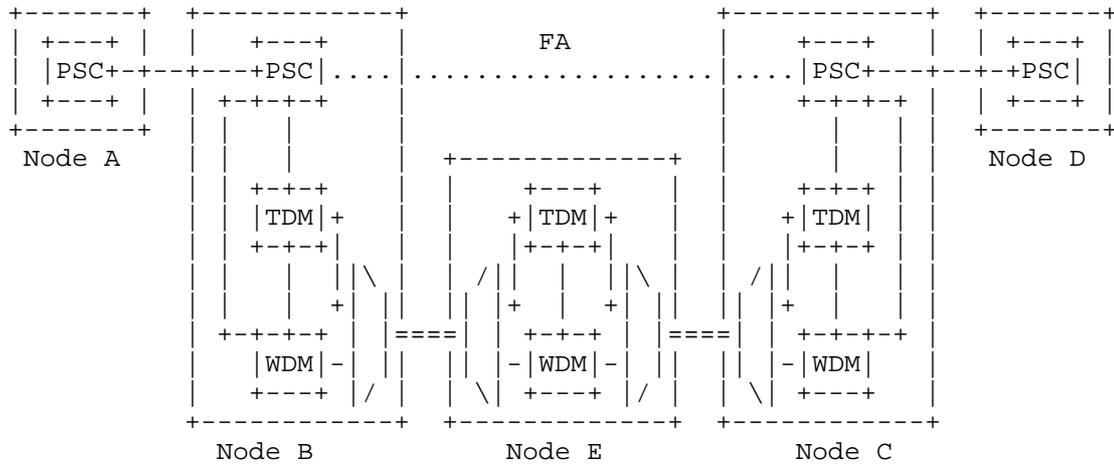


Figure 2 - MLN with multiple ISCs at edge node

As can be seen in Figure 2, there are two choices when providing FA in the PSC layer network between node B and C: one is creating FA-LSP with TDM switching matrix through node B, E and C, the other is creating FA-LSP with WDM switching matrix through node B, E and C.

[RFC6001] introduces a new SC (Switching Capability) sub-object into the XRO (ref. to [RFC4874]), which is used to indicate which switching capability is not expected to be used. When one of the switching capabilities is selected, the SC sub-object can be included in the message to exclude all other SCs.

3.2. Selection of Switching Granularity Levels

Even in the case that the edge node only has one switching capability in the server layer, there may be still multiple choices for the server layer network to set up FA-LSP to provide new FA in the client layer network. This is because the server layer network may have the

capability of providing different switching granularity levels for the FA-LSP.

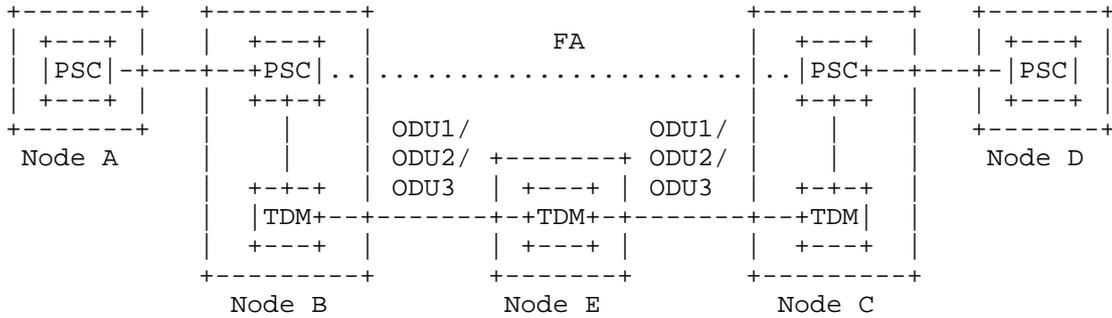


Figure 3a - Multiple switching granularities in server layer

Figure 3a shows an example multi-region network, where the edge node B and C have PSC and TDM switching matrices, and where the TDM switching matrix supports ODU1, ODU2 and ODU3 switching levels. Therefore, when an FA between node B and C in the PSC layer network is needed, either of ODU1, ODU2 or ODU3 connection (FA-LSP) can be created in the TDM layer network.

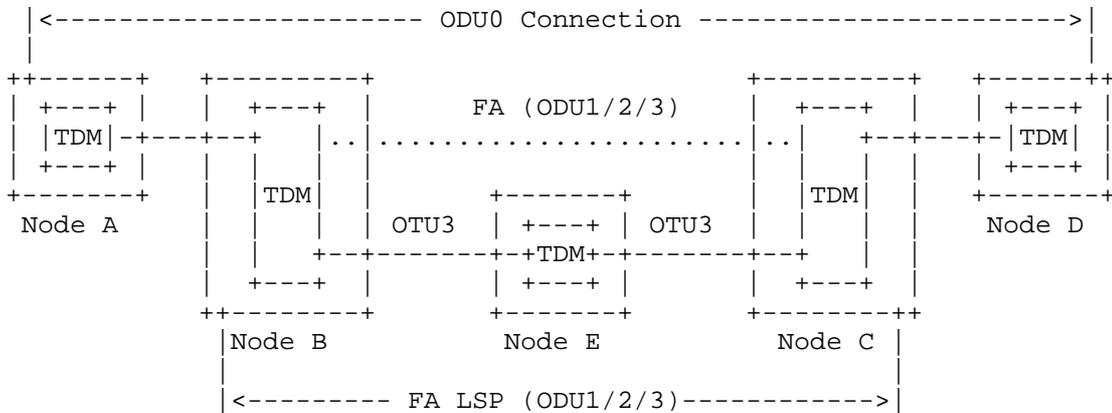
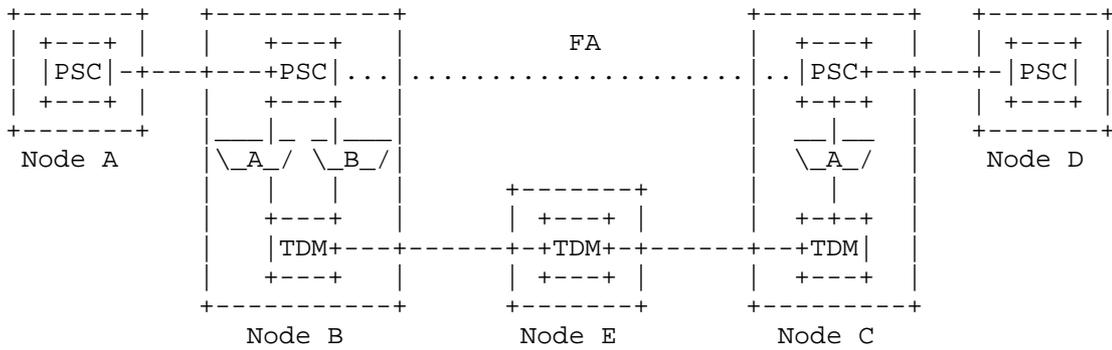


Figure 3b - TDM nested LSP provisioning



A /: Adaptation_Function_A; _B_ /: Adaptation_Function_B;

Figure 4 - Selection of adaptation function

For example, in figure 4, the edge node B supports two adaptation functions, i.e., adaptation_function_A and adaptation_function_B, while the edge node C only supports adaptation_function_A. In this case, only adaptation_function_A can be used for the server layer connection.

The Call procedure ([RFC4974]) between edge node B and C may be used to negotiate and determine the adaptation function for the server layer if the call function is supported.

4. Signaling Extension Requirements for Server Layer Selection

[RFC5623], the framework of PCE-based MLN, provides the models of cross-layer LSP path computation and creation, which are listed below:

- Inter-Layer Path Computation Models:
 - o Single PCE
 - o Multiple PCE with inter-PCE
 - o Multiple PCE without inter-PCE
- Inter-Layer Path Control Models:
 - o PCE-VNTM cooperation

- o Higher-layer signaling trigger
- o NMS-VNTM cooperation (integrated flavor)
- o NMS-VNTM cooperation (separate flavor)

This session keeps align with [RFC5623] except that the restriction of using PCE for path computation is not necessary (i.e., other element, such as network node, may also have path computation capability).

In this document, those models in [RFC4206] are reclassified into 3 models on the viewpoint of signaling:

- Model 1: Pre-provisioning of FA-LSP
- Model 2: Signaling trigger server layer path computation
- Model 3: Full path computation at source node

4.1. Model 1: Pre-provisioning of FA-LSP

In this model, the FA-LSP in the server layer is created before initiating the signaling of the client layer LSP. Two typical scenarios using this model are:

- Network planning and building at the stage of client network initialization.
- NMS/VNTM triggering the creation of FA-LSP when computing the path of client layer LSP. The path control models of PCE-VNTM cooperation and NMS-VNTM cooperation (both integrated and separate flavor) in [RFC5623] belong to this scenario.

In such case, the server layer selection and path computation is performed by planning tool or NMS/PCE/VNTM or the edge node. The signaling of client layer LSP and server layer FA-LSP are separated. The normal LSP creation procedures ([RFC3471] and [RFC3473]) are performed for these two LSPs and no new extension is required.

4.2. Model 2: Signaling trigger server layer path computation

In this model, the source node of client layer LSP only computes the route in its layer network. When the signaling of the client layer LSP reaches at the region edge node, the edge node performs server layer FA-LSP path computation and then creates the FA-LSP. When PCE is introduced to perform path computation in the multi-layer network, this model is the same as the model of "Higher-layer signaling trigger with Multiple PCE without inter-PCE" in [RFC5623].

In such case, the edge node will receive the client layer PATH message with a loose ERO indicating an FA is requested, and may perform the server layer selection (e.g., through the server layer PCE or the VNTM) and then compute and set up the path of the FA-LSP. The signaling procedure of client layer LSP and server layer FA-LSP is described detailedly in [RFC4206] and [RFC6107].

It's possible that the source node of the client layer LSP selects the server layer SC and/or granularity and/or adaptation function when performing path computation in the client layer, and requests or suggests the edge node to use an appointed server layer to create the FA-LSP.

The XRO including SC sub-object ([RFC6001]) is adopted for the server layer SC exclusion, which can be used indirectly to select server layer SC. Such solution is not straightforward enough and further more cannot be used for the selection of server layer granularity and adaptation function.

Therefore, in this case, new extensions for the selection of server layer SC, switching granularity and adaptation function are required.

4.3. Model 3: Full path computation at source node

In this model, the source node of the client layer LSP performs a full path computation including the client layer and the server layer routes. The server layer FA-LSP creation is triggered at the edge node by the client layer LSP signaling. When PCE is introduced to perform path computation in the multi-layer network, this model is the same as the model of "Higher-layer signaling trigger with Single PCE" or "Higher-layer signaling trigger with Multiple PCE with inter-PCE" in [RFC5623].

In such case, the server layer selection and server layer path computation is performed at the source node of the client layer LSP (e.g., through VNTM or PCE), but not at the edge node.

In [RFC4206], the ERO which contains the list of nodes and links (including the client layer and server layer) along the path is used in the client layer PATH message. The edge node can find out the tail end of the FA-LSP based on the switching capability of the node using the IGP database (see session 6.2 of [RFC 4206]).

Similar to the problem of model 2, the edge node is not aware of which switching granularity and which adaptation function to be selected for the FA-LSP because the ERO and/or XRO do not contain such information. Therefore, the edge node may not be able to create the FA-LSP, or may select another switching granularity by itself which is different from the one selected previously at the source node, which makes the creation of hierarchy LSP out of control.

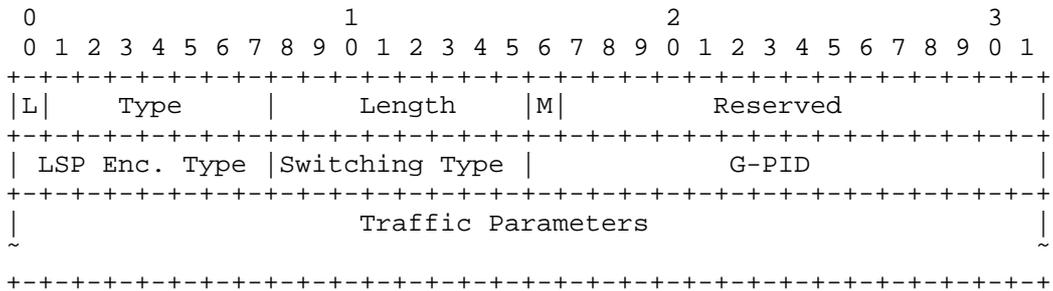
Therefore, new extensions for the selection of server layer SC, switching granularity and adaptation function are also required in this model.

5. ERO Sub-Object

In order to solve the problems described in the previous sessions, a new sub-object named SERVER_LAYER_INFO sub-object is introduced in this document, which is carried in the ERO and is used to indicate which server layer to create the FA-LSP.

The SERVER_LAYER_INFO sub-object is put immediately behind the node or link (interface) address sub-object, indicating the related node is a region edge node on the LSP in the ERO.

The format of the SERVER_LAYER_INFO sub-object is shown below:



- L bit: MUST be zero and MUST be ignored when received.
- Type: The SERVER_LAYER_INFO sub-object has a type of xx (TBD).
- Length: The total length of the sub-object in bytes, including the Type and Length fields. The value of this field is always a multiple of 4.
- M (Mandatory) bit: When set, it means the edge node MUST set up the FA-LSP in the appointed server layer; otherwise, the appointed server layer is suggested and the edge node may select other server layer by local policy.
- LSP Encoding Type, Switching Type and G-PID: These 3 fields are used to point out which switching layer is requested to set up the FA-LSP. The values of these 3 fields are inherited from the Generalized Label Request Object in GMPLS signaling, referring to [RFC3471], [RFC3473] and other related standards and drafts. Note that G-PID can be used to indicate the payload type of the server layer (i.e., the client signal) as well as the adaptation function for adapting the client signal into the server layer FA-LSP.
- Traffic Parameters: The traffic parameters field is used to indicate the switching granularity of the FA-LSP. The format of this field depends on the switching technology of the server layer (which can be deduced from the LSP Encoding Type and Switching Type fields in this sub-object) and is consistent with the existing standards and drafts. For example, the Traffic Parameters of Ethernet, SONET/SDH and OTN are defined by the [RFC6003], [RFC4606] and [OTN-ctrl] respectively.

5.1. Application of SERVER_LAYER_INFO sub-object

When a node receives a PATH message containing ERO and finds that there is a SERVER_LAYER_INFO sub-object immediately behind the node or link address sub-object related to itself, the node determines that it's a region edge node. Then, the edge node finds out the server layer selection information from the sub-object:

- Determine the switching layer by the LSP Encoding Type and Switching Type fields;
- Determine the switching granularity of the FA-LSP by the Traffic Parameters field;

- Determine the adaptation function for adapting the client signal into the server layer FA-LSP by the G-PID field.

The edge node MUST then determine the other edge of the region, i.e., the tail end of the FA-LSP, with respect to the subsequence of hops of the ERO. The node that satisfies the following conditions will be treated as the tail end of the FA-LSP:

- There is a SERVER_LAYER_INFO sub-object that immediately behind the node or link address sub-object which is related to that node;
- The LSP Encoding Type, Switching Type, G-PID and the Traffic Parameters fields of this SERVER_LAYER_INFO sub-object is the same as the SERVER_LAYER_INFO sub-object corresponding to the head end;
- The node is the first one that satisfies the two conditions above in the subsequence of hops of the ERO.

If a match of tail end is found, the head end now has the clear server layer information of the FA-LSP and then initiates an RSVP-TE session to create the FA-LSP in the appointed server layer between the head end and the tail end.

6. Security Considerations

TBD.

7. IANA Considerations

TBD.

8. Acknowledgments

TBD.

9. References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC3945] Mannie, E., "Generalized Multi-Protocol Label Switching (GMPLS) Architecture", RFC 3945, October 2004.

- [RFC3209] D. Awduche et al, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC3209, December 2001.
- [RFC3471] Berger, L., Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Functional Description", RFC 3471, January 2003.
- [RFC3473] L. Berger, Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions", RFC 3473, January 2003.
- [RFC5212] K. Shiomoto et al, "Requirements for GMPLS-Based Multi-Region and Multi-Layer Networks (MRN/MLN)", RFC5212, July 2008.
- [RFC5339] JL. Le Roux et al, "Evaluation of Existing GMPLS Protocols against Multi-Layer and Multi-Region Networks (MLN/MRN)", RFC5339, September 2008.
- [RFC4206] K. Kompella et al, "Label Switched Paths (LSP) Hierarchy with Generalized Multi-Protocol Label Switching (GMPLS) Traffic Engineering (TE)", RFC4206, October 2005.
- [RFC6107] K. Shiomoto, A. Farrel, "Procedures for Dynamically Signaled Hierarchical Label Switched Paths", RFC6107, February 2011.
- [RFC4974] D. Papadimitriou and A. Farrel, "Generalized MPLS (GMPLS) RSVP-TE Signaling Extensions in Support of Calls", RFC4974, August 2007.
- [RFC6001] Dimitri Papadimitriou et al, "Generalized Multi-Protocol Label Switching (GMPLS) Protocol Extensions for Multi-Layer and Multi-Region Networks (MLN/MRN)", RFC6001, October, 2010.
- [RFC5623] E. Oki et al, "Framework for PCE-Based Inter-Layer MPLS and GMPLS Traffic Engineering", RFC 5623, September 2009.
- [RFC4606] E. Mannie, D. Papadimitriou, "Generalized Multi-Protocol Label Switching (GMPLS) Extensions for Synchronous Optical Network (SONET) and Synchronous Digital Hierarchy (SDH) Control", RFC 4606, August 2006.

- [OTN-ctrl] Fatai Zhang et al, "Generalized Multi-Protocol Label switching (GMPLS) Signaling Extensions for the evolving G.709 Optical Transport Networks Control", draft-zhang-ccamp-gmpls-evolving-g709-04.txt, February 27, 2010.
- [RFC6003] D. Papadimitriou, "Ethernet Traffic Parameters", RFC6003, October, 2010.
- [IEEE] "IEEE Standard for Binary Floating-Point Arithmetic", ANSI/IEEE Standard 754-1985, Institute of Electrical and Electronics Engineers, August 1985.

10. Authors' Addresses

Fatai Zhang
Huawei Technologies
F3-5-B R&D Center, Huawei Base
Bantian, Longgang District
Shenzhen 518129 P.R.China

Phone: +86-755-28972912
Email: zhangfatai@huawei.com

Dan Li
Huawei Technologies
F3-5-B R&D Center, Huawei Base
Bantian, Longgang District
Shenzhen 518129 P.R.China

Phone: +86-755-28970230
Email: huawei.danli@huawei.com

Yi Lin
Huawei Technologies Co., Ltd.
F3-5-B R&D Center, Huawei Base
Bantian, Longgang District
Shenzhen 518129 P.R.China

Phone: +86-755-28972914
Email: yi.lin@huawei.com

Francisco Javier Jimenez Chico
Telefonica Investigacion y Desarrollo
Emilio Vargas 6
Madrid, 28043 Spain

Phone: +34 913379037
Email: fjjc@tid.es

Oscar Gonzalez de Dios
Telefonica Investigacion y Desarrollo
Emilio Vargas 6
Madrid, 28045 Spain

Phone: +34 913374013
Email: ogondio@tid.es

Cyril Margaria
Nokia Siemens Networks
St Martin Strasse 76
Munich, 81541
Germany

Phone: +49 89 5159 16934
Email: cyril.margaria@nsn.com

Intellectual Property

The IETF Trust takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in any IETF Document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights.

Copies of Intellectual Property disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a 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 any standard or specification contained in an IETF Document. Please address the information to the IETF at ietf-ipr@ietf.org.

The definitive version of an IETF Document is that published by, or under the auspices of, the IETF. Versions of IETF Documents that are published by third parties, including those that are translated into other languages, should not be considered to be definitive versions of IETF Documents. The definitive version of these Legal Provisions is that published by, or under the auspices of, the IETF. Versions of these Legal Provisions that are published by third parties, including those that are translated into other languages, should not be considered to be definitive versions of these Legal Provisions.

For the avoidance of doubt, each Contributor to the IETF Standards Process licenses each Contribution that he or she makes as part of the IETF Standards Process to the IETF Trust pursuant to the provisions of RFC 5378. No language to the contrary, or terms, conditions or rights that differ from or are inconsistent with the rights and licenses granted under RFC 5378, shall have any effect and shall be null and void, whether published or posted by such Contributor, or included with or in such Contribution.

Disclaimer of Validity

All IETF Documents and the information contained therein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY, THE IETF TRUST AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION THEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Copyright Notice

Copyright (c) 2010 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents

(<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Network Working Group
Internet Draft
Category: Informational

Fatai Zhang
Huawei
O. Gonzalez de Dios
Telefonica Investigacion y Desarrollo
D. Ceccarelli
Ericsson
G. Bernstein
Grotto Networking
A. Farrel
Old Dog Consulting
March 14, 2011

Expires: September 14, 2011

Applicability of Generalized Multiprotocol Label Switching (GMPLS)
User-Network Interface (UNI)

draft-zhang-ccamp-gmpls-uni-app-01.txt

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at <http://www.ietf.org/ietf/lid-abstracts.txt>.

The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>.

This Internet-Draft will expire on September 14, 2011.

Abstract

Generalized Multiprotocol Label Switching (GMPLS) defines a series of protocols for the creation of Label Switched Paths (LSPs) in various switching technologies. The GMPLS User-Network Interface (UNI) was

developed in RFC4208 in order to be applied to an overlay network architectural model.

This document examines a number of GMPLS UNI application scenarios. It shows how techniques developed after the GMPLS UNI can be applied to automate or enable critical processes for these applications. This document also suggested simple extensions to existing technologies to further enable the UNI and points out some existing unresolved issues.

Conventions used in this document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

Table of Contents

1. Introduction	3
2. UNI Addressing	5
3. UNI Auto Discovery	6
4. UNI Path Computation	7
4.1. UNI Link Selection	7
5. UNI Path Provisioning	9
5.1. Flat Model	10
5.2. Stitching Model	10
5.3. Session Shuffling Model	11
5.4. Hierarchy Model	11
6. UNI Recovery	12
6.1. End-to-end Recovery	12
6.1.1. Serial Provisioning of Working & Protection Path ...	12
6.1.2. Concurrent Computation of Working & Protection Path.	13
6.2. Segment Recovery	13
7. UNI Call	14
7.1. Exchange of UNI Link Information	15
7.2. Control of Call Route	15
8. UNI Multicast	16
8.1. UNI Multicast Connection Model	16
8.2. UNI Multicast Connection Provisioning	17
9. Security Considerations	18
10. IANA Considerations	18
11. Acknowledgments	19
12. References	19
12.1. Normative References	19

12.2. Informative References 21
 13. Authors' Addresses 22

1. Introduction

Generalized Multiprotocol Label Switching (GMPLS) defines a series of protocols, including Open Shortest Path First - Traffic Engineering (OSPF-TE) [RFC4203] and Resource ReserVation Protocol - Traffic Engineering (RSVP-TE) [RFC3473], which can be used to create Label Switched Paths (LSPs) in a number of deployment scenarios with various transport technologies.

The User-Network Interface (UNI) reference point is defined in the Automatically Switched Optical Network (ASON) [G.8080]. According to [G.8080], the UNI may be implemented as a peering between a client-side entity (UNI-C) and a network-side entity (UNI-N). End-to-end connectivity between UNI-C nodes is achieved across the core network by three components: a UNI request from source UNI-C to source UNI-N; a core network connection from source UNI-N to destination UNI-N; and a UNI request from destination UNI-N to destination UNI-C.

The GMPLS overlay model, as per [RFC4208], can be applied at the UNI, as shown in Figure 1.

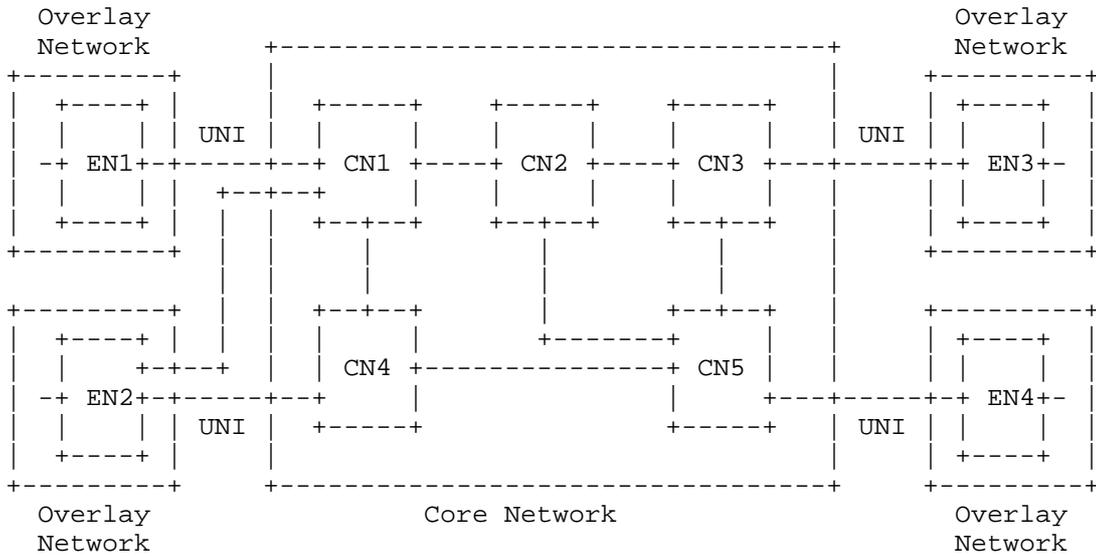


Figure 1 - Applying GMPLS overlay model at UNI

In Figure 1, assume that there is an end-to-end UNI connection passing through EN1-CN1-CN2-CN3-EN3. For convenience, some terms used in this document are defined below:

- "source EN" refers to the edge-node who initiates the connection (e.g., EN1);
- "destination EN" refers to the edge-node where the connection is terminated (e.g., EN3);
- "ingress CN" refers to the core-node to which the source EN is attached (e.g., CN1);
- "egress CN" refers to the core-node to which the destination EN is attached (e.g., CN3).

[RFC4208] provides mechanisms for UNI signaling, which are compatible with GMPLS RSVP-TE signaling ([RFC3471] and [RFC3473]). A single end-to-end RSVP session between source EN and destination EN is used for the user connection, just as it would be for connection creation between two core nodes. However, when considering the isolation of topology information between core network and the overlay network, additional processing of the RSVP-TE Explicit Route Object (ERO) and Record Route Object (RRO) is required. For example, the ingress CN should verify the ERO it received against its topology database before forwarding the PATH message. And the ingress/egress CN may edit or remove the RRO in order to hide the path segment used inside the core network from the EN.

The UNI can be used in many application scenarios. For example, in a multi-layer network [RFC6001], the interface between client layer node and server layer node can be seen as a UNI. Or, when deploying VPN services such as Layer One Virtual Private Networks (L1VPNs) [RFC4847], [RFC5253], users can connect to a service provider network via a UNI.

This document examines a number of current and future GMPLS application scenarios. It shows how techniques developed after the GMPLS UNI was developed can be used to automate or enable critical aspects of these application scenarios. It points out some potential technology extensions that could improve UNI operation, and highlights some existing unresolved issues.

2. UNI Addressing

In [RFC4208], the GMPLS overlay model is applied at the UNI reference point, and it is required that the edge-node and its attached core-node of the overlay network share the same address space that is used by GMPLS to signal between the edge-nodes across the core network. Under this condition, the user connection can be created using a single end-to-end RSVP session, which is consistent with the RSVP model. Therefore, RSVP-TE defined in [RFC3473] can be used for support GMPLS UNI without any extensions.

However, in the practical deployment of GMPLS UNI, the requirement of sharing the same address space between EN and its attached CN may not be satisfied if the core network and the overlay network are designed and deployed separately, especially if the two networks belong to different carriers. For example, the core network may use IPv6 addresses, while the overlay network uses IPv4 addresses. Or, since the core network is a closed system, the assignment of the IP addresses of the CNs is independent of other IP addresses outside the core network. This implies that the nodes in the core network may use addresses which collide with the edge nodes in the overlay network.

Thus, the addressing deployment for the GMPLS UNI can be divided into three scenarios:

1. Overlay network and core network share a common addressing policy. As noted above, there are many situations where this may be impractical, but it might be quite feasible in a multi-layer network operated by a single carrier. In this scenario, end-to-end UNI connectivity may use a single RSVP session, and the core routing information (assuming it is shared and not stripped for confidentiality reasons) will be meaningful to the ENs. Note, however, that the overlay model examined by this document assumes that there is some separation between the overlay and core networks, and this might mean that the overlay network is not able to see the topology or routing information of the core network even when they share a common address space.
2. ENs have visibility into the core network, but overlay and core networks have different address spaces. This is the more common model envisaged by [RFC4208] and for basic mode L1VPN deployments ([RFC5251]), and the previous scenario can be seen to be a special case of this scenario where the two address spaces are complementary. In this deployment, the source EN is aware of the addresses for itself, the ingress CN, the egress CN, and the destination EN in the address space of the core network. It may also have full visibility into the core network, but this is not a

requirement. In this scenario, the ENs are responsible for performing address mapping between the overlay network's addresses for the ENs, and the core network's addresses for the same nodes. Additionally, the source EN must use core network addresses to identify the CNs. In this deployment, a single end-to-end RSVP-TE session can still be utilized from source EN to destination EN.

3. ENs do not have any knowledge of the core address space and has no visibility into the core network. The source EN must still know an address for the ingress CN, but in this scenario, the EN uses an overlay address to reach the CN and to identify the destination EN. The ingress CN is responsible for mapping addresses to the core address space and for filling in any additional routing information. In this deployment the end-to-end connectivity must be created either using "session stitching" (see Section 5.2) or "session shuffling" (see Section 5.3).

3. UNI Auto Discovery

When the end-to-end connection is set up across the core network it must be targeted at the destination CN so that it can be extended to the destination EN. This means that either the source EN must know the identity of the destination CN to which the destination EN is attached, or the source CN must know this information. This requires some form of "discovery" (possibly including configuration), and depending on the addressing scheme in use (see Section 2) will require address mapping to be performed by the source EN or the source CN.

The discovery problem may be exacerbated when the a variety of services may be requested since the source EN will need to know the capabilities and available resources on the link between the destination CN and the destination EN. It could discover this by attempting to set up a connection and by drawing conclusions from the connection setup failures, but this is not efficient. Furthermore, in the case of a dual-homed destination EN (such as EN2 in Figure 1), a choice of destination CN must be made, and that choice may be influenced by the capabilities and available resources on the CN-EN links leading to the destination EN.

If the UNI is applied in L1VPN scenario, the auto discovery of UNI using OSPFv2 is provided in [RFC5252]. A new L1VPN LSA is introduced to advertise the L1VPN information via the L1VPN info TLV and the TE information of the CE-PE link (in the language of UNI, it's the EN-CN link) via the TE link TLV.

4. UNI Path Computation

End-to-end UNI path computation includes three parts: the selection of the source UNI link, the path computation inside the core network and the selection of the destination UNI link.

The selection of UNI links may not necessary in some scenarios. One example is in case of single-homing with only one UNI link between EN and CN, and another example is manual selection of UNI link when the service is requested. In such cases, the CN to which the source EN is attached, or the path Computation Element (PCE) ([RFC4655]) which is responsible for the core network, can perform the path computation across the core network when the UNI signaling request is sent from the source EN to the source CN.

4.1. UNI Link Selection

This document is specific to the overlay architectural model to the source EN which does not have the topology and TE information of the core network. Therefore, in the case of multi-homing (i.e., the source EN is connected to more than one CN), the source EN does not have enough information to make a correct choice among all the UNI links between itself and the core network for an optimal end-to-end connection.

In this case, a PCE whose computation domain covers both the core network and the ENs attached to it can be used. Note that the GMPLS UNI predates PCE and hence a PCE was not available to solve this problem in early GMPLS UNI deployments. The PCE can use the UNI discovery mechanism described in Section 3 to learn the EN-CN relationship and the TE information of the UNI links, and therefore has the ability to select the optimal UNI link for the connection.

Figure 2 shows an example of UNI path computation using a single PCE with visibility into both networks. The PCE can help the source EN to compute the end-to-end connection when the UNI path computation request is received, so that the source EN can learn which UNI link to be selected.

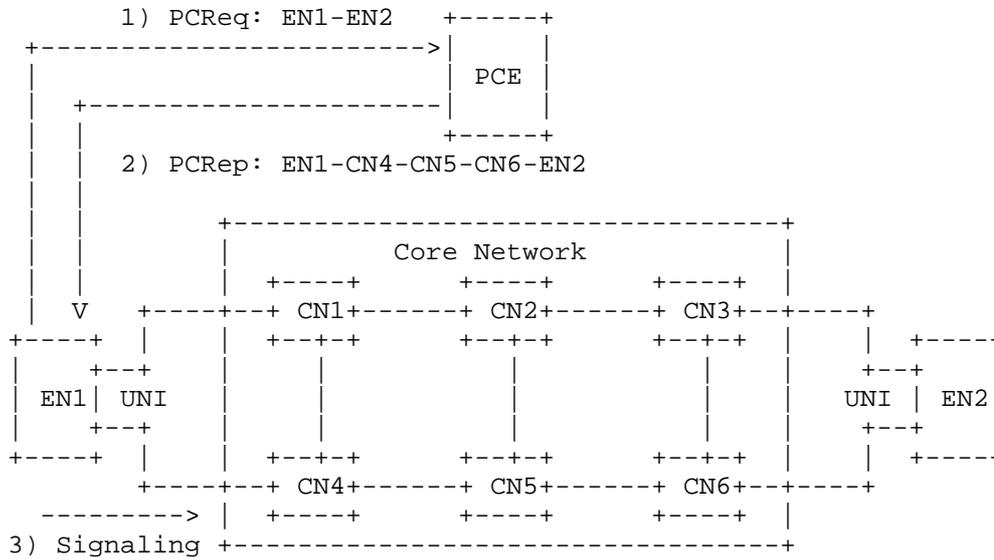
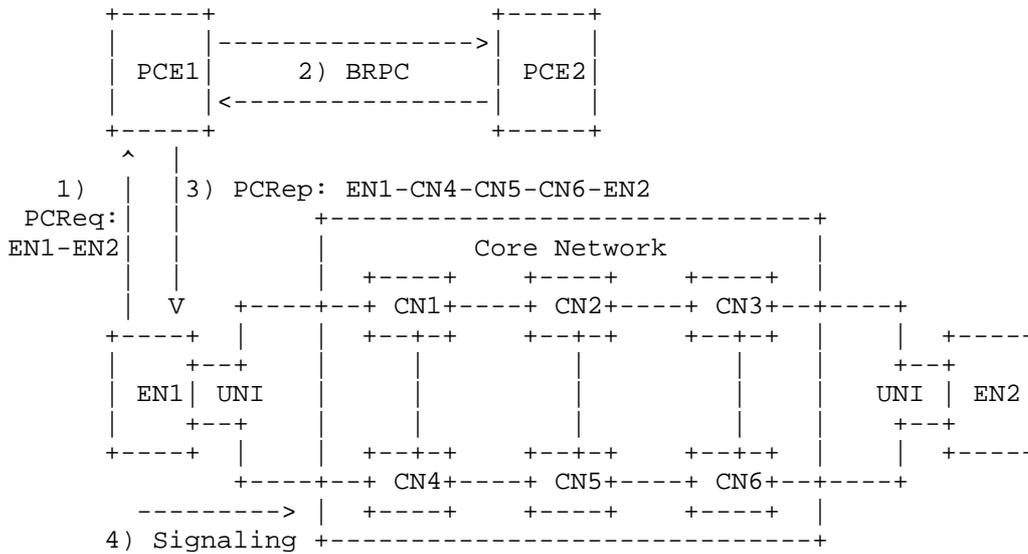


Figure 2 - PCE for UNI path computation (1)

Alternatively, the path can be computed by cooperating PCEs, as shown in Figure 3. The source EN does not experience any difference in behavior in that it sends its computation request to its local PCE, and receives a response telling it what path to use. However, the local PCE may not be aware of the topology of the core network and may need to contact a second PCE to supply the missing information.



(BRPC: Backward-Recursive PCE-Based Computation, see [RFC5441])

Figure 3 - PCE for UNI path computation (2)

If confidentiality of the topology within the core network needs to be preserved, the Path Key Subobject (PKS) can be used for either approach outlined here (see [RFC5520] and [RFC5553]). In the PCRep message returned to EN1, the Confidential Path Segment (CPS) (i.e., CN4-CN5-CN6) is encoded as a PKS by the PCE. Therefore, the EN1 only learns the selected UNI link from PCE. When receiving the UNI signaling carrying the PKS from EN1, CN4 can request the PCE to decode the PKS and then continue to create the connection.

Note that in both cases the PCE should be visible to the ENs and there should be control channel between PCE and EN for the transmission of PCEP messages. An alternative implementation could be that the PCE is located inside each CN to which the source EN is attached, so that the source EN can use the UNI control channel to send and receive the PCEP messages.

5. UNI Path Provisioning

The basic GMPLS UNI application is to provide end-to-end connections between edge-nodes through a core network via the overlay model.

6. UNI Recovery

One of the significant uses of GMPLS is to provide recovery mechanisms for connections, which is also needed in many UNI scenarios.

6.1. End-to-end Recovery

In the case of multi-homing, UNI end-to-end recovery is possible. As shown in Figure 7, the working path (W) and the protection path (P) are disjoint from each other not only inside the core network, but also at both the source and destination sides of the UNI. Mechanisms need to be provided to ensure the selection of disjoint working and backup paths.

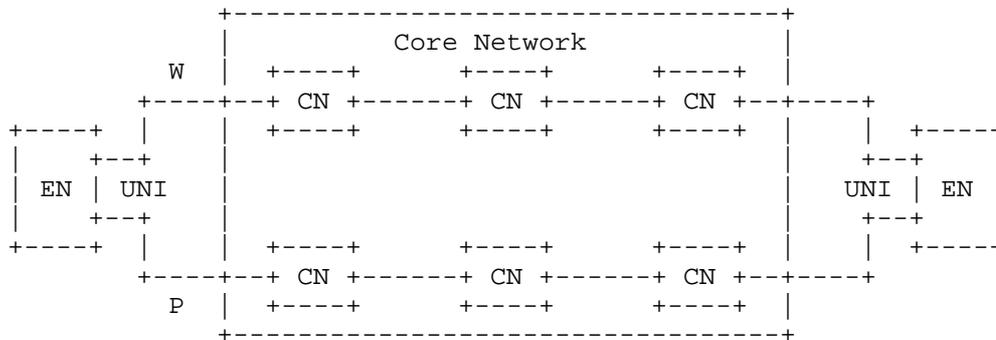


Figure 7 - UNI end-to-end recovery

6.1.1. Serial Provisioning of Working & Protection Path

In the case that the working path is computed and created before the protection path, path computation needs to compute a disjoint (or maximally disjoint) protection path given this existing working path.

If the information concerning the working path segment traversing the core network is known by the EN without considering the confidentiality, then the EN can easily use the RRO to collect the working path information, and use the XRO to exclude the working path when creating the protection path, as described in [RFC4874].

But in most cases, in order to preserve the confidentiality of topology within the core network, the information of path segment traversing the core network should be hidden from the EN. In such

case, the RRO & XRO mechanism in [RFC4874] cannot be used. An alternative would be to only collect the Shared Risk Group (SRG) information but not the full path information. This is because the SRG information is normally less confidential than the information of node ID and link ID.

In an application scenario where a PCE is involved inside the core network, then the Path Key mechanism can be used. The confidential path segment, i.e., the working path segment traversing the core network, is encoded as a PKS by the PCE when computing the working path. This PKS can be brought to the source EN, so when it request that the PCE compute a protection path, the PKS can be used to exclude the working path segment inside the core network.

[RFC5520] provides a mechanism to hide the CPS using PKS in the PCEP message, while [RFC5553] makes extensions to RSVP-TE to carry the PKS in ERO and RRO objects. It is required that the PKS should also be allowed to be carried in the XRO in both PCEP message and RSVP-TE signaling.

6.1.2. Concurrent Computation of Working & Protection Path

Alternatively, the working and protection path can be computed at the same time (e.g., by PCE or by one of the CNs to which the source EN is attached).

[PCE-GMPLS] allows requesting the PCE for path computation with specified protection type defined in [RFC4872]. Therefore, it's possible that the source EN requests the edge CN or PCE to compute both the working and the protection path at the same time. At this time, the disjunction problem can be resolved inside the path computation server.

Same as described in the previous section, the path segment traversing the core network can be encoded as a PKS if confidentiality is requested.

6.2. Segment Recovery

The UNI connection may only request protection inside the core network, especially in case of single-homing. One UNI segment protection example is shown in Figure 8.

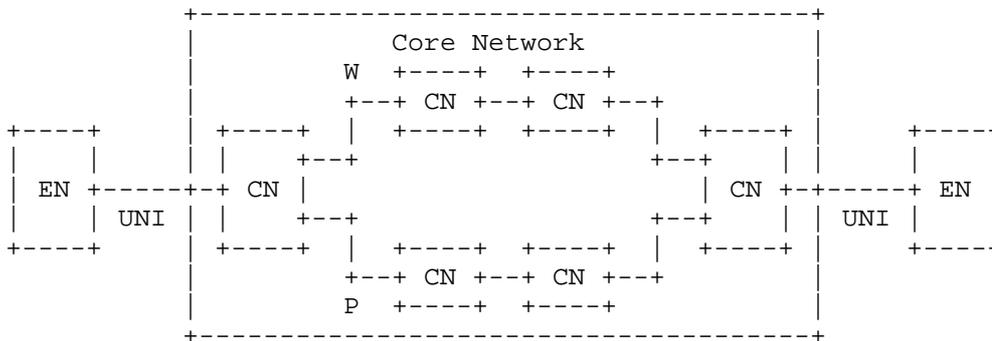


Figure 8 - UNI segment recovery

[RFC4873] provides the mechanism of segment recovery, in which the PROTECTION Object is extended to indicate the segment recovery, and the SERO object is introduced for the explicit control of the protection LSP between the branch node and the merge node.

However, due to the overlay model, the source EN may not have the information concerning the CN to which the destination EN is attached. In other words, the source EN does not know which node is the merge node of the UNI segment protection, so the SERO object cannot be used to request the edge CN for the UNI segment recovery. Therefore, segment recovery may not be controlled explicitly by the source EN.

7. UNI Call

The Call is a fundamental component of the ASON model [G.8080]. It is used to maintain the association between one or more user applications and the network to control the set-up, release, modification and maintenance of sets of connections. In simple cases, the Call and Connection can be established at the same time and in a strict one-to-one ratio. In this case, Call signaling is simple and requires only minor extensions to connection signaling. However, if Calls are to be handled separately from Connections, or if more than one Connection can be associated with a single Call, additional Call signaling is required.

The GMPLS Call, defined in [RFC4974], provides a mechanism to negotiate agreement between endpoints possibly in cooperation with the nodes that provide access to the network. Typically the GMPLS Call can be applied in the UNI scenario for access link capability exchange, policy, authorization, security, and so on.

7.1. Exchange of UNI Link Information

It is possible that the TE attributes of the access link (i.e., the UNI link) are not shared across the core network. So the source EN may not have the TE information of the destination access link as well as the capability of the destination EN. For example, in case of TDM network, the Virtual Concatenation (VCAT) and Link Capacity Adjustment Scheme (LCAS) capability of the destination EN may not be known.

In this case, the source EN can raise a Call carrying the LINK_CAPABILITY object to have a capability exchange with the destination EN, as described in [RFC4974].

7.2. Control of Call Route

When applying the Call, it's possible that there are multiple core network domains between the source EN (Call initiator) and the destination EN (Call terminator), or there is more than one Call manager in the core network (e.g., in the multi-homing scenario where the CNs to which the ENs are attached act as the Call managers).

In the both cases, when establishing the Call, there may be multiple alternative routes for the Call message to reach the destination EN. One can simply use the hop-by-hop manner (i.e., each Call manager determines the next Call manager to which the Call message will be sent by itself) to control the path of the Call.

However, in the practical deployment of UNI Call, commercial and policy motivations normally play an important role in selecting the Call route, especially in the multi-domain scenario. In this case, the hop-by-hop manner is not practical because the route of the Call needs to be pre-determined in consideration of commercial and policy factors before establishing the Call.

Therefore, it is desirable to allow full control of the Call by the source EN. That is, the source EN can identify the full Call route and signal it explicitly, so that the Call message can be forwarded along the desired route. Moreover, the management plane needs to be able to identify the Call route explicitly as an instruction to the source EN.

For the flat model, one end-to-end P2MP session as described in [RFC4875] can be used directly to create the P2MP LSP from source EN to leaf ENs.

For the stitching model, multiple P2P LSP segments or one P2MP LSP segment between the ingress CN and each egress CNs needs to be created and then stitched to the UNI P2MP LSP. GMPLS UNI signaling should have the capability to convey the multicast information by using stitching model.

For the session shuffling model, one end-to-end P2MP session can be used to create the P2MP LSP, with an address mapping performed at both ingress and egress CNs.

For the hierarchy model, multiple P2P LSP tunnels or one P2MP LSP tunnel between the ingress CN and each egress CNs needs be triggered by the UNI signaling for creating P2MP LSP. GMPLS UNI signaling should have the capability to convey the multicast information by using hierarchy model.

9. Security Considerations

[RFC5920] provides an overview of security vulnerabilities and protection mechanisms for the GMPLS control plane, which is applicable to this document.

The details of the specific security measures of the overlay network architectural model are provided in [RFC4208], which permits the core network to filter out specific RSVP objects to hide its topology from the EN.

Furthermore, if PCE is used, the security issues described in [RFC4655] and other related standards should also be considered.

Additionally, when the PKS mechanism is applied, the security issues can be dealt with using [RFC5520] and [RFC5553].

10. IANA Considerations

This informational document does not make any requests for IANA action.

11. Acknowledgments

TBD.

12. References

12.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC3209] D. Awduche et al, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC3209, December 2001.
- [RFC3471] Berger, L., Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Functional Description", RFC 3471, January 2003.
- [RFC3473] L. Berger, Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions", RFC 3473, January 2003.
- [RFC3945] Mannie, E., "Generalized Multi-Protocol Label Switching (GMPLS) Architecture", RFC 3945, October 2004.
- [RFC4203] Kompella, K., and Rekhter, Y., "OSPF Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4203, October 2005.
- [RFC4206] K. Kompella et al, "Label Switched Paths (LSP) Hierarchy with Generalized Multi-Protocol Label Switching (GMPLS) Traffic Engineering (TE)", RFC4206, October 2005.
- [RFC4208] G. Swallow et al, "Generalized Multiprotocol Label Switching (GMPLS) User-Network Interface (UNI): Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Support for the Overlay Model", RFC4208, October 2005.
- [RFC4655] A. Farrel et al, "A Path Computation Element (PCE)-Based Architecture", RFC4655, August 2006.
- [RFC4847] T. Takeda, Ed., "Framework and Requirements for Layer 1 Virtual Private Networks", RFC4847, April 2007.

- [RFC4872] J.P. Lang et al, "RSVP-TE Extensions in Support of End-to-End Generalized Multi-Protocol Label Switching (GMPLS) Recovery", RFC4872, May 2007.
- [RFC4873] L. Berger et al, "GMPLS Segment Recovery", RFC4873, May 2007.
- [RFC4874] CY. Lee et al, "Exclude Routes - Extension to Resource ReserVation Protocol-Traffic Engineering (RSVP-TE)", RFC4874, April 2007.
- [RFC4875] R. Aggarwal et al, "Extensions to Resource Reservation Protocol - Traffic Engineering (RSVP-TE) for Point-to-Multipoint TE Label Switched Paths (LSPs)", RFC4875, May 2007.
- [RFC4974] D. Papadimitriou and A. Farrel, Ed., "Generalized MPLS (GMPLS) RSVP-TE Signaling Extensions in Support of Calls", RFC4974, August 2007.
- [RFC5150] A. Ayyangar et al, "Label Switched Path Stitching with Generalized Multiprotocol Label Switching Traffic Engineering (GMPLS TE)", RFC5150, February 2008.
- [RFC5251] D. Fedyk and Y. Rekhter, Ed., "Layer 1 VPN Basic Mode", RFC5251, July 2008.
- [RFC5252] I. Bryskin and L. Berger Ed., "OSPF-Based Layer 1 VPN Auto-Discovery", RFC5252, July 2008.
- [RFC5520] R. Bradford, Ed., "Preserving Topology Confidentiality in Inter-Domain Path Computation Using a Path-Key-Based Mechanism", RFC5520, April 2009.
- [RFC5553] A. Farrel, Ed., "Resource Reservation Protocol (RSVP) Extensions for Path Key Support", RFC5553, May 2009.
- [RFC6001] Dimitri Papadimitriou et al, "Generalized Multi-Protocol Label Switching (GMPLS) Protocol Extensions for Multi-Layer and Multi-Region Networks (MLN/MRN)", RFC6001, October, 2010.
- [RFC6107] K. Shiomoto, A. Farrel, "Procedures for Dynamically Signaled Hierarchical Label Switched Paths", RFC6107, February 2011.

- [G.8080] ITU-T Rec. G.8080/Y.1304, "Architecture for the Automatically Switched Optical Network (ASON)," June 2006 (and Amend.2, September 2010).

12.2. Informative References

- [RFC4461] S. Yasukawa, Ed., "Signaling Requirements for Point-to-Multipoint Traffic-Engineered MPLS Label Switched Paths (LSPs)", RFC4461, April 2006.
- [RFC5212] K. Shiomoto et al, "Requirements for GMPLS-Based Multi-Region and Multi-Layer Networks (MRN/MLN)", RFC5212, July 2008.
- [RFC5253] T. Takeda, Ed., "Applicability Statement for Layer 1 Virtual Private Network (L1VPN) Basic Mode", RFC 5253, July 2008.
- [RFC5339] JL. Le Roux et al, "Evaluation of Existing GMPLS Protocols against Multi-Layer and Multi-Region Networks (MLN/MRN)", RFC5339, September 2008.
- [RFC5441] JP. Vasseur et al, "A Backward-Recursive PCE-Based Computation (BRPC) Procedure to Compute Shortest Constrained Inter-Domain Traffic Engineering Label Switched Paths", RFC5441, April 2009.
- [RFC5623] Oki, E., Takeda, T., Le Roux, J.L., and Farrel, A., "Framework for PCE-Based Inter-Layer MPLS and GMPLS Traffic Engineering", RFC 5623, September 2009.
- [RFC5920] L. Fang, Ed., "Security Framework for MPLS and GMPLS Networks", RFC5920, July 2010.
- [Call-ext] Fatai Zhang et al, "RSVP-TE extensions to GMPLS Calls", draft-zhang-ccamp-gmpls-call-extensions-01.txt, July 08, 2009.
- [PCE-GMPLS] C. Margaria et al, "PCEP extensions for GMPLS", draft-ietf-pce-gmpls-pcep-extensions-01.txt, October 24, 2010
- [SRLG-FA] Fatai Zhang et al, "RSVP-TE Extensions for Configuration SRLG of an FA", draft-zhang-ccamp-srlg-fa-configuration-01.txt, October 20, 2010.

[VCAT] G. Bernstein et al, "Operating Virtual Concatenation (VCAT) and the Link Capacity Adjustment Scheme (LCAS) with Generalized Multi-Protocol Label Switching (GMPLS)", draft-ietf-ccamp-gmpls-vcat-lcas-11.txt, March 9, 2011.

13. Authors' Addresses

Fatai Zhang
Huawei Technologies
F3-5-B R&D Center, Huawei Base
Bantian, Longgang District
Shenzhen 518129 P.R.China

Phone: +86-755-28972912
Email: zhangfatai@huawei.com

Oscar Gonzalez de Dios
Telefonica Investigacion y Desarrollo
Emilio Vargas 6
Madrid, 28045
Spain

Phone: +34 913374013
Email: ogondio@tid.es

Daniele Ceccarelli
Ericsson
Via A. Negrone 1/A
Genova - Sestri Ponente
Italy

Email: daniele.ceccarelli@ericsson.com

Greg M. Bernstein
Grotto Networking
Fremont California, USA

Phone: (510) 573-2237
Email: gregb@grotto-networking.com

Adrian Farrel
Old Dog Consulting

EMail: adrian@olddog.co.uk

Yi Lin
Huawei Technologies
F3-5-B R&D Center, Huawei Base
Bantian, Longgang District
Shenzhen 518129 P.R.China

Phone: +86-755-28972914
Email: yi.lin@huawei.com

Young Lee
Huawei Technologies
1700 Alma Drive, Suite 100
Plano, TX 75075
USA

Phone: (972) 509-5599 (x2240)
Email: leeyoung@huawei.com

Dan Li
Huawei Technologies
F3-5-B R&D Center, Huawei Base
Bantian, Longgang District
Shenzhen 518129 P.R.China

Phone: +86-755-28973237
Email: huawei.danli@huawei.com

Intellectual Property

The IETF Trust takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in any IETF Document or the extent to which any license under such rights might or might not be available; nor does it

represent that it has made any independent effort to identify any such rights.

Copies of Intellectual Property disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a 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 any standard or specification contained in an IETF Document. Please address the information to the IETF at ietf-ipr@ietf.org.

The definitive version of an IETF Document is that published by, or under the auspices of, the IETF. Versions of IETF Documents that are published by third parties, including those that are translated into other languages, should not be considered to be definitive versions of IETF Documents. The definitive version of these Legal Provisions is that published by, or under the auspices of, the IETF. Versions of these Legal Provisions that are published by third parties, including those that are translated into other languages, should not be considered to be definitive versions of these Legal Provisions.

For the avoidance of doubt, each Contributor to the IETF Standards Process licenses each Contribution that he or she makes as part of the IETF Standards Process to the IETF Trust pursuant to the provisions of RFC 5378. No language to the contrary, or terms, conditions or rights that differ from or are inconsistent with the rights and licenses granted under RFC 5378, shall have any effect and shall be null and void, whether published or posted by such Contributor, or included with or in such Contribution.

Disclaimer of Validity

All IETF Documents and the information contained therein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY, THE IETF TRUST AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION THEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Copyright Notice

Copyright (c) 2010 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Network Working Group
Internet-Draft
Intended status: Standards Track

Fatai Zhang
Dan Li
Huawei
O. Gonzalez de Dios
Telefonica Investigacion y Desarrollo
C. Margaria. C
Nokia Siemens Networks
March 11, 2011

Expires: September 11, 2011

RSVP-TE Extensions for Configuration SRLG of an FA
draft-zhang-ccamp-srlg-fa-configuration-02.txt

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at <http://www.ietf.org/ietf/lid-abstracts.txt>.

The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>.

This Internet-Draft will expire on September 11, 2011.

Abstract

This memo provides extensions for the Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) for the support of the automatic discovery of SRLG of an LSP.

Table of Contents

1. Introduction.....	2
2. RSVP-TE Requirements.....	4
2.1. SRLG Collection Indication.....	4
2.2. SRLG Collecting.....	4
2.3. SRLG Update.....	4
3. RSVP-TE Extensions.....	4
3.1. SRLG Collection Flag.....	4
3.2. SRLG sub-object.....	5
3.3. Signaling Procedures.....	6
4. Manageability Considerations.....	6
5. IANA Considerations.....	7
5.1. RSVP Attribute Bit Flags.....	7
5.2. ROUTE_RECORD Object.....	7
6. Security Considerations.....	7
7. References.....	7

1. Introduction

As described in [RFC4206], H-LSP (Hierarchical LSP) can be used for carrying one or more other LSPs. [RFC6107] further mentions the implementation of H-LSP. In packet networks, e.g. MPLS networks, H-LSP mechanism can be implemented by MPLS label stack. In non-packet networks where the label is implicit, label stacks are not possible, and H-LSPs rely on the ability to nest switching technologies. Thus, for example, a lambda switch capable (LSC) LSP can carry a time division multiplexing (TDM) LSP, but cannot carry another LSC LSP.

S-LSP (LSP Stitching), which is defined in [RFC5150], is an LSP that represents a segment of another LSP, i.e., the S-LSP is viewed as one hop by another LSP. As described in [RFC6107], in the data plane the LSPs are stitched so that there is no label stacking or nesting. Thus, an S-LSP must be of the same switching technology as the end-to-end LSP that it facilitates.

Therefore, H-LSP mechanism can be used in both multi-domain and multi-layer scenarios and S-LSP mechanism can only be used in multi-domain scenario.

Both of the H-LSP and S-LSP can be advertised as a TE link in a GMPLS routing instance for path computation purpose. As described in [RFC6107], if the LSP (H-LSP or S-LSP) is advertised in the same instance of the control plane that advertises the TE links from which the LSP is constructed, the LSP is called an FA.

In multi-domain or multi-layer context, the path information of an LSP may not be provided to the ingress node for confidential reasons and the ingress node may not run the same routing instance with the intermediate nodes traversed by the path. In such scenarios, the ingress node can not get the SRLG information of the path information which the LSP traverse.

Even if the ingress node has the same routing instance with the intermediate nodes traversed by the path, the path information of the H-LSP or S-LSP may not be provided to the ingress node. Hence the ingress node may also not know the SRLG of the path the LSP traverses.

In the case that the ingress node does not get the SRLG of the path the LSP traverses (i.e. H-LSP or S-LSP), there are disadvantages as follows:

- o SRLG-disjoint path, for instance in case of end-to-end path protection, cannot be calculated
- o Intermediate nodes of a pre-planned shared restoration LSP cannot correctly decide on the SRLG-disjointness between two PPRO (PRIMARY_PATH_ROUTE Object)
- o In case that an LSP is advertised as a TE-Link, the ingress node cannot provide the correct SRLG for the TE-Link automatically

In case that an LSP is advertised as a TE-Link, the SRLG information of the TE link needs to be configured manually or automatically. However, for manually configuration, there are some disadvantages (e.g., require configuration coordination and additional management; manual errors may be introduced) mentioned in Section 1.3.4 of [RFC6107].

In addition, Section 1.2 of [RFC6107] describes it is desirable to have a kind of automatic mechanism to advertise the FA (i.e., to signal an LSP and automatically coordinate its use and

advertisement in any of the ways with minimum involvement from an operator).

Thus, in order to provide the SRLG information to the TE link automatically when an LSP (H-LSP or S-LSP) is advertised as a TE link, allow disjoint path calculation at ingress and allow correct pre-planned shared LSP to correctly share resource, this document provides an automatic mechanism to collect the SRLG used by a LSP automatically.

2. RSVP-TE Requirements

2.1. SRLG Collection Indication

The head nodes of the LSP must be capable of indicating whether the SRLG information of the LSP should be collected during the signaling procedure of setting up an LSP.

2.2. SRLG Collecting

The SRLG information can be collected during the setup of an LSP. Then the endpoints of the LSP can get the SRLG information and use it for routing, sharing and TE link configuration purposes.

2.3. SRLG Update

When the SRLG information changes, the endpoints of the LSP need to be capable of updating the SRLG information of the path. It means that the signaling needs to be capable of updating the newly SRLG information to the endpoints.

3. RSVP-TE Extensions

3.1. SRLG Collection Flag

In order to indicate nodes that SRLG collection is desired, a new flag in the Attribute Flags TLV which can be carried in an LSP_REQUIRED_ATTRIBUTES Object is needed:

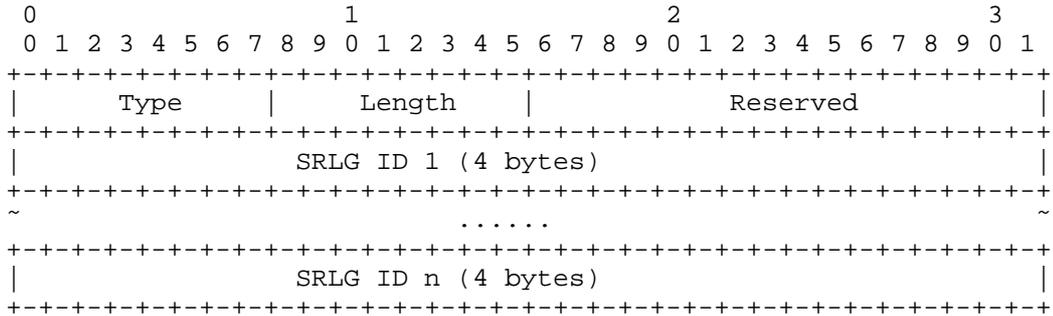
SRLG Collection flag (to be assigned by IANA, recommended bit zero)

The SRLG Collection flag is meaningful on a Path message. If the SRLG Collection flag is set to 1, it means that the SRLG information should be reported to the head and tail node along the setup of the LSP.

The rules of the processing of the Attribute Flags TLV are not changed.

3.2. SRLG sub-object

A new SRLG sub-object is defined for RRO(ROUTE_RECORD Object) to record the SRLG information of the LSP. Its format is modeled on the RRO sub-objects defined in [RFC3209].



Type

The type of the sub-object, to be assigned by IANA, which is recommended 34.

Length

The Length contains the total length of the sub-object in bytes, including the Type and Length fields. The Length depends on the number of SRLG IDs.

SRLG Id

The 32-bit identifier of the SRLG.

Reserved

This field is reserved. It SHOULD be set to zero on transmission and MUST be ignored on receipt.

The rules of the processing of the LSP_REQUIRED_ATTRIBUTES Object and ROUTE_RECORD Object are not changed.

3.3. Signaling Procedures

Typically, the head node gets the route information of an LSP by adding a RRO which contains the sender's IP addresses in the Path message. If a head node also desires SRLG recording, it sets the SRLG Collection Flag in the Attribute Flags TLV which can be carried in an LSP_REQUIRED_ATTRIBUTES Object.

When a node receives a Path message which carries an LSP_REQUIRED_ATTRIBUTES Object and the SRLG Collection Flag is set, if local policy determines that the SRLG information should not be provided to the endpoints, it must return a PathErr message to reject the Path message. Otherwise, it must add an SRLG sub-object to the RRO to carry the local SRLG information. Then it forwards the Path message to the next node in the downstream direction.

Following the steps described above, the intermediate nodes of the LSP can collect the SRLG information in the RRO during the forwarding of the Path message hop by hop. When the Path message arrives at the tail node, the tail node can get the SRLG information from the RRO.

Before the Resv message is sent to the upstream node, the tail node adds an SRLG sub-object to the RRO. The collected SRLG information can be carried in the SRLG sub-object. Therefore, during the forwarding of the Resv message in the upstream direction, the SRLG information is not needed to be collected hop by hop.

Based on the above procedure, the endpoints can get the SRLG information automatically. Then the endpoints can for instance advertise it as a TE link to the routing instance based on the procedure described in [RFC6107] and configure the SRLG information of the FA automatically.

It is noted that a node (e.g. the edge node of a domain) may edit the RRO to remove the route information (e.g. node, interface identifier information) before forwarding it due to some reasons (e.g. confidentiality or reduce the size of RRO), but the SRLG information should be retained if it is desirable for the endpoints of the LSP.

4. Manageability Considerations

TBD.

5. IANA Considerations

5.1. RSVP Attribute Bit Flags

The IANA has created a registry and manages the space of attributes bit flags of Attribute Flags TLV as described in section 11.3 of [RFC5420]. It is requested that the IANA makes assignments from the Attribute Bit Flags.

This document introduces a new Attribute Bit Flag:

- Bit number: TBD (0)
- Defining RFC: this I-D
- Name of bit: SRLG Collection Flag
- The meaning of the Attribute Flags TLV on a Path is defined in this I-D

5.2. ROUTE_RECORD Object

IANA has made the following assignments in the "Class Names, Class Numbers, and Class Types" section of the "RSVP PARAMETERS" registry located at <http://www.iana.org/assignments/rsvp-parameters>. We request that IANA make assignments from the ROUTE_RECORD [RFC3209] portions of this registry.

This document introduces a new RRO sub-object:

Type	Name	Reference
-----	-----	-----
TBD (34)	SRLG sub-object	This I-D

6. Security Considerations

TBD.

7. References

- [RFC3477] K. Kompella, Y. Rekhter, " Signalling Unnumbered Links in Resource ReSerVation Protocol - Traffic Engineering (RSVP-TE) ", rfc3477, January 2003.

- [RFC4206] K. Kompella, Y. Rekhter, " Label Switched Paths (LSP) Hierarchy with Generalized Multi-Protocol Label Switching (GMPLS) Traffic Engineering (TE) ", rfc4206, October 2005.
- [RFC4208] G. Swallow, J. Drake, Boeing, H. Ishimatsu, and Y. Rekhter, "Generalized Multiprotocol Label Switching (GMPLS) User-Network Interface (UNI): Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Support for the Overlay Model", RFC 4208, October 2005.
- [RFC4874] CY. Lee, A. Farrel, S. De Cnodder, " Exclude Routes - Extension to Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) ", rfc4874, April 2007.
- [RFC5150] Ayyangar, A., Vasseur, J.P, and Farrel, A., "Label Switched Path Stitching with Generalized Multiprotocol Label Switching Traffic Engineering (GMPLS TE)", RFC 5150, February 2008.
- [RFC5420] A. Farrel, D. Papadimitriou, J.P, and A. Ayyangar, "Encoding of Attributes for MPLS LSP Establishment Using Resource Reservation Protocol Traffic Engineering (RSVP-TE)", RFC 5420, February 2009.
- [RFC6107] K. Shiomoto, A. Farrel, " Procedures for Dynamically Signaled Hierarchical Label Switched Paths ", draft-ietf-ccamp-lsp-hierarchy-bis-08, August 2010.

Authors' Addresses

Fatai Zhang
Huawei Technologies
F3-5-B R&D Center, Huawei Base
Bantian, Longgang District
Shenzhen 518129 P.R.China

Phone: +86-755-28972912
Email: zhangfatai@huawei.com

Dan Li
Huawei Technologies
F3-5-B R&D Center, Huawei Base
Bantian, Longgang District
Shenzhen 518129 P.R.China

Phone: +86-755-28970230
Email: danli@huawei.com

Oscar Gonzalez de Dios
Telefonica Investigacion y Desarrollo
Emilio Vargas 6
Madrid, 28045
Spain

Phone: +34 913374013
Email: ogondio@tid.es

Cyril Margaria
Nokia Siemens Networks
St Martin Strasse 76
Munich, 81541
Germany

Phone: +49 89 5159 16934
Email: cyril.margaria@nsn.com

Xiaobing Zi
Huawei Technologies
F3-5-B R&D Center, Huawei Base
Bantian, Longgang District
Shenzhen 518129 P.R.China

Phone: +86-755-28973229
Email: zixiaobing@huawei.com

Intellectual Property

The IETF Trust takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in any IETF Document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights.

Copies of Intellectual Property disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a 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 any standard or specification contained in an IETF Document. Please address the information to the IETF at ietf-ipr@ietf.org.

The definitive version of an IETF Document is that published by, or under the auspices of, the IETF. Versions of IETF Documents that are published by third parties, including those that are translated into other languages, should not be considered to be definitive versions of IETF Documents. The definitive version of these Legal Provisions is that published by, or under the auspices of, the IETF. Versions of these Legal Provisions that are published by third parties, including those that are translated into other languages, should not be considered to be definitive versions of these Legal Provisions.

For the avoidance of doubt, each Contributor to the IETF Standards Process licenses each Contribution that he or she makes as part of the IETF Standards Process to the IETF Trust pursuant to the provisions of RFC 5378. No language to the contrary, or terms, conditions or rights that differ from or are inconsistent with the rights and licenses granted under RFC 5378, shall have any effect and shall be null and void, whether published or posted by such Contributor, or included with or in such Contribution.

Disclaimer of Validity

All IETF Documents and the information contained therein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY, THE IETF TRUST AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION THEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Copyright Notice

Copyright (c) 2010 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

MPLS Working Group
Internet-Draft
Intended status: Standards Track
Expires: August 29, 2011

F. Zhang, Ed.
ZTE
R. Jing
China Telecom
February 25, 2011

RSVP-TE Extensions to Establish Associated Bidirectional LSP
draft-zhang-mpls-tp-rsvp-te-ext-associated-lsp-03

Abstract

This document provides a method to bind two unidirectional LSPs into an associated bidirectional LSP, by extending the Extended ASSOCIATION object defined in [I-D.ietf-ccamp-assoc-info].

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on August 29, 2011.

Copyright Notice

Copyright (c) 2011 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

- 1. Introduction 3
- 2. Conventions used in this document 3
- 3. Association of Two Reverse Unidirectional LSPs 4
 - 3.1. Provisioning Model 4
 - 3.2. Signaling Procedure 4
 - 3.2.1. Asymmetric Bandwidth LSPs 5
 - 3.3. Recovery Considerations 6
- 4. Extensions to the Extended ASSOCIATION object 6
- 5. REVERSE_TSPEC Object 8
- 6. IANA Considerations 9
 - 6.1. Association Type of Two Reverse Unidirectional LSPs Association 9
 - 6.2. REVERSE_TSPEC Object 9
- 7. Security Considerations 9
- 8. Acknowledgement 10
- 9. References 10
 - 9.1. Normative references 10
 - 9.2. Informative References 11
- Authors' Addresses 11

1. Introduction

The associated bidirectional LSP, defined in [RFC5654], is constructed from a pair of unidirectional LSPs that are associated with each other at the LSP's ingress/egress points. It is useful for protection switching, for Operations, Administrations and Maintenance (OAM) messages that require a reply path. The corresponding requirements are also specified in as follow:

7 MPLS-TP MUST support associated bidirectional point-to-point LSPs.

11 The end points of an associated bidirectional LSP MUST be aware of the pairing relationship of the forward and reverse LSPs used to support the bidirectional service.

12 Nodes on the LSP of an associated bidirectional LSP where both the forward and backward directions transit the same node in the same (sub)layer as the LSP SHOULD be aware of the pairing relationship of the forward and the backward directions of the LSP.

50 The MPLS-TP control plane MUST support establishing associated bidirectional P2P LSP including configuration of protection functions and any associated maintenance functions.

Furthermore, these requirements are repeated in [I-D.ietf-ccamp-mpls-tp-cp-framework].

The notion of association as well as the corresponding RSVP ASSOCIATION object is defined in [RFC4872] and [RFC4873]. In that context, the object is used to associate recovery LSPs with the LSP they are protecting. This object also has broader applicability as a mechanism to associate RSVP state, and [I-D.ietf-ccamp-assoc-info] defines the Extended ASSOCIATION object that can be more generally applied.

This document extends the Extended ASSOCIATION object to establish associated bidirectional LSPs.

2. Conventions used in this document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC2119.

3. Association of Two Reverse Unidirectional LSPs

3.1. Provisioning Model

The associated bidirectional LSP's forward and backward directions are set up, monitored, and protected independently [RFC5654], so the configurations about it can be sent to one end or two ends. Depending on this, there are two models of signaling associated bidirectional LSP, one is the single sided provisioning and the other is the double sided provisioning.

For the single sided provisioning, the configurations are sent to one end. Firstly, the associated bidirectional TE tunnel is configured on this end, then a LSP under this tunnel is initiated with the Extended ASSOCIATION object carried in the Path message to trigger the peer end to set up the corresponding associated TE tunnel and LSP.

For the double sided provisioning, the commands are sent to two end points. They establish their associated bidirectional TE tunnels independently, then each one starts to set up the LSP using the Extended ASSOCIATION objects carried in the Path message to indicate each other to associate the two LSPs together to be an associated bidirectional LSP.

It can happen to bind two reverse unidirectional LSPs to be an associated bidirectional LSP not only when they are being created, but also when they have existed; or when one LSP exists, but the other end needs to be established. To all these scenarios, the provisioning models discussed above are applicable.

3.2. Signaling Procedure

Consider the following example, two reverse unidirectional LSPs are being established or have been established, the forward LSP1 is from A to B over [A,D,B], and the associated backward LSP2 is from B to A over [B,D,C,A].

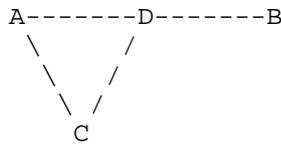


Figure 1: An example of associated bidirectional LSP

For single sided provisioning model, LSP1 is triggered by LSP2 or LSP2 is triggered by LSP1. When LSP2 is triggered by LSP1, the Extended ASSOCIATION object of LSP1 is initialized before the Extended ASSOCIATION objects of LSP2, The Extended ASSOCIATION object of LSP1 and LSP2 will carry the same value and this value SHOULD be LSP1'tunnel ID, LSP ID and tunnel sender address. When LSP1 is triggered by LSP2, the Extended ASSOCIATION object of LSP2 is initialized before the Extended ASSOCIATION objects of LSP1, The Extended ASSOCIATION object of LSP1 and LSP2 will carry the same value and this value SHOULD be LSP2'tunnel ID, LSP ID and tunnel sender address.

For double sided provisioning model, LSP1 and LSP2 are concurrently initialized, the values of the Extended ASSOCIATION object carried in LSP1's Path message are LSP1's tunnel ID, LSP ID and tunnel sender address; the values of the Extended ASSOCIATION object carried in LSP2's Path message are LSP1's tunnel ID, LSP ID and tunnel sender address. According to the general rules defined in [I-D.ietf-ccamp-assoc-info], the two LSPs cannot be bound together to be an associated bidirectional LSP because of the different values. In this case, the two edge nodes should firstly compare their router ID, then the bigger one sends Path refresh message, carrying the Extended ASSOCIATION object of the reverse LSP. Based on this Path refresh message, the two LSPs can be bounded together to be an associated bidirectional LSP also.

3.2.1. Asymmetric Bandwidth LSPs

There are some kind of applications, such as internet services and the return paths of OAM messages, which MAY have different bandwidth requirements for each direction. [RFC5654] specifies the requirements as follow:

14 MPLS-TP MUST support bidirectional LSPs with asymmetric bandwidth requirements, i.e., the amount of reserved bandwidth differs between the forward and backward directions.

The approach for supporting asymmetric bandwidth co-routed bidirectional LSPs is defined in [I-D.ietf-ccamp-asymm-bw-bidir-lsps-bis], which introduces three new objects named UPSTREAM_FLOWSPEC object, UPSTREAM_TSPEC object and UPSTREAM_ADSPEC object to represent the asymmetric upstream traffic flow. For the asymmetric bandwidth associated bidirectional LSPs, the existing SENDER_TSPEC, ADSPEC, and FLOWSPEC are complemented with the addition of new REVERSE_TSPEC object, which is used in exactly the same fashion as the old SENDER_TSPEC object, but refers to set up the reverse unidirectional LSP.

In the context of asymmetric associated bidirectional LSP, the REVERSE_TSPEC object MUST be carried in the LSP's Path message together with the Extended ASSOCIATION object whose Association Type is "Association of two reverse unidirectional LSPs" to trigger the peer end to set up the reverse LSP with the corresponding asymmetric bandwidth. For the single sided provisioning, the peer end just copies the value of the REVERSE_TSPEC object into the SENDER_TSPEC object in the Path message. For the double sided provisioning, the ends need to compare the values of the SENDER_TSPEC and REVERSE_TSPEC objects in the two Path messages. If match, the end with the bigger router ID sends Path refresh message, carrying the Extended ASSOCIATION object of the reverse LSP.

Nodes not supporting this extension will not recognize the new class number and should respond with an "Unknown Object Class".

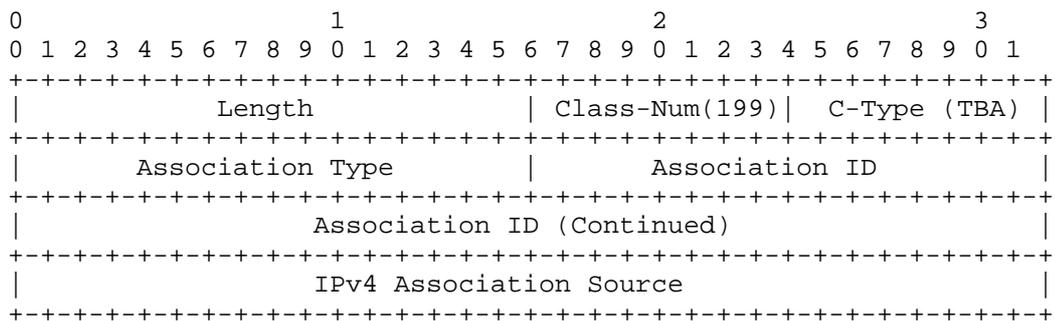
3.3. Recovery Considerations

Assume that LSP3 is used to protect LSP1, which can be established before or after the failure occurs, can share the same TE tunnel with LSP1 or not. LSP3 SHOULD inherits the associated bidirectional attributes between LSP1 and LSP2 when the traffic is switched from LSP1 to LSP3. This can be done by inserting the Extended ASSOCIATION object in LSP3's Path message with the same value as in LSP1's Path message.

4. Extensions to the Extended ASSOCIATION object

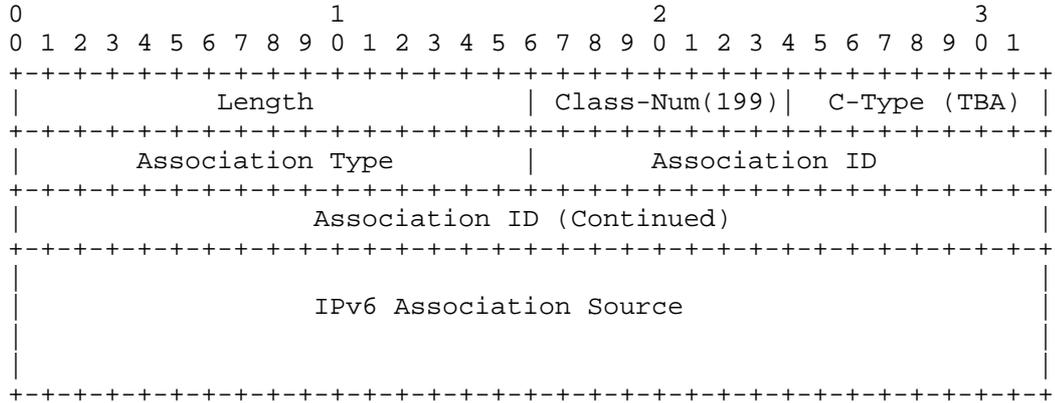
The Extended ASSOCIATION object is defined in [I-D.ietf-ccamp-assoc-info].

The Extended IPv4 ASSOCIATION object (Class-Num of the form 11bbbbbb with value = 199, C-Type = TBA) has the format:



Extended IPv4 ASSOCIATION object

The Extended IPv6 ASSOCIATION object (Class-Num of the form 11bbbbbb with value = 199, C-Type = TBA) has the format:



Extended IPv6 ASSOCIATION object

o Association Type:

Now, the following values of the Association Type have been defined.

Value	Type
0	Reserved
1	Recovery (R)
2	Resource sharing (S)

In order to bind two reverse unidirectional LSPs to be an associated bidirectional LSP, this document defined a new value:

Value	Type
4	Association of two reverse unidirectional LSPs (A)

If the downstream nodes do not know this Association Type, MUST return a PathErr message with error code/sub-code "LSP Admission Failure/Bad Association Type".

- o Association Source:

The general rule is that the address is "associated to the node that originate the association" and provide global scope (within the address space) to identified association, see [RFC4872] and [I-D.ietf-ccamp-assoc-info]. This document adds specific rules: the Association source MUST be set to the tunnel sender address of the initiating node.

- o Association ID:

The association ID is set to a value that uniquely identifies the set of LSPs to be associated and the generic definition does not provide any specific rules on how matching is to be done. In order to provide global Association ID based on MPLS-TP identifiers, this document adds specific rules: the first 16 bits MUST be set to the tunnel ID of the initiating node and the following 16 bits MUST be set to the LSP ID of the corresponding tunnel, the rest MUST be set to zero on transmission and MUST be ignored on receipt.

As described in [I-D.ietf-ccamp-assoc-info], association is always done based on matching Path state or Resv state. Upstream initialized association is represented in Extended ASSOCIATION objects carried in Path message and downstream initialized association is represented in Extended ASSOCIATION objects carried in Resv messages. The new defined association type in this document is only defined for use in upstream initialized association. Thus it can only appear in Extended ASSOCIATION objects signaled in Path message.

The rules associated with the processing of the Extended ASSOCIATION objects in RSVP message are discussed in [I-D.ietf-ccamp-assoc-info]. It said that in the absence of Association Type-specific rules for identifying association, the included Extended ASSOCIATION objects MUST be identical. This document adds no specific rules, the association will always operate based on the same Extended ASSOCIATION objects.

5. REVERSE_TSPEC Object

The REVERSE_TSPEC object is used in Path, PathTear, PathErr, and Notify message (via sender descriptor). This includes the definition of class type and format. It's class number is TBD (of the form 0bbbbbbb), and class type and format is the same as the SENDER_TSPEC object.

This object modifies the RSVP message-related formats defined in

[RFC2205], [RFC3209] and [RFC3473]. See [RFC5511] for the syntax used by RSVP. The format of the sender description for asymmetric associated bidirectional LSPs is:

```
<sender descriptor> ::= <SENDER_TEMPLATE> <SENDER_TSPEC>
                        [<ADSPEC>]
                        [<RCEORD_ROUTE>]
                        [<SUGGESTED_LABEL>]
                        [<RECOVERY_LABEL>]
                        <REVERSE_TSPEC>
```

6. IANA Considerations

IANA is requested to administer assignment of new values for namespace defined in this document and summarized in this section.

6.1. Association Type of Two Reverse Unidirectional LSPs Association

Within the current document, a new Association Type is defined in the Extended ASSOCIATION object.

Value	Type
TBD	Association of two reverse unidirectional LSPs (A)

6.2. REVERSE_TSPEC Object

A new class named REVERSE_TSPEC has been created in the 0bbbbbbb rang (TBD) with the following definition:

Class Types or C-types:

Same values as SENDER_TPSCE object (C-Num 12)

There are no other IANA considerations introduced by this document.

7. Security Considerations

This document introduces a new association type, and except this, there are no security issues about the Extended ASSOCIATION object are introduced here.

Furthermore, this document introduces the REVERSE_TSPEC object for use in GMPLS signaling [RFC3473], which is parallel the existing SENDER_TSPEC object. As such, any vulnerabilities that are due to the use of the old SENDER_TSPEC object now apply here also.

Otherwise, this document introduces no additional security considerations. For a general discussion on MPLS and GMPLS related security issues, see the MPLS/GMPLS security framework [RFC5920].

8. Acknowledgement

The authors would like to thank Lou Berger for his great guidance in this work, George Swallow for the discussion of recovery, Lamberto Sterling for his valuable comments on the section of asymmetric bandwidths. At the same time, the authors would also like to acknowledge the contributions of Bo Wu, Xihua Fu, Lizhong Jin for the initial discussions.

9. References

9.1. Normative references

- [I-D.ietf-ccamp-assoc-info]
Berger, L., Faucheur, F., and A. Narayanan, "Usage of The RSVP Association Object", draft-ietf-ccamp-assoc-info-00 (work in progress), October 2010.
- [I-D.ietf-ccamp-mpls-tp-cp-framework]
Andersson, L., Berger, L., Fang, L., Bitar, N., Gray, E., Takacs, A., Vigoureux, M., and E. Bellagamba, "MPLS-TP Control Plane Framework", draft-ietf-ccamp-mpls-tp-cp-framework-06 (work in progress), February 2011.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC4872] Lang, J., Rekhter, Y., and D. Papadimitriou, "RSVP-TE Extensions in Support of End-to-End Generalized Multi-Protocol Label Switching (GMPLS) Recovery", RFC 4872, May 2007.
- [RFC4873] Berger, L., Bryskin, I., Papadimitriou, D., and A. Farrel, "GMPLS Segment Recovery", RFC 4873, May 2007.
- [RFC5654] Niven-Jenkins, B., Brungard, D., Betts, M., Sprecher, N.,

and S. Ueno, "Requirements of an MPLS Transport Profile", RFC 5654, September 2009.

9.2. Informative References

- [I-D.ietf-ccamp-asymm-bw-bidir-lsps-bis]
Takacs, A., Berger, L., Caviglia, D., Fedyk, D., and J. Meuric, "GMPLS Asymmetric Bandwidth Bidirectional Label Switched Paths (LSPs)", draft-ietf-ccamp-asymm-bw-bidir-lsps-bis-01 (work in progress), January 2011.
- [RFC2205] Braden, B., Zhang, L., Berson, S., Herzog, S., and S. Jamin, "Resource ReSerVation Protocol (RSVP) -- Version 1 Functional Specification", RFC 2205, September 1997.
- [RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, December 2001.
- [RFC3473] Berger, L., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions", RFC 3473, January 2003.
- [RFC5511] Farrel, A., "Routing Backus-Naur Form (RBNF): A Syntax Used to Form Encoding Rules in Various Routing Protocol Specifications", RFC 5511, April 2009.
- [RFC5920] Fang, L., "Security Framework for MPLS and GMPLS Networks", RFC 5920, July 2010.

Authors' Addresses

Fei Zhang (editor)
ZTE

Email: zhang.fei3@zte.com.cn

Ruiquan Jing
China Telecom

Email: jingrq@ctbri.com.cn

Fan Yang
ZTE

Email: yang.fan5@zte.com.cn

Weilian Jiang
ZTE

Email: jiang.weilian@zte.com.cn

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: August 23, 2011

Z. Zheng
ZTE Corporation
February 19, 2011

RSVP-TE extensions for dynamic hostname traversing OSPF routing areas
draft-zheng-ccamp-rsvp-te-dynamic-hostname-00

Abstract

RFC 5642 defines an OSPF Router Information TLV that allows OSPF Routers to flood their hostname-to-Router-ID mapping information. Sometimes, when the operators create an inter-area MPLS LSP tunnel with Resource ReSerVation Protocol-Traffic Engineering (RSVP-TE), they need the hostname display on the CLI at the ingress node for management and operational reasons. This document describes extensions to RSVP-TE to support hostname-to-Router-ID mapping information traversing areas in an inter-area MPLS LSP tunnel situation.

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on August 23, 2011.

Copyright Notice

Copyright (c) 2011 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect

to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	3
1.1. Conventions Used in This Document	3
2. Implementation	3
2.1. Subobjects	4
2.1.1. Subobject 1: IPv4 address	4
2.1.2. Subobject 2: IPv6 address	4
2.2. Procedures	5
3. Security Considerations	6
4. IANA Considerations	6
5. Normative References	6
Author's Address	6

1. Introduction

RFC 5642 defines an OSPF Router Information TLV that allows OSPF Routers to flood their hostname-to-Router-ID mapping information. The flooding scope of the Dynamic Hostname TLV is controlled by the Opaque LSA type. Because of the constraint of the OSPF LSA flooding scope, routers in an area cannot get the hostname-to-Router-ID mapping information of the routers other than ASBRs in another area.

Sometimes, when the operators create an inter-area MPLS LSP tunnel with RSVP-TE, they need the hostname display on the CLI at the ingress node for management and operational reasons. However, as mentioned above, the ingress node may not have the hostname-to-Router-ID mapping information of the other nodes in the MPLS LSP tunnel.

This document describes extensions to RSVP-TE to support hostname-to-Router-ID mapping information traversing OSPF areas in an inter-area MPLS LSP tunnel situation.

1.1. Conventions Used in This Document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

2. Implementation

These extensions make use of the Notify message described in [RFC3473], by defining a new Dynamic Hostname Object. These extensions are OPTIONAL. In this implementation, Record Route Object MUST be contained in both Path and Resv message.

Dynamic Hostname Object is defined for the Notify message described in [RFC3473], to carry the hostname-to-Router-ID mapping information. The Class-Num needs to be assigned by the IANA. The suggested C-type is 1.

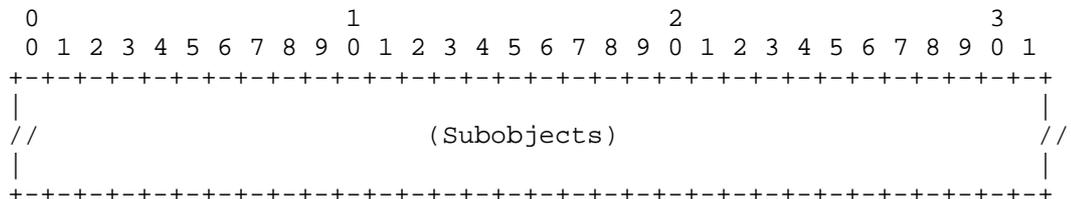


Figure 1: Subobjects

- Subobjects: The contents of a Dynamic Hostname object are a series of variable-length data items called subobjects. The subobjects are defined below.

The Dynamic Hostname Object SHOULD be presented in Notify messages.

2.1. Subobjects

2.1.1. Subobject 1: IPv4 address

The suggested Type is 1.



Figure 2: IPv4 address

- Type: 0x01 IPv4 address
- Length: The Length contains the total length of the subobject in bytes, including the Type and Length fields.
- IPv4 address: Router ID
- Hostname: See [RFC5642] section 3.1

2.1.2. Subobject 2: IPv6 address

The suggested Type is 1.

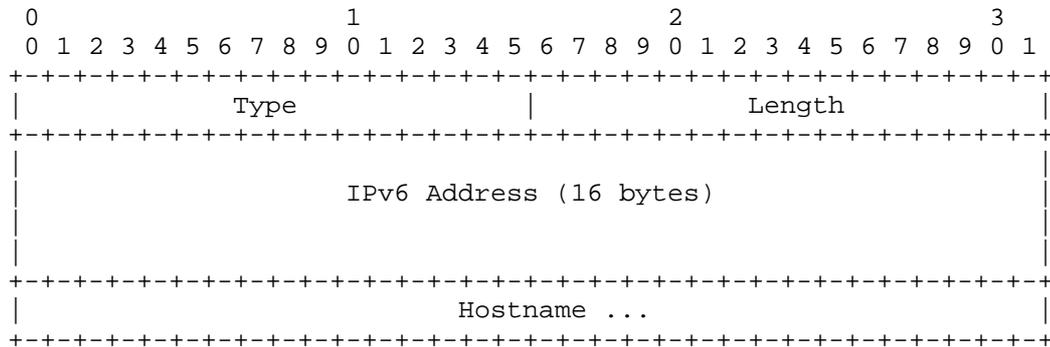


Figure 3: IPv6 address

- Type: 0x02 IPv6 address
- Length: The Length contains the total length of the subobject in bytes, including the Type and Length fields.
- IPv6 address: Router ID
- Hostname: See [RFC5642] section 3.1

2.2. Procedures

The node as Area Border Routers in OSPF routing area, can gain the hostname-to-Router-ID mapping information of the nodes in their attached areas, as described in [RFC 5642]. Thus, ABR can be used to generate Notify messages with Dynamic Hostname Object containing the hostname-to-Router-ID mapping information of the nodes in any area it attaches. The nodes other than ABR in the LSP tunnel, would never generate Notify messages with Dynamic Hostname Object.

An ABR in the LSP tunnel receives a Resv message from downstream, and could know from the Record Route Object which nodes of the LSP tunnel are in the same area that the interface of the ABR received Resv message belongs to. Then the ABR generates the Notify messages to ingress node carrying the Dynamic Hostname Object with the hostname-to-Router-ID mapping information of those nodes, which can be obtained from its local mapping table.

The ingress node will have the hostname-to-Router-ID mapping information of all nodes in the LSP tunnel, as it has obtained the mapping information of the nodes in other areas from the Notify messages sending by the ABRs.

If the mapping information of a node in another area changed, the ingress node MUST be notified immediately by the corresponding ABR using Notify message only containing the changed mapping information.

3. Security Considerations

TBD

4. IANA Considerations

TBD

5. Normative References

- [RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, December 2001.
- [RFC3473] Berger, L., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions", RFC 3473, January 2003.
- [RFC5642] Venkata, S., Harwani, S., Pignataro, C., and D. McPherson, "Dynamic Hostname Exchange Mechanism for OSPF", RFC 5642, August 2009.

Author's Address

Zhi Zheng
ZTE Corporation
No.68 ZiJingHua Road,Yuhuatai District
Nanjing 210012
P.R.China

Email: zheng.zhi@zte.com.cn

