

CCAMP Working Group
Internet-Draft
Intended status: Standards Track
Expires: April 18, 2011

D. Ceccarelli
D. Caviglia
Ericsson
S. Belotti
P. Grandi
Alcatel-Lucent
F. Zhang
D. Li
Huawei Technologies
J. Drake
Juniper
October 15, 2010

Technology Agnostic OSPF Traffic Engineering Extensions for Generalized
MPLS (GMPLS)
draft-bccdg-ccamp-gmpls-ospf-agnostic-00

Abstract

This document defines a new approach to Generalized Multiprotocol Label Switching (GMPLS) bandwidth advertisement aiming at providing the Network Elements (NEs) and Path Computation Elements (PCEs) with all the data required for crank-backs minimization and scalability optimization.

A new Open Shortest Path First - Traffic Engineering (OSPF-TE) routing protocol sub-tlv is defined for bandwidth advertisement per service type.

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 18, 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. OSPF Extensions	3
2.1. Bandwidth Accounting sub-TLV	4
3. LSA composition	8
4. Examples	8
5. Applicability	12
6. Compatibility Considerations	16
7. Security Considerations	16
8. IANA Considerations	16
9. Contributors	16
10. Acknowledgements	17
11. References	17
11.1. Normative References	17
11.2. Informative References	17
Authors' Addresses	17

1. Introduction

An Opaque OSPF (Open Shortest Path First) LSA (Link State Advertisements) carrying application-specific information can be generated and advertised to other nodes following the flooding procedures 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.

Traffic Engineering (TE) LSA using type 10 opaque LSA is defined in [RFC3630] for TE purposes. This type of LSA is composed of a standard LSA header and a payload including one top-level TLV (Type/Length/Value triplet) 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 GMPLS networks, routing serves as the foundation for automatically establishing Label Switched Paths (LSPs) through GMPLS RSVP-TE signaling.

This document describes technology agnostic OSPF LSA extensions to support connection oriented transport networks under the control of GMPLS (e.g. OTN, SDH, MPLS-TP). In particular a new OSPF-TE LSP is defined for bandwidth advertisement per service type tanking into account priorities and technology specific capabilities.

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. OSPF Extensions

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].

This document defines a new link sub-TLV, called Bandwidth Accounting (BA) sub-TLV (Sub-tlv value TBA by IANA, suggested 26).

One or more component links can be bundled as a TE link. In case of link bundling a single BA sub-TLV will be used to describe several component links.

2.1. Bandwidth Accounting sub-TLV

The BA sub-TLV has a so generic format that it can be used for the advertisement of any type of transport technology, from SDH/SONET to OTN, from L2SC to PSC etc. The main difference from the ISCD defined in [RFC4202] is the fact that unreserved bandwidth is advertised per service type per priority. The format of the BA sub-TLV is based on "8 bytes" data blocks repeated for each service type/priority/technology specific capability combination as is illustrated in Figure 1.

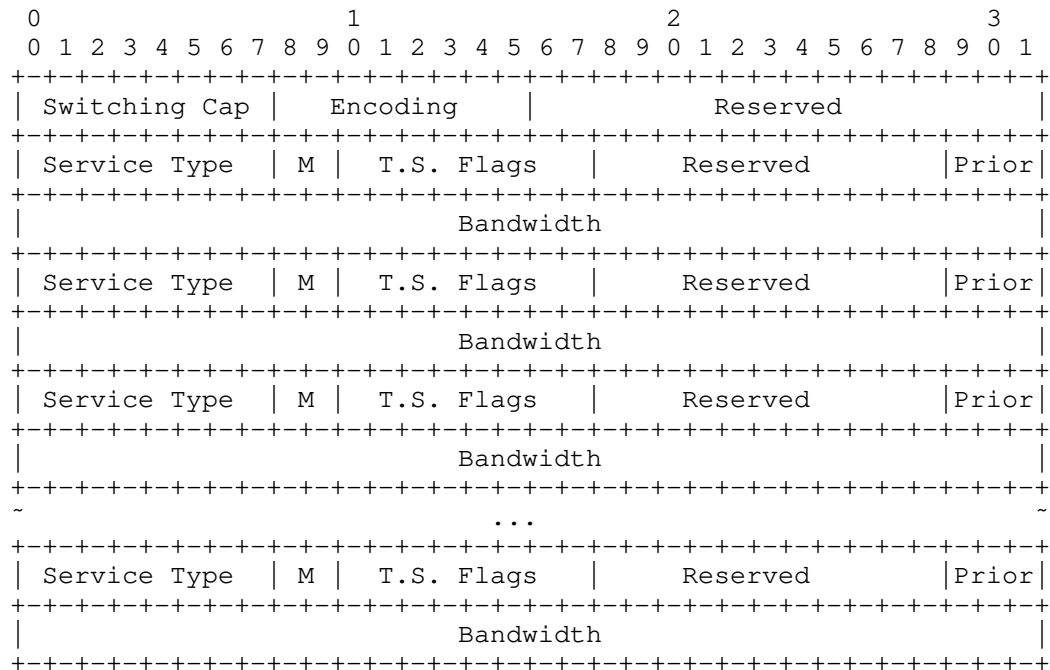


Figure 1: Bandwidth Accounting sub-TLV format

Where:

o Switching Capability (8 bits): the values for this field are defined in [RFC4203] section 1.4.

o Encoding (8 bits): the values for this field are defined in [RFC3471] section 3.1.1 and [RFC4328] section 3.1.1

- Data Blocks: Data blocks are composed by 64 bits and contain Service Type, M field, Technology Specific Flags, Priority and Bandwidth. For the definition of each field refer below. The number of data blocks depends on the number of service types, priority and technology specific features supported. Blocks declared in the LSA MUST contain a supported service type. Blocks declaring bandwidth at priority P_i , MUST NOT be declared in case priority P_i is not supported by the network element. Data blocks SHOULD be ordered from the highest to the lowest priority. If no priority is supported, just the 0 priority MUST be advertised. Please see the Example section for further details.

o Service Type (8 bits): Indicates the type of service supported by

TE link (e.g. STMx in an SDH network, ODUx in an OTN network). Each Service Type in a TE-link can be advertised only once for each supported priority.

o M field (2 bits): This field defines the meaning of the Bandwidth field. It states that the Bandwidth field is indicating Unreserved Bandwidth, Max LSP Bandwidth or Available Bandwidth. Possible values are:

0 - Unreserved bandwidth at priority P_i

1 - Max LSP bandwidth at priority P_i

For the service types where the advertisement of more than one of the previous values needs to be advertised (e.g. OTN ODUflex, MPLS-TP interface), a data block for each value MUST be advertised. For example, when advertising an ODUflex service type in an OTN network, both Unreserved bandwidth and MAX LSDP bandwidth are advertised as illustrated in Figure 2 (assuming supported priorities: P_1 and P_5).

[EDITOR NOTE]: Under Discussion - M=2 - Available bandwidth at priority P_i , Where Available bandwidth is defined as the unused link bandwidth available for additional non-traffic engineered IP/LDP forwarding and can be used as input to a node equal cost multipath load balancing function

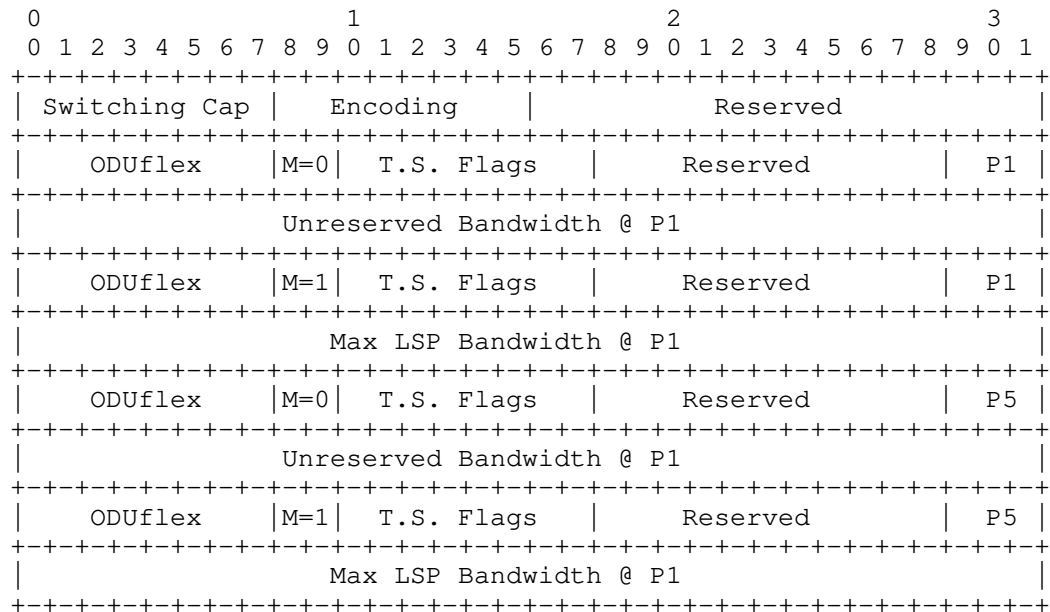


Figure 2: M field utilization example

o Technology Specific Flags (8 bits): These bits are used for the advertisement of technology specific interface capabilities and are defined in companion technology specific IDs. Depending on the technology it could be possible to have different data block advertised for different capability flags.

o Reserved (11 bits): Reserved bits MUST be set to zero.

o Priority (3 bits): Indicates the priority related to the advertised service type. Only supported priorities MUST be advertised.

o Bandwidth (32 bits): Independently on the type of bandwidth being advertised (see M field), this field is expressed in Bytes/sec in IEEE floating point format unless differently stated in technology specific documents.

The maximum bandwidth that an LSP can occupy in a TE link is determined by the component link with the maximum unreserved bandwidth in such TE link. For example, if two OTN OTU3 component links are bundled in a TE link, the unreserved bandwidth of the first component link is 20*1.25 Gbps, and the unreserved bandwidth of the second component link is 24*1.25Gbps, then the unreserved bandwidth of this TE link is 44*1.25Gbps, but the maximum bandwidth an LSP can

occupy in this TE link is 24*1,25Gbps, not 44*1,25Gbps.

All the reserved fields MUST be set to zero and SHOULD be ignored when received.

3. LSA composition

Each NE generates an LSA to describe the attributes of each TE link. If we suppose to have unnumbered link IDs, the LSA should carry a link TLV with the following nested minimal sub-TLVs:

```
< Link > ::= < Link Type > < Link ID > < Link
Local/Remote Identifiers > < Generalized-ISCD >
```

- o Link Type sub-TLV: Defined in [RFC 3630].
- o Link ID sub-TLV: Defined in [RFC 3630], for point-to-point link, indicates the remote router ID.
- o Link Local/Remote Identifiers sub-TLV: Defined in [RFC 4203], indicates the local link ID and the remote link ID.
- o Bandwidth Accounting sub-TLV: Defined in this document, carries the Bandwidth related information of the advertised TE-link.

4. Examples

The examples in the following pages are not normative and are not intended to infer or mandate any specific implementation. Moreover they aim at giving a general idea of the utilization of the BA sub-TLV in a technology agnostic scenario.

Figure 3 shows the case of a TE-link composed of two component links.

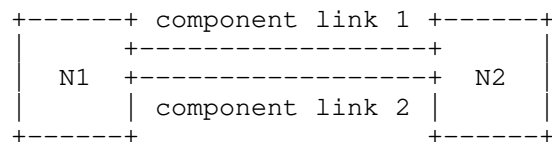


Figure 3: Example

The nominal bandwidth of the two component links is 10Gbps and 40Gbps respectively. The former has the capability of carrying service types A and B, while the latter, service types B and C, where A and C are fixed bandwidth service types (just unreserved bandwidth is advertised) and B variable bandwidth service types (unreserved bandwidth and Max LSP bandwidth advertised). The supported priorities are:0 and 3.

In this example the two component links are bundled as a TE link but it could also be possible to consider each of them as separate TE links.

If the two component links are bundled together, N1 and N2 should assign a link local ID to the TE link and then N1 can get the link remote ID automatically or manually.

Just after the creation of the TE Link comprising the two component links, the BA sub-TLV would be advertised 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								
Switching Cap										Encoding										Reserved																			
S.Type (A)										M=0	T.S. Flags										Reserved										P0								
Unreserved Bandwidth = 10 Gbps																																							
S.Type (A)										M=0	T.S. Flags										Reserved										P3								
Unreserved Bandwidth = 10 Gbps																																							
S.Type (B)										M=0	T.S. Flags										Reserved										P0								
Unreserved Bandwidth = 50 Gbps																																							
S.Type (B)										M=1	T.S. Flags										Reserved										P0								
Max LSP Bandwidth = 40 Gbps																																							
S.Type (B)										M=0	T.S. Flags										Reserved										P3								
Unreserved Bandwidth = 50 Gbps																																							
S.Type (B)										M=1	T.S. Flags										Reserved										P3								
Max LSP Bandwidth = 40 Gbps																																							
S.Type (C)										M=0	T.S. Flags										Reserved										P0								
Unreserved Bandwidth = 40 Gbps																																							
S.Type (C)										M=0	T.S. Flags										Reserved										P3								
Unreserved Bandwidth = 40 Gbps																																							

Figure 4: Example - BA sub-TLV(to)

Suppose that at time t1 an service type B LSP is created allocating 35 Gbps at priority 3. The BA sub-TLV will be modified 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								
Switching Cap										Encoding										Reserved																			
S.Type (A)										M=0	T.S. Flags										Reserved										P0								
Unreserved Bandwidth = 10 Gbps																																							
S.Type (A)										M=0	T.S. Flags										Reserved										P3								
Unreserved Bandwidth = 10 Gbps																																							
S.Type (B)										M=0	T.S. Flags										Reserved										P0								
Unreserved Bandwidth = 50 Gbps																																							
S.Type (B)										M=1	T.S. Flags										Reserved										P0								
Max LSP Bandwidth = 40 Gbps																																							
S.Type (B)										M=0	T.S. Flags										Reserved										P3								
Unreserved Bandwidth = 15 Gbps																																							
S.Type (B)										M=1	T.S. Flags										Reserved										P3								
Max LSP Bandwidth = 10 Gbps																																							
S.Type (C)										M=0	T.S. Flags										Reserved										P0								
Unreserved Bandwidth = 40 Gbps																																							
S.Type (C)										M=0	T.S. Flags										Reserved										P3								
Unreserved Bandwidth = 5 Gbps																																							

Figure 5: Example - BA sub-TLV(t1)

The last example shows how the preemption is managed. In particular, if at time t2 a new 15 GBps service type B LSP with priority 0 is created, the LSP with priority 3 is pre-empted and its resources (or part of them) are allocated to the LSP with higher priority. The BA sub-TLV is updated accordingly to Figure 6:

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								
Switching Cap										Encoding										Reserved																			
S.Type (A)										M=0	T.S. Flags										Reserved										P0								
Unreserved Bandwidth = 10 Gbps																																							
S.Type (A)										M=0	T.S. Flags										Reserved										P3								
Unreserved Bandwidth = 10 Gbps																																							
S.Type (B)										M=0	T.S. Flags										Reserved										P0								
Unreserved Bandwidth = 35 Gbps																																							
S.Type (B)										M=1	T.S. Flags										Reserved										P0								
Max LSP Bandwidth = 25 Gbps																																							
S.Type (B)										M=0	T.S. Flags										Reserved										P3								
Unreserved Bandwidth = 35 Gbps																																							
S.Type (B)										M=1	T.S. Flags										Reserved										P3								
Max LSP Bandwidth = 25 Gbps																																							
S.Type (C)										M=0	T.S. Flags										Reserved										P0								
Unreserved Bandwidth = 15 Gbps																																							
S.Type (C)										M=0	T.S. Flags										Reserved										P3								
Unreserved Bandwidth = 15 Gbps																																							

Figure 6: Example - BA sub-TLV (t2)

5. Applicability

The goal of this section is providing a comparison in term of bandwidth utilization between the BA sub-TLV based advertisement and the [RFC4203] based one. In order to provide a meaningful comparison between the two solutions (i.e. with same type and quantity of

information carried) it is necessary to assume [RFC4203] tools properly extended.

In other words it is assumed that both unreserved bandwidth and max LSP bandwidth are advertised per signal type. The unreserved bandwidth per signal type could be advertised by means of an unreserved bandwidth sub-tlv per signal type (1 header word + 8 body words) or using the technology specific part of the ISCD (8 words). In this example the utilization of the technology specific part of the ISCD is considered in order to take into account the most optimized option.

The following example is based on the advertisement of a simple link supporting six different types of fixed bandwidth service types (A,B,C,D,E,F) and a variable length service type (G).

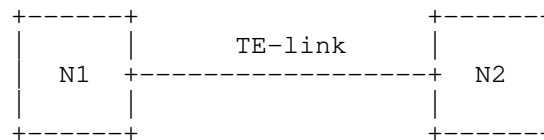


Figure 7: Example

Three different cases are analyzed:

- 8 priorities supported
- 5 priorities supported
- 1 priorities supported

In the first case, [RFC4203] approach would use 1 ISCD per signal type. The ISCD would need to be extended 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								
Type										Length																													
Switching Cap										Encoding										Reserved																			
Max LSP Bandwidth at priority 0																																							
Max LSP Bandwidth at priority 1																																							
Max LSP Bandwidth at priority 2																																							
Max LSP Bandwidth at priority 3																																							
Max LSP Bandwidth at priority 4																																							
Max LSP Bandwidth at priority 5																																							
Max LSP Bandwidth at priority 6																																							
Max LSP Bandwidth at priority 7																																							
Technology Specific Part																																							
Technology Specific Part																																							
Unreserved Bandwidth at priority 0																																							
Unreserved Bandwidth at priority 1																																							
Unreserved Bandwidth at priority 2																																							
Unreserved Bandwidth at priority 3																																							
Unreserved Bandwidth at priority 4																																							
Unreserved Bandwidth at priority 5																																							
Unreserved Bandwidth at priority 6																																							
Unreserved Bandwidth at priority 7																																							

Figure 8: Example

The amount of words used per ISCD is 20 for a total amount of 140

words. On the other side, using the BA sub-TLV these words would be used:

- 1 word for type/length declaration
- 1 word for sub-tlv header
- 2 words per (fixed) service type per priority = $2*6*8 = 96$
- 4 words per (variable) service type per priority = $4*1*8 = 32$

Total words used with 8 priorities: 140 (RFC4203) vs 130 (BA sub-TLV).

Performing the same computation in a scenario where 5 priorities are supported, the number of words used in the [RFC4203] approach would be the same (140), while in the BA sub-TLV would be:

- 1 word for type/length declaration
- 1 word for sub-tlv header
- 2 words per (fixed) service type per priority = $2*6*5 = 60$
- 4 words per (variable) service type per priority = $4*1*5 = 20$

Total words used with 5 priorities: 140 (RFC4203) vs 82 (BA sub-TLV).

The difference is significantly higher as the number of supported priorities decreases. Considering the case of single priority, the number of words used by the BA sub-TLV approach would be:

- 1 word for type/length declaration
- 1 word for sub-tlv header
- 2 words per (fixed) service type per priority = $2*6*1 = 12$
- 4 words per (variable) service type per priority = $4*1*1 = 4$

Total words used with 1 priority: 140 (RFC4203) vs 18 (BA sub-TLV).

It is worth considering that using the Unreserved bandwidth sub-TLV for unreserved bandwidth advertisement would increase the difference between the two solutions due to the fact that a higher number of headers is needed and at least a new word per sub-TLV would be required for the identification of the service type.

6. Compatibility Considerations

Backward compatibility issues are addressed in technology specific documents.

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

Francesco Fondelli, Ericsson

Email: francesco.fondelli@ericsson.com

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

10. Acknowledgements

TBD

11. References

11.1. Normative References

- [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

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

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

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

John E Drake
Juniper

Email: jdrake@juniper.net

Network Working Group
Internet Draft
Intended status: Informational

Y. Lee (Ed.)
Huawei
G. Bernstein (Ed.)
Grotto Networking
Moustafa Kattan
Cisco
October 22, 2010

Expires: April 2011

Information Model for Impaired Optical Path Validation
draft-bernstein-wson-impairment-info-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 April 22, 2010.

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

This document provides an information model for the optical impairment characteristics of optical network elements for use in GMPLS/PCE control plane protocols and mechanisms. This information model supports Impairment Aware Routing and Wavelength Assignment (IA-RWA) in optical networks in which path computation and optical path validation are essential components. This is not a general network management information model.

This model is based on ITU-T defined optical network element characteristics as given in ITU-T recommendation G.680 and related specifications. This model is intentionally compatible with a previous impairment free optical information model used in optical path computations and wavelength assignment.

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
2. Properties of an Impairment Information Model.....	3
3. Optical Impairment Information Model.....	4
3.1. Network Element Wide Parameters.....	5
3.2. Per Port Parameters.....	6
3.3. Port to Port Parameters.....	6
3.4. Frequency Dependent Parameters.....	6

4. Encoding Considerations.....	7
5. Usage of Parameters in Optical Path Validation.....	8
5.1. Centralized Computation.....	8
5.2. Distributed Computation.....	8
6. Security Considerations.....	9
7. IANA Considerations.....	9
8. Conclusions.....	9
9. Acknowledgments.....	9
APPENDIX A: Distributed Impairment Accumulation Model.....	10
A.1. Distributed Computation of OSNR.....	11
A.2. Distributed Computation of Residual Dispersion.....	12
A.3. Distributed Computation of PMD.....	12
A.4. Distributed Computation of PDL.....	13
APPENDIX B: Optical Parameters.....	14
B.1. Parameters for NEs without optical amplifiers.....	14
B.2. Additional parameters for NEs with optical amplifiers....	16
References.....	18
9.1. Normative References.....	18
9.2. Informative References.....	18
Author's Addresses.....	19
Intellectual Property Statement.....	19
Disclaimer of Validity.....	20

1. Introduction

Impairments in optical networks can be accounted for in a number of ways as discussed in reference [Imp-Frame]. This document provides an information model for path validation in optical networks utilizing approximate computations. The definitions, characteristics and usage of the optical parameters that form this model are based on ITU-T recommendation G.680 [G.680]. This impairment related model is intentionally compatible with the impairment free model of reference [RWA-Info]. Although this document focuses on the optical impairment parameters from a control plane point of view, Appendix B provides a list of optical parameter definitions from ITU-T G.680 and related documents.

This document only covers the links and network elements. The end system models (i.e., transmitter and receiver models based on the interfaces defined in G.698.1 and G.698.2) are subject to further study.

2. Properties of an Impairment Information Model

In term of information model there are properties that needs to be defined for each optical parameter available within the control plane. The properties will help to determine how the control plane

can deal with it depending architectural options defined in [Imp-Frame]. In some case properties value will help to indentify the level of approximation supported by the IV process.

- o Time Dependency. This will identify how the impairment may vary along the time. There could be cases where there's no time dependency, while in other cases there is need of an impairment re-evaluation after a certain time. In some cases a level of approximation will consider an impairment that has time dependency as constant.
- o Wavelength Dependency. This property will identify if an impairment value can be considered as constant over all the wavelength spectrum of interest or if it has different values. Also in this case a detailed impairment evaluation might lead to consider the exact value while an approximation IV might take a constant value for all wavelengths.
- o Linearity. As impairments are representation of physical effects there are some that have a linear behavior while other are non linear. Linear impairments are in general easy to consider while a non linear will require the knowledge of the full path to be evaluated. An approximation level could only consider linear effects or approximate non-linear impairments in linear ones.
- o Multi-Channel. There are cases where an impairments take different values depending on the aside wavelengths already in place. In this case a dependency among different LSP is introduced. An approximation level can neglect or not the effects on neighbor LSPs.
- o Value range. An impairment that has to be considered by a computational element will needs a representation in bits. So depending on the impairments different types can be considered form integer to real numbers as well as a fixed set of values. This information is important in term of protocol definition and level of approximation introduced by the number representation.

3. Optical Impairment Information Model

The definitions of optical impairment parameters of network elements and examples of their use can be found in [G.680] and related documents (also see Appendix B). From an information modeling and control plane perspective, one basic aspect of a given parameter is the scope of its applicability within a network element. In

particular we need to know which parameters will (a) apply to the network element as a whole, (b) can vary on a per port basis for a network element, and (c) can vary based on ingress to egress port pairs. A second orthogonal aspect of impairment parameters is whether a parameter exhibits a strong frequency variation over the optical frequencies supported by the subnetwork.

3.1. Network Element Wide Parameters

Based on the definitions in [G.680] and related documents the following parameters apply to the network element as a whole. At most one of these parameters is required per network element.

1. Channel frequency range (GHz, Max, Min)
2. Channel insertion loss deviation (dB, Max)
3. Ripple (dB, Max)
4. Channel chromatic dispersion (ps/nm, Max, Min)
5. Differential group delay (ps, Max)
6. Polarization dependent loss (dB, Max)
7. Reflectance (passive component) (dB, Max)
8. Reconfigure time/Switching time (ms, Max, Min)
9. Channel uniformity (dB, Max)
10. Channel addition/removal (steady-state) gain response (dB, Max, Min)
11. Transient duration (ms, Max)
12. Transient gain increase (dB, Max)
13. Transient gain reduction (dB, Max)
14. Multichannel gain-change difference (inter-channel gain-change difference) (dB, Max)
15. Multichannel gain tilt (inter-channel gain-change ratio) (dB, Max)

3.2. Per Port Parameters

The following optical parameters may exhibit per port dependence, hence may be specified at most once for each port of the network element.

1. Total input power range (dBm, Max, Min)
2. Channel input power range (dBm, Max, Min)
3. Channel output power range (dBm, Max, Min)
4. Input reflectance (dB, Max) (with amplifiers)
5. Output reflectance (dB, Max) (with amplifiers)
6. Maximum reflectance tolerable at input (dB, Min)
7. Maximum reflectance tolerable at output (dB, Min)
8. Maximum total output power (dBm, Max)

3.3. Port to Port Parameters

The following optical parameters may exhibit a port-to-port dependence and hence may be specified at most once for each ingress/egress port pair of the network element.

1. Insertion loss (dB, Max, Min)
2. Isolation, adjacent channel (dB, Min)
3. Isolation, non-adjacent channel (dB, Min)
4. Channel extinction (dB, Min)
5. Channel signal-spontaneous noise figure (dB, Max)
6. Channel gain (dB, Max, Min)

3.4. Frequency Dependent Parameters

Many of the previously mentioned parameters can exhibit significant frequency dependence over the range of wavelength supported by a subnetwork. In reference [G.680] parameters denoted as related to "channel" could exhibit significant frequency variation that would need to be encoded efficiently. These parameters may include:

1. Channel insertion loss deviation (dB, Max)
2. Channel chromatic dispersion (ps/nm, Max, Min)
3. Channel uniformity (dB, Max)
4. Insertion loss (dB, Max, Min)
5. Channel extinction (dB, Min)
6. Channel signal-spontaneous noise figure (dB, Max)
7. Channel gain (dB, Max, Min)

Finalization of this list is TBD and will need liaison with ITU-T.

4. Encoding Considerations

The units for the various parameters include GHz, dB, dBm, ms, ps, and ps/nm. These are typically expressed as floating point numbers. Due to the measurement limitations inherent in these parameters single precision floating point, e.g., 32 bit IEEE floating point, numbers should be sufficient. For this purpose the guideline is provided by [G.697] Appendix V that lists parameters and defines a suitable encoding.

For realistic optical network elements per port and port-to-port parameters typically only assume a few values. For example, the channel gain of a ROADM is usually specified in terms of input to drop, add to output, and input to output. This implies that many port and port-to-port parameters could be efficiently specified, stored and transported by making use of the Link Set Sub-TLV and Connectivity Matrix Sub-TLV of reference [Encode].

For parameters that vary with frequency we have the following options:

1. Explicit parameter list with associated frequencies: Here we would give the parameter and frequencies it applies to. We would need as many of these parameter/frequency pairs as necessary to cover all the frequencies and parameters. This could get large for a high channel count system with strong frequency dependencies in some parameters.

2. Provide "standardized" general interpolation formulas and parameters for use over an entire frequency range or sub-range.
3. Use parameter specific interpolation formulas based on ITU-T and other standards. For example in reference [G.650.1] Annex A equations and fitting coefficients are given for chromatic dispersion interpolation. Such formulas may be valid over an entire frequency range or a sub-range.

5. Usage of Parameters in Optical Path Validation

Given an optical path and the optical characteristics of each network element along the path we can then use these characteristics to validate the path. We envision that these parameters will be made available via some mechanism to the entity which validates optical paths. Refer to [Imp-Frame] for architectural options in which impairment validation for an optical path is defined.

Sections 9 and 10 of G.680 give techniques and formulas for use in calculating the impact of a cascade of network elements such as occurs along an optical signal path. These range from relatively simple bounds on the sum of uncompensated chromatic dispersion (residual dispersion) to more elaborate formulas for overall optical signal to noise ration (OSNR) computations based on multiple parameters including noise factor.

To further aid understanding and use of these optical parameters Appendix I of [G.680] provides example parameter values for different network element types and appendix II provides examples of computations involving the cascades of network elements along a path.

5.1. Centralized Computation

[TBD]

5.2. Distributed Computation

This section lists the parameters required for a distributed computation according to [G.680] model. Details about the formula are reported in the appendix. This section here lists only the parameters that need to be exchanged among nodes.

- o OSNR
 - o Power Input (required by OSNR)
- o Chromatic Dispersion

- o Differential Group Delay

6. Security Considerations

This document defines an information model for impairments in optical networks. If such a model 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. Conclusions

The state of standardization of optical device characteristics has matured from when initial IETF work concerning optical impairments was investigated in [RFC4054]. Relatively recent ITU-T recommendations provide a standardized based of optical characteristic definitions and parameters that control plane technologies such as GMPLS and PCE can make use of in performing optical path validation. The enclosed information model shows how readily such ITU-T optical work can be utilized within the control plane.

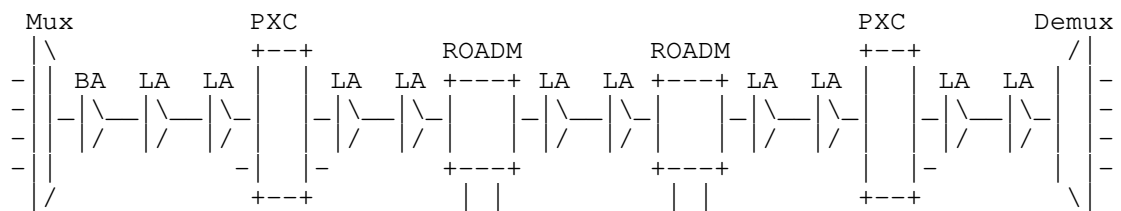
9. Acknowledgments

This document was prepared using 2-Word-v2.0.template.dot.

APPENDIX A: Distributed Impairment Accumulation Model

In reference [Imp-Frame] an alternative impairment aware RWA control plane based on distributed impairment validation was discussed. In such a scheme the preceding impairment information model would not be distributed via a link state IGP, instead a set of impairment parameters would be computed along the proposed path and a final decision on whether the path is viable would be made based on these accumulated impairment parameters. It should be noted that these accumulated impairment parameters are estimated at each node along the path and not measured.

When signaling a path we think of the "nodes" as being the switching nodes along the path. In the case of optical impairments the properties of the links (WDM line systems) are just as important as the properties of the nodes. In the following we will assume that the switching nodes (GMPLS nodes) will act on behalf of all the line systems corresponding to their egress ports. In particular this implies that some how these nodes will obtain the line system impairment information.



<---- NE1 ----><---- NE2 --><---- NE3 ----><---- NE4 ----><---- NE5 -->

Figure 1 A path through an optical network with line systems, PXCs, ROADMs, and multiplexers.

In Figure 1 we show an example system from appendix II of [G.680]. This diagram shows the DWDM line systems including amplifiers, BA = booster amplifier, LA = line amplifier. For distributed impairment validation we would group the line systems with their preceding nodes as shown for computational purposes.

Section 9 of ITU-T G.680 [G.680] shows how various impairment parameters accumulate and this suggests that the following parameters or subset thereof could be used in distributed impairment estimation:

- o Optical Signal to Noise Ratio (OSNR)
- o Residual Dispersion (chromatic)

- o Polarization Mode Dispersion (PMD)
- o Polarization Dependent Loss (PDL)
- o Ripple
- o Channel Uniformity

For each of the above the units and accumulation procedure needs to be defined. In the following we suggest units and procedures for the above for which computation of cascaded elements are suitably defined in [G.680]. Note: ONE = Optical Network Element.

A.1. Distributed Computation of OSNR

Section 9.1 of ITU-T G.680 gives several equivalent formulas for the estimation of OSNR. For distributed impairment validation the following formula from [G.680] is convenient:

$$\text{OSNR}_{\text{out}} = -10 \cdot \log(\text{Term1} + \text{Term2})$$

Where

$$\text{Term1} = 10^{-(\text{OSNR}_{\text{in}}/10)}, \text{ and}$$

$$\text{Term2} = 10^{-((\text{P}_{\text{in}} - \text{NF} - 10 \cdot \log(h \cdot v \cdot v_r))/10)}$$

and we have the following additional definitions:

OSNR_{out} is the output optical signal to noise ratio in dB of the ONE

OSNR_{in} is the input optical signal to noise ratio in dB of the ONE

P_{in} is the channel power (dBm) at the input port of the ONE

NF is the noise figure (dB) of the relevant path through the ONE

h is Planck's constant (in mJ*s to be consistent with P_{in} in dBm)

v is the optical frequency in Hz

v_r is the reference bandwidth in Hz (usually the frequency equivalent of 0.1nm)

From the previous formula, a distributed computation of OSNR requires knowing the OSNR_{in} and the P_{in} based on computations from the previous node along the path. The noise figure, F, is something that

the current node performing the computation would know along with the frequency, ν , and the reference bandwidth ν_r (TBD: confirm with ITU-T).

The control plane will need to distribute the following information from node to node along the path:

- o OSNR_in (this is the accumulated OSNR along the path) (dB)
- o P_in (this is the estimated power into the next node) (dBm)

The input power would be calculated by the previous node by taking into account gain and attenuation on the link between the nodes.

A.2. Distributed Computation of Residual Dispersion

The residual dispersion for a path is required to be bounded, in particular from [G.680] equation 9-4:

$$\text{Min RD} < \text{Residual Dispersion} < \text{Max RD}$$

Where Min RD and Max RD are the minimum and maximum tolerable residual dispersion for a particular transmitter/receiver combination.

The residual dispersion for a cascade of network elements can be computed by [G.680] equation 9-5:

$$\text{Residual dispersion} = \text{sum}(\text{fiber dispersion}) + \text{sum}(\text{DCM dispersion}) + \text{sum}(\text{ONE dispersion})$$

Where DCM dispersion is from Dispersion Compensation Modules (DCM), and ONE dispersion is due to optical network elements.

Although the residual dispersion formula is a relatively simple linear formula [G.680] indicates two possible methods for its evaluation (a) Worst-case upper and lower bounds, or (b) Statistical approach. In case (a) two parameters would need to be accumulated along the path a worst case upper and lower bound. In case (b) some type of statistical information would be needed in [G.680] mean and standard deviation are used under a Gaussian assumption.

A.3. Distributed Computation of PMD

The accumulated impact of line system and ONE polarization mode dispersion can be estimated via the formula [G.680] equation (9-6):

$$\text{DGDmax_link} = \{\text{DGDmaxf}^2 + S^2 \cdot \sum_i (\text{PMDc}_i^2)\}^{(1/2)}$$

where

DGDmax_link is the max link DGD (ps)

DGDmxf is the max concatenated optical fiber cable DGD (ps)

S is the Maxwell adjustment factor (Table 9-2 of [G.680])

PMDc_i is the PMD value for the ith component (ps)

Under a distributed computation approach the above could be computed by keeping track of DGDmxf and the running sum of PMDc_i^2. The Maxwell adjustment factor and final square root can be applied at the final node in the path. [Question for Q6: does DGDMaxf^2 need to be accumulated over the different link segments?]

A.4. Distributed Computation of PDL

See section 9.3.2 of [G.680]

APPENDIX B: Optical Parameters

The following provides an annotated list of optical characteristics from ITU-T recommendation G.680 [G.680] for use in optical path impairment computations. For each parameter we specify the units to be used, whether minimum or maximum values are used, and whether the parameters applies to the optical network element as a whole, on a per port basis or on a port-to-port pair basis.

Not all these parameters will apply to all devices. The main differentiation in G.680 comes from those network elements that include or do not include optical amplifiers.

B.1. Parameters for NEs without optical amplifiers

Channel frequency range (GHz, Max, Min): [G.671] The frequency range within which a DWDM device is required to operate with a specified performance. For a particular nominal channel central frequency, f_{nomi} , this frequency range is from $f_{imin} = (f_{nomi} - df_{max})$ to $f_{imax} = (f_{nomi} + df_{max})$, where df_{max} is the maximum channel central frequency deviation. Nominal channel central frequency and maximum channel central frequency deviation are defined in ITU-T Rec. G.692.

Insertion loss (dB, Port-Port, Max, Min): [G.671] It is the reduction in optical power between an input and output port of a WDM device in decibels (dB).

Channel insertion loss deviation (dB, Max): [G.671] This is the maximum variation of insertion loss at any frequency within the channel frequency range (DWDM devices) or channel wavelength range (CWDM and WWDM devices).

Ripple (dB, Max): [G.671] For WDM devices and tuneable filters, the peak-to-peak difference in insertion loss within a channel frequency (or wavelength) range.

Channel chromatic dispersion (ps/nm, Max, Min): [G.650.1] Change of the group delay of a light pulse for a unit fibre length caused by a unit wavelength change.

Differential group delay (ps, Max): [G.671] Polarization Mode Dispersion (PMD) is usually described in terms of a Differential Group Delay (DGD), which is the time difference between the principal States of Polarization (SOPs) of an optical signal at a particular wavelength and time.

Polarization dependent loss (dB, Max): [G.671] Maximum variation of insertion loss due to a variation of the state of polarization (SOP) over all SOPs.

Reflectance (dB, Max): [G.671] The ratio of reflected power P_r to incident power, P_i at a given port of a passive component, for given conditions of spectral composition, polarization and geometrical distribution.

Isolation, adjacent channel (dB, Min, Port-Port): [G.671] The adjacent channel isolation (of a WDM device) is defined to be equal to the unidirectional (far-end) isolation of that device with the restriction that x , the isolation wavelength number, is restricted to the channels immediately adjacent to the (channel) wavelength number associated with port o .

Isolation, non-adjacent channel (dB, Min, Port-Port): [G.671] The non-adjacent channel isolation (of a WDM device) is defined to be equal to the unidirectional (far-end) isolation of that device with the restriction that x , the isolation wavelength number, is restricted to each of the channels not immediately adjacent to the (channel) wavelength number associated with port o .

Note: [G.671] In a WDM device able to separate k wavelengths (w_1, w_2, \dots, w_k) radiation coming from one input port into k output ports, each one nominally passing radiation at one specific wavelength only. The unidirectional (far-end) isolation is a measure of the part of the optical power at each wavelength exiting from the port at wavelengths different from the nominal wavelength relative to the power at the nominal wavelength.

Channel extinction (dB, Min, Port-Port): [G.671] Within the operating wavelength range, the difference (in dB) between the maximum insertion loss for the non-extinguished (non-blocked) channels and the minimum insertion loss for the extinguished (blocked) channels.

Reconfigure time (ms, Max, Min): [G.680] The reconfigure time (of an ROADM) is the elapsed time measured from the earliest point that the actuation energy is applied to reconfigure the ONE to the time when the channel insertion loss for all wanted channels has settled to within 0.5 dB of its final steady state value and all other parameters of the device (e.g., isolation and channel extinction) are within the allowed limits.

Switching time (for PXC) (ms, Max, Min): [G.671] The elapsed time it takes the switch to turn path io on or off from a particular initial

state, measured from the time the actuation energy is applied or removed.

Channel uniformity (dB, Max): [G.671] The difference (in dB) between the powers of the channel with the most power (in dBm) and the channel with the least power (in dBm). This applies to a multichannel signal across the operating wavelength range.

B.2. Additional parameters for NEs with optical amplifiers

Total input power range (dBm, Max, Min, Port): [G.661] The range of optical power levels at the input for which the corresponding output signal optical power lies in the specified output power range, where the OA performance is ensured.

Channel input power range (dBm, Max, Min, Port): see above.

Channel output power range (dBm, Max, Min, Port): [G.661] The range of optical power levels at the output of the OA for which the corresponding input signal power lies in the specified input power range, where the OA performance is ensured.

Channel signal-spontaneous noise figure (dB, Max, Port-Port) [G.661] The signal-spontaneous beat noise contribution to the noise figure, expressed in dB.

Input reflectance (dB, Max, Port): [G.661] The maximum fraction of incident optical power, at the operating wavelength and over all states of input light polarization, reflected by the OA from the input port, under nominal specified operating conditions, expressed in dB.

Output reflectance (dB, Max, Port): [G.661] The fraction of incident optical power at the operating wavelength reflected by the OA from the output port, under nominal operating conditions, expressed in dB.

Maximum reflectance tolerable at input (dB, Min, Port): [G.661] The maximum fraction of power, expressed in dB, exiting the optical input port of the OA which, when reflected back into the OA, allows the device to still meet its specifications.

Maximum reflectance tolerable at output (dB, Min, Port): [G.661] The maximum fraction of power, expressed in dB, exiting the optical output port of the OA which, when reflected back into the OA, allows the device to still meet its specifications.

Maximum total output power (dBm, Max, Port): [G.661] The highest signal optical power at the output that can be obtained from the OA under nominal operating conditions.

Channel addition/removal (steady-state) gain response (dB, Max, Min): [G.661] For a specified multichannel configuration, the steady-state change in channel gain of any one of the channels due to the addition/removal of one or more other channels, expressed in dB.

Transient duration (ms, Max): [G.661] The time period from the addition/removal of a channel to the time when the output power level of that or another channel reaches and remains within $\pm N$ dB from its steady-state value.

Transient gain increase (dB, Max): [G.661] For a specified multichannel configuration, the maximum change in channel gain of any one of the channels due to the addition/removal of one or more other channels during the transient period after channel addition/removal, expressed in dB.

Transient gain reduction (dB, Max): see above.

Channel gain (dB, Max, Min, Port-Port): [G.661] Gain for each channel (at wavelength w_j) in a specified multichannel configuration, expressed in dB.

Multichannel gain-change difference (inter-channel gain-change difference) (dB, Max): [G.661] For a specified channel allocation, the difference of change in gain in one channel with respect to the change in gain of another channel for two specified sets of channel input powers, expressed in dB.

Multichannel gain tilt (inter-channel gain-change ratio) (dB, Max): [G.661] The ratio of the changes in gain in each channel to the change in gain at a reference channel as the input conditions are varied from one set of input channel powers to a second set of input channel powers, expressed in dB per dB.

References

9.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [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.661] ITU-T Recommendation G.661, Definition and test methods for the relevant generic parameters of optical amplifier devices and subsystems, March 2006.
- [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.697] ITU-T Recommendation G.697, Optical Monitoring for dense wavelength division multiplexing system, November 2009.
- [Imp-Frame] G. Bernstein, Y. Lee, D. Li, G. Martinelli, "A Framework for the Control and Measurement of Wavelength Switched Optical Networks (WSON) with Impairments", Work in Progress, draft-bernstein-ccamp-wson-impairments-05.txt
- [RWA-Info] Y. Lee, G. Bernstein, D. Li, W. Imajuku, "Routing and Wavelength Assignment Information Model for Wavelength Switched Optical Networks", Work in Progress, draft-ietf-ccamp-rwa-info-02.txt.

9.2. Informative References

- [RFC4054] Strand, J., Ed., and A. Chiu, Ed., "Impairments and Other Constraints on Optical Layer Routing", RFC 4054, May 2005.
- [Encode] G. Bernstein, Y. Lee, D. Li, W. Imajuku, "Routing and Wavelength Assignment Information Encoding for Wavelength Switched Optical Networks" Work in progress, draft-bernstein-ccamp-wson-encode-01.txt.

Author's 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 Bernstein (ed.)
Grotto Networking
Fremont CA, USA

Phone: (510) 573-2237
Email: gregb@grotto-networking.com

Moustafa Kattan
Cisco Systems,
Dubai Internet City # 10,
Dubai, UAE

Phone (408) 527-5101
Email: mkattan@cisco.com

Giovanni Martinelli
Cisco Systems, Inc.
20052 Monza, Italy

Email: giomarti@cisco.com

Andrea Zanardi
Create-Net
Via della Cascata 56/D Povo,
38123 Trento, Italy

Email: andrea.zanardi@create-net.org

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.

CCAMP Working Group
Internet-Draft
Intended status: Standards Track
Expires: April 25, 2011

D. Ceccarelli
D. Caviglia
Ericsson
F. Zhang
D. Li
Huawei Technologies
Y. Xu
CATR
S. Belotti
P. Grandi
Alcatel-Lucent
J. Drake
Juniper
October 22, 2010

Traffic Engineering Extensions to OSPF for Generalized MPLS (GMPLS)
Control of Evolving G.709 OTN Networks
draft-ceccarelli-ccamp-gmpls-ospf-g709-04

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 April 25, 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. Terminology	3
2. OSPF Extensions	3
2.1. Bandwidth Accounting sub-TLV	4
2.2. Example using BA sub-TLV	7
2.3. Example of T and S bits usage	13
3. Scalability Improvement	14
4. Compatibility Considerations	14
5. Security Considerations	14
6. IANA Considerations	15
7. Contributors	15
8. Acknowledgements	16
9. References	16
9.1. Normative References	16
9.2. Informative References	17
Authors' Addresses	17

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 of the Bandwidth Accounting sub-TLV defined in [OSPF-AGN].

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 [WSN-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].

This document defines OTN specific encoding for the Bandwidth Accounting sub-TLV defined in [OSPF-AGN].

One or more component links can be bundled as a TE link. In case of link bundling a Generalized-ISCD will be used to describe several component links.

As discussed in [OTN-FWK] and [OTN-INFO], usage of multi-stage

multiplexing implies the advertisement of cascaded adaptation capabilities together with matrix access constraints. Modifications to ISCD/IACD [RFC4202][RFC5339] and [MLN-EXT], if needed, are for further study.

2.1. Bandwidth Accounting sub-TLV

The Bandwidth Accounting (BA) sub-TLV format as defined in [OSPF-AGN] is shown in Figure 1.

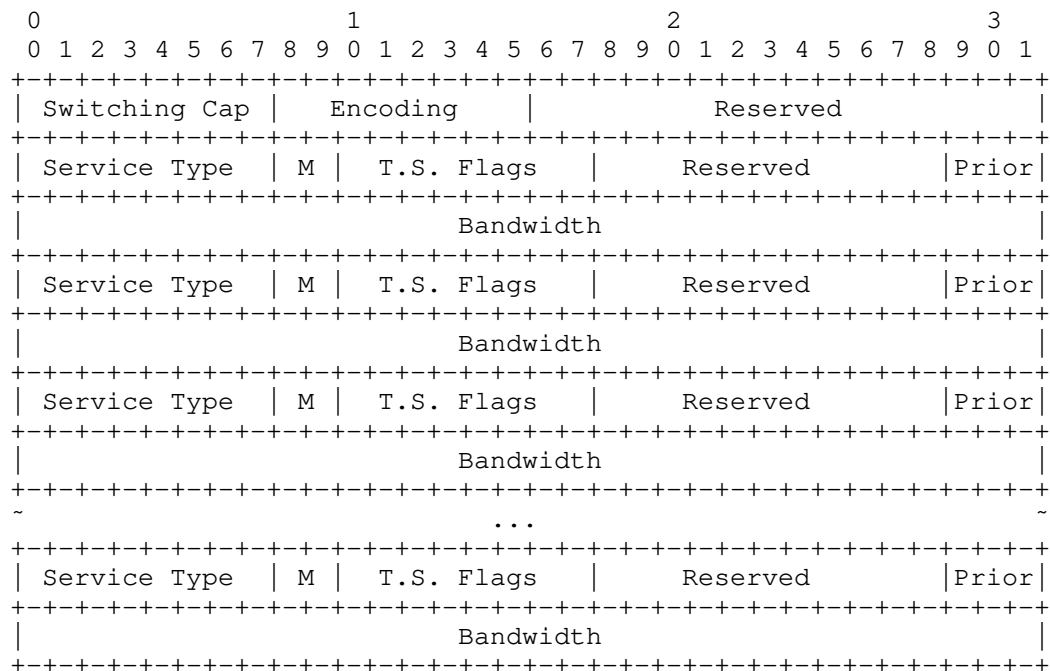


Figure 1: BA sub-TLV format

Common fields values are defined in [OSPF-AGN] while OTN specific values are:

- o Service Type (8 bits): Indicates the type of ODUk/ODUflex supported by the TE link. Possible values are:

- o 0 - ODU0

- o 1 - ODU1
 - o 2 - ODU2
 - o 3 - ODU3
 - o 4 - ODU4
 - o 10 - ODU2e
 - o 20 - ODUFlex non resizable
 - o 21 - ODUFlex resizable
 - o Technology Specific Flags (8 bits): Indicate OTN specific capabilities and are defined as follows:
 - o G field (merge of bit 11 and 12): Indicates the granularity of the Tributary Slot used on the advertised TE link. Possible values are:
 - 0 - 1.25 Gbps
 - 1 - 2.5 Gbps
 - 2-3 for future uses
 - o T Flag (bit 13): Indicates whether the advertised bandwidth can be terminated on the related interface or not:
 - 0 - Advertised service type cannot be terminated
 - 1 - Advertised service type can be terminated
 - o S Flag (bit 14): Indicates whether the advertised bandwidth can be switched on the related interface or not:
 - 0 - Advertised service type cannot be switched
 - 1 - Advertised service type can be switched
- For further details on the definition of the terminating and switching capabilities please refer to [OTN-INFO].
- Non defined flags MUST be set to zero.
- o Bandwidth (32 bit): Depending on the M field value, indicates:

- Unreserved Bandwidth for fixed dimension containers: expressed in number of containers (M=0)
- Unreserved Bandwidth for variable dimension containers: expressed in Bytes/Second in IEEE floating point format (M=0)
- MAX LSP Bandwidth for variable dimension containers: expressed in Bytes/Second in IEEE floating point format (M=1)

Please see the Example section for further details.

Each ODUk in a TE-link can be advertised only once for each supported priority.

In case an ODUFlex service types is advertised, two data block per priority are advertised, the first one with M bit clear (indicating the unreserved bandwidth available for such combination of service type and priority) and the second one with the M bit set (indicating the max LSP bandwidth). Please see Figure 2 for further clarification.

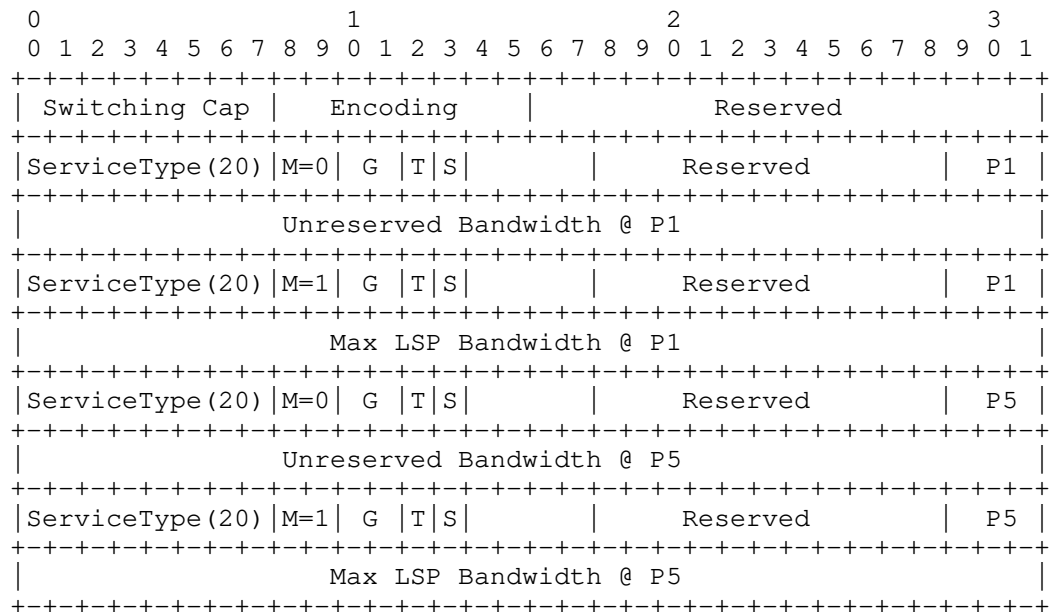


Figure 2: BA sub-TLV for ODUFlex

The Maximum Bandwidth that an LSP can occupy in a TE link is determined by the component link with the maximum unreserved bandwidth in such TE link. For example, if two OTU3 component links are bundled in a TE link, the unreserved bandwidth of the first component link is 20*1.25G TSs, and the unreserved bandwidth of the second component link is 24*1.25G TSs, then the unreserved bandwidth of this TE link is 44*1.25G TSs, but the maximum TSs that a LSP can occupy in this TE link is 24 TSs, not 44 TSs.

2.2. Example using BA sub-TLV

The examples in the following pages are not normative and are not intended to infer or mandate any specific implementation. Figure 3 shows the case of a TE-link composed of two component links.

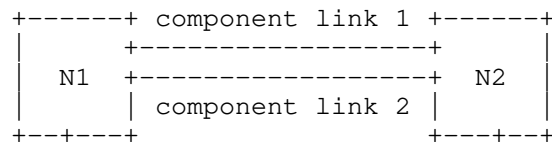


Figure 3: Example

The link type of the two component links are OTU2 and OTU3 respectively. The former has the capability of carrying ODU0, ODU1 and ODUflex client signals, while the latter, ODU1, ODU3 and ODUflex. All the service types can be switched and only the ODU2 on LC1 and ODU3 on LC2 can be both switched and terminated. The TS type is 1.25Gbps and the supported priorities are:0, 3 and 7.

In this example the two component links are bundled as a TE link but it could also be possible to consider each of them as a separate TE link.

Just after the creation of the TE Link comprising the two component links, the Generalized-ISCD sub-TLV would be advertised 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				
SC = TDM									Enc = G709									Reserved																	
S.Type (ODU0)									M=0 G=0 0 1									Reserved									P0								

Unreserved Bandwidth 8 ODU0			
S.Type (ODU0)	M=0 G=0 0 1	Reserved	P3
Unreserved Bandwidth 8 ODU0			
S.Type (ODU0)	M=0 G=0 0 1	Reserved	P7
Unreserved Bandwidth 8 ODU0			
S.Type (ODU1)	M=0 G=0 0 1	Reserved	P0
Unreserved Bandwidth 4+16=20 ODU1			
S.Type (ODU1)	M=0 G=0 0 1	Reserved	P3
Unreserved Bandwidth 4+16=20 ODU1			
S.Type (ODU1)	M=0 G=0 0 1	Reserved	P7
Unreserved Bandwidth 4+16=20 ODU1			
S.Type (ODU2)	M=0 G=0 1 0	Reserved	P0
Unreserved Bandwidth 4 ODU2			
S.Type (ODU2)	M=0 G=0 1 0	Reserved	P3
Unreserved Bandwidth 4 ODU2			
S.Type (ODU2)	M=0 G=0 1 0	Reserved	P7
Unreserved Bandwidth 4 ODU2			
S.Type (ODU3)	M=0 G=0 1 1	Reserved	P0
Unreserved Bandwidth 1 ODU3			
S.Type (ODU3)	M=0 G=0 1 1	Reserved	P3
Unreserved Bandwidth 1 ODU3			
S.Type (ODU3)	M=0 G=0 1 1	Reserved	P7
Unreserved Bandwidth 1 ODU3			
S.Type (ODUFlex)	M=0 G=0 0 1	Reserved	P0

Unreserved Bandwidth 50 Gbps																													
S.Type (ODUFlex) M=1 G=0 0 1										Reserved										P0									
Max LSP Bandwidth 40 Gbps																													
S.Type (ODUFlex) M=0 G=0 0 1										Reserved										P3									
Unreserved Bandwidth 50 Gbps																													
S.Type (ODUFlex) M=1 G=0 0 1										Reserved										P3									
Max LSP Bandwidth 40 Gbps																													
S.Type (ODUFlex) M=0 G=0 0 1										Reserved										P7									
Unreserved Bandwidth 50 Gbps																													
S.Type (ODUFlex) M=1 G=0 0 1										Reserved										P7									
Max LSP Bandwidth 40 Gbps																													

Figure 4: Example - BA sub-TLV(to)

Suppose that at time t1 an ODUflex LSP is created allocating 35 Gbps at priority 3. The BA sub-TLV will be modified 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								
SC = TDM										Enc = G709										Reserved																			
S.Type (ODU0)										M=0	G=0	0	1	Reserved										P0															
Unreserved Bandwidth 8 ODU0																																							
S.Type (ODU0)										M=0	G=0	0	1	Reserved										P3															
Unreserved Bandwidth 8 ODU0																																							
S.Type (ODU0)										M=0	G=0	0	1	Reserved										P7															
Unreserved Bandwidth 8 ODU0																																							

S.Type (ODU1)	M=0	G=0	0	1		Reserved	P0
Unreserved Bandwidth 4+16=20 ODU1							
S.Type (ODU1)	M=0	G=0	0	1		Reserved	P3
Unreserved Bandwidth 4+1=5 ODU1							
S.Type (ODU1)	M=0	G=0	0	1		Reserved	P7
Unreserved Bandwidth 4+1=5 ODU1							
S.Type (ODU2)	M=0	G=0	1	0		Reserved	P0
Unreserved Bandwidth 4 ODU2							
S.Type (ODU2)	M=0	G=0	1	0		Reserved	P3
Unreserved Bandwidth 4 ODU2							
S.Type (ODU2)	M=0	G=0	1	0		Reserved	P7
Unreserved Bandwidth 4 ODU2							
S.Type (ODU3)	M=0	G=0	1	1		Reserved	P0
Unreserved Bandwidth 1 ODU3							
S.Type (ODU3)	M=0	G=0	1	1		Reserved	P3
Unreserved Bandwidth 0 ODU3							
S.Type (ODU3)	M=0	G=0	1	1		Reserved	P7
Unreserved Bandwidth 0 ODU3							
S.Type (ODUFlex)	M=0	G=0	0	1		Reserved	P0
Unreserved Bandwidth 50 Gbps							
S.Type (ODUFlex)	M=1	G=0	0	1		Reserved	P0
Max LSP Bandwidth 40 Gbps							
S.Type (ODUFlex)	M=0	G=0	0	1		Reserved	P3
Unreserved Bandwidth 15 Gbps							

S.Type (ODUFlex)	M=1	G=0	0	1		Reserved	P3
Max LSP Bandwidth 10 Gbps							
S.Type (ODUFlex)	M=0	G=0	0	1		Reserved	P7
Unreserved Bandwidth 15 Gbps							
S.Type (ODUFlex)	M=1	G=0	0	1		Reserved	P7
Max LSP Bandwidth 10 Gbps							

Figure 5: Example - BA sub-TLV(t1)

The last example shows how the preemption is managed. In particular, if at time t2 a new 15 Gbps ODUflex LSP with priority 0 is created, the LSP with priority 3 is pre-empted and its resources (or part of them) are allocated to the LSP with higher priority. The BA sub-TLV is updated accordingly to Figure 6:

0										1										2										3									
SC = TDM										Enc = G709										Reserved																			
S.Type (ODU0)										M=0 G=0 0 1										Reserved										P0									
Unreserved Bandwidth 8 ODU0																																							
S.Type (ODU0)										M=0 G=0 0 1										Reserved										P3									
Unreserved Bandwidth 8 ODU0																																							
S.Type (ODU0)										M=0 G=0 0 1										Reserved										P7									
Unreserved Bandwidth 8 ODU0																																							
S.Type (ODU1)										M=0 G=0 0 1										Reserved										P0									
Unreserved Bandwidth 4+10=14 ODU1																																							
S.Type (ODU1)										M=0 G=0 0 1										Reserved										P3									
Unreserved Bandwidth 4+10=14 ODU1																																							

S.Type (ODU1)	M=0	G=0	0	1		Reserved	P7
Unreserved Bandwidth 4+10=14 ODU1							
S.Type (ODU2)	M=0	G=0	1	0		Reserved	P0
Unreserved Bandwidth 1 ODU2							
S.Type (ODU2)	M=0	G=0	1	0		Reserved	P3
Unreserved Bandwidth 1 ODU2							
S.Type (ODU2)	M=0	G=0	1	0		Reserved	P7
Unreserved Bandwidth 1 ODU2							
S.Type (ODU3)	M=0	G=0	1	1		Reserved	P0
Unreserved Bandwidth 0 ODU3							
S.Type (ODU3)	M=0	G=0	1	1		Reserved	P3
Unreserved Bandwidth 0 ODU3							
S.Type (ODU3)	M=0	G=0	1	1		Reserved	P7
Unreserved Bandwidth 0 ODU3							
S.Type (ODUFlex)	M=0	G=0	0	1		Reserved	P0
Unreserved Bandwidth 25 Gbps							
S.Type (ODUFlex)	M=1	G=0	0	1		Reserved	P0
Max LSP Bandwidth 25 Gbps							
S.Type (ODUFlex)	M=0	G=0	0	1		Reserved	P3
Unreserved Bandwidth 25 Gbps							
S.Type (ODUFlex)	M=1	G=0	0	1		Reserved	P3
Max LSP Bandwidth 25 Gbps							
S.Type (ODUFlex)	M=0	G=0	0	1		Reserved	P7
Unreserved Bandwidth 25 Gbps							

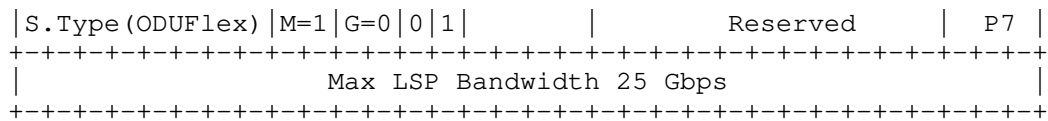


Figure 6: Example - Generalized-ISCD (t2)

2.3. Example of T and S bits usage

This section provides an example of T and S bits utilization in a sample network depicted in the following figure:

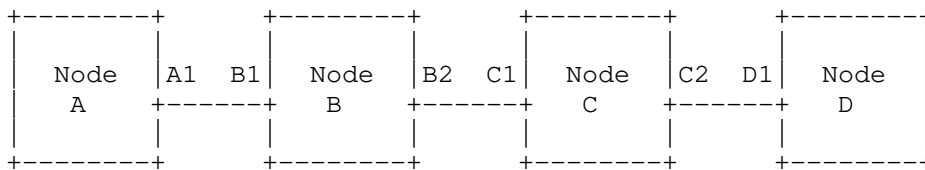


Figure 7: Example T-S Scenario

The amount of available resources of the sample network is summarized in this table:

ODU (T,S) /Link	A1-B1	B2-C1	C2-D1
ODU-4 (1,1)	1	1	0
ODU-3 (1,1)	2	1	2
ODU-3 (0,1)	2	2	0
ODU-2 (1,1)	2	0	2
ODU-2 (0,1)	18	14	8
ODU-1 (0,1)	80	56	40

Figure 8: Example T-S Resources

The bits in brackets indicate the T and S flags respectively, while each entry indicates the number of related containers available on corresponding link. The following table summarizes the number of different LSPs that can be set-up between node A and D depending on the type of constraint on T and S flags.

ODU (T,S) /LSP	FLR	EFLR	ANY	SWCO
ODU-4	0	0	0	0
ODU-3	1	2	2	0
ODU-2	0	2	10	8
ODU-1	0	0	40	40

FLR = Full Lambda rate on all links

(T=1, S=1 on intermediate nodes, T=1, S=any, on end nodes)

EFLR = Full Lambda rate on end nodes only (T=1, S=any on end nodes only)

ANY = Use any available resource. (No constraint on S and T)

SWCO = use only switched capacity. (T=0, S=1)

Figure 9: Example - Types of LSP

It is possible to see that the number of LSPs that can be set-up varies depending on the type of resources that T and S bit impose to use.

3. Scalability Improvement

TBD

4. Compatibility Considerations

No OTN specific extension to GMPLS routing has been defined up to now. As a consequence there should be no backward compatibility issue.

5. 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.

6. IANA Considerations

TBD

7. 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

8. Acknowledgements

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

9. References

9.1. 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.
- [OSPF-AGN] D.Ceccarelli, D.Caviglia, S.Belotti, P.Grandi, F.Zhang, D.Li, J.Drake, "Technology Agnostic OSPF Traffic Engineering Extensions for Generalized MPLS (GMPLS), work in progress draft-bccgd-ccamp-gmpls-ospf-agnostic-00", October 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.

9.2. 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

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.

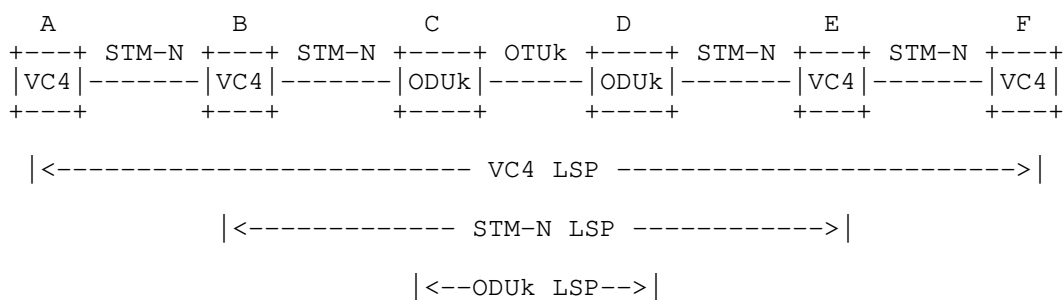


Figure 1: Example of Server Layer/Sub-Layer Selection

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.

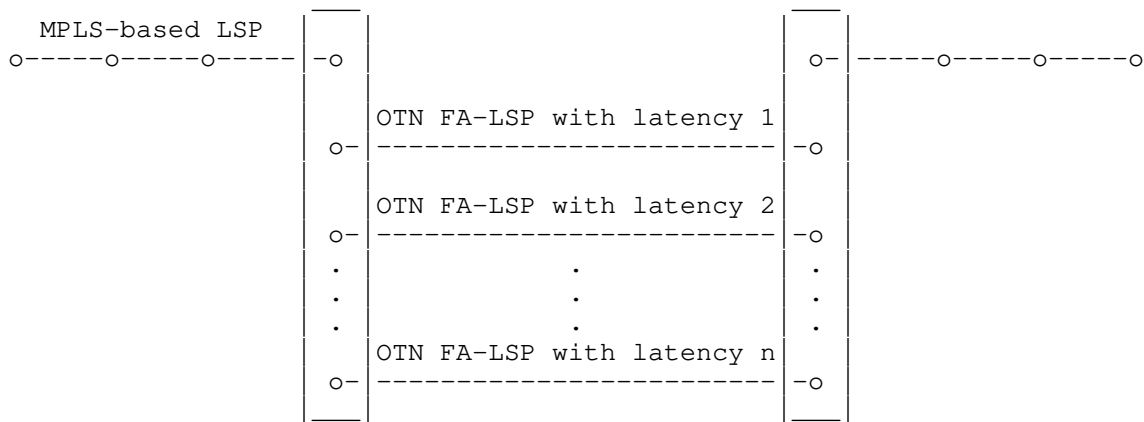


Figure 2: Example of FA-LSP Selection/Creation based on Latency

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 Multiplexing 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 hierarchy 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.

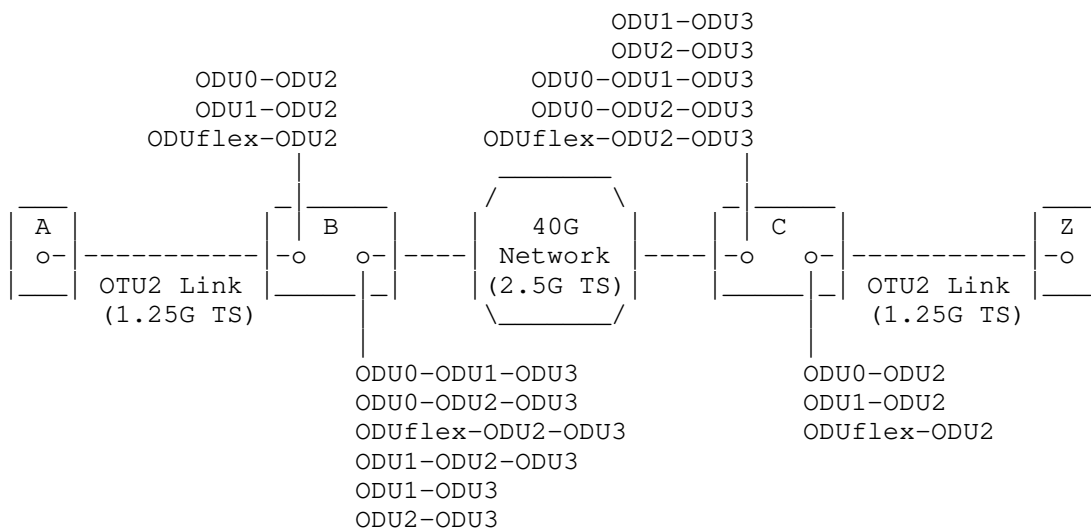


Figure 3 Example of Multi-Stages Multiplexing Hierarchy Selection

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.

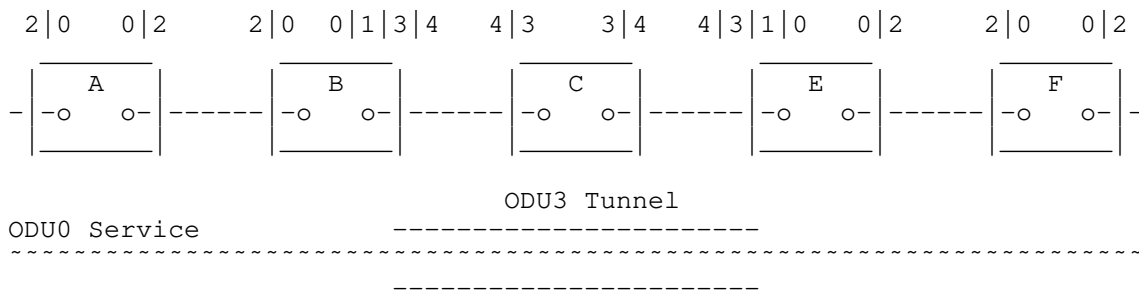


Figure 4 Example of Multi-Stages Multiplexing Hierarchy Selection

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.

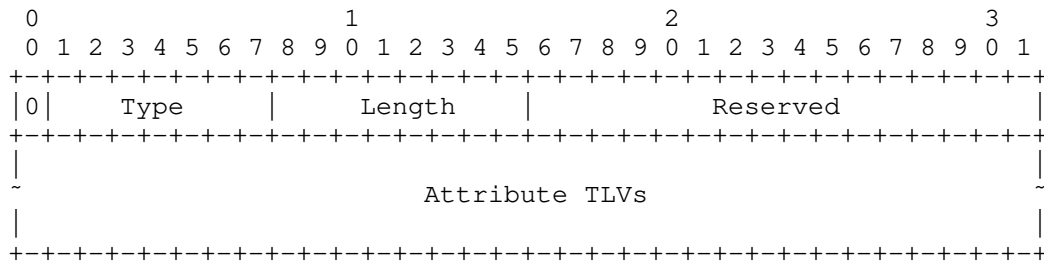


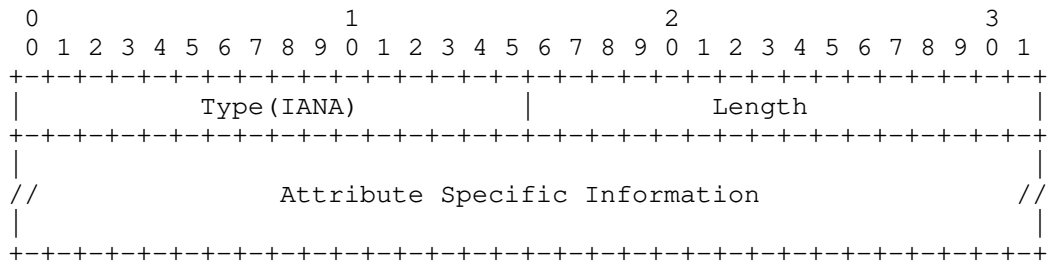
Figure 5 Format of BOUNDARY_ATTRIBUTES

- 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:

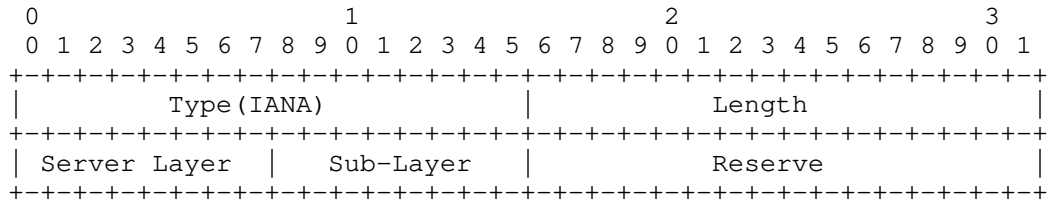


The following types are supported.

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.



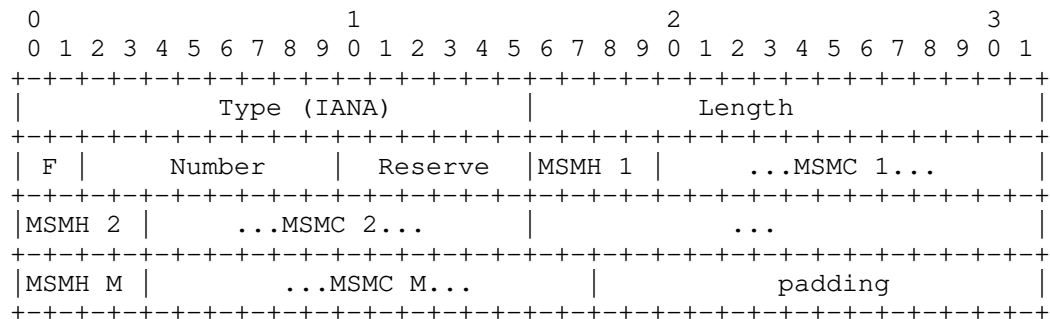
- o Type: indicates different values of Attribute TLV.
- o Length: indicates the total length of this Attribute TLV value.
- o Server Layer: Indicates which corresponding server layer should be triggered by the boundary node. The value of server layer is the same as the switching capability [RFC3471].
- 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 hierarchies constraint information of the link port.



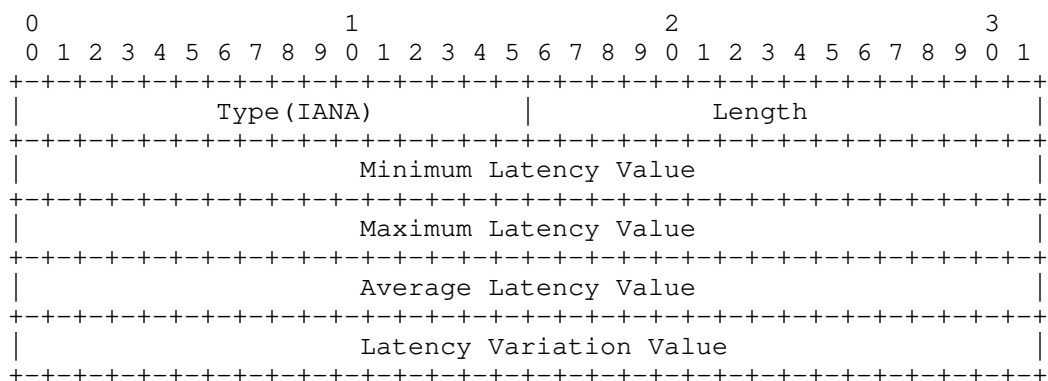
- o F (2 bits): Indicates the multi stages multiplexing hierarchies are included or excluded.
 - * 0 - Inclusive Multiplexing Hierarchies: Indicates that the object/TLV contains one or more multi stages multiplexing hierarchies which can be supported.
 - * 1 - Exclusive Multiplexing Hierarchies: Indicates 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 ODUi is mapped into ODUK (k > i) by single stage multiplexing (e.g., ODU0-ODU3).
 - + MSMH > 1: It indicates ODUi is mapped into ODUK (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 ODUK (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] 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.
- [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., Shiimoto, 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 Draft
Updates: 2205, 3209, 3473
Category: Standards Track
Expiration Date: April 14, 2011

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

October 14, 2010

Usage of The RSVP Association Object

draft-ietf-ccamp-assoc-info-00.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.

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 April 14, 2011

Copyright and License 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	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-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	Extended IPv4 and IPv6 ASSOCIATION Objects	16
5.1	Extended IPv4 and IPv6 ASSOCIATION Object Format	16
6	Security Considerations	18
7	IANA Considerations	18
7.1	Extended IPv4 and IPv6 ASSOCIATION Objects	18
7.2	Resource Sharing Association Type	19
8	Acknowledgments	19
9	References	19
9.1	Normative References	19
9.2	Informative References	20
10	Authors' Addresses	20

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 this context, this document formalizes the explanation provided in an e-mail to the Common Control and Measurement Plane (CCAMP) working group authored by Adrian Farrel, see [AF-EMAIL]. This document in no way modifies the normative definitions of end-to-end and segment recovery, see [RFC4872] or [RFC4873].

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.

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 three fields. The three fields taken together enable the identification of which LSPs are association with one another. The three 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 ($N \geq 1$) 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:

- The Association Type value is set to "Resource Sharing".
- Association Source is set to the originating node's router address.
- 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 in no way does this section modify 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 three 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-Recovery Usage

While the ASSOCIATION object, [RFC4872], is defined in the context of 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.

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 three 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 three 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. Extended IPv4 and IPv6 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.)

This sections defines new ASSOCIATION objects to address this limitation. Specifically, the Extended IPv4 ASSOCIATION object and Extended IPv6 ASSOCIATION object are defined below. Both new objects include an extended association ID field, which allows identification of a larger number of associations scoped within a given association source IP address.

The Extended IPv4 ASSOCIATION object and Extended IPv6 ASSOCIATION object SHOULD be supported by an implementation compliant with this document. The processing rules for the Extended IPv4 and IPv6 ASSOCIATION object are identical to those of the existing IPv4 and IPv6 ASSOCIATION objects.

5.1. Extended IPv4 and IPv6 ASSOCIATION Object Format

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

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								
Length										Class-Num(199)										C-Type (TBA)																			
Association Type										Association ID																													
Association ID (Continued)																																							
IPv4 Association Source																																							

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

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								
Length										Class-Num(199)										C-Type (TBA)																			
Association Type										Association ID																													
Association ID (Continued)																																							
IPv6 Association Source																																							

Association Type: 16 bits

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

Association ID: 48 bits

Same as for IPv4 and IPv6 ASSOCIATION objects, see [RFC4872].
(Only the size of this field differs from the [RFC4872] definition.)

Association Source: 4 or 16 bytes

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

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

This document formalizes the explanation provided in an e-mail to the working group authored by Adrian Farrel, see [AF-EMAIL]. 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.
- [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.

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: Thu, Oct 14, 2010 3:20:05 PM

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: April 12, 2011

W. Sun, Ed.
SJTU
G. Zhang, Ed.
CATR
October 9, 2010

Label Switched Path (LSP) Data Path Delay Metrics in Generalized MPLS/
MPLS-TE Networks
draft-ietf-ccamp-dpm-01.txt

Abstract

When setting up a label switched path (LSP) in Generalized MPLS and MPLS/TE networks, the completion of the signaling process does not necessarily mean that the cross connection along the LSP have been programmed accordingly and in a timely manner. Meanwhile, the completion of signaling process may be used by applications as indication that data path has become usable. The existence of this delay and the possible failure of cross connection programming, if not properly treated, will result in data loss or even application failure. Characterization of this performance can thus help designers to improve the application model and to build more robust applications. This document defines a series of performance metrics to evaluate the availability of data path in the signaling process.

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 12, 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.

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.

Table of Contents

1. Introduction	5
2. Conventions Used in This Document	6
3. Overview of Performance Metrics	7
4. Terms used in this document	8
5. A singleton Definition for RRFD	9
5.1. Motivation	9
5.2. Metric Name	9
5.3. Metric Parameters	9
5.4. Metric Units	9
5.5. Definition	10
5.6. Discussion	10
5.7. Methodologies	11
6. A singleton Definition for RSRD	12
6.1. Motivation	12
6.2. Metric Name	12
6.3. Metric Parameters	12
6.4. Metric Units	12
6.5. Definition	13
6.6. Discussion	13
6.7. Methodologies	14
7. A singleton Definition for PRFD	15
7.1. Motivation	15
7.2. Metric Name	15
7.3. Metric Parameters	15
7.4. Metric Units	15
7.5. Definition	15
7.6. Discussion	16
7.7. Methodologies	16
8. A singleton Definition for PSFD	18
8.1. Motivation	18
8.2. Metric Name	18
8.3. Metric Parameters	18
8.4. Metric Units	18
8.5. Definition	19
8.6. Discussion	19
8.7. Methodologies	20
9. A singleton Definition for PSRD	21
9.1. Motivation	21

9.2. Metric Name	21
9.3. Metric Parameters	21
9.4. Metric Units	21
9.5. Definition	21
9.6. Discussion	22
9.7. Methodologies	22
10. A Definition for Samples of Data Path Delay	24
10.1. Metric Name	24
10.2. Metric Parameters	24
10.3. Metric Units	24
10.4. Definition	24
10.5. Discussion	25
10.6. Methodologies	25
10.7. Typical testing cases	25
10.7.1. With No LSP in the Network	25
10.7.2. With a Number of LSPs in the Network	25
11. Some Statistics Definitions for Metrics to Report	27
11.1. The Minimum of Metric	27
11.2. The Median of Metric	27
11.3. The percentile of Metric	27
11.4. The Failure Probability	27
11.4.1. Failure Count	28
11.4.2. Failure Ratio	28
12. Security Considerations	29
13. IANA Considerations	30
14. Acknowledgements	31
15. References	32
15.1. Normative References	32
15.2. Informative References	32
Authors' Addresses	33

1. Introduction

Ideally, the completion of the signaling process means that the signaled label switched path (LSP) is available and is ready to carry traffic. However, in actual implementations, vendors may choose to program the cross connection in a pipelined manner, so that the overall LSP provisioning delay can be reduced. In such situations, the data path may not be available instantly after the signaling process completes. Implementation deficiency may also cause the inconsistency in between the signaling process and data path provisioning. For example, if the data plane fails to program the cross connection accordingly but does not manage to report this to the control plane, the signaling process may complete successfully while the corresponding data path will never become functional at all.

On the other hand, the completion of the signaling process may be used in many cases as indication of data path availability. For example, when invoking through User Network Interface (UNI), a client device or an application may use the reception of the correct RESV message as indication that data path is fully functional and start to transmit traffic. This will results in data loss or even application failure.

Although RSVP(-TE) specifications have suggested that the cross connections are programmed before signaling messages are propagated upstream, it is still worthwhile to verify the conformance of an implementation and measure the delay, when necessary.

This document defines a series of performance metrics to evaluate the availability of data path when the signaling process completes. The metrics defined in this document complements the control plane metrics defined in [RFC5814]. These metrics can be used to verify the conformance of implementations against related specifications, as elaborated in [I-D.shiomoto-ccamp-switch-programming]. They also can be used to build more robust applications.

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. Overview of Performance Metrics

In this memo, we define five performance metrics to characterize the performance of data path provisioning with GMPLS/MPLS-TE signaling. These metrics complement the metrics defined in [RFC5814], in the sense that the completion of the signaling process for a Label Switched Path (LSP) and the programming of cross connections along the LSP may not be consistent. The performance metrics in [RFC5814] characterize the performance of LSP provisioning from the pure signaling point of view, while the metric in this document takes into account the validity of the data path.

The five metrics are:

- o RRFD - the delay between RESV message received by ingress node and forward data path becomes available.
- o RSRD - the delay between RESV message sent by egress node and reverse data path becomes available.
- o PRFD - the delay between PATH message received by egress node and forward data path becomes available.
- o PSFD - the delay between PATH message sent by ingress and forward data path becomes available.
- o PSRD - the delay between PATH message sent by ingress and reverse data path becomes available.

As in [RFC5814], we continue to use the structures and notions introduced and discussed in the IPPM Framework document, [RFC2330] [RFC2679] [RFC2681]. The reader is assumed to be familiar with the notions in those documents. The readers are assumed to be familiar with the definitions in [RFC5814] as well.

4. Terms used in this document

- o Forward data path - the data path from the ingress to the egress. Instances of forward data path include the data path of a uni-directional LSP and data path from the ingress node to the egress node in a bidirectional LSP.
- o Reverse data path - the data path from the egress to the ingress in a bidirectional LSP.
- o Data path delay - the time needed to complete the data path configuration, in relation to the signaling process. five types of data path delay are defined in this document, namely RRFD, RSRD and PRFD. Data path delay used in this document must be distinguished from the data path transmission delay.
- o Error free signal - data plane specific indication of availability of the data path. For example, for packet switching capable interfaces, the reception of the first error free packet from one side of the LSP to the other can be used as the error free signal. For SDH/SONET cross connects, the disappearance of alarm can be used as the error free signal. Through out this document, we will use the "error free signal" as a general term. An implementations must choose a proper data path signal that is specific to the data path technology being tested.
- o Ingress/egress node - in this memo, an ingress/egress node means a measurement endpoint with both control plane and data plane features. Typically, the control plane part on an ingress/egress node interact with the control plane of the network under test. The data plane part of an ingress/egress node will generate data path signals and send the signal to the data plane of the network under test, or receive data path signals from the network under test.

5. A singleton Definition for RRFD

This part defines a metric for forward data path delay when an LSP is setup.

As described in [I-D.shiomoto-ccamp-switch-programming], the completion of the RSVP-TE signaling process does not necessarily mean that the cross connections along the LSP being setup are in place and ready to carry traffic. This metric defines the time difference between the reception of RESV message by the ingress node and the completion of the cross connection programming along the forward data path.

5.1. Motivation

RRFD is useful for several reasons:

- o For the reasons described in [I-D.shiomoto-ccamp-switch-programming], the data path may not be available instantly after the completion of the RSVP-TE signaling process. The delay itself is part of the implementation performance.
- o The completion of the signaling process may be used by application designers as indication of data path availability. The existence of this delay and the potential failure of cross connection programming, if not properly treated, will result in data loss or application failure. The typical value of this delay can thus help designers to improve the application model.

5.2. Metric Name

RRFD

5.3. Metric Parameters

- o ID0, the ingress LSR ID
- o ID1, the egress LSR ID
- o T, a time when the setup is attempted

5.4. Metric Units

Either a real number of milli-seconds or undefined.

5.5. Definition

For a real number dT , RRFD from ingress node ID0 to egress node ID1 at T is dT means that ingress node ID0 send a PATH message to egress node ID1 and the last bit of the corresponding RESV message is received by ingress node ID0 at T , and an error free signal is received by egress node ID1 by using a data plane specific test pattern at $T+dT$.

5.6. Discussion

The following issues are likely to come up in practice:

- o The accuracy of RRFD depends on the clock resolution of both the ingress node and egress node. Clock synchronization between the ingress node and egress node is required.
- o The accuracy of RRFD is also dependent on how the error free signal is received and may differ significantly when the underline data plane technology is different. For instance, for an LSP between a pair of Ethernet interfaces, the ingress node may use a rate based method to verify the availability of the data path and use the reception of the first error free frame as the error free signal. In this case, the interval between two successive frames has a significant impact on accuracy. It is RECOMMENDED that the ingress node uses small intervals, under the condition that the injected traffic does not exceed the capacity of the forward data path. The value of the interval MUST be reported.
- o The accuracy of RRFD is also dependent on the time needed to propagate the error free signal from the ingress node to the egress node. A typical value of propagating the error free signal from the ingress node to the egress node under the same measurement setup MAY be reported. The methodology to obtain such values is outside the scope of this document.
- o It is possible that under some implementations, a node may program the cross connection before it sends PATH message further downstream and the data path may be available before a RESV message reaches the ingress node. In such cases, RRFD can be a negative value. It is RECOMMENDED that PRFD measurement is carried out to further characterize the forward data path delay when a negative RRFD value is observed.
- o If error free signal is received by the egress node before PATH message is sent on the ingress node, an error MUST be reported and the measurement SHOULD terminate.

- o If the corresponding RESV message is received, but no error free signal is received by the egress node within a reasonable period of time, i.e., a threshold, RRFD MUST be treated as undefined. The value of the threshold MUST be reported.
- o If the LSP setup fails, the metric value MUST NOT be counted.

5.7. Methodologies

Generally the methodology would proceed as follows:

- o Make sure that the network has enough resource to set up the requested LSP.
- o Start the data path measurement and/or monitoring procedures on the ingress node and egress node. If error free signal is received by the egress node before PATH message is sent, report an error and terminate the measurement.
- o At the ingress node, form the PATH message according to the LSP requirements and send the message towards the egress node.
- o Upon receiving the last bit of the corresponding RESV message, take the time stamp (T1) on the ingress node as soon as possible.
- o When an error free signal is observed on the egress node, take the time stamp (T2) as soon as possible. An estimate of RRFD ($T2 - T1$) can be computed.
- o If the corresponding RESV message arrives, but no error free signal is received within a reasonable period of time by the ingress node, RRFD is deemed to be undefined.
- o If the LSP setup fails, RRFD is not counted.

6. A singleton Definition for RSRD

This part defines a metric for reverse data path delay when an LSP is setup.

As described in [I-D.shiomoto-ccamp-switch-programming], the completion of the RSVP-TE signaling process does not necessarily mean that the cross connections along the LSP being setup are in place and ready to carry traffic. This metric defines the time difference between the completion of the signaling process and the completion of the cross connection programming along the reverse data path. This metric MAY be used together with RRFD to characterize the data path delay of a bidirectional LSP.

6.1. Motivation

RSRD is useful for several reasons:

- o For the reasons described in [I-D.shiomoto-ccamp-switch-programming], the data path may not be available instantly after the completion of the RSVP-TE signaling process. The delay itself is part of the implementation performance.
- o The completion of the signaling process may be used by application designers as indication of data path availability. The existence of this delay and the possible failure of cross connection programming, if not properly treated, will result in data loss or application failure. The typical value of this delay can thus help designers to improve the application model.

6.2. Metric Name

RSRD

6.3. Metric Parameters

- o ID0, the ingress LSR ID
- o ID1, the egress LSR ID
- o T, a time when the setup is attempted

6.4. Metric Units

Either a real number of milli-seconds or undefined.

6.5. Definition

For a real number dT , RSRD from ingress node ID0 to egress node ID1 at T is dT means that ingress node ID0 send a PATH message to egress node ID1 and the last bit of the corresponding RESV message is sent by egress node ID1 at T , and an error free signal is received by the ingress node ID0 using a data plane specific test pattern at $T+dT$.

6.6. Discussion

The following issues are likely to come up in practice:

- o The accuracy of RSRD depends on the clock resolution of both the ingress node and egress node. And clock synchronization between the ingress node and egress node is required.
- o The accuracy of RSRD is also dependent on how the error free signal is received and may differ significantly when the underline data plane technology is different. For instance, for an LSP between a pair of Ethernet interfaces, the egress node (sometimes the tester) may use a rate based method to verify the availability of the data path and use the reception of the first error free frame as the error free signal. In this case, the interval between two successive frames has a significant impact on accuracy. It is RECOMMENDED that in this case the egress node uses small intervals, under the condition that the injected traffic does not exceed the capacity of the reverse data path. The value of the interval MUST be reported.
- o The accuracy of RSRD is also dependent on the time needed to propagate the error free signal from the egress node to the ingress node. A typical value of propagating the error free signal from the egress node to the ingress node under the same measurement setup MAY be reported. The methodology to obtain such values is outside the scope of this document.
- o If the corresponding RESV message is sent, but no error free signal is received by the ingress node within a reasonable period of time, i.e., a threshold, RSRD MUST be treated as undefined. The value of the threshold MUST be reported.
- o If error free signal is received before PATH message is sent on the ingress node, an error MUST be reported and the measurement SHOULD terminate.
- o If the LSP setup fails, the metric value MUST NOT be counted.

6.7. Methodologies

Generally the methodology would proceed as follows:

- o Make sure that the network has enough resource to set up the requested LSP.
- o Start the data path measurement and/or monitoring procedures on the ingress node and egress node. If error free signal is received by the ingress node before PATH message is sent, report an error and terminate the measurement.
- o At the ingress node, form the PATH message according to the LSP requirements and send the message towards the egress node.
- o Upon sending the last bit of the corresponding RESV message, take the time stamp (T1) on the egress node as soon as possible.
- o When an error free signal is observed on the ingress node, take the time stamp (T2) as soon as possible. An estimate of RSRD (T2-T1) can be computed.
- o If the LSP setup fails, RSRD is not counted.
- o If no error free signal is received within a reasonable period of time by the ingress node, RSRD is deemed to be undefined.

7. A singleton Definition for PRFD

This part defines a metric for forward data path delay when an LSP is setup.

In an RSVP-TE implementation, when setting up an LSP, each node may choose to program the cross connection before it sends PATH message further downstream. In this case, the forward data path may become available before the signaling process completes, ie. before the RESV reaches the ingress node. This metric can be used to identify such implementation practice and give useful information to application designers.

7.1. Motivation

PRFD is useful for the following reasons:

- o PRFD can be used to identify an RSVP-TE implementation practice, in which cross connections are programmed before PATH message is sent downstream.
- o The value of PRFD may also help application designers to fine tune their application model.

7.2. Metric Name

PRFD

7.3. Metric Parameters

- o ID0, the ingress LSR ID
- o ID1, the egress LSR ID
- o T, a time when the setup is attempted

7.4. Metric Units

Either a real number of milli-seconds or undefined.

7.5. Definition

For a real number dT , PRFD from ingress node ID0 to egress node ID1 at T is dT means that ingress node ID0 send a PATH message to egress node ID1 and the last bit of the PATH message is received by egress node ID1 at T, and an error free signal is received by the egress node ID1 using a data plane specific test pattern at $T+dT$.

7.6. Discussion

The following issues are likely to come up in practice:

- o The accuracy of PRFD depends on the clock resolution of the egress node. And clock synchronization between the ingress node and egress node is not required.
- o The accuracy of PRFD is also dependent on how the error free signal is received and may differ significantly when the underline data plane technology is different. For instance, for an LSP between a pair of Ethernet interfaces, the egress node (sometimes the tester) may use a rate based method to verify the availability of the data path and use the reception of the first error free frame as the error free signal. In this case, the interval between two successive frames has a significant impact on accuracy. It is RECOMMENDED that in this case the ingress node uses small intervals, under the condition that the injected traffic does not exceed the capacity of the forward data path. The value of the interval MUST be reported.
- o The accuracy of PRFD is also dependent on the time needed to propagate the error free signal from the ingress node to the egress node. A typical value of propagating the error free signal from the ingress node to the egress node under the same measurement setup MAY be reported. The methodology to obtain such values is outside the scope of this document.
- o If error free signal is received before PATH message is sent, an error MUST be reported and the measurement SHOULD terminate.
- o If the LSP setup fails, the metric value MUST NOT be counted.
- o This metric SHOULD be used together with RRFD. It is RECOMMENDED that PRFD measurement is carried out after a negative RRFD value has already been observed.

7.7. Methodologies

Generally the methodology would proceed as follows:

- o Make sure that the network has enough resource to set up the requested LSP.
- o Start the data path measurement and/or monitoring procedures on the ingress node and egress node. If error free signal is received by the egress node before PATH message is sent, report an error and terminate the measurement.

- o At the ingress node, form the PATH message according to the LSP requirements and send the message towards the egress node.
- o Upon receiving the last bit of the PATH message, take the time stamp (T1) on the egress node as soon as possible.
- o When an error free signal is observed on the egress node, take the time stamp (T2) as soon as possible. An estimate of PRFD (T2-T1) can be computed.
- o If the LSP setup fails, PRFD is not counted.
- o If no error free signal is received within a reasonable period of time by the egress node, PRFD is deemed to be undefined.

8. A singleton Definition for PSFD

This part defines a metric for forward data path delay when an LSP is setup.

As described in [I-D.shiomoto-ccamp-switch-programming], the completion of the RSVP-TE signaling process does not necessarily mean that the cross connections along the LSP being setup are in place and ready to carry traffic. This metric defines the time from the PATH message sent by the ingress node, till the completion of the cross connection programming along the LSP forward data path.

8.1. Motivation

PSFD is useful for the following reasons:

- o For the reasons described in [I-D.shiomoto-ccamp-switch-programming], the data path setup delay may not be consistent with the control plane LSP setup delay. The data path setup delay metric is more precise for LSP setup performance measurement.
- o The completion of the signaling process may be used by application designers as indication of data path availability. The difference between the control plane setup delay and data path delay, and the potential failure of cross connection programming, if not properly treated, will result in data loss or application failure. This metric can thus help designers to improve the application model.

8.2. Metric Name

PSFD

8.3. Metric Parameters

- o ID0, the ingress LSR ID
- o ID1, the egress LSR ID
- o T, a time when the setup is attempted

8.4. Metric Units

Either a real number of milli-seconds or undefined.

8.5. Definition

For a real number dT , PSFD from ingress node ID0 to egress node ID1 at T is dT means that ingress node ID0 sends the first bit of a PATH message to egress node ID1 at T , and an error free signal is received by the egress node ID1 using a data plane specific test pattern at $T+dT$.

8.6. Discussion

The following issues are likely to come up in practice:

- o The accuracy of PSFD depends on the clock resolution of both the ingress node and egress node. And clock synchronization between the ingress node and egress node is required.
- o The accuracy of this metric is also dependent on how the error free signal is received and may differ significantly when the underlying data plane technology is different. For instance, for an LSP between a pair of Ethernet interfaces, the ingress node may use a rate based method to verify the availability of the data path and use the reception of the first error free frame as the error free signal. In this case, the interval between two successive frames has a significant impact on accuracy. It is RECOMMENDED that the ingress node uses small intervals, under the condition that the injected traffic does not exceed the capacity of the forward data path. The value of the interval MUST be reported.
- o The accuracy of this metric is also dependent on the time needed to propagate the error free signal from the ingress node to the egress node. A typical value of propagating the error free signal from the ingress node to the egress node under the same measurement setup MAY be reported. The methodology to obtain such values is outside the scope of this document.
- o If error free signal is received before PATH message is sent, an error MUST be reported and the measurement SHOULD terminate.
- o If the LSP setup fails, the metric value MUST NOT be counted.
- o If the PATH message is sent by the ingress node, but no error free signal is received by the egress node within a reasonable period of time, i.e., a threshold, the metric value MUST be treated as undefined. The value of the threshold MUST be reported.

8.7. Methodologies

Generally the methodology would proceed as follows:

- o Make sure that the network has enough resource to set up the requested LSP.
- o Start the data path measurement and/or monitoring procedures on the ingress node and egress node. If error free signal is received by the egress node before PATH message is sent, report an error and terminate the measurement.
- o At the ingress node, form the PATH message according to the LSP requirements and send the message towards the egress node. A timestamp (T1) may be stored locally in the ingress node when the PATH message packet is sent towards the egress node.
- o When an error free signal is observed on the egress node, take the time stamp (T2) as soon as possible. An estimate of PSFD ($T2-T1$) can be computed.
- o If the LSP setup fails, this metric is not counted.
- o If no error free signal is received within a reasonable period of time by the egress node, PSFD is deemed to be undefined.

9. A singleton Definition for PSRD

This part defines a metric for reverse data path delay when an LSP is setup.

This metric defines the time from the ingress node sends the PATH message, till the completion of the cross connection programming along the LSP reverse data path. This metric MAY be used together with PSFD to characterize the data path delay of a bidirectional LSP.

9.1. Motivation

PSRD is useful for the following reasons:

- o For the reasons described in [I-D.shiomoto-ccamp-switch-programming], the data path setup delay may not be consistent with the control plane LSP setup delay. The data path setup delay metric is more precise for LSP setup performance measurement.
- o The completion of the signaling process may be used by application designers as indication of data path availability. The difference between the control plane setup delay and data path delay, and the potential failure of cross connection programming, if not properly treated, will result in data loss or application failure. This metric can thus help designers to improve the application model.

9.2. Metric Name

PSRD

9.3. Metric Parameters

- o ID0, the ingress LSR ID
- o ID1, the egress LSR ID
- o T, a time when the setup is attempted

9.4. Metric Units

Either a real number of milli-seconds or undefined.

9.5. Definition

For a real number dT , PSRD from ingress node ID0 to egress node ID1 at T is dT means that ingress node ID0 sends the first bit of a PATH message to egress node ID1 at T, and an error free signal is received

through the reverse data path by the ingress node ID0 using a data plane specific test pattern at T+dT.

9.6. Discussion

The following issues are likely to come up in practice:

- o The accuracy of PSRD depends on the clock resolution of the ingress node. And clock synchronization between the ingress node and egress node is not required.
- o The accuracy of this metric is also dependent on how the error free signal is received and may differ significantly when the underlying data plane technology is different. For instance, for an LSP between a pair of Ethernet interfaces, the egress node may use a rate based method to verify the availability of the data path and use the reception of the first error free frame as the error free signal. In this case, the interval between two successive frames has a significant impact on accuracy. It is RECOMMENDED that the egress node uses small intervals, under the condition that the injected traffic does not exceed the capacity of the forward data path. The value of the interval MUST be reported.
- o The accuracy of this metric is also dependent on the time needed to propagate the error free signal from the egress node to the ingress node. A typical value of propagating the error free signal from the egress node to the ingress node under the same measurement setup MAY be reported. The methodology to obtain such values is outside the scope of this document.
- o If error free signal is received before PATH message is sent, an error MUST be reported and the measurement SHOULD terminate.
- o If the LSP setup fails, this metric value MUST NOT be counted.
- o If the PATH message is sent by the ingress node, but no error free signal is received by the ingress node within a reasonable period of time, i.e., a threshold, the metric value MUST be treated as undefined. The value of the threshold MUST be reported.

9.7. Methodologies

Generally the methodology would proceed as follows:

- o Make sure that the network has enough resource to set up the requested LSP.

- o Start the data path measurement and/or monitoring procedures on the ingress node and egress node. If error free signal is received by the egress node before PATH message is sent, report an error and terminate the measurement.
- o At the ingress node, form the PATH message according to the LSP requirements and send the message towards the egress node. A timestamp (T1) may be stored locally in the ingress node when the PATH message packet is sent towards the egress node.
- o When an error free signal is observed on the ingress node, take the time stamp (T2) as soon as possible. An estimate of PSFD (T2-T1) can be computed.
- o If the LSP setup fails, this metric is not counted.
- o If no error free signal is received within a reasonable period of time by the ingress node, the metric value is deemed to be undefined.

10. A Definition for Samples of Data Path Delay

In Section 5, Section 6, Section 7, Section 8 and Section 9, we define the singleton metrics of data path delay. Now we define how to get one particular sample of such delay. Sampling is to select a particular portion of singleton values of the given parameters. Like in [RFC2330], we use Poisson sampling as an example.

10.1. Metric Name

Type <X> Data path delay sample, where X is either RRFD, RSRD, PRFD, PSFD and PSRD.

10.2. Metric Parameters

- o ID0, the ingress LSR ID
- o ID1, the egress LSR ID
- o T0, a time
- o Tf, a time
- o Lambda, a rate in the reciprocal seconds
- o Th, LSP holding time
- o Td, the maximum waiting time for successful LSP setup
- o Ts, the maximum waiting time for error free signal

10.3. Metric Units

A sequence of pairs; the elements of each pair are:

- o T, a time when setup is attempted
- o dT, either a real number of milli-seconds or undefined

10.4. Definition

Given T0, Tf, and Lambda, compute a pseudo-random Poisson process beginning at or before T0, with average arrival rate Lambda, and ending at or after Tf. Those time values greater than or equal to T0 and less than or equal to Tf are then selected. At each of the times in this process, we obtain the value of data path delay sample of type <X> at this time. The value of the sample is the sequence made up of the resulting <time, type <X> data path delay> pairs. If there

are no such pairs, the sequence is of length zero and the sample is said to be empty.

10.5. Discussion

The following issues are likely to come up in practice:

- o The parameters Lambda, Th and Td should be carefully chosen, as explained in the discussions for LSP setup delay (see [RFC5814]).
- o The parameter Ts should be carefully chosen and MUST be reported along with the LSP forward/reverse data path delay sample.

10.6. Methodologies

Generally the methodology would proceed as follows:

- o The selection of specific times, using the specified Poisson arrival process, and
- o Set up the LSP and obtain the value of type <X> data path delay
- o Release the LSP after Th, and wait for the next Poisson arrival process

10.7. Typical testing cases

10.7.1. With No LSP in the Network

10.7.1.1. Motivation

Data path delay with no LSP in the network is important because this reflects the inherent delay of a device implementation. The minimum value provides an indication of the delay that will likely be experienced when an LSP data path is configured under light traffic load.

10.7.1.2. Methodologies

Make sure that there is no LSP in the network, and proceed with the methodologies described in Section 10.6.

10.7.2. With a Number of LSPs in the Network

10.7.2.1. Motivation

Data path delay with a number of LSPs in the network is important because it reflects the performance of an operational network with

considerable load. This delay may vary significantly as the number of existing LSPs varies. It can be used as a scalability metric of a device implementation.

10.7.2.2. Methodologies

Setup the required number of LSPs, and wait until the network reaches a stable state, and then proceed with the methodologies described in Section 10.6.

11. Some Statistics Definitions for Metrics to Report

Given the samples of the performance metric, we now offer several statistics of these samples to report. From these statistics, we can draw some useful conclusions of a GMPLS network. The value of these metrics is either a real number, or an undefined number of milliseconds. In the following discussion, we only consider the finite values.

11.1. The Minimum of Metric

The minimum of metric is the minimum of all the dT values in the sample. In computing this, undefined values SHOULD be treated as infinitely large. Note that this means that the minimum could thus be undefined if all the dT values are undefined. In addition, the metric minimum SHOULD be set to undefined if the sample is empty.

11.2. The Median of Metric

Metric median is the median of the dT values in the given sample. In computing the median, the undefined values MUST NOT be counted in. The Median SHOULD be set to undefined if all the dT values are undefined, or if the sample is empty.

11.3. The percentile of Metric

The "empirical distribution function" (EDF) of a set of scalar measurements is a function $F(x)$ which for any x gives the fractional proportion of the total measurements that were $\leq x$.

Given a percentage X , the X -th percentile of Metric means the smallest value of x for which $F(x) \geq X$. In computing the percentile, undefined values MUST NOT be included.

See [RFC2330] for further details.

11.4. The Failure Probability

Given the samples of the performance metric, we now offer two statistics of failure events of these samples to report. The two statistics can be applied to both forward data path and reverse data path. For example, when a sample of RRFD has been obtained the forward data path failure statistics can be obtained, while when a sample of RSRD can be used to calculate the reverse data path failure statistics. Detailed definitions of the Failure Count and Failure Ratio are given below.

11.4.1. Failure Count

Failure Count is defined as the number of the undefined value of the corresponding performance metric in a sample. The value of Failure Count is an integer.

11.4.2. Failure Ratio

Failure Ratio is the percentage of the number of failure events to the total number of requests in a sample. Here an failure event means that the signaling completes with no error, while no error free signal is observed. The calculation for Failure Ratio is defined as follows:

Failure Ratio = Number of undefined value / (Number of valid metric values + Number of undefined value) * 100%.

12. Security Considerations

In the control plane, since the measurement endpoints must be conformant to signaling specifications and behave as normal signaling endpoints, it will not incur other security issues than normal LSP provisioning. However, the measurement parameters must be carefully selected so that the measurements inject trivial amounts of additional traffic into the networks they measure. If they inject "too much" traffic, they can skew the results of the measurement, and in extreme cases cause congestion and denial of service.

In the data plane, the measurement endpoint MUST use a signal that is consistent with what is specified in the control plane. For example, in a packet switched case, the traffic injected into the data plane MUST NOT exceed the specified rate in the corresponding LSP setup request. In a wavelength switched case, the measurement endpoint MUST use the specified or negotiated lambda with appropriate power.

The security considerations pertaining to the original RSVP protocol [RFC2205] and its TE extensions [RFC3209] also remain relevant.

13. IANA Considerations

This document makes no requests for IANA action.

14. Acknowledgements

We wish to thank Adrian Farrel and Lou Berger for their comments and helps.

This document contains ideas as well as text that have appeared in existing IETF documents. The authors wish to thank G. Almes, S. Kalidindi and M. Zekauskas.

We also wish to thank Weisheng Hu, Yaohui Jin and Wei Guo in the state key laboratory of advanced optical communication systems and networks for the valuable comments. We also wish to thank the support from NSFC and 863 program of China.

15. References

15.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC2205] Braden, B., Zhang, L., Berson, S., Herzog, S., and S. Jamin, "Resource ReSerVation Protocol (RSVP) -- Version 1 Functional Specification", RFC 2205, September 1997.
- [RFC2679] Almes, G., Kalidindi, S., and M. Zekauskas, "A One-way Delay Metric for IPPM", RFC 2679, September 1999.
- [RFC2681] Almes, G., Kalidindi, S., and M. Zekauskas, "A Round-trip Delay Metric for IPPM", RFC 2681, September 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.

15.2. Informative References

- [I-D.shiomoto-ccamp-switch-programming] Shiomoto, K. and A. Farrel, "Advice on When It is Safe to Start Sending Data on Label Switched Paths Established Using RSVP-TE", draft-shiomoto-ccamp-switch-programming-01 (work in progress), October 2009.
- [RFC2330] Paxson, V., Almes, G., Mahdavi, J., and M. Mathis, "Framework for IP Performance Metrics", RFC 2330, May 1998.
- [RFC5814] Sun, W. and G. Zhang, "Label Switched Path (LSP) Dynamic Provisioning Performance Metrics in Generalized MPLS Networks", RFC 5814, March 2010.

Authors' Addresses

Weiqiang Sun, Editor
Shanghai Jiao Tong University
800 Dongchuan Road
Shanghai 200240
China

Phone: +86 21 3420 5359
Email: sunwq@mit.edu

Guoying Zhang, Editor
China Academy of Telecommunication Research, MIIT, China.
No.52 Hua Yuan Bei Lu, Haidian District
Beijing 100083
China

Phone: +86 1062300103
Email: zhangguoying@mail.ritt.com.cn

Jianhua Gao
Huawei Technologies Co., LTD.
China

Phone: +86 755 28973237
Email: gjhhit@huawei.com

Guowu Xie
University of California, Riverside
900 University Ave.
Riverside, CA 92521
USA

Phone: +1 951 237 8825
Email: xieg@cs.ucr.edu

Rajiv Papneja
Isocore
12359 Sunrise Valley Drive, STE 100
Reston, VA 20190
USA

Phone: +1 703 860 9273
Email: rpapneja@isocore.com

Contributors

Bin Gu
IXIA
Oriental Kenzo Plaza 8M, 48 Dongzhimen Wai Street, Dongcheng District
Beijing 200240
China

Phone: +86 13611590766
Email: BGu@ixiacom.com

Xueqin Wei
Fiberhome Telecommunication Technology Co., Ltd.
Wuhan
China

Phone: +86 13871127882
Email: xqwei@fiberhome.com.cn

Tomohiro Otani
KDDI R&D Laboratories, Inc.
2-1-15 Ohara Kamifukuoka Saitama
356-8502
Japan

Phone: +81-49-278-7357
Email: otani@kddilabs.jp

Ruiquan Jing
China Telecom Beijing Research Institute
118 Xizhimenwai Avenue
Beijing 100035
China

Phone: +86-10-58552000
Email: jingrq@ctbri.com.cn

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

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

October 13, 2010

General Network Element Constraint Encoding for GMPLS Controlled
Networks

draft-ietf-ccamp-general-constraint-encode-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 March 13, 2007.

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.....	4
1.2. Non-Local Label Assignment Constraints.....	4
2. Extension Encoding Usage Recommendations.....	5
2.1. Extension Node TLV.....	6
2.2. Extension Link TLV.....	6
2.3. Extension Dynamic Link TLV.....	6
3. Encoding.....	6
3.1. Link Set Field.....	6
3.2. Label Set Field.....	8
3.2.1. Inclusive/Exclusive Label Lists.....	9

3.2.2. Inclusive/Exclusive Label Ranges.....	10
3.2.3. Bitmap Label Set.....	10
3.3. Available Labels Sub-TLV.....	11
3.4. Shared Backup Labels Sub-TLV.....	12
3.5. Connectivity Matrix Sub-TLV.....	12
3.6. Port Label Restriction sub-TLV.....	13
3.6.1. SIMPLE_LABEL.....	14
3.6.2. CHANNEL_COUNT.....	15
3.6.3. LABEL_RANGE1.....	15
3.6.4. SIMPLE_LABEL & CHANNEL_COUNT.....	16
3.6.5. Link Label Exclusivity.....	16
4. Security Considerations.....	16
5. IANA Considerations.....	17
6. Acknowledgments.....	17
APPENDIX A: Encoding Examples.....	18
A.1. Link Set Field.....	18
A.2. Label Set Field.....	18
A.3. Connectivity Matrix Sub-TLV.....	19
A.4. Connectivity Matrix with Bi-directional Symmetry.....	22
7. References.....	25
7.1. Normative References.....	25
7.2. Informative References.....	25
8. Contributors.....	26
Authors' Addresses.....	27
Intellectual Property Statement.....	28
Disclaimer of Validity.....	28

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.

ConnectivityMatrix(i, j) ::= <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 Matrix(i, j) = 0 or 1 depending on whether ingress port i can connect to egress port j for one or more labels.

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.

```
<PortLabelRestriction> ::= [<GeneralPortRestrictions>...]  
[<MatrixSpecificRestrictions>...]  
  
<GeneralPortRestrictions> ::= <RestrictionType>  
[<RestrictionParameters>]  
  
<MatrixSpecificRestriction> ::= <MatrixID> <RestrictionType>  
[<RestrictionParameters>]
```

Where

MatrixID is the ID of the corresponding connectivity matrix

The RestrictionType parameter is used to specify general port restrictions and matrix specific restrictions.

2. Extension Encoding Usage Recommendations

In this section we give recommendations of typical usage of the sub-TLVs and composite TLVs.

2.1. Extension Node TLV

The Extension Node TLV could consist of the following list of sub-TLVs:

```
<Node_Info> ::= <Node_ID>[Other GMPLS sub-TLVs]
[<ConnectivityMatrix>...]
```

2.2. Extension Link TLV

The new link related sub-TLVs could be incorporated into a composite link TLV as follows:

```
<LinkInfo> ::= <LinkID> [Other GMPLS sub-TLVs]
[<PortLabelRestriction>...][<AvailableLabels>] [<SharedBackupLabels>]
```

2.3. Extension Dynamic Link TLV

If the protocol supports the separation of dynamic information from relatively static information then the available wavelength and shared backup status can be separated from the general link TLV into a TLV for dynamic link information.

```
<DynamicLinkInfo> ::= <LinkID> <AvailableLabels>
[<SharedBackupLabels>]
```

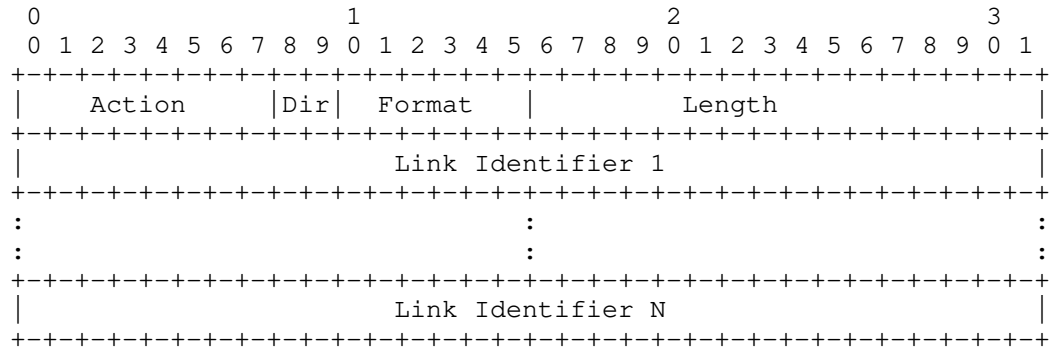
3. Encoding

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.

3.1. 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 3.5.
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.

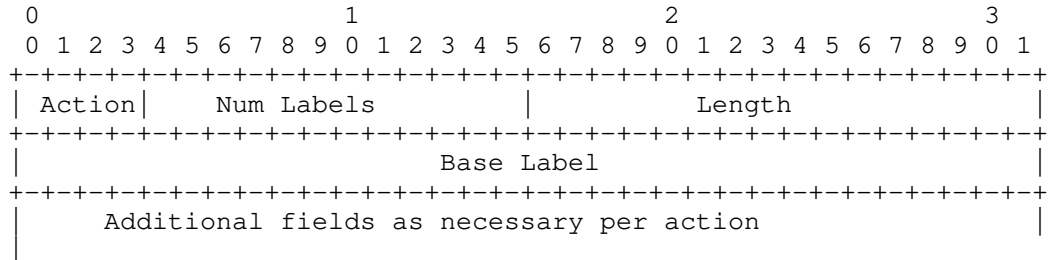
Link Identifier: length is dependent on the link format

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.

3.2. Label Set Field

Label Set Field is used within the <AvailableLabels> sub-TLV or the <SharedBackupLabels> sub-TLV, which is defined in Section 3.3. and 3.4. , 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:

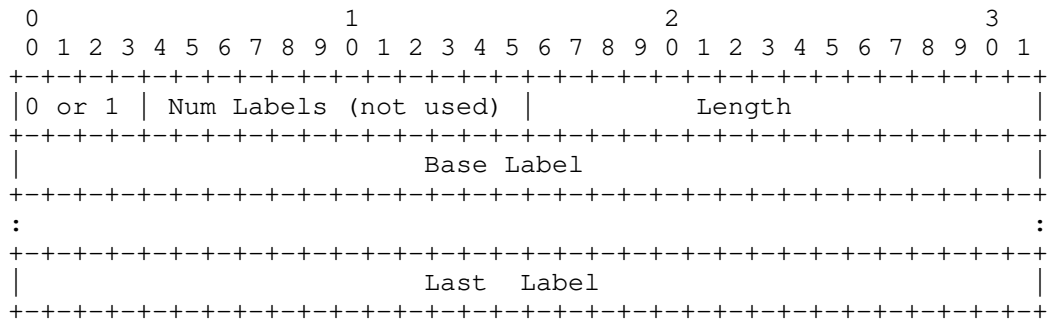
- 0 - Inclusive List
- 1 - Exclusive List
- 2 - Inclusive Range
- 3 - Exclusive Range
- 4 - Bitmap Set

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.

3.2.1. Inclusive/Exclusive Label Lists

In the case of the inclusive/exclusive lists the wavelength set format is given by:

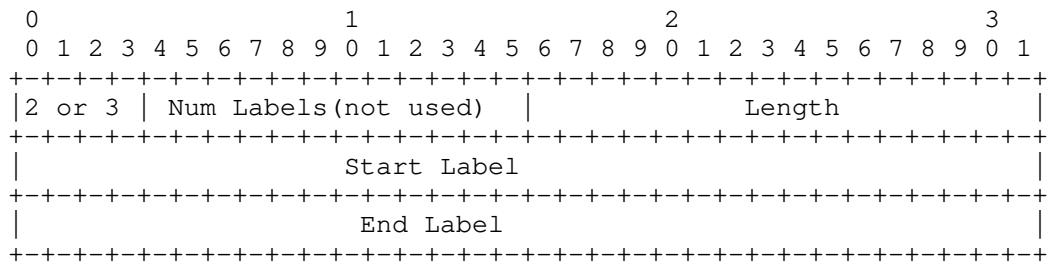


Where:

Num Labels is not used in this particular format since the Length parameter is sufficient to determine the number of labels in the list.

3.2.2. Inclusive/Exclusive Label Ranges

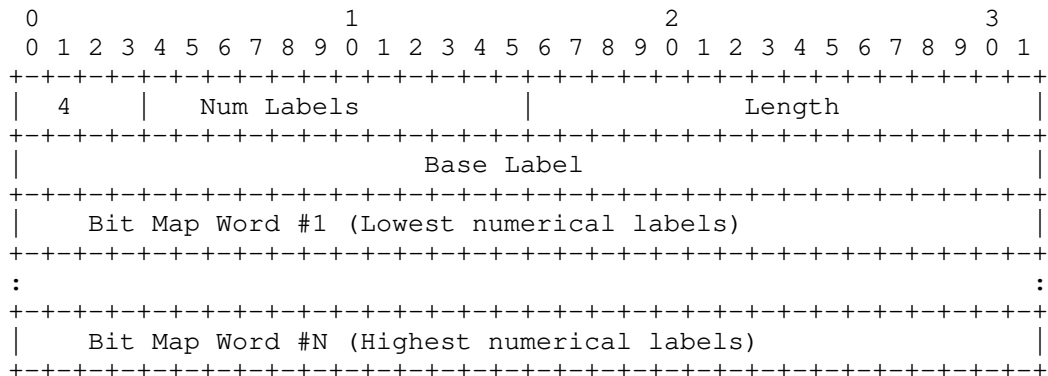
In the case of inclusive/exclusive ranges the label set format is given by:



Note that the start and end label must in some sense "compatible" in the technology being used.

3.2.3. Bitmap Label Set

In the case of Action = 4, the bitmap the label set format is given by:

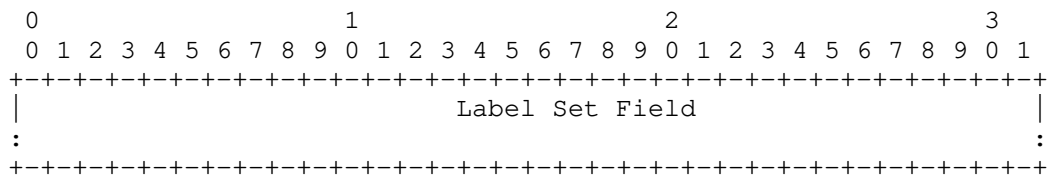


Where Num Labels in this case tells us the number of labels represented by the bit map. Each bit in the bit map represents a particular label with a value of 1/0 indicating whether the label is in the set or not. Bit position zero represents the lowest label and corresponds to the base label, while each 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.

3.3. 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.

3.4. 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:

```

      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
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|                               Label Set Field                               |
+-----+-----+-----+-----+-----+-----+-----+-----+
|                               :                               |
+-----+-----+-----+-----+-----+-----+-----+-----+

```

3.5. 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:

```

      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 | MatrixID | Reserved |
+-----+-----+-----+-----+-----+-----+-----+-----+
|                               Link Set A #1                               |
|                               :                               |
+-----+-----+-----+-----+-----+-----+-----+-----+
|                               Link Set B #1                               |
|                               :                               |
+-----+-----+-----+-----+-----+-----+-----+-----+
|                               Additional Link set pairs as needed         |
|                               to specify connectivity                     |
+-----+-----+-----+-----+-----+-----+-----+-----+

```

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.

3.6. 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.

```

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
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|   MatrixID   | RestrictionType |   Reserved/Parameter   |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
| Additional Restriction Parameters per RestrictionType |
:                                                         :
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

```

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)

3.6.1. SIMPLE_LABEL

In the case of the SIMPLE_LABEL the GeneralPortRestrictions (or MatrixSpecificRestrictions) format is given by:

```

      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
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
| MatrixID      | RstType = 0  |           Reserved           |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|                                     Label Set Field          |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

```

In this case the accompanying label set indicates the labels permitted on the port.

3.6.2. CHANNEL_COUNT

In the case of the CHANNEL_COUNT the format is given by:

```

      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
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
| MatrixID      | RstType = 1  |           MaxNumChannels           |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

```

In this case the accompanying MaxNumChannels indicates the maximum number of channels (labels) that can be simultaneously used on the port/matrix.

3.6.3. LABEL_RANGE1

In the case of the LABEL_RANGE1 the GeneralPortRestrictions (or MatrixSpecificRestrictions) format is given by:

```

      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
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
| MatrixID      | RstType = 2  |           MaxLabelRange           |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|                                     Label Set Field          |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

```

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.

3.6.4. SIMPLE_LABEL & CHANNEL_COUNT

In the case of the SIMPLE_LABEL & CHANNEL_COUNT the format is given by:

```

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
+-----+-----+-----+-----+-----+-----+-----+-----+
| MatrixID      | RstType = 3  |           MaxNumChannels          |
+-----+-----+-----+-----+-----+-----+-----+-----+
|                                     Label Set Field                 |
+-----+-----+-----+-----+-----+-----+-----+-----+

```

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.

3.6.5. Link Label Exclusivity

In the case of the SIMPLE_LABEL & CHANNEL_COUNT the format is given by:

```

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
+-----+-----+-----+-----+-----+-----+-----+-----+
| MatrixID      | RstType = 4  |           Reserved              |
+-----+-----+-----+-----+-----+-----+-----+-----+
|                                     Link Set Field                 |
+-----+-----+-----+-----+-----+-----+-----+-----+

```

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.

4. 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.

5. IANA Considerations

TBD. Once our approach is finalized we may need identifiers for the various TLVs and sub-TLVs.

6. 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	2	3	4	5	6	7	8	9
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 0										0 1 1 0 0 0 0 0 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)																													

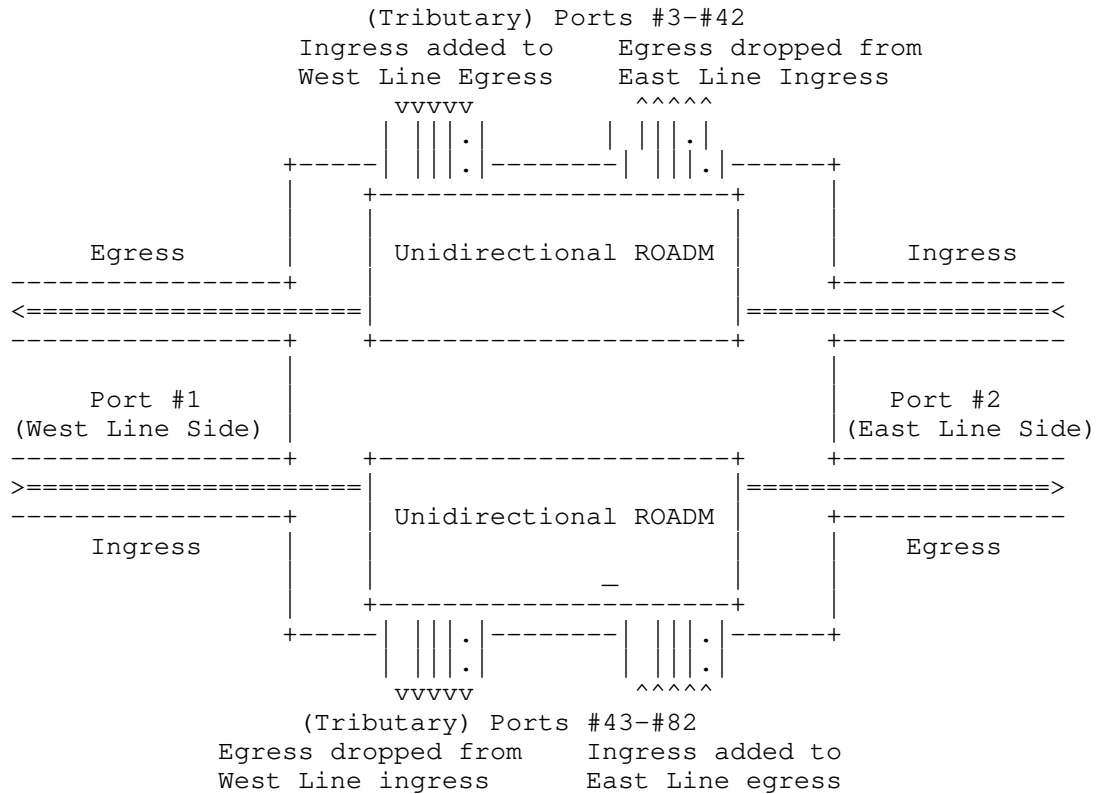
To encode this same set as an inclusive list we would have:

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	2	3	4	5	6	7	8	9
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.



Referring to the figure we see that the ingress direction of ports #3-#42 (add ports) can only connect to the egress on port #1. While the ingress side of port #2 (line side) can only connect to the egress on ports #3-#42 (drop) and to the egress on port #1 (pass through). Similarly, the ingress direction of ports #43-#82 can only connect to the egress on port #2 (line). While the ingress direction of port #1 can only connect to the egress on ports #43-#82 (drop) or port #2 (pass through). We can now represent this potential connectivity matrix as follows. This representation uses only 30 32-bit words.


```

      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
+-----+-----+-----+-----+-----+-----+-----+-----+

```

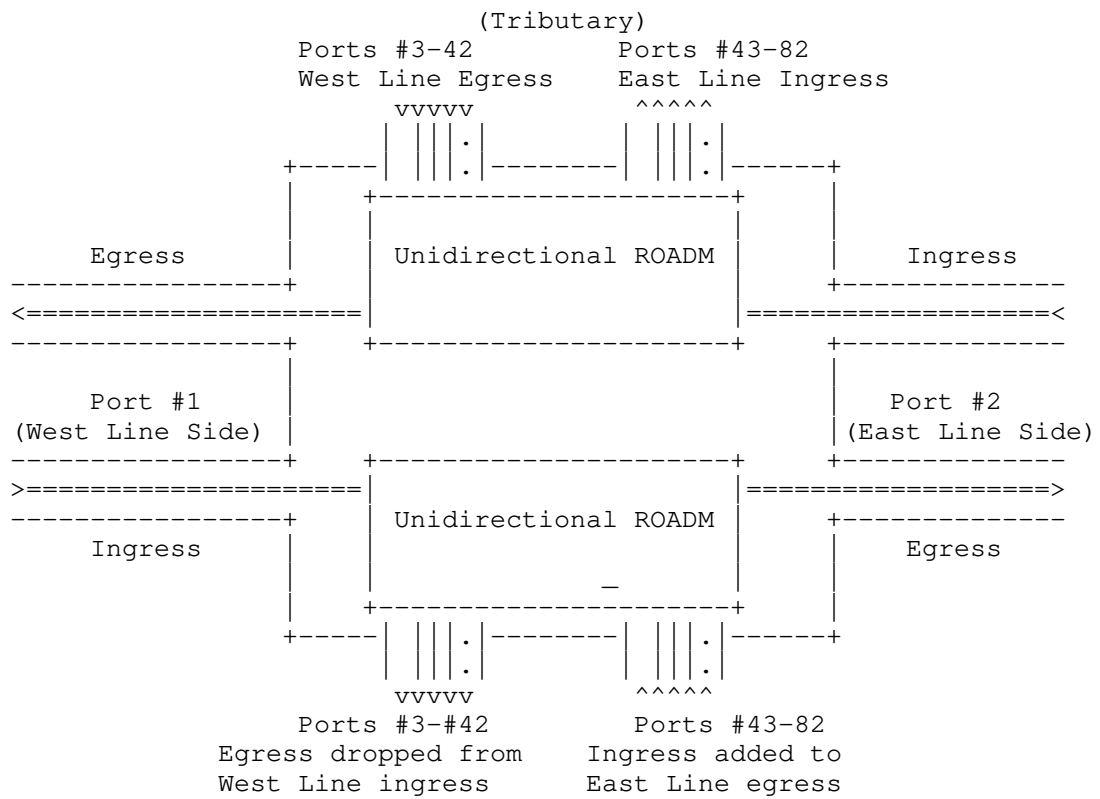
```

|                                     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
+-----+-----+-----+-----+-----+-----+-----+-----+

```

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.
- [RFC2863] McCloghrie, K. and F. Kastenholtz, "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.

7.2. 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.

8. 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
October 21, 2010

Expires: April 21, 2011

Framework for GMPLS and PCE Control of
G.709 Optical Transport Networks

draft-ietf-ccamp-gmpls-g709-framework-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 April 21, 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.....	15
5.4. Implications for Link Management Protocol (LMP).....	18
5.4.1. Correlating the Granularity of the TS.....	18
5.4.2. Correlating the Supported LO ODU Signal Types.....	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.....	20
10.1. Normative References.....	20
10.2. Informative References.....	21
11. Authors' Addresses.....	22
12. Contributors.....	23
APPENDIX A: ODU connection examples.....	24

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 ODUj (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, ODU2e, 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].

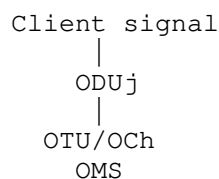


Figure 1 - Basic OTN signal hierarchy

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	client signal bit rate tolerance, with a maximum of +/-100 ppm
ODUflex for GFP-F Mapped client signal	+/- 20 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 +/- 20ppm, 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 OPU. Since all of these signals have a 1:1:1 relationship, we only refer to the OTU for clarity. The ODUj's are mapped into the TS of the OTUk. 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 [WSO-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 ODUj connections independent of the value of j. Note that when the bit rate of ODUj is less than the server bit rate, ODUj 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 TSes available, and nodes that are ODXCs. Lo ODU connections can be setup based on the network topology.

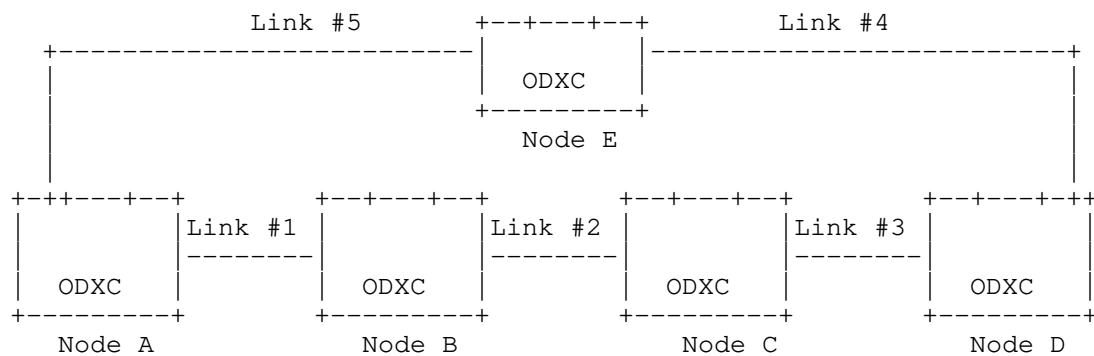


Figure 2 - Example Topology for LO ODU connection management

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.

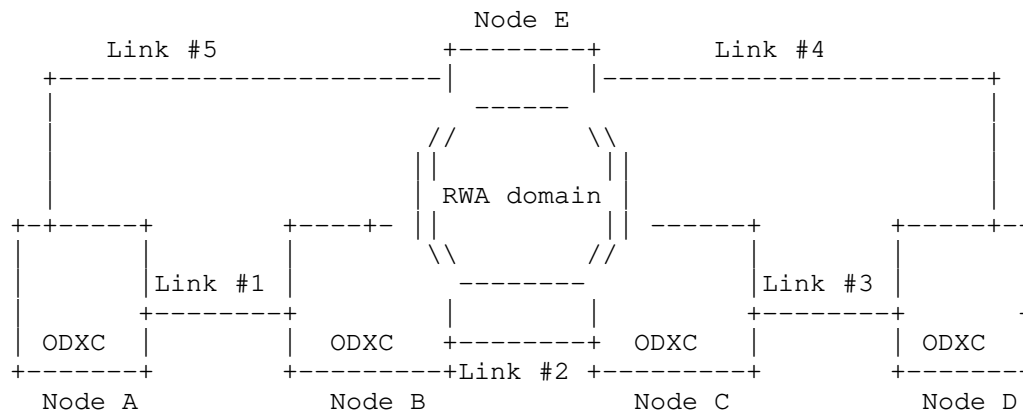


Figure 3 - RWA Hidden Topology for LO ODU connection management

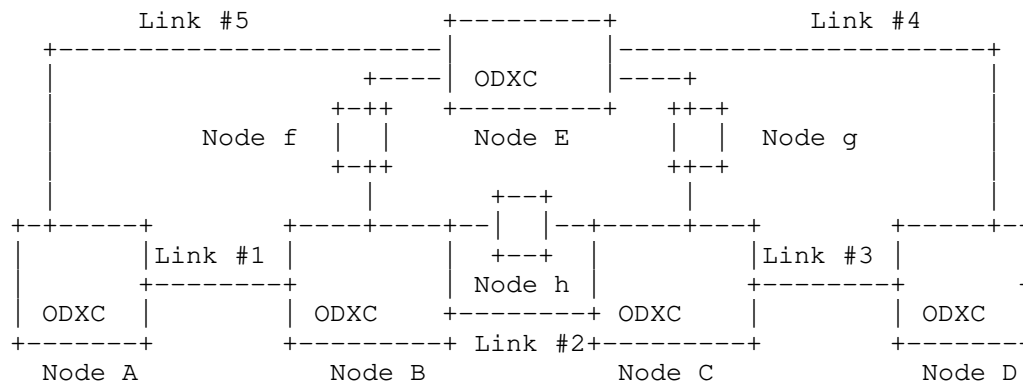


Figure 4 - RWA Visible Topology for LO ODUj connection management

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 to 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 [RFC4206bis] 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], [RFC4206bis] 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.

For the ODU layers, in order 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), it may be needed to consider introducing multi-stage multiplexing capability in specific network transition scenarios. One method for enabling multi-stage multiplexing 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 ODU_j connection can be nested into another ODU_k connection, which forms the LSP hierarchy in ODU layer. Here, [RFC4206], [RFC4206bis] and [MLN-EXT] (including related modifications, if needed) are relevant to connection set up.

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 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 ODUj 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 for carrying the TS granularity that the interface can support

One type of ODU_j can be multiplexed to an OTU_k using different TS granularity. For example, ODU₁ can be multiplexed into ODU₂ with either 2.5Gbps TS granularity or 1.25G TS granularity. The routing protocol should be capable of carrying the TS granularity supported by the ODU interface.

- 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 TSs when multiplexed into different HO ODUs. For example, for ODU_{2e}, 9 TSs are needed when multiplexed into an ODU₃, while only 8 TSs 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 link multiplexing capacity and link rate capacity

When a network operator receives a request for a particular ODU connection service, the operator governs the manner in which the request is fulfilled in their network. Considerations include deployed network infrastructure capabilities, associated policies (e.g., at what link fill threshold should a particular higher-rate ODUk be utilized), etc. Thus, for example, an ODU2 connection service request could be supported by: OTU2 links (here the connection service rate is the same as the link rate), a combination of OTU2 and OTU3 links, OTU3 links, etc.

Therefore, to allow the required flexibility, the routing protocol should clearly distinguish the capacity that is multiplexed in an ODUk that in turn is adapted in an OTUk from the ODUk capacity that is switched in matrix and directly adapted in an OTUk without further multiplexing.

- 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 multi-stage multiplexing 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.

5.4.1. 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 that both ends of the link advertise consistent capabilities (for routing) and ensures that viable connections are established.

5.4.2. Correlating the Supported LO ODU Signal Types

Many new ODU signal types have been introduced [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. If one end of a link can not support a certain LO ODU signal type, the link cannot be selected to carry such type of LO ODU connection.

Therefore, it is necessary for the two ends of an HO ODU link to correlate which types of LO ODU can be supported by the link. After correlating, the capability information can be flooded by IGP, so that the correct path for an ODU connection can be calculated.

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.

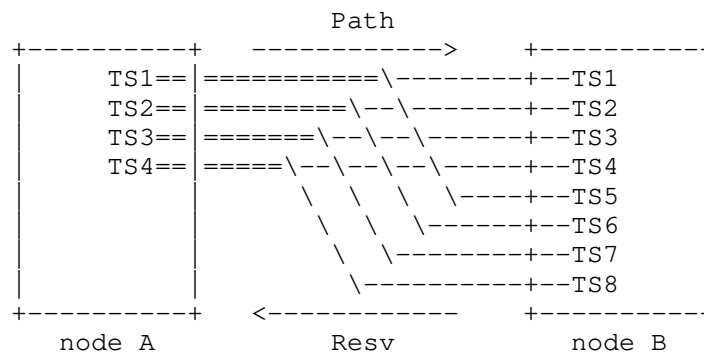


Figure 5 - Interworking between 1.25Gbps TS and 2.5Gbps TS

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.
- [RFC4206bis] K. Shiimoto, A. Farrel, "Procedures for Dynamically Signaled Hierarchical Label Switched Paths", draft-ietf-ccamp-lsp-hierarchy-bis-08.txt, February 2010.
- [MLN-EXT] Dimitri Papadimitriou et al, "Generalized Multi-Protocol Label Switching (GMPLS) Protocol Extensions for Multi-Layer and Multi-Region Networks (MLN/MRN)", draft-ietf-ccamp-gmpls-mln-extensions-12.txt, February 21, 2010.
- [RFC5440] JP. Vasseur, JL. Le Roux, Ed., " Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, March 2009.
- [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.
- [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: 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.

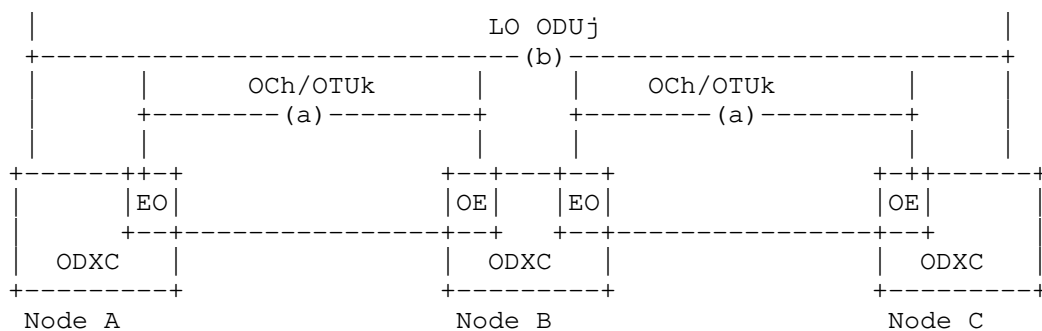


Figure 6 - Connection of LO ODU_j (1)

In the case of LO ODU_j multiplexing into HO ODU_k, Figure 7 gives an example of this kind of LO ODU connection.

In Figure 7, OCh, OTU_k, HO ODU_k are associated with each other. The LO ODU_j 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 ODU_k and OTU_k are similar to OCh. LO ODU_j and OCh/OTU_k/HO ODU_k have client/server relationships. When a LO ODU connection is setup, it will be using the existing HO ODU_k (/OTU_k/OCh) connections which have been set up. Those HO ODU_k 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.

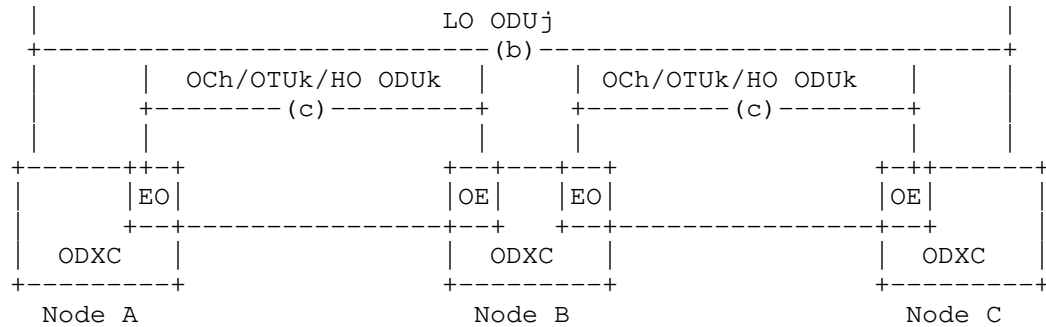


Figure 7 - Connection of LO ODUj (2)

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: April 2011

October 11, 2010

Behavior Negotiation in The Link Management Protocol

draft-ietf-ccamp-lmp-behavior-negotiation-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 April 11, 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 negotiated capabilities 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. Backwards Compatibility	5
4. Security Considerations	5
5. IANA Considerations	6
5.1. New LMP Class Type	6
5.2. New Capabilities Registry	6
6. Contributors	7
7. Acknowledgments	7
8. References	7
8.1. Normative References	7
8.2. Informative References	8
9. Authors' Addresses	8

1. Introduction

The Link Management Protocol (LMP) [RFC4204] is being successfully deployed in Generalized Multiprotocol Label Switching (GMPLS)-controlled networks. New LMP behaviors and protocol extensions are being introduced in a number of IETF documents.

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. There is no such procedure 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:

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
Length								B S D C O								Reserved															

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].

O: 1 bit

This bit indicates support for the TEST behavior of OTN technology-specific defined in [LMP TEST].

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 ConfigAck message MUST be sent with a <CONFIG> object, the BEHAVIOR_CONFIG C-Type representing the supported LMP behaviors. An LSR receiving such a ConfigAck 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.

3. Backwards Compatibility

An LSR that receives a Config message containing a <CONFIG> object with a C-Type that it does not recognize MUST respond with a ConfigAck 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 ConfigAck message.

It's not explicitly stated in [RFC4204] that a Config Message could include multiple <CONFIG> objects. But with new CONFIG C-Types are defined, multiple <CONFIG> objects (each with a different Class Type) MAY be present on a Config message in which case all of the objects MUST be processed.

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]
4	O	OTN TEST behavior	[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: April 28, 2011

A. Takacs
Ericsson
D. Fedyk
Alcatel-Lucent
J. He
Huawei
October 25, 2010

GMPLS RSVP-TE extensions for OAM Configuration
draft-ietf-ccamp-oam-configuration-fwk-04

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 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	4
2. Requirements	6
3. RSVP-TE based OAM Configuration	8
3.1. Establishment of OAM Entities and Functions	8
3.2. Adjustment of OAM Parameters	10
3.3. Deleting OAM Entities	10
4. RSVP-TE Extensions	12
4.1. LSP Attributes Flags	12
4.2. OAM Configuration TLV	13
4.2.1. OAM Function Flags Sub-TLV	14
4.2.2. Technology Specific sub-TLVs	14
4.3. Administrative Status Information	15
4.4. Handling OAM Configuration Errors	15
4.5. Considerations on Point-to-Multipoint OAM Configuration	16
5. IANA Considerations	18
6. Security Considerations	19
7. Acknowledgements	20
8. References	21
8.1. Normative References	21
8.2. Informative References	21
Authors' Addresses	23

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 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 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 will 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 [GMPLS-PBBTE]) work is extending the GMPLS control plane to support the establishment of point-to-point 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) it may be required that measurement points agree on a common probing rate to avoid measurement problems.
- 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 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" 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.

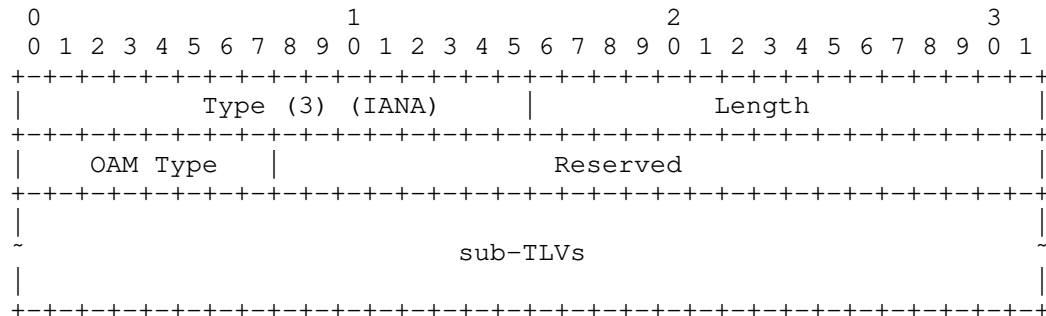
One bit (10 IANA to assign): "OAM MEP entities desired" is allocated in the LSP Attributes Flags TLV. If the "OAM MEP entities desired" bit is set it is indicating that the establishment of OAM MEP entities are required at the endpoints of the signaled LSP. If the establishment of MEPs is not supported 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 (11 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".

4.2. OAM Configuration TLV

This TLV provides information about which OAM technology/method should be used and carries sub-TLVs for any additional OAM configuration information. The OAM Configuration TLV may be carried in the LSP_ATTRIBUTES or LSP_REQUIRED_ATTRIBUTES object in Path and Resv messages.



Type: indicates a new type: the OAM Configuration TLV (3) (IANA to assign).

OAM Type: specifies the technology specific OAM method. If the requested OAM method is not supported an error must be generated: "OAM Problem/Unsupported OAM Type".

OAM Type	Description
-----	-----
0-255	Reserved

This document defines no types. IANA is requested to maintain the values in a new "RSVP-TE OAM Configuration Registry".

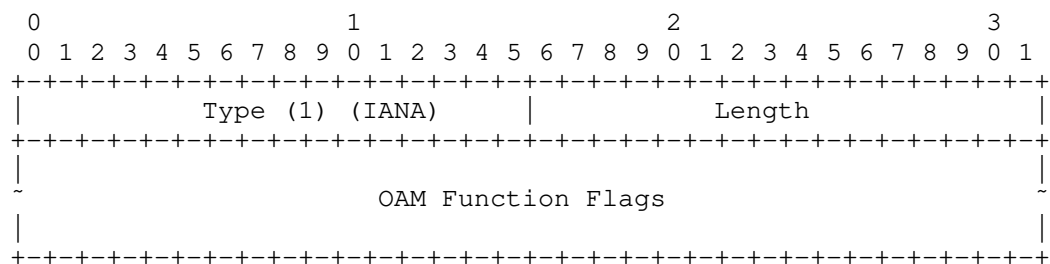
The receiving node based on the OAM Type will check if a corresponding technology specific OAM configuration sub-TLV is included. If the included technology specific OAM configuration sub-TLV is different than what is specified in the OAM Type an error must be generated: "OAM Problem/OAM Type Mismatch".

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 one "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 as shown in the figure. 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 SHOULD 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.

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. In addition to the Reflect (R) bit, 7 bits are currently occupied (assigned by IANA or temporarily blocked by work in progress Internet drafts). As the 24th and 25th bits (IANA to assign) this draft introduces 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

- [GMPLS-PBBTE] "Generalized Multiprotocol Label Switching (GMPLS) control of Ethernet Provider Backbone Traffic Engineering (PBB-TE)", Internet Draft, work in progress.
- [IEEE-CFM] "IEEE 802.1ag, 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.

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

CCAMP
Internet-Draft
Intended status: Standards Track
Expires: April 21, 2011

F. Le Faucheur
A. Narayanan
S. Dhesikan
Cisco
October 18, 2010

RSVP Resource Sharing Remote Identification Association
draft-ietf-ccamp-rsvp-resource-sharing-00.txt

Abstract

The RSVP ASSOCIATION object allows to create association across RSVP path states or across Resv states. Two association types are currently defined: recovery and resource sharing. This document defines a new association type called "Resource Sharing Remote Identification". It can be used by the sender to convey to the receiver the information that can then be used by the receiver to identify a downstream initiated resource sharing association.

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 21, 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.

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.

Table of Contents

1. Introduction	3
1.1. Conventions Used in This Document	4
2. Resource Sharing Remote Identification Association	5
3. Security Considerations	7
4. IANA Considerations	8
4.1. Resource Sharing Remote Identification Association Type	8
5. Acknowledgments	9
6. References	10
6.1. Normative References	10
6.2. Informative References	10
Authors' Addresses	11

1. Introduction

The notion of association as well as the corresponding RSVP ASSOCIATION object are defined in [RFC4872] and [RFC4873] 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 [I-D.ietf-ccamp-assoc-info] defines how the ASSOCIATION object can be more generally applied. [I-D.ietf-ccamp-assoc-info] also reviews how the association is to be provided in the context of GMPLS recovery.

[RFC4872] defines the IPv4 ASSOCIATION object and the IPv6 ASSOCIATION object. In addition, [I-D.ietf-ccamp-assoc-info] defines the Extended IPv4 ASSOCIATION object and the Extended IPv6 ASSOCIATION object. These four forms of the ASSOCIATION object contain an Association Type field that indicates the type of association being identified by the ASSOCIATION object. For example, Figure 1 illustrates the format of the IPv4 ASSOCIATION object.

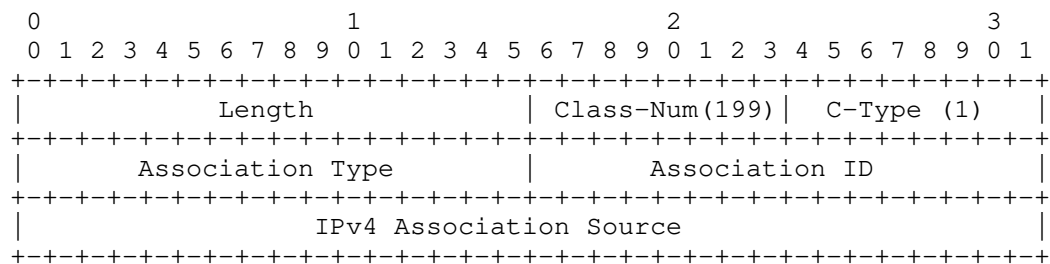


Figure 1: IPv4 ASSOCIATION object format

[RFC4872] and [RFC4873] define two association types: 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 section 4.3.1 of [I-D.ietf-ccamp-assoc-info]. For non-recovery Usage (for example for resource sharing), [I-D.ietf-ccamp-assoc-info] defines, in section 4, the notion of upstream initiated association and downstream 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. 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.

This document defines a new association type called "Resource Sharing Remote Identification".

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. Resource Sharing Remote Identification Association

We define here a new association type called the Resource Sharing Remote Identification.

The Resource Sharing Remote Identification association is only defined for use in upstream initiated association. Thus it can only appear in ASSOCIATION objects signaled in Path messages.

The Resource Sharing Remote Identification association can be used by the sender to convey to the receiver (inside the Association Source and Association ID fields), information that can then be used by the receiver to identify an upstream initiated resource sharing association. This is useful in upstream initiated resource sharing applications where the identification of the resource sharing association is not known a priori by the receiver, and instead is known by the sender (for example because the sender is in a better position to assign the association identification necessary to implement the desired resource sharing across RSVP sessions).

[I-D.ietf-ccamp-assoc-info] discusses the rules associated with the processing of ASSOCIATION objects in RSVP messages. In addition to generic rules applicable to all association types, a given association type may define type-specific processing rules. The following type-specific association rule is defined for the Resource Sharing Remote Identification association type:

- o The Resource Sharing Remote Identification association does not create any association across Path states.

This is because the purpose of signaling an Resource Sharing Remote Identification association in the downstream direction is purely to convey identification information from the sender to the receiver that can be used by the receiver to establish an upstream initiated resource sharing association.

Any implementation of the present specification MUST support the Resource Sharing Remote Identification association.

On receipt of an ASSOCIATION object whose association type is Resource Sharing Remote Identification, the receiver MAY use the association identification information contained in the received ASSOCIATION object as the association identification information in an upstream initiated resource sharing association.

On receipt of an ASSOCIATION object whose association type is Resource Sharing Remote Identification, an RSVP receiver proxy as defined in [RFC5945], SHOULD initiate an upstream initiated Resource

Sharing association whose association identification information is copied from the received ASSOCIATION object. This behavior MAY be overridden by local policy on the receiver proxy.

3. Security Considerations

TBD.

4. IANA Considerations

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

4.1. Resource Sharing Remote Identification Association Type

This document defines, in Section 2, a new association type. Thus, IANA is requested to allocate the following entry in the Association Type registry found at <http://www.iana.org/assignments/gmpls-sig-parameters/> :

3 Resource Sharing Remote Identification (I) [this-document]

There are no other IANA considerations introduced by this document.

5. Acknowledgments

We thank Lou Berger for his guidance in this work and in particular with respect to aligning it with the related CCAMP work on Association .

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.
- [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.

6.2. Informative References

- [RFC5945] Le Faucheur, F., Manner, J., Wing, D., and A. Guillou, "Resource Reservation Protocol (RSVP) Proxy Approaches", RFC 5945, October 2010.

Authors' Addresses

Francois Le Faucheur
Cisco Systems
Greenside, 400 Avenue de Roumanille
Sophia Antipolis 06410
France

Phone: +33 4 97 23 26 19
Email: flefauch@cisco.com

Ashok Narayanan
Cisco Systems
300 Beaver Brook Road
Boxborough, MAS 01719
United States

Email: ashokn@cisco.com

Subha Dhesikan
Cisco Systems

Phone:
Email: sdhesika@cisco.com

Network Working Group
Internet Draft
Intended status: Informational
Expires: March 2011

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

September 3, 2010

Routing and Wavelength Assignment Information Model for Wavelength
Switched Optical Networks

draft-ietf-ccamp-rwa-info-09.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 March 3, 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

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 lightpath 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
2. Terminology.....	5
3. Routing and Wavelength Assignment Information Model.....	6
3.1. Dynamic and Relatively Static Information.....	6
4. Node Information (General).....	7
4.1. Connectivity Matrix.....	7
4.2. Shared Risk Node Group.....	8
5. Node Information (WSON specific).....	8
5.1. Resource Accessibility/Availability.....	9
5.2. Resource Signal Constraints and Processing Capabilities..	13

5.3. Compatibility and Capability Details.....	14
5.3.1. Shared Ingress or Egress Indication.....	14
5.3.2. Modulation Type List.....	14
5.3.3. FEC Type List.....	14
5.3.4. Bit Rate Range List.....	14
5.3.5. Acceptable Client Signal List.....	15
5.3.6. Processing Capability List.....	15
6. Link Information (General).....	15
6.1. Administrative Group.....	16
6.2. Interface Switching Capability Descriptor.....	16
6.3. Link Protection Type (for this link).....	16
6.4. Shared Risk Link Group Information.....	16
6.5. Traffic Engineering Metric.....	16
6.6. Port Label (Wavelength) Restrictions.....	16
6.6.1. Port-Wavelength Exclusivity Example.....	18
7. Dynamic Components of the Information Model.....	19
7.1. Dynamic Link Information (General).....	20
7.2. Dynamic Node Information (WSON Specific).....	20
8. Security Considerations.....	20
9. IANA Considerations.....	21
10. Acknowledgments.....	21
11. References.....	22
11.1. Normative References.....	22
11.2. Informative References.....	23
12. Contributors.....	24
Author's Addresses.....	24
Intellectual Property Statement.....	25
Disclaimer of Validity.....	26

1. Introduction

The purpose of the following information model for WSONs is to facilitate constrained lightpath 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 lightpath 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. We refer to the information model applicable to

other technologies beyond WSON as "general" to distinguish from the "WSON-specific" model 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 we added signal compatibility modeling.

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.

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 (WSN): 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

We group the following WSON RWA information model 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 we use, where applicable, the reduced Backus-Naur form (RBNF) syntax of [RBNF] 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.

In this document we will separate, 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. We include these connectivity matrices with our node information the switched and fixed port wavelength constraints 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. We say this is a "conceptual" matrix

since this 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(i, j) ::= <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 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 we need to know the following:

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 we group identical processing resources such as regenerators or wavelength converters into "blocks". The accessibility to and from any resource within a block must be the same. The resource pool is composed of one or more blocks.

This leads to the following formal high level model:

```
<Node_Information> ::= <Node_ID> [<ConnectivityMatrix>...]
[<ResourcePool>]
```

Where

```
<ResourcePool> ::= <ResourceBlockInfo>...
[<ResourceBlockAccessibility>...] [<ResourceWaveConstraints>...]
[<RBPoolState>]
```

First we will address the accessibility of resource blocks then we will discuss their properties.

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 [WSO-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 we will also want to our model to include as a minimum 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 as shown schematically in Figure 1 and Figure 2. The difference between the two figures is that in Figure 1 we assume that each signal that can get to a resource block may do so, while in Figure 2 the access to the resource blocks is via a shared fiber which imposes its own wavelength collision constraint. In the representation of Figure 1 we can have more than one ingress to each resource block since each ingress represents a single wavelength signal, while in Figure 2 we show a single multiplexed WDM ingress, e.g., a fiber, to each block.

In this model we assume 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 we have a set of relatively static ingress port constraints for each resource. In addition, if the access to a resource block is via a shared fiber (Figure 2) this would impose a dynamic wavelength availability constraint on that shared fiber. We can model each resource block ingress port constraint via a static wavelength set mechanism and in the case of shared access to a block via another dynamic wavelength set mechanism.

Next we have a state vector $RA(j) = \{0, \dots, k\}$ which tells us 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, we have a set of static resource egress wavelength constraints and possibly dynamic shared egress fiber constraints. 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 λ . The dynamic constraints would be used in the case where a single shared fiber is used to egress the resource block (Figure 2).

Finally, we have a resource pool egress matrix $RE(p,k) = \{0,1\}$ depending on whether the output from resource block p can reach egress port k . Examples of this method being used to model wavelength converter pools for several switch architectures from the literature are given in reference [WC-Pool].

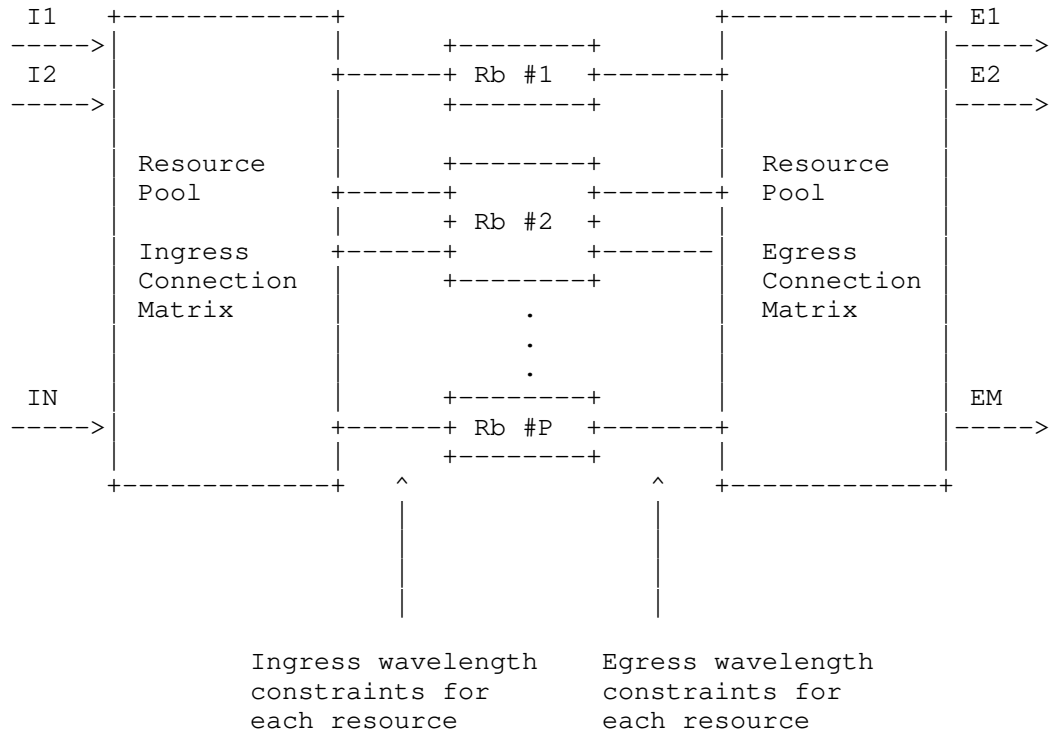


Figure 1 Schematic diagram of resource pool model.

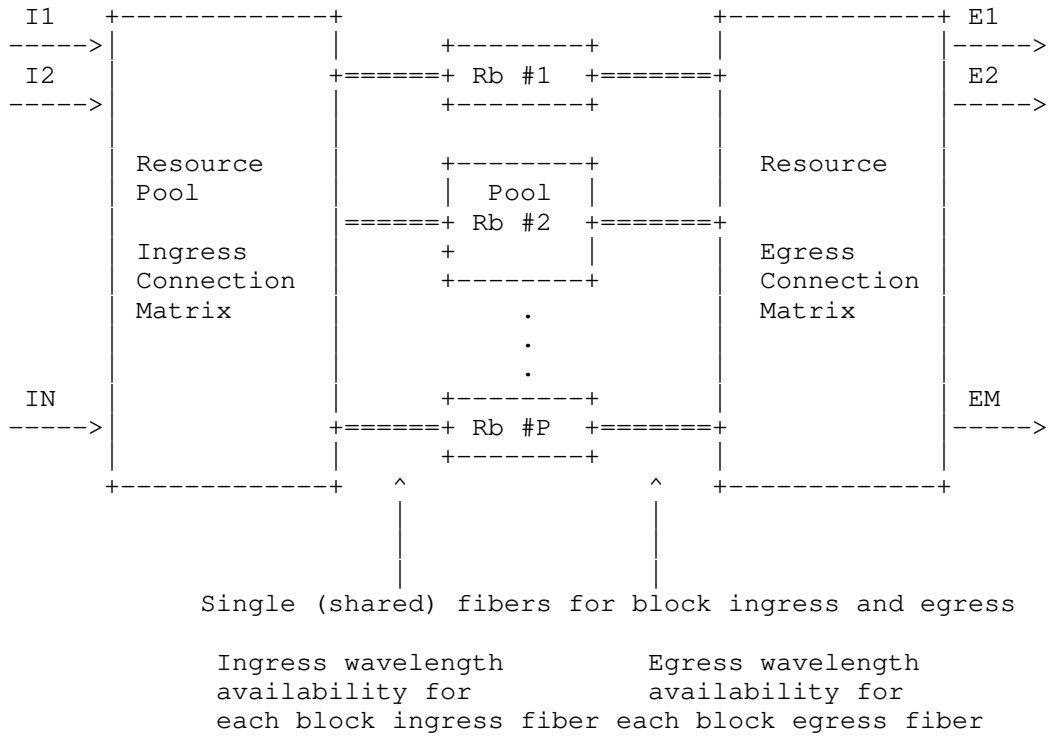


Figure 2 Schematic diagram of resource pool model with shared block accessibility.

Formally we can specify the model as:

```
<ResourceBlockAccessibility >::= <PoolIngressMatrix>
<PoolEgressMatrix>

[<ResourceWaveConstraints> ::= <IngressWaveConstraints>
<EgressWaveConstraints>

<ResourcePoolState>
::= (<ResourceBlockID><NumResourcesInUse><InAvailableWavelengths><OutA
vailableWavelengths>)...

```


Note that except for <ResourcePoolState> 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 [WSN-Frame] we can model the constraints on an electro-optical resource in terms of input constraints, processing capabilities, and output constraints:

```
<ResourceBlockInfo> ::=
([<ResourceSet>]<InputConstraints><ProcessingCapabilities><OutputCon-
straints>)*
```

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

We have defined ProcessingCapabilities 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 ROADM 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 we have 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 we see that egress ports 5-8 have a similar constraint due to the internal structure.

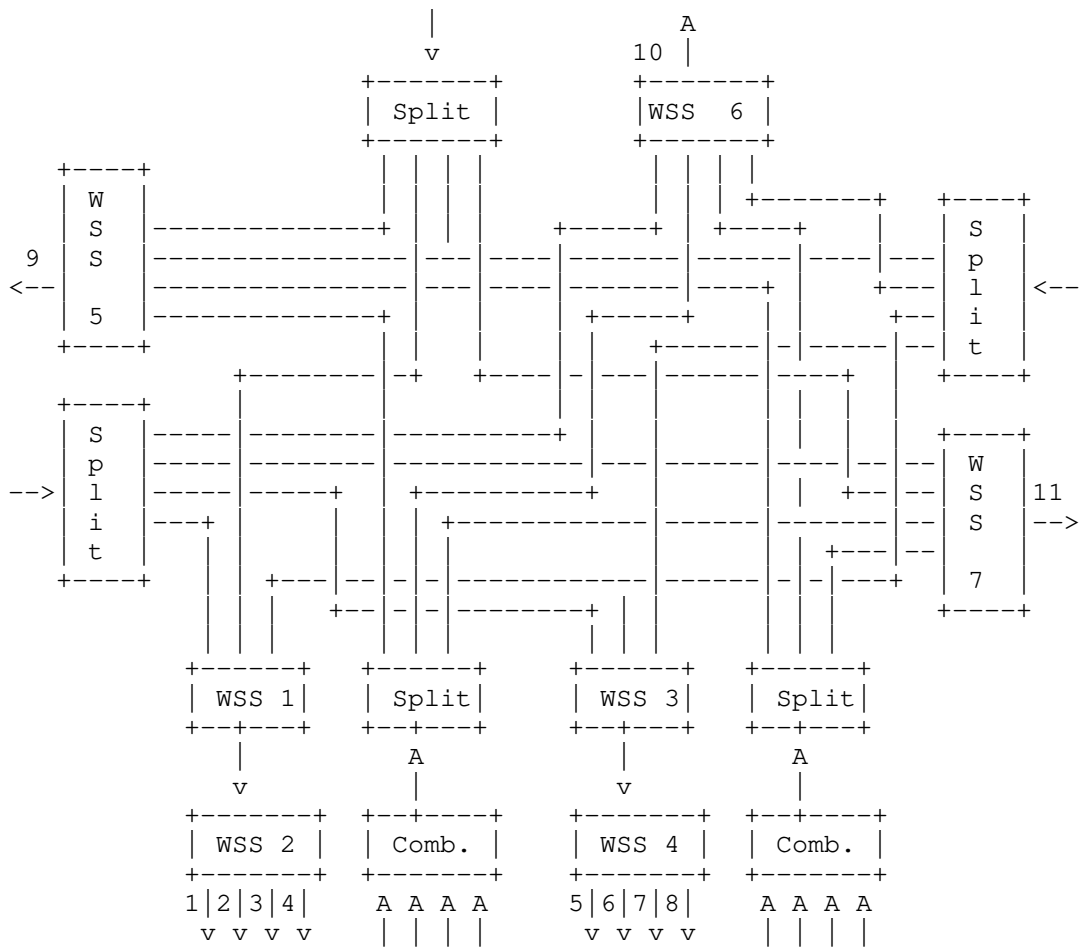


Figure 3 A ROADM composed from splitter, combiners, and WSSs.

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 we can isolate 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 we 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.

[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.

11.2. Informative References

[OFC08] P. Roorda and B. Collings, "Evolution to Colorless and Directionless ROADMs 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.

[WC-Pool] G. Bernstein, Y. Lee, "Modeling WDM Switching Systems including Wavelength Converters" to appear www.grotto-networking.com, 2008.

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: April 2011

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

October 13, 2010

Routing and Wavelength Assignment Information Encoding for
Wavelength Switched Optical Networks

draft-ietf-ccamp-rwa-wson-encode-06.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 April 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

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.....	3
1.1. Revision History.....	4
1.1.1. Changes from 00 draft.....	4
1.1.2. Changes from 01 draft.....	4
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
2. Terminology.....	5
3. WSON Encoding Usage Recommendations.....	6
3.1. WSON Node TLV.....	6
3.2. WSON Dynamic Node TLV.....	6
4. Resource Accessibility/Availability.....	7
4.1. Block Accessibility Sub-TLV.....	8
4.2. Wavelength Constraints Sub-TLV.....	10
4.3. Block Pool State Sub-TLV.....	10
4.4. Block Shared Access Wavelength Availability sub-TLV.....	12
5. Resource Properties Encoding.....	13
5.1. Resource Block Information Sub-TLV.....	13
5.2. Input Modulation Format List Sub-Sub-TLV.....	14
5.2.1. Modulation Format Field.....	15
5.3. Input FEC Type List Sub-Sub-TLV.....	16
5.3.1. FEC Type Field.....	17
5.4. Input Bit Range List Sub-Sub-TLV.....	19
5.4.1. Bit Range Field.....	19
5.5. Input Client Signal List Sub-Sub-TLV.....	20
5.6. Processing Capability List Sub-Sub-TLV.....	21
5.6.1. Processing Capabilities Field.....	21
5.7. Output Modulation Format List Sub-Sub-TLV.....	23
5.8. Output FEC Type List Sub-Sub-TLV.....	23
6. Security Considerations.....	23
7. IANA Considerations.....	24
8. Acknowledgments.....	24
APPENDIX A: Encoding Examples.....	25
A.1. Wavelength Converter Accessibility Sub-TLV.....	25
A.2. Wavelength Conversion Range Sub-TLV.....	26
A.3. An OEO Switch with DWDM Optics.....	27
9. References.....	31
9.1. Normative References.....	31
9.2. Informative References.....	31
10. Contributors.....	32
Authors' Addresses.....	33
Intellectual Property Statement.....	34
Disclaimer of Validity.....	34

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.

1.1.5. 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.

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. WSON Encoding Usage Recommendations

In this section we give recommendations of typical usage of the sub-TLVs and composite TLVs which are based on the high level information bundles of [WSON-Info].

3.1. WSON Node TLV

The WSON Node TLV would consist of the following list of sub-TLVs:

```
<Node_Info> ::= <Node_ID>[Other GMPLS sub-TLVs]
[<ResourcePool>][<RBPoolState>]
```

Where

```
<ResourcePool> ::= <ResourceBlockInfo>...
[<ResourceBlockAccessibility>...] [<ResourceWaveConstraints>...]
```

The encoding of structure and properties of a general resource pool utilizes a resource block info sub-TLV (<ResourceBlockInfo> in section 5.), an accessibility sub-TLV (<ResourceBlockAccessibility> in section 4.1.), and a resource pool wavelength constraint sub-TLV (<ResourceWaveConstraints> in section 4.2.).

3.2. WSON Dynamic Node TLV

If the protocol supports the separation of dynamic information from relatively static information then the wavelength converter pool state can be separated from the general Node TLV into a dynamic Node TLV as follows.

```
<NodeInfoDynamic> ::= <NodeID>
[<RBPoolState>][<BlockSharedAccessWavelengthAvailability>...]
```

Where the resource pool state sub-TLV <RBPoolState> is defined in section 4.3. Note that currently the only dynamic information modeled

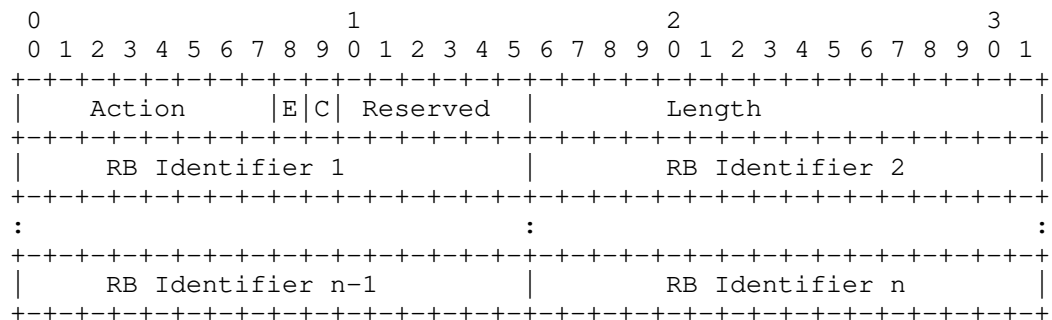
with a node is associated with the status of the wavelength converter pool.

4. Resource Accessibility/Availability

In this section we define the sub-TLVs for dealing with accessibility and availability of resource blocks. These include the ResourceBlockAccessibility, ResourceWaveConstraints, and RBPoolState sub-TLVs. All these sub-TLVs are concerned with sets of resources.

In a WSON node that includes resource blocks (RB) we will want to denote subsets these blocks 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.

4.1. Block 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 [WSO-Info].

The resource block accessibility sub-TLV is defined by:

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										Reserved																														
										Ingress Link Set Field A #1																														
:																																								:
										RB Set Field A #1																														
:																																								:
										Additional Link set and RB set pairs as needed to																														
:										specify PoolIngressMatrix																														:
										Egress Link Set Field B #1																														
:																																								:
										RB Set B Field #1 (for egress connectivity)																														
:																																								:
										Additional Link Set and RB set pairs as needed to																														
:										specify PoolEgressMatrix																														:

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)

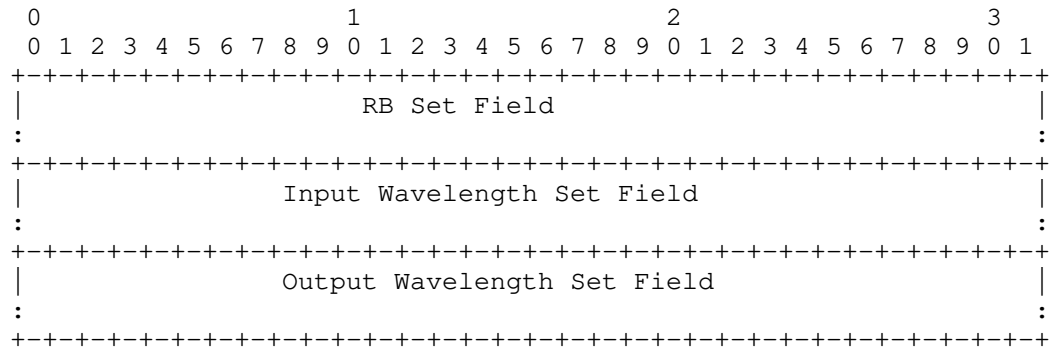
The Link Set Field is defined in [Gen-Encode].

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.

4.2. 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:

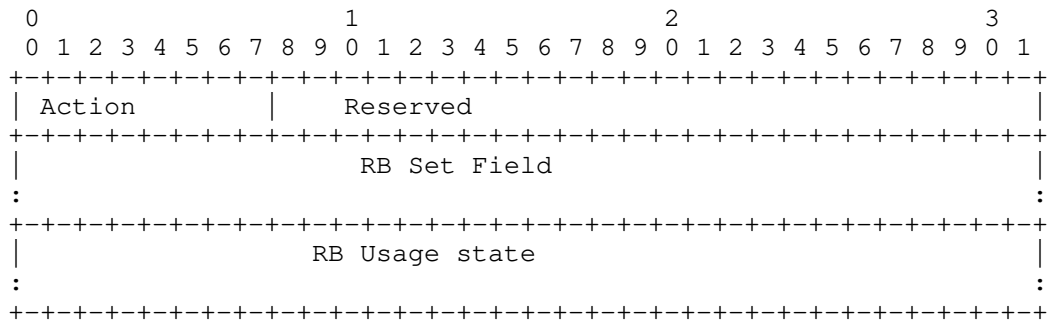
Indicates the wavelength input restrictions of the RBs in the corresponding RB set.

Output Wavelength Set Field:

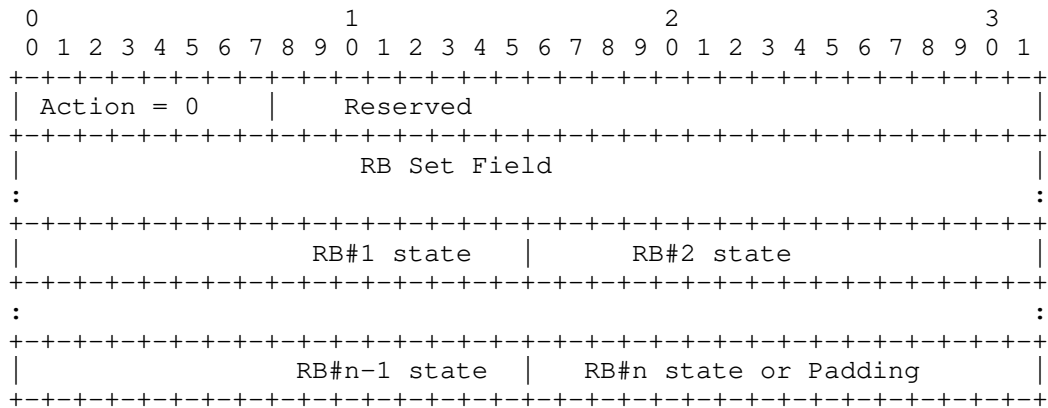
Indicates the wavelength output restrictions of RBs in the corresponding RB set.

4.3. Block Pool State Sub-TLV

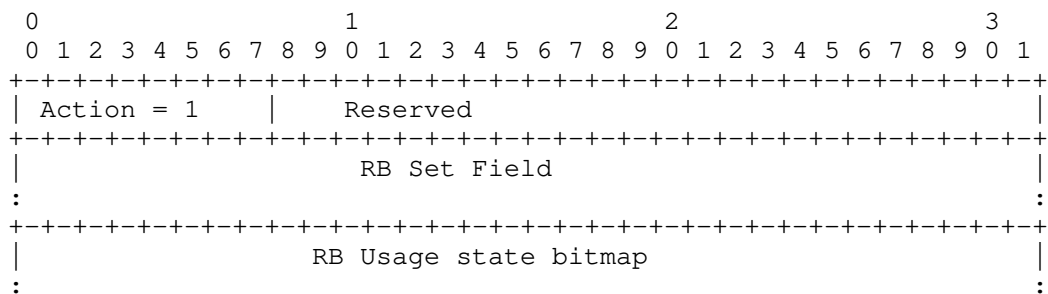
The usage state of a resource is encoded as either a list of 16 bit integer values or a bit map indicating whether a single resource is available or in use. This information can be relatively dynamic, i.e., can change when a connection is established or torn down.



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.



```

+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|               .....               |           Padding bits           |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

```

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

4.4. 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 fiber is needed to understand resource availability.

```

      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
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|   Resource Block ID   | I | E |           Reserved           |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|           Ingress Available Wavelength Set Field           |
:                                                               :
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|           Egress Available Wavelength Set Field           |
:                                                               :
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

```

Resource Block ID:

The 16 bit integer used to identify a particular resource block.

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).

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.

5. 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. In this section we primarily focus on the signal compatibility and processing properties of such a resource block, i.e., <ResourceBlockInfo> of section 3.1. The accessibility aspects of a resource in a shared pool, except for the shared access indicators, were encoded in the previous section.

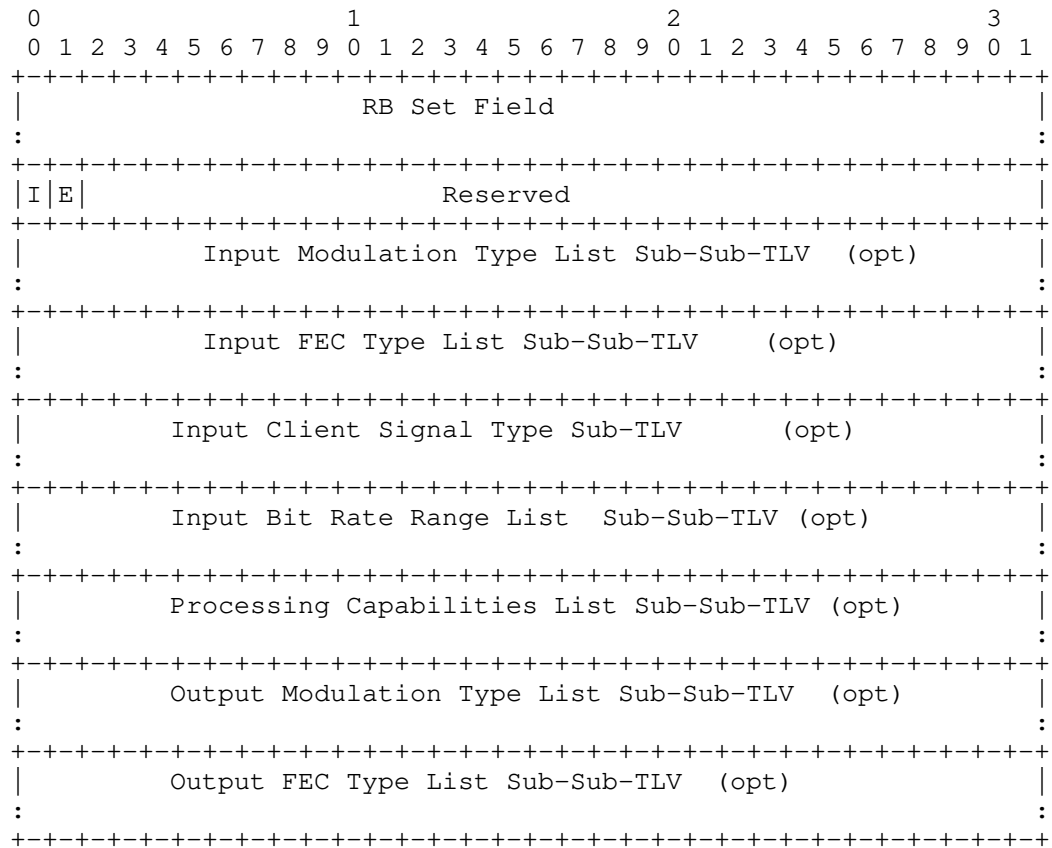
The fundamental properties of a resource block, such as a regenerator or wavelength converter, are:

- (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)

5.1. Resource Block Information Sub-TLV

Resource Block descriptor sub-TLVs are used to convey relatively static information about individual resource blocks including the resource block properties of section 3. and the number of resources in a block.

This sub-TLV has the following format:



Where I and E, the shared ingress/egress indicator, is set to 1 if the resource blocks identified in the RB set field utilized a shared fiber for ingress/egress access and set to 0 otherwise.

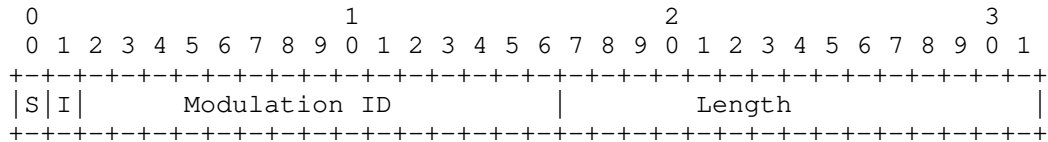
5.2. Input Modulation Format List Sub-Sub-TLV

This sub-TLV contains a list of acceptable input modulation formats.

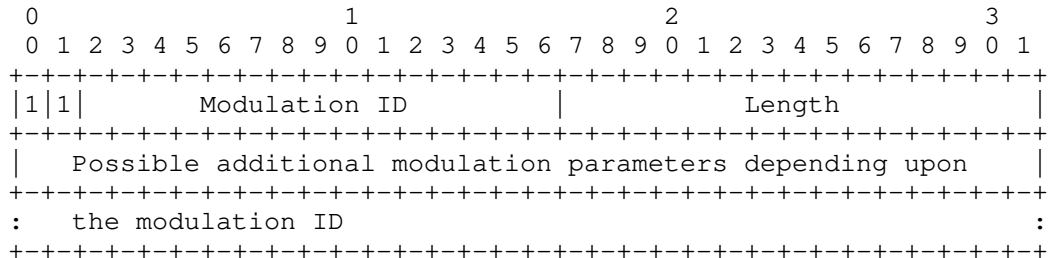
Type := Input Modulation Format List

Value:= A list of Modulation Format Fields

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.



The format for the standardized type for the input modulation is given by:



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:

```

      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|1|  Vendor Modulation ID   |               Length               |
+-----+-----+-----+-----+-----+-----+-----+-----+
|               Enterprise Number               |
+-----+-----+-----+-----+-----+-----+-----+-----+
:   Any vendor specific additional modulation parameters   :
+-----+-----+-----+-----+-----+-----+-----+-----+

```

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.

5.3. Input FEC Type List Sub-Sub-TLV

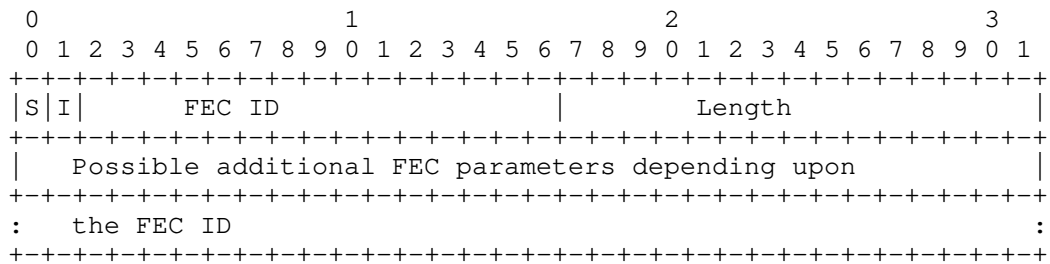
This sub-TLV contains a list of acceptable FEC types.

Type := Input FEC Type field List

Value:= A list of FEC type Fields

5.3.1. 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.



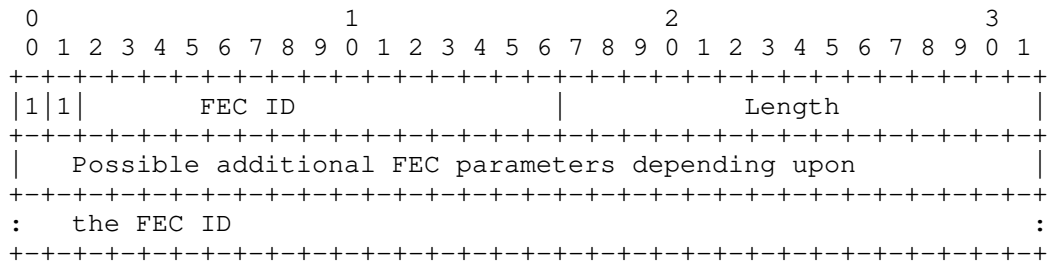
Where S bit set to 1 indicates a standardized FEC format and S bit set to 0 indicates a vendor specific FEC format. The length is the length in bytes of the entire FEC type field.

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:

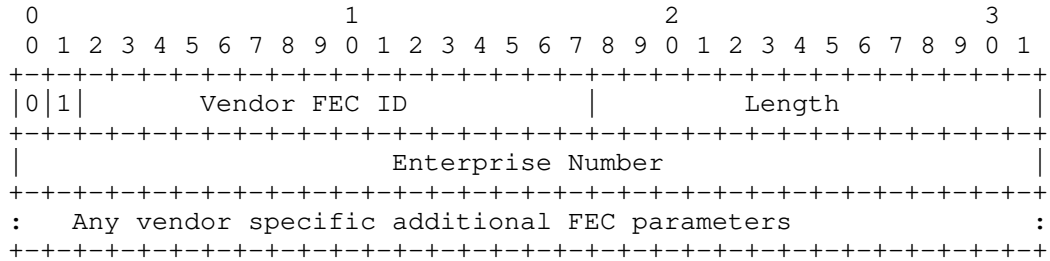


Takes on the following currently defined values for the standard FEC ID:

- | | |
|----|---|
| 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.

5.4. Input Bit Range List Sub-Sub-TLV

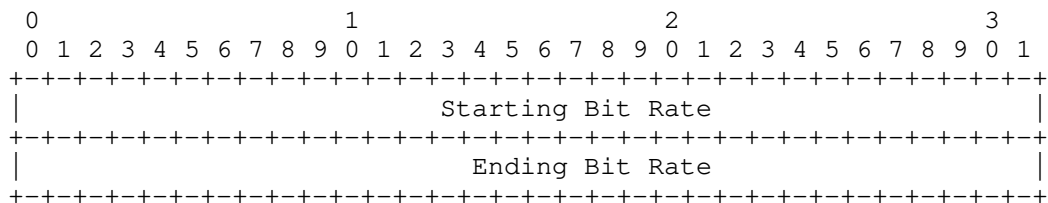
This sub-TLV contains a list of acceptable input bit rate ranges.

Type := Input Bit Range List

Value:= A list of Bit Range Fields

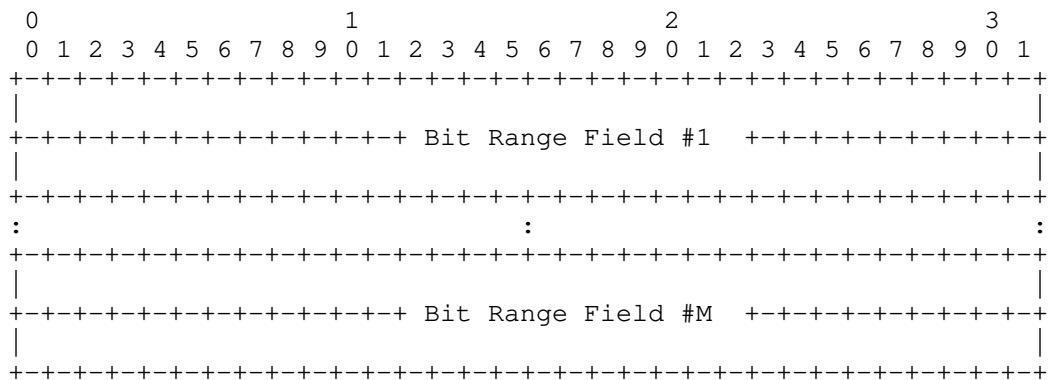
5.4.1. Bit Range Field

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:



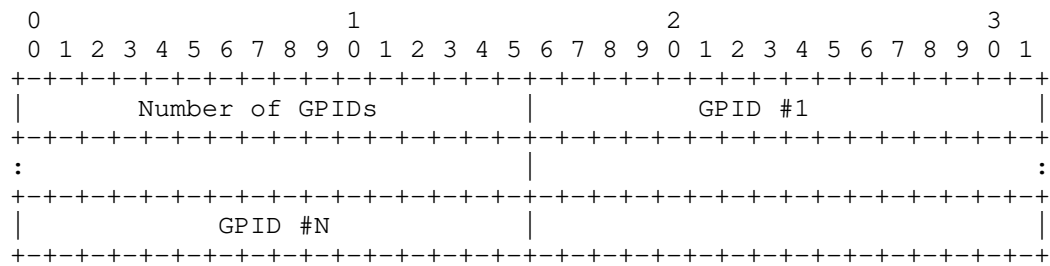
5.5. Input Client Signal List Sub-Sub-TLV

This 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.

5.6. Processing Capability List Sub-Sub-TLV

This 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.

5.6.1. Processing Capabilities Field

The processing capability field is then given by:

```

      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
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|           Processing Cap ID           |           Length           |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
| Possible additional capability parameters depending upon           |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
:   the processing ID   :
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

```

When the processing Cap ID is "number of resources" the format is simply:

```

      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
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|           Processing Cap ID           |           Length = 8           |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|           Number of resources per block           |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

```

When the processing Cap ID is "regeneration capability", the following additional capability parameters are provided in the sub-TLV:

```

      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
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|   T   |   C   |           Reserved           |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

```

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.

5.7. Output Modulation Format List Sub-Sub-TLV

This sub-TLV contains a list of available output modulation formats.

Type := Output Modulation Format List

Value:= A list of Modulation Format Fields

5.8. Output FEC Type List Sub-Sub-TLV

This sub-TLV contains a list of output FEC types.

Type := Output FEC Type field List

Value:= A list of FEC type Fields

6. 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.

7. IANA Considerations

TBD. Once our approach is finalized we may need identifiers for the various sub-sub-TLVs.

8. 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:

$$WI = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad WE = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

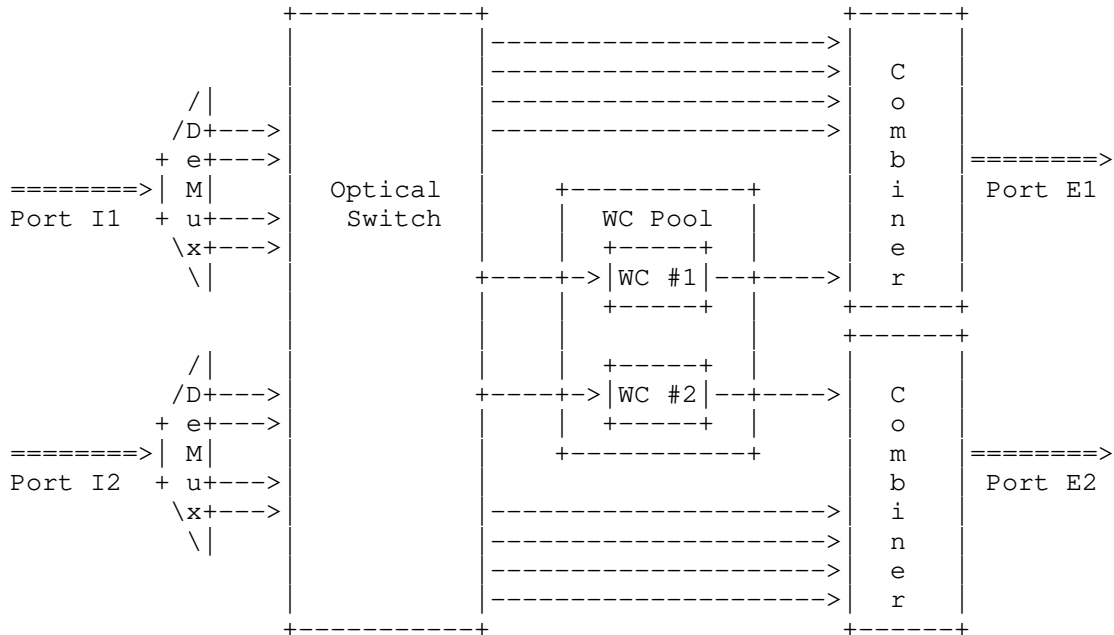


Figure 1 An optical switch featuring a shared per fiber wavelength converter pool architecture.

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      | 0 |   Reserved   |               Length = 8 |
+-----+-----+-----+-----+
|               | RB ID = #1 |               zero padding |
+-----+-----+-----+-----+
| Action=0      | 1 0 | 0 0 0 0 0 0 |               Length = 8 |
+-----+-----+-----+-----+
|               | Link Local Identifier = #1 |
+-----+-----+-----+-----+
|               | Note: WC2 can only connect to E2 |
+-----+-----+-----+-----+
| Action=0      | 0 |               |               Length = 8 |
+-----+-----+-----+-----+
|               | RB ID = #2 |               zero padding |
+-----+-----+-----+-----+
| Action=0      | 1 0 | 0 0 0 0 0 0 |               Length = 8 |
+-----+-----+-----+-----+
|               | Link Local Identifier = #2 |
+-----+-----+-----+-----+

```

A.2. Wavelength Conversion Range Sub-TLV

Example:

We give an example based on figure 1 about 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

In Figure 2 we show 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 we have the potential information:

```

<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 we will not need the <RBPoolState> element.

Hence our attention will be focused on the <ResourcePool> sub-TLV:

```

<ResourcePool> ::=
<ResourceBlockInfo>[<ResourceBlockAccessibility>...][<ResourceWaveCon
straints>...]

```

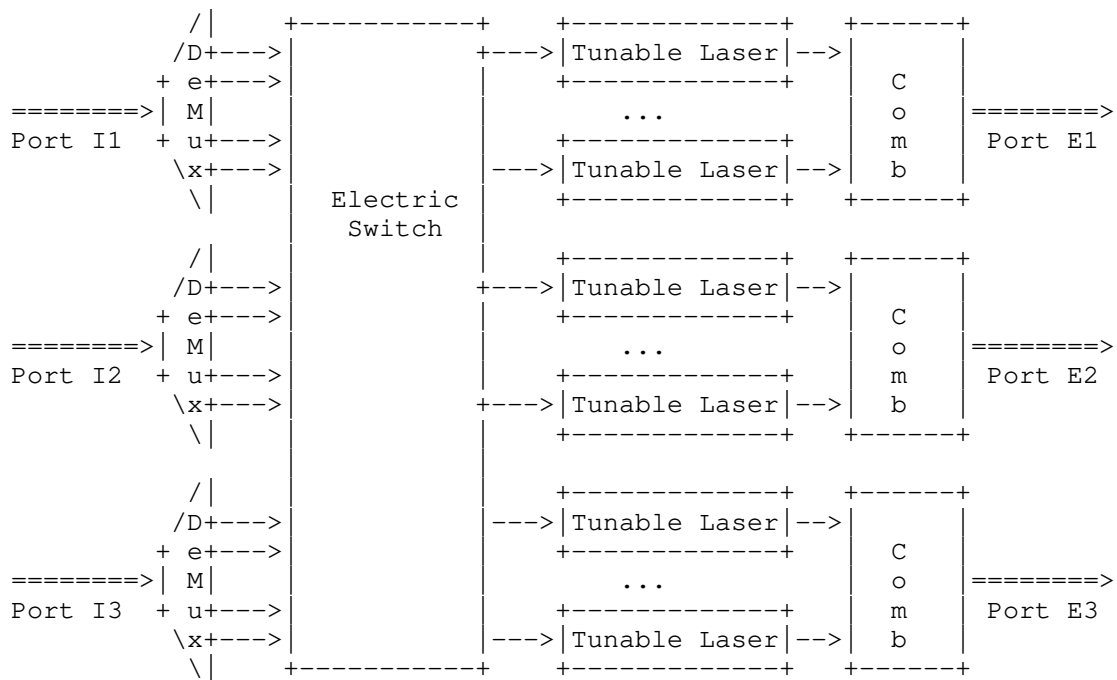


Figure 2 An optical switch built around an electronic switching fabric.

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>:

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
RB Set Field																																							
: (only one resource block in this example with shared input/output case)																																							
0 0 Reserved																																							
Input Modulation Type List Sub-Sub-TLV																																							
: (The receivers can only process NRZ)																																							
Input FEC Type List Sub-Sub-TLV																																							
: (Only Standard SDH and G.709 FECs)																																							
Input Client Signal Type Sub-TLV																																							
: (GPIDs for SDH and G.709)																																							
Input Bit Rate Range List Sub-Sub-TLV																																							
: (2.5Gbps, 10Gbps)																																							
Processing Capabilities List Sub-Sub-TLV																																							
: Fixed (non optional) 3R regeneration																																							
Output Modulation Type List Sub-Sub-TLV																																							
: NRZ																																							
Output FEC Type List Sub-Sub-TLV																																							
: Standard SDH, G.709 FECs																																							

Since we have fixed connectivity to resource block (the electronic switch) we get <ResourceBlockAccessibility>:

```

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) :
+-----+-----+-----+-----+-----+-----+-----+-----+

```

9. References

9.1. 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.

9.2. 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-ext-encode-00.txt.
- [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.
- [WSO-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, Marh 2009.

[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, March 2009.

10. 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

Y. Lee
Huawei
G. Bernstein
Grotto Networking
D. Li
Huawei
G. Martinelli
Cisco
October 21, 2010

Intended status: Informational
Expires: April 2011

A Framework for the Control of Wavelength Switched Optical Networks
(WSO) with Impairments
draft-ietf-ccamp-wson-impairments-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 November 21, 2010.

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 operation of optical networks requires information on the physical characterization of optical network elements, subsystems, devices, and cabling. These physical characteristics may be important to consider when using a GMPLS control plane to support path setup and maintenance. This document discusses how the definition and characterization of optical fiber, devices, subsystems, and network elements contained in various ITU-T recommendations can be combined with GMPLS control plane protocols and mechanisms to support Impairment Aware Routing and Wavelength Assignment (IA-RWA) in optical networks.

Table of Contents

1. Introduction.....	4
1.1. Revision History.....	5
2. Motivation.....	5
3. Impairment Aware Optical Path Computation.....	6
3.1. Optical Network Requirements and Constraints.....	7
3.1.1. Impairment Aware Computation Scenarios.....	7
3.1.2. Impairment Computation and Information Sharing Constraints.....	8
3.1.3. Impairment Estimation Process.....	10
3.2. IA-RWA Computation and Control Plane Architectures.....	11
3.2.1. Combined Routing, WA, and IV.....	13
3.2.2. Separate Routing, WA, or IV.....	13
3.2.3. Distributed WA and/or IV.....	13
3.3. Mapping Network Requirements to Architectures.....	14

4. Protocol Implications.....	17
4.1. Information Model for Impairments.....	17
4.2. Routing.....	18
4.3. Signaling.....	18
4.4. PCE.....	19
4.4.1. Combined IV & RWA.....	19
4.4.2. IV-Candidates + RWA.....	19
4.4.3. Approximate IA-RWA + Separate Detailed IV.....	21
5. Security Considerations.....	23
6. IANA Considerations.....	23
7. Acknowledgments.....	23
8. References.....	31
8.1. Normative References.....	31
8.2. Informative References.....	33

1. Introduction

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, we usually refer to these processes as "impairments". An overview of some critical optical impairments and their routing (path selection) implications can be found in [RFC4054]. Roughly speaking, 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. Therefore, path selection in any WSON requires consideration of optical impairments so that the signal will be propagated from the network ingress point to the egress point with an acceptable signal quality.

Some optical subnetworks are designed such that over any path the degradation to an optical signal due to impairments never exceeds prescribed bounds. This may be due to the limited geographic extent of the network, the network topology, and/or the quality of the fiber and devices employed. In such networks the path selection problem reduces to determining a continuous wavelength from source to destination (the Routing and Wavelength Assignment problem). These networks are discussed in [WSON-Frame]. In other optical networks, impairments are important and the path selection process must be impairment-aware.

Although [RFC4054] describes a number of key optical impairments, a more complete description of optical impairments and processes can be found in the ITU-T Recommendations. Appendix A of this document provides an overview of the extensive ITU-T documentation in this area.

The benefits of operating networks using the Generalized Multiprotocol Label Switching (GMPLS) control plane is described in [RFC3945]. The advantages of using a path computation element (PCE) to perform complex path computations are discussed in [RFC4655].

Based on the existing ITU-T standards covering optical characteristics (impairments) and the knowledge of how the impact of impairments may be estimated along a path, this document provides a framework for impairment aware path computation and establishment utilizing GMPLS protocols and the PCE architecture. As in the impairment free case covered in [WSON-Frame], a number of different control plane architectural options are described.

1.1. Revision History

Changes from 00 to 01:

Added discussion of regenerators to section 3.

Added to discussion of interface parameters in section 3.1.3.

Added to discussion of IV Candidates function in section 3.2.

Changes from 01 to 02:

Correct and refine use of "black link" concept based on liaison with ITU-T and WG feedback.

Changes from 02 to 03:

Insert additional information on use and considerations for regenerators in section 3.

2. Motivation

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 a full 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. Although the same considerations may apply to other network equipment upgrades, the optical physical interfaces are a typical case because they 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 some impairment considerations 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, since the beginning, takes into account evolving network status in term of equipments and traffic. This could result in over-engineering the DWDM network with additional regenerators nodes and optical amplifiers. Optical impairment awareness allows for the concept of photonic switching where possible and provides regeneration when it is a must. In addition, network operations such as path establishment, will require significant pre-design via non-control plane processes resulting in significantly slower network provisioning.

3. 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. This could include parameters such as the Q factor to correlate both linear and non-linear parameters into one value.

The impairment-aware path calculation needs also to take into account when regeneration happens along the path. [WSON-Frame] introduces the concept of Optical translucent network that contains transparent elements and electro-optical elements such as OEO regenerations. In such networks a generic light path can go through a certain number of regeneration points.

Regeneration points could happen for two reasons:

- (i) wavelength conversion to assist the RWA process to avoid wavelength blocking. This is the impairment free case covered by [WSON-Frame].

- (ii) the optical signal is too degraded. This is the case when the RWA take into consideration impairment estimation covered by this document.

In the latter case a light 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.

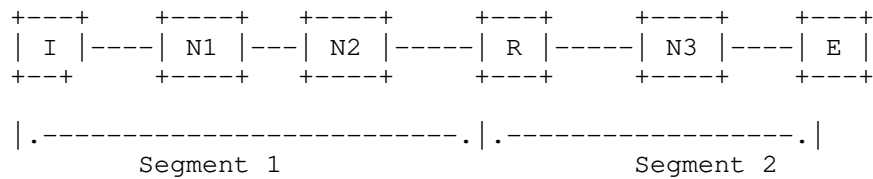


Figure 1 Light path as a set of transparent segments

For example, Figure 1 represents a Light path from node I to node E with a regeneration point R in between. It is feasible from an impairment validation perspective if both segments (I, N1, N2, R) and (R, N3, E) are feasible.

3.1. 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. We can break the different optical networks contexts up along 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.

3.1.1. 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 [WSN-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. Adding or removing an optical signal on the path will 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 those existing signals.

Much work at ITU-T has gone into developing impairment models at this and more detailed levels. Impairment characterization of network elements could then may be used to calculate which paths are conformant with a specified BER for a particular signal type. In such a case, we can combine the impairment aware (IA) path computation with the RWA process to permit more optimal IA-RWA computations. Note, 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.

3.1.2. 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 we will use the term "black links" to describe these computation and information sharing constraints in optical networks. From the control plane perspective we have the following options:

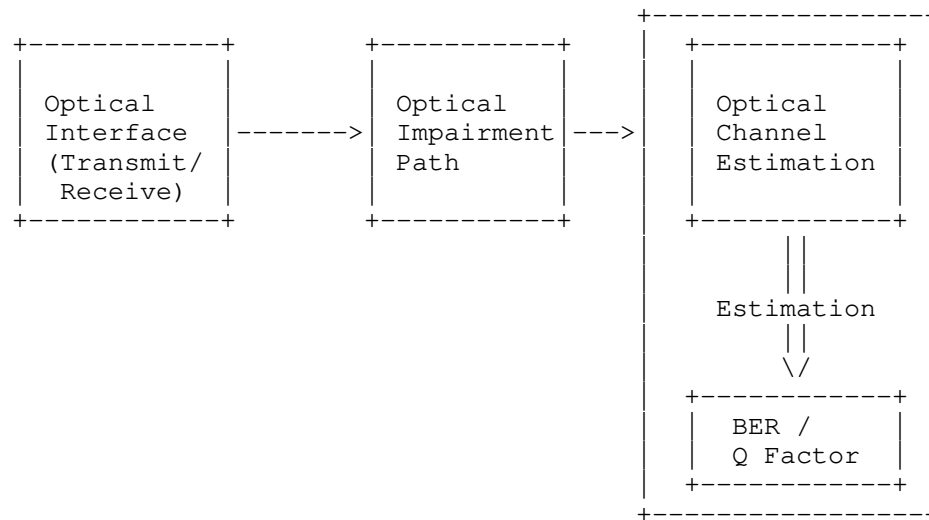
- A. The authority in control of the "black links" can furnish a list of all viable paths between all viable node pairs to a computational entity. This information would be particularly useful as an input to RWA optimization to be performed by another computation entity. The difficulty here is for larger networks such a list of paths along with any wavelength constraints could get unmanageably large.
- B. 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.
- C. 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.

3.1.3. Impairment Estimation Process

The Impairment Estimation Process can be modeled through the following functional blocks. These blocks are independent from 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 3.1.1 needs to consider a minimum set of interface characteristics. In such case only few parameters to assess the signal compatibility will be taken into account (see [WSN-Frame]). For the impairment-awareness case signal compatibility 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 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 worth 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.

3.2. 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 we will consider the following three forms of impairment validation:

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 would 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 may still be gathered from network elements however. Depending how information are gathered this may put requirements on routing protocols. This will be detailed in following sections.

- o IV-Distributed

In the distributed IV process, impairment 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 spreading optical parameters at network level.

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-validates (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 their respective advantages and disadvantages.

3.2.1. Combined Routing, WA, and IV

From the point of view of optimality, the "best" 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.

3.2.2. 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 will be discussed in the section on network contexts some impairment information may not be shared and this may lead to the need to separate IV from RWA. In addition, as also discussed in the section on network contexts, 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.

3.2.3. 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 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 route 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.

A distributed impairment validation for a prescribed network path requires that the effects of impairments can be calculated by approximate models with cumulative quality measures such as those in [G.680]. For such a system to be interoperable the various impairment measures to be accumulated would need to be agreed according to [G.680].

If distributed WA is being done at the same time as distributed IV then we may need 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.

3.3. Mapping Network Requirements to Architectures

In Figure 2 we show process flows for three main architectural alternatives to IA-RWA when approximate impairment validation suffices. In Figure 3 we show process flows for two main architectural alternatives when detailed impairment verification is required.

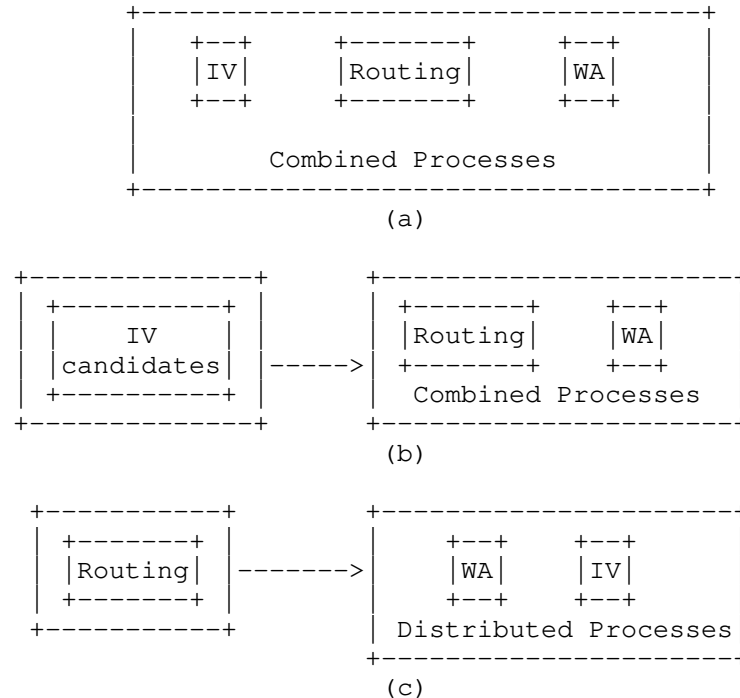


Figure 2 Process flows for the three main approximate impairment architectural alternatives.

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".

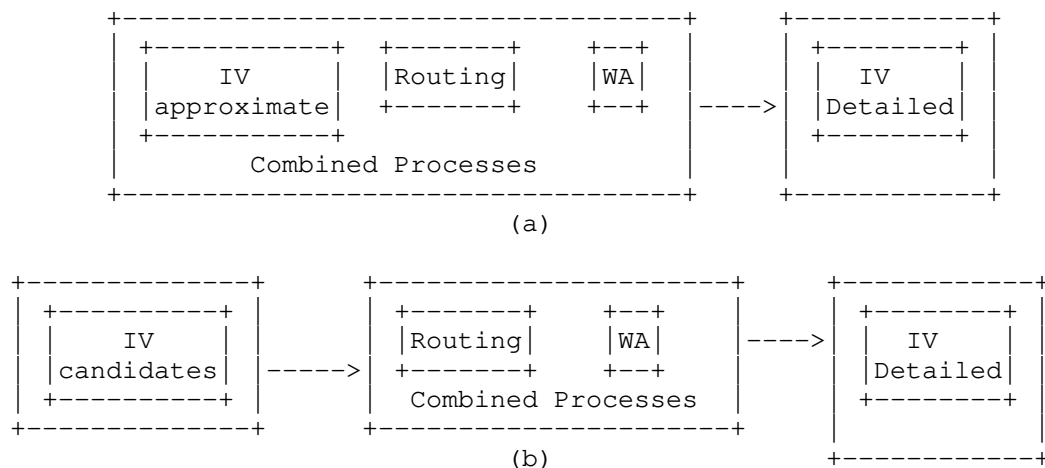


Figure 3 Process flows for the two main detailed impairment validation architectural options.

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.

4. Protocol Implications

The previous IA-RWA architectural alternatives and process flows make differing demands on a GMPLS/PCE based control plane. In this section we discuss 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
Detailed IV	Yes	No	Yes	Yes

Figure 4 IA-RWA architectural options and control plane impacts.

4.1. 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 PXCs 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.

4.2. 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 3.1.1.) 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.

4.3. Signaling

The largest impacts on signaling occur in the cases where distributed impairment validation is performed. In this we need 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.

4.4. PCE

In section 3.3. we gave a number of computation architectural alternatives that could be used to meet the various requirements and constraints of section 3.1. Here we look at 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.

4.4.1. 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 we'd like 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.

4.4.2. IV-Candidates + RWA

In this situation, shown in Figure 2(b), we have two separate processes involved in the IA-RWA computation. This requires at least two cooperating PCEs: 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 we can add this functionality to one of previously defined PCEs. We choose this later option and require the RWA PCE to also

act as the overall process coordinator. The roles, responsibilities and information requirements for these two 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:

h 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 PCC.

- o TE Database Requirements

- WSON Topology and switching capabilities and WSON impairment information (no information link wavelength utilization required).

In Figure 5 we show a sequence diagram for the interactions between the PCC, RWA-PCE and IV-Candidates-PCE.

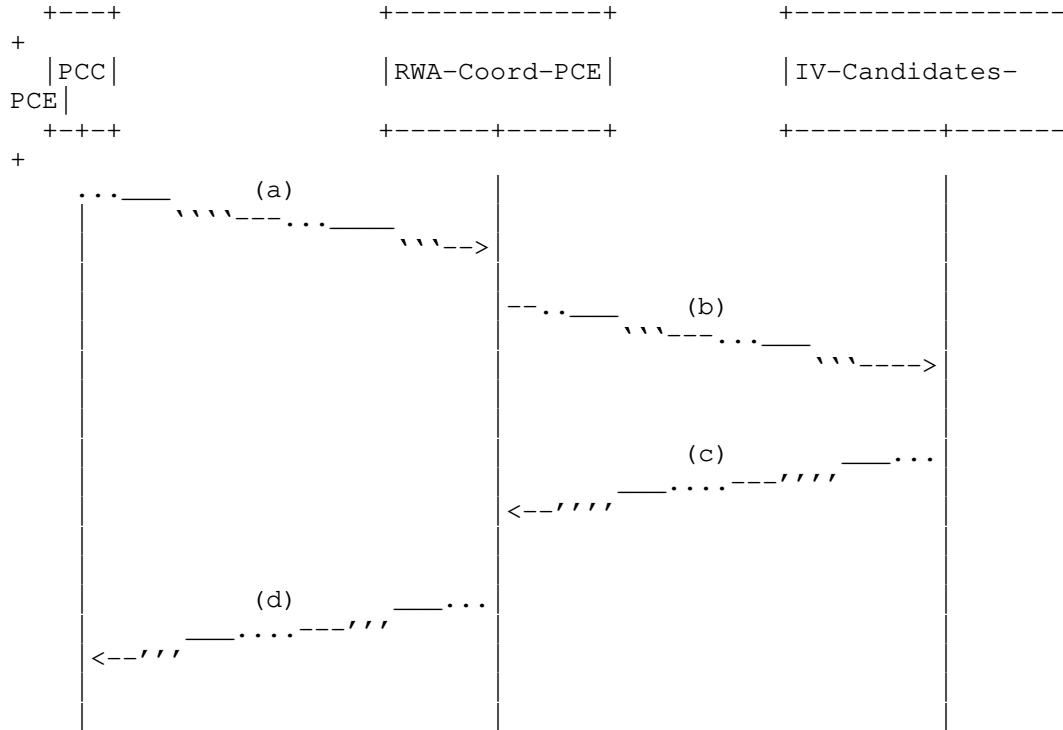


Figure 5 Sequence diagram for the interactions between PCC, RWA- Coordinating-PCE and the IV-Candidates-PCE.

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.

4.4.3. Approximate IA-RWA + Separate Detailed IV

In Figure 3 we showed two cases where a separate detailed impairment validation process could be utilized. We can place the detailed validation process into a separate PCE. Assuming that a different PCE assumes a coordinating role and interacts with the PCC we can 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.

In Figure 6 we show 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.

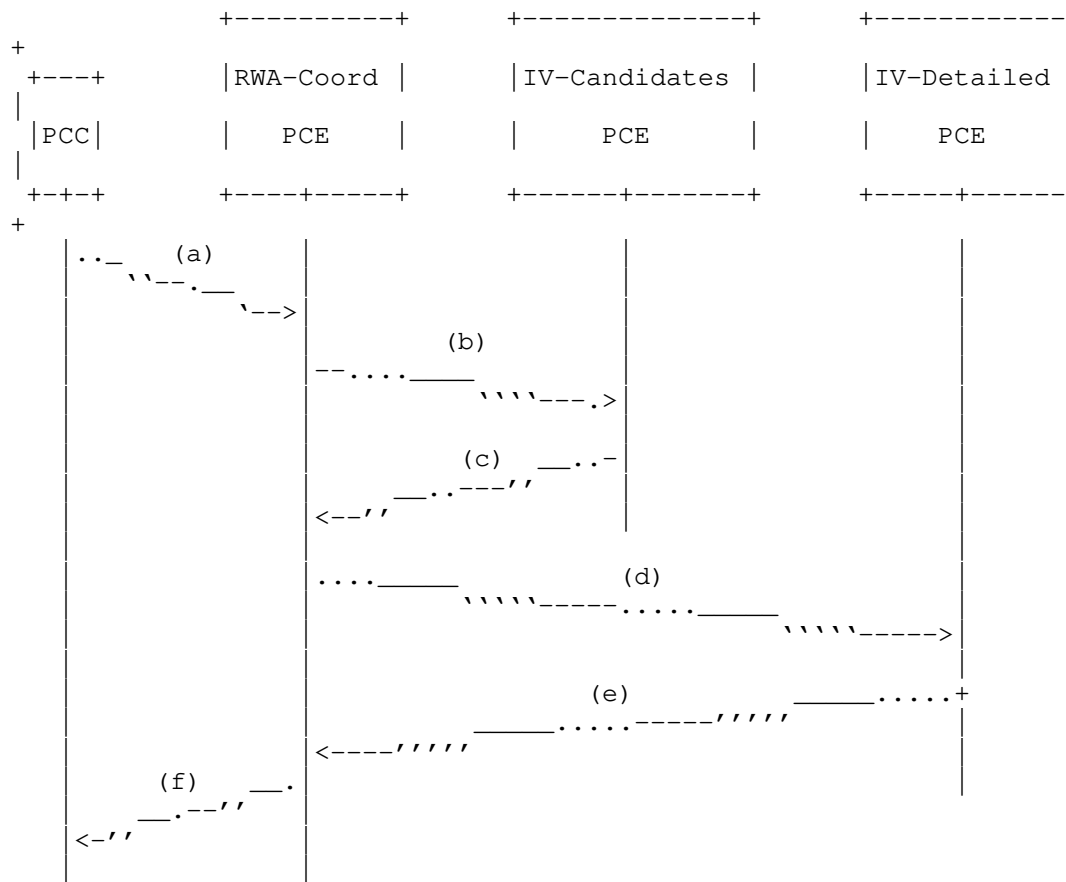


Figure 6 Sequence diagram for the interactions between PCC, RWA-Coordinating-PCE, IV-Candidates-PCE and IV-Detailed-PCE.

5. 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.

6. IANA Considerations

This draft does not currently require any consideration from IANA.

7. Acknowledgments

This document was prepared using 2-Word-v2.0.template.dot.

APPENDIX A: Overview of Optical Layer ITU-T Recommendations

For optical fiber, devices, subsystems and network elements the ITU-T has a variety of recommendations that include definitions, characterization parameters and test methods. In the following we take a bottom up survey to emphasize the breadth and depth of the existing recommendations. We focus on digital communications over single mode optical fiber.

A.1. Fiber and Cables

Fibers and cables form a key component of what from the control plane perspective could be termed an optical link. Due to the wide range of uses of optical networks a fairly wide range of fiber types are used in practice. The ITU-T has three main recommendations covering the definition of attributes and test methods for single mode fiber:

- o Definitions and test methods for linear, deterministic attributes of single-mode fibre and cable [G.650.1]
- o Definitions and test methods for statistical and non-linear related attributes of single-mode fibre and cable [G.650.2]
- o Test methods for installed single-mode fibre cable sections [G.650.3]

General Definitions[G.650.1]: Mechanical Characteristics (numerous), Mode field characteristics(mode field, mode field diameter, mode field centre, mode field concentricity error, mode field non-circularity), Glass geometry characteristics, Chromatic dispersion definitions (chromatic dispersion, group delay, chromatic dispersion coefficient, chromatic dispersion slope, zero-dispersion wavelength, zero-dispersion slope), cut-off wavelength, attenuation. Definition of equations and fitting coefficients for chromatic dispersion (Annex A). [G.650.2] polarization mode dispersion (PMD) - phenomenon of PMD, principal states of polarization (PSP), differential group delay (DGD), PMD value, PMD coefficient, random mode coupling, negligible mode coupling, mathematical definitions in terms of Stokes or Jones vectors. Nonlinear attributes: Effective area, correction factor k, non-linear coefficient (refractive index dependent on intensity), Stimulated Brillouin scattering.

Tests defined [G.650.1]: Mode field diameter, cladding diameter, core concentricity error, cut-off wavelength, attenuation, chromatic dispersion. [G.650.2]: test methods for polarization mode dispersion. [G.650.3] Test methods for characteristics of fibre cable sections following installation: attenuation, splice loss, splice location, fibre uniformity and length of cable sections (these are OTDR based), PMD, Chromatic dispersion.

With these definitions a variety of single mode fiber types are defined as shown in the table below:

ITU-T Standard		Common Name

G.652	[G.652]	Standard SMF
G.653	[G.653]	Dispersion shifted SMF
G.654	[G.654]	Cut-off shifted SMF
G.655	[G.655]	Non-zero dispersion shifted SMF
G.656	[G.656]	Wideband non-zero dispersion shifted SMF

A.2. Devices

A.2.1. Optical Amplifiers

Optical amplifiers greatly extend the transmission distance of optical signals in both single channel and multi channel (WDM) subsystems. The ITU-T has the following recommendations:

- o Definition and test methods for the relevant generic parameters of optical amplifier devices and subsystems [G.661]
- o Generic characteristics of optical amplifier devices and subsystems [G.662]
- o Application related aspects of optical amplifier devices and subsystems [G.663]
- o Generic characteristics of Raman amplifiers and Raman amplified subsystems [G.665]

Reference [G.661] starts with general classifications of optical amplifiers based on technology and usage, and include a near exhaustive list of over 60 definitions for optical amplifier device attributes and parameters. In references [G.662] and [G.665] we have characterization of specific devices, e.g., semiconductor optical amplifier, used in a particular setting, e.g., line amplifier. For example reference [G.662] gives the following minimum list of relevant parameters for the specification of an optical amplifier device used as line amplifier in a multichannel application:

- a) Channel allocation.
- b) Total input power range.
- c) Channel input power range.

- d) Channel output power range.
- e) Channel signal-spontaneous noise figure.
- f) Input reflectance.
- g) Output reflectance.
- h) Maximum reflectance tolerable at input.
- i) Maximum reflectance tolerable at output.
- j) Maximum total output power.
- k) Channel addition/removal (steady-state) gain response.
- l) Channel addition/removal (transient) gain response.
- m) Channel gain.
- n) Multichannel gain variation (inter-channel gain difference).
- o) Multichannel gain-change difference (inter-channel gain-change difference).
- p) Multichannel gain tilt (inter-channel gain-change ratio).
- q) Polarization Mode Dispersion (PMD).

A.2.2. Dispersion Compensation

In optical systems two forms of dispersion are commonly encountered [RFC4054] chromatic dispersion and polarization mode dispersion (PMD). There are a number of techniques and devices used for compensating for these effects. The following ITU-T recommendations characterize such devices:

- o Characteristics of PMD compensators and PMD compensating receivers [G.666]
- o Characteristics of Adaptive Chromatic Dispersion Compensators [G.667]

The above furnish definitions as well as parameters and characteristics. For example in [G.667] adaptive chromatic dispersion compensators are classified as being receiver, transmitter or line based, while in [G.666] PMD compensators are only defined for line and receiver configurations. Parameters that are common to both PMD and chromatic dispersion compensators

include: line fiber type, maximum and minimum input power, maximum and minimum bit rate, and modulation type. In addition there are a great many parameters that apply to each type of device and configuration.

A.2.3. Optical Transmitters

The definitions of the characteristics of optical transmitters can be found in references [G.957], [G.691], [G.692] and [G.959.1]. In addition references [G.957], [G.691], and [G.959.1] define specific parameter values or parameter ranges for these characteristics for interfaces for use in particular situations.

We generally have the following types of parameters

Wavelength related: Central frequency, Channel spacing, Central frequency deviation [G.692].

Spectral characteristics of the transmitter: Nominal source type (LED, MLM lasers, SLM lasers) [G.957], Maximum spectral width, Chirp parameter, Side mode suppression ratio, Maximum spectral power density [G.691].

Power related: Mean launched power, Extinction ration, Eye pattern mask [G.691], Maximum and minimum mean channel output power [G.959.1].

A.2.4. Optical Receivers

References [G.959.1], [G.691], [G.692] and [G.957], define optical receiver characteristics and [G.959.1], [G.691] and [G.957] give specific values of these parameters for particular interface types and network contexts.

The receiver parameters include:

Receiver sensitivity: minimum value of average received power to achieve a 1×10^{-10} BER [G.957] or 1×10^{-12} BER [G.691]. See [G.957] and [G.691] for assumptions on signal condition.

Receiver overload: Receiver overload is the maximum acceptable value of the received average power for a 1×10^{-10} BER [G.957] or a 1×10^{-12} BER [G.691].

Receiver reflectance: "Reflections from the receiver back to the cable plant are specified by the maximum permissible reflectance of the receiver measured at reference point R."

Optical path power penalty: "The receiver is required to tolerate an optical path penalty not exceeding X dB to account for total

degradations due to reflections, intersymbol interference, mode partition noise, and laser chirp."

When dealing with multi-channel systems or systems with optical amplifiers we may also need:

Optical signal-to-noise ratio: "The minimum value of optical SNR required to obtain a 1×10^{-12} BER." [G.692]

Receiver wavelength range: "The receiver wavelength range is defined as the acceptable range of wavelengths at point Rn. This range must be wide enough to cover the entire range of central frequencies over the OA passband." [G.692]

Minimum equivalent sensitivity: "This is the minimum sensitivity that would be required of a receiver placed at MPI-RM in multichannel applications to achieve the specified maximum BER of the application code if all except one of the channels were to be removed (with an ideal loss-less filter) at point MPI-RM." [G.959.1]

A.3. Components and Subsystems

Reference [G.671] "Transmission characteristics of optical components and subsystems" covers the following components:

- o optical add drop multiplexer (OADM) subsystem;
- o asymmetric branching component;
- o optical attenuator;
- o optical branching component (wavelength non-selective);
- o optical connector;
- o dynamic channel equalizer (DCE);
- o optical filter;
- o optical isolator;
- o passive dispersion compensator;
- o optical splice;
- o optical switch;
- o optical termination;
- o tuneable filter;

- o optical wavelength multiplexer (MUX)/demultiplexer (DMUX);
 - coarse WDM device;
 - dense WDM device;
 - wide WDM device.

Reference [G.671] then specifies applicable parameters for these components. For example an OADM subsystem will have parameters such as: insertion loss (input to output, input to drop, add to output), number of add, drop and through channels, polarization dependent loss, adjacent channel isolation, allowable input power, polarization mode dispersion, etc...

A.4. Network Elements

The previously cited ITU-T recommendations provide a plethora of definitions and characterizations of optical fiber, devices, components and subsystems. Reference [G.Sup39] "Optical system design and engineering considerations" provides useful guidance on the use of such parameters.

In many situations the previous models while good don't encompass the higher level network structures that one typically deals with in the control plane, i.e, "links" and "nodes". In addition such models include the full range of network applications from planning, installation, and possibly day to day network operations, while with the control plane we are generally concerned with a subset of the later. In particular for many control plane applications we are interested in formulating the total degradation to an optical signal as it travels through multiple optical subsystems, devices and fiber segments.

In reference [G.680] "Physical transfer functions of optical networks elements", a degradation function is currently defined for the following optical network elements: (a) DWDM Line segment, (b) Optical Add/Drop Multiplexers (OADM), and (c) Photonic cross-connect (PXC). The scope of [G.680] is currently for optical networks consisting of one vendors DWDM line systems along with another vendors OADMs or PXC's.

The DWDM line system of [G.680] consists of the optical fiber, line amplifiers and any embedded dispersion compensators. Similarly the OADM/PXC network element may consist of the basic OADM component and optionally included optical amplifiers. The parameters for these optical network elements (ONE) are given under the following circumstances:

- o General ONE without optical amplifiers

- o General ONE with optical amplifiers
- o OADM without optical amplifiers
- o OADM with optical amplifiers
- o Reconfigurable OADM (ROADM) without optical amplifiers
- o ROADM with optical amplifiers
- o PXC without optical amplifiers
- o PXC with optical amplifiers

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.
- [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.
- [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.
- [WSON-Frame] G. Bernstein, Y. Lee, W. Imajuku, "Framework for GMPLS and PCE Control of Wavelength Switched Optical Networks", work in progress: draft-ietf-ccamp-wavelength-switched-framework-02.txt, March 2009.

8.2. Informative References

- [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-02.txt, February 2008.
- [WSON-Comp] G. Bernstein, Y. Lee, Ben Mack-Crane, "WSON Signal Characteristics and Network Element Compatibility Constraints for GMPLS", work in progress: draft-bernstein-ccamp-wson-signal.

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

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@ericcson.com

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

Email: diego.caviglia@ericcson.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

We thank Chen Ming of DICONNET Project who provided valuable information relevant to this document.

We'd also like to thank Deborah Brungard for editorial and technical assistance.

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

Y. Lee
Huawei
G. Bernstein
Grotto Networking

September 2, 2010

OSPF Enhancement for Signal and Network Element Compatibility for
Wavelength Switched Optical Networks

draft-ietf-ccamp-wson-signal-compatibility-ospf-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 2, 2010.

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

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. Compatibility and Accessibility Sub-TLVs.....	3
2.1. Resource Block Information sub-TLV.....	4
3. Security Considerations.....	5
4. IANA Considerations.....	5
4.1. Node Information.....	5
5. References.....	6
5.1. Normative References.....	6
6. Contributors.....	7
Authors' Addresses.....	7
Intellectual Property Statement.....	8

Disclaimer of Validity.....	8
-----------------------------	---

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.

2. Compatibility and Accessibility Sub-TLVs

The encodings described in [RWA-Encode] involve node level properties, rather than link level, and hence belong in an appropriate node oriented top level TLV. The OSPF TE LSA node attribute TLV of [OSPF-Node] is used for this purpose.

This document defines four OSPF TE LSA node attribute sub-TLVs based on the encodings in [RWA-Encode]:

Sub-TLV	Type	Length	Name
---------	------	--------	------

TBA	variable	Resource Block Information
TBA	variable	Resource Block Accessibility
TBA	variable	Resource Block Wavelength Constraints
TBA	variable	Resource Block Pool State

The detail encodings of these sub-TLVs are found in [RWA-Encode] as indicated in the table below.

Name	Section[RWA-Encode]
Resource Block Information	5.1
Resource Block Accessibility	4.1
Resource Block Wavelength Constraints	4.2
Resource Block Pool State	4.3

Among the sub-TLVs defined above, the Resource Block 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. Resource Block Information sub-TLV

There are seven nested sub-TLVs defined in the Resource Block Information sub-TLV.

Sub-TLV Type	Length	Name
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	5.2
Input FEC Type List	5.3
Input Bit Range List	5.4
Input Client Signal List	5.5

Processing Capability List	5.6
Output Modulation Format List	5.7
Output FEC Type List	5.8

3. Security Considerations

This document does not introduce any further security issues other than those discussed in [RFC 3630], [RFC 4203].

4. IANA Considerations

According to [RFC3630], the OSPF TE LSA and Types for sub-TLVs for each top level Types must be assigned by Expert Review, and must be registered with IANA.

IANA is requested to allocate new Types for the sub-TLVs as defined in Sections 2 and 2.1 as follows:

4.1. Node Information

This document introduces the following sub-TLVs of Node Attribute TLV (Value TBD, see [OSPF-Node])

Sub-TLV Type	Length	Name
TBA	variable	Resource Block Information
TBA	variable	Resource Block Accessibility
TBA	variable	Resource Block Wavelength Constraints
TBA	variable	Resource Block Pool State

Sub-TLV Type	Length	Name
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

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. 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: April 2011

G. Bernstein
Grotto Networking
Sugang Xu
NICT
Y.Lee
Huawei
Hiroaki Harai
NICT

October 7, 2010

Signaling Extensions for Wavelength Switched Optical Networks
draft-ietf-ccamp-wson-signaling-00.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 7, 2010.

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

Expires April 7, 2011

[Page 1]

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 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.

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.....	7
4.2.1. Modulation Type sub-TLV.....	8
4.2.2. FEC Type sub-TLV.....	10
4.2.3. Regeneration Processing TLV.....	12
5. Bidirectional Lightpath using Same Wavelength.....	13
5.1. Using LSP_ATTRIBUTES Object.....	13
5.2. Bidirectional Lightpath Signaling Procedure.....	14

5.3. Backward Compatibility Considerations.....	15
6. Bidirectional Lightpath using Different Wavelengths.....	15
7. RWA Related.....	15
7.1. Wavelength Assignment Method Selection.....	15
8. Security Considerations.....	16
9. IANA Considerations.....	16
10. Acknowledgments.....	16
11. References.....	17
11.1. Normative References.....	17
11.2. Informative References.....	17
Author's Addresses.....	19
Intellectual Property Statement.....	20
Disclaimer of Validity.....	20

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 (WSO): 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 lambda) 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

WSO 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 [WSO-Frame] we have the following list of WSO 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 two cases MAY be separately supported: (a) Where the same wavelength is used for both upstream

and downstream directions, and (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 AWGs (i.e. multiplexer/de-multiplexer) ports directly and fixedly, or (2) transponders are connected to the add/drop ports of ROADM and each port is mapped to a dedicated wavelength fixedly.

On the other hand, with the 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 an easy management 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.

As discussed in the [WSON-Frame] a variety of different wavelength assignment algorithms have been developed. A number of these are suitable for use in distributed wavelength assignment. This feature would allow the specification of a particular approach when more than one are 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].

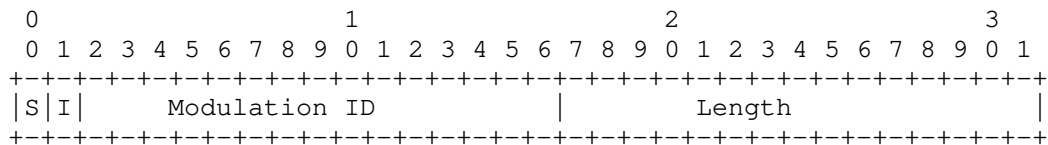
The contents of this new "processing" subobject would be a list of TLVs that could include:

- 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. Possible encodings and values for these TLV are given in below.

4.2.1. Modulation Type sub-TLV

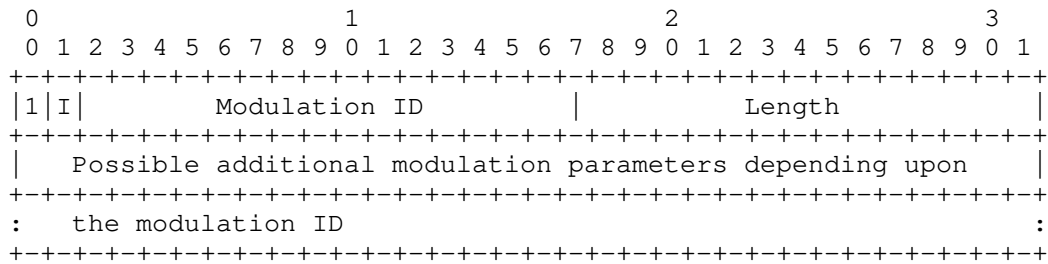
The modulation type sub-TLV may come in two different formats: a standard modulation field or a vendor specific modulation field. Both start with the same 32 bit header shown below.



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:



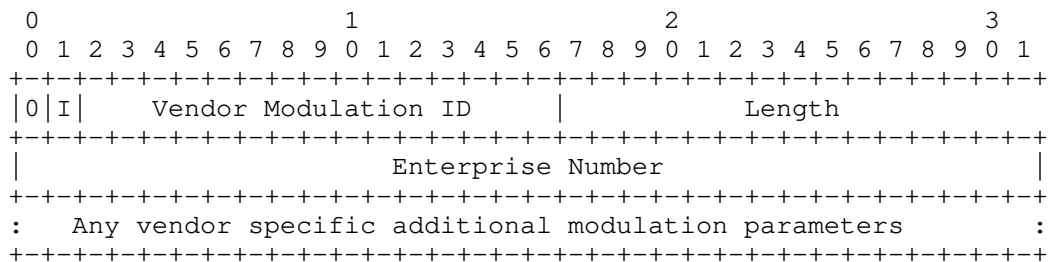
Modulation ID

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 FEC Type TLV 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.

```

      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
+-----+-----+-----+-----+-----+-----+-----+-----+
|S|I|          FEC ID          |          Length          |
+-----+-----+-----+-----+-----+-----+-----+-----+
| Possible additional FEC parameters depending upon          |
+-----+-----+-----+-----+-----+-----+-----+-----+
:   the FEC ID   :
+-----+-----+-----+-----+-----+-----+-----+-----+

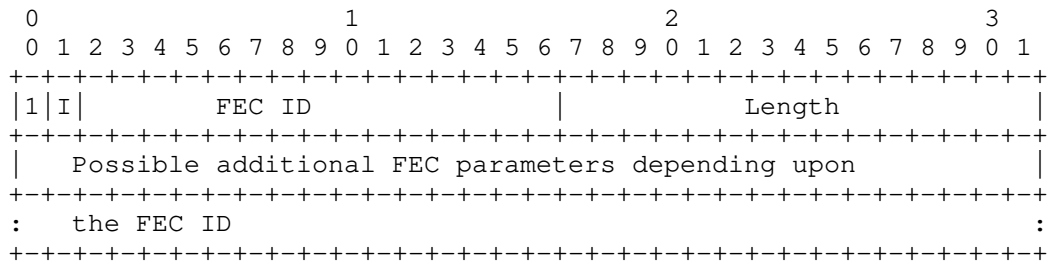
```

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:

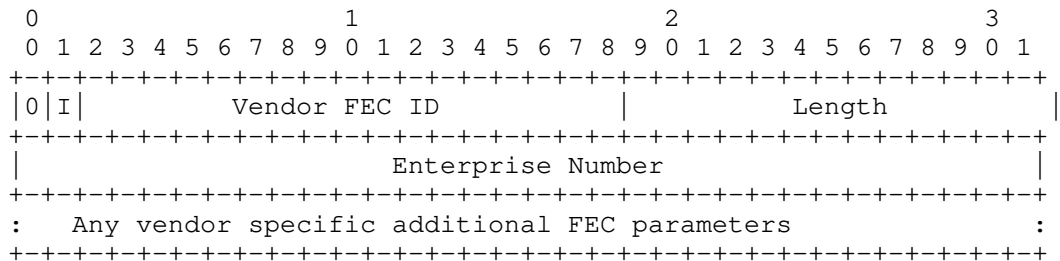


Takes on the following currently defined values for the standard FEC ID:

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

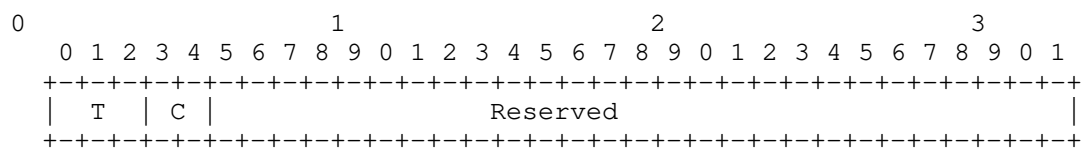
A unique identifier of an organization encoded as a 32-bit integer. Enterprise Numbers are assigned by IANA and managed through an IANA registry [RFC2578].

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 using Same Wavelength

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. Per the definition in [RFC3471], a bidirectional lightpath can be seen as a pair of unidirectional lightpaths, which are provisioned along the same route simultaneously by the RSVP-TE signaling with Upstream Label and Label Set Objects in the messages [RFC3473]. This does not necessarily require the same wavelength in both directions.

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.

The simplest and efficient way is to only define an extension to the processing of Label Set [RFC3473], and leave the other processes untouched. The issues related to this new functionality including an LSP_ATTRIBUTES object defined in [RFC5420] and the new procedure are described in the following sections. This approach would have a lower blocking probability and a shorter provisioning time. In cases of equipment failure, etc., fast provisioning used in quick recovery is critical to protect Carriers/Users against system loss.

5.1. Using LSP_ATTRIBUTES Object

To trigger the new functionality at each GMPLS node, it is necessary to notify the receiver the new type lightpath request. One multi-purpose flag/attribute parameter container object called

LSP_ATTRIBUTES object and related mechanism defined in [RFC5420] meet this requirement. One bit in Attributes Flags TLV which indicates the new type lightpath, say, the bidirectional same wavelength lightpath will be present in an LSP_ATTRIBUTES object. Please refer to [RFC5420] for detailed descriptions of the Flag and related issues.

5.2. Bidirectional Lightpath Signaling Procedure

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. Bidirectional Lightpath using Different Wavelengths

TBD

7. RWA Related

7.1. Wavelength Assignment Method Selection

As discussed in [HZang00] a number of different wavelength assignment algorithms maybe employed. In addition as discussed in [WSON-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]

Directionality: 0 unidirectional, 1 bidirectional

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
+++++																																							
Direction										WA Method										Reserved																			
+++++																																							

8. 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.

9. IANA Considerations

TBD. Once finalized in our approach we will need identifiers for such things and modulation types, modulation parameters, wavelength assignment methods, etc...

10. Acknowledgments

This document was prepared using 2-Word-v2.0.template.dot.

11. References

11.1. 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.

11.2. 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] G. Bernstein, Y. Lee, 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.

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

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: Informational
Expires: April 21, 2011

YF. Ji
WW. Bian
HX. Wang
SG. Huang
BUPT
GY. Zhang
CATR
October 18, 2010

Performance Measurement Metrics of Label Switched Path (LSP)
Establishment in Multi-Layer and Multi-Domain Networks
draft-jiyf-ccamp-lsp-00

Abstract

As the increment of network scale and the variety of user request, traditional networks are to be partitioned into multi-layer and multi-domain networks for the purpose of better management. In multi-layer and multi-domain networks, various user requests are mapped into different LSPs, and the performance of a LSP is of great importance for the users. Therefore, the LSP is necessary to be evaluated as soon as it is established. For the purpose of judging whether a LSP establishment meets a user requirement or not, typical performance measurement metrics need to be proposed. In this document, LSP establishment delay and bit error ratio (BER) which are serving as the typical performance measurement metrics are illustrated, and the definition and methodologies are proposed.

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 21, 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	4
1.1. Requirements Language	4
1.2. Terminology	4
2. Overview of the Performance Measurement Metrics of LSP Establishment	4
2.1. Overview of LSP Establishment Delay	4
2.2. Overview of LSP Establishment BER	5
3. Motivation	5
4. LSP Establishment Delay in Multi-Layer and Multi-Domain Networks	6
4.1. Measurement Metric Parameters	6
4.2. Definition	7
4.2.1. A Definition in Single Layer and Multi-Domain Networks	7
4.2.2. A Definition in Multi-Layer and Multi-Domain Networks	8
4.2.3. A Definition in Other Networks	9
4.3. Discussion	10
5. LSP Establishment BER in Multi-Domain Networks	10
5.1. General Assumptions	10
5.2. Definition	10
6. Methodologies	11
6.1. Definition	11
6.2. Methodologies	11
6.2.1. LSP Establishment Delay	11
6.2.2. LSP Establishment BER	12
7. Protocol Extension Requirements	12
8. Security Considerations	13
9. Acknowledgments	13
10. References	13
10.1. Normative References	13
10.2. Informative References	14

Appendix A. Other Authors	14
Authors' Addresses	15

1. Introduction

As the increment of network scale and the variety of user request, traditional networks are to be partitioned into multi-layer and multi-domain networks for the purpose of better management. User requests are mapped into various LSPs in multi-layer and multi-domain networks. Different users have different requirements, thus, LSP establishment is also different in order to satisfy different user requirements. To measure whether a LSP establishment meets a user requirement or not, objective performance measurement metrics and methodologies should be proposed. In this document, LSP establishment delay and BER are considered as the objective performance measurement metrics.

This document defines the performance measurement metrics and methodologies that can be used to measure the LSP establishment quality in multi-layer and multi-domain networks.

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].

1.2. Terminology

BER: Bit Error Ratio.

BRPC: Backward-Recursive PCE-Based Computation.

GMPLS: Generalized Multiprotocol Label Switching.

LID: Local Information Database.

LSP: Label Switched Path.

PCE: Path Computation Element.

VSPT: Virtual Shortest Path Tree.

2. Overview of the Performance Measurement Metrics of LSP Establishment

2.1. Overview of LSP Establishment Delay

In the process of LSP establishment, delay is considered as one of the main performance measurement metrics. In the background of GMPLS networks, LSPs that have different granularities are established.

Two typical LSP establishment methods are explained here: LSP nesting and LSP stitching. LSP nesting corresponds to the LSP establishment in the multi-layer networks, while LSP stitching corresponds to the LSP establishment in the same layer networks. LSP establishment delay in above two methods is divided into two parts: path computation delay and LSP setup delay.

In multi-layer and multi-domain networks, owing to the complexity of path computation, PCE-based path computation scheme is considered. Furthermore, optimal inter-domain LSP can not be got on a per-domain basis, so BRPC method is considered to complete inter-domain path computation in this document. Path computation delay is approximately defined from the time that source node sends the path computation request to the time that source node receives the optimal path computation result.

In multi-layer and multi-domain networks, end-to-end LSP setup is mainly considered in this document. LSP setup delay is approximately defined from the time that source node sends the LSP setup message to the time that source node receives the confirm message of successful reservation.

2.2. Overview of LSP Establishment BER

To measure the performance of LSP establishment, physical impairment parameter is one of the main performance measurement metrics, and BER is the main embodiment among all of physical impairment parameters, so BER is considered as one of the performance measurement metrics in the process of LSP establishment.

In the measurement process of LSP establishment BER, BRPC method is used for the path computation and end-to-end way is used for the LSP setup, and BER is evaluated in the LSP setup process. The approximate procedure is as follows: the signaling collects some physical parameter information from source node to destination node, and destination node evaluates the LSP performance combining corresponding physical parameter information, then destination node returns Resv message to establish LSP if the LSP performance meets the user request, otherwise, LSP establishment fails.

3. Motivation

LSP establishment delay in multi-layer and multi-domain networks is useful for several reasons:

- o Average LSP establishment delay is an important performance measurement metric that MAY reflect the scalability ability of a

multi-layer and multi-domain network. Longer LSP establishment delay with the increasing numbers of domains and nodes or traffic volumes will most likely show that the network scalability is not good, especially when the LSP establishment delay surpasses linearity curve.

- o LSP establishment delay is an important performance measurement metric that MAY reflect the quality of LSP establishment in multi-layer and multi-domain networks. Longer LSP establishment delay will most likely show that the quality of LSP establishment is not good.
- o The values in the samples of LSP establishment delay MAY serve as an early indicator to provide references on whether to accept a service request that has stringent establishment delay requirement or not.

LSP establishment BER in multi-domain networks is useful for several reasons:

- o LSP establishment BER is an important performance measurement metric that MAY reflect the quality of LSP establishment in multi-domain networks. Higher LSP establishment BER will most likely show that the quality of LSP establishment is not good.
- o The values in the samples of LSP establishment BER MAY serve as an early indicator to provide references on whether to accept a service request that has stringent establishment BER requirement or not.

4. LSP Establishment Delay in Multi-Layer and Multi-Domain Networks

This section integrally defines a performance measurement metric named LSP establishment delay in multi-layer and multi-domain networks.

4.1. Measurement Metric Parameters

- o ID0, the source node ID.
- o ID1, the destination node ID.
- o T0, a time when the path computation is attempted.
- o T1, a time when the LSP setup is attempted.

4.2. Definition

4.2.1. A Definition in Single Layer and Multi-Domain Networks

In single layer and multi-domain networks, LSP can be established using LSP stitching method. In this method, the LSP establishment delay is collected from two parts: path computation delay and LSP setup delay.

The path computation from source node ID0 to destination node ID1 mainly includes following process: source node ID0 sends a path computation Req message to the PCE responsible for the source domain. This request is forwarded between PCEs, domain-by-domain, to the PCE responsible for the destination domain. The PCE in the destination domain creates a set of optimal paths from all of the domain ingress nodes to the destination node. This set is represented as a tree of potential paths called a VSPT, and the PCE passes it back to the previous PCE in a Rep message. Each PCE in turn adds to the VSPT and passes it back until the PCE in the source domain uses the VSPT to select an optimal end-to-end path from the tree, and returns the path to the source node. The BRPC procedure above makes an assumption that the sequence of domains is known in advance. The path computation delay from source node ID0 to destination node ID1 at T0 is dT means that source node ID0 sends the path computation Req message to the PCE responsible for the source domain at time T0, and that source node receives the path computation results from the PCE responsible for the source domain at time T0+dT.

The LSP setup from source node ID0 to destination node ID1 mainly includes following process: source node ID0 firstly sends the LSP setup message, which includes two procedures: determining if service layer exist and sending Path message. The detailed procedures are as follows: source node ID0 firstly determine if service layers exist. If service layer exists, source node needs to finishes the switch reversing function, then sends Path message to the next node to reserve resource, and the next node carries out the same function like source node until Path message arrives at destination node ID1. Subsequently, destination node returns Resv message to the previous node until source node receives the Resv message. If service layer does not exist, source node firstly establishes a service layer using signaling, then source node sends Path message to determine an available wavelength set until Path message arrives at destination node. If the available wavelength set exists, then destination node sends Resv message to source node to reserve available resources, and the switch reversing function of corresponding nodes are also finished simultaneously, otherwise, PathErr message is returned to the source node. In the circumstance of service layer exists, any node which Path message traverses detects the unavailable service

layer, then PathErr message is also returned to the source node. The LSP setup delay from source node ID0 to destination node ID1 at T1 is dT means that source node ID0 sends the LSP setup message at time T1, and that source node receives the corresponding Resv message from destination node ID1 at time T1+dT.

The value of LSP establishment delay in single layer and multi-domain networks is a real number of milliseconds.

There is another case in which source node does not receive the optimal path computation result or the LSP confirm message of successful reservation within a reasonable period of time, and the value of LSP establishment delay in this case is marked undefined.

4.2.2. A Definition in Multi-Layer and Multi-Domain Networks

In multi-layer and multi-domain networks, LSP can be established using LSP nesting method. In this method, the LSP establishment delay is collected from two parts: path computation delay and LSP setup delay.

The path computation from source node ID0 to destination node ID1 mainly includes following process: source node ID0 sends a path computation Req message to the PCE responsible for the source domain. This request is forwarded between PCEs, domain-by-domain, to the PCE responsible for the destination domain. The PCE in the destination domain creates a set of optimal paths from all of the domain ingress nodes to the destination node. This set is represented as a tree of potential paths called a VSPT, and the PCE passes it back to the previous PCE in a Rep message. Each PCE in turn adds to the VSPT and passes it back until the PCE in the source domain uses the VSPT to select an optimal end-to-end path from the tree, and returns the path to the source node. The BRPC procedure above makes an assumption that the sequence of domains is known in advance. The path computation delay from source node ID0 to destination node ID1 at T0 is dT means that source node ID0 sends the path computation Req message to the PCE responsible for the source domain at time T0, and that source node receives the path computation results from the PCE responsible for the source domain at time T0+dT.

The LSP setup from source node ID0 to destination node ID1 mainly includes following process: source node ID0 firstly sends the LSP setup message, which includes two procedures: determining if service layer exist and sending Path message. The detailed procedures are as follows: source node ID0 firstly determine if service layers exist. If service layer exists, source node needs to finishes the switch reversing function, then sends Path message to the next node to reserve resource, and the next node carries out the same function

like source node until Path message arrives at destination node ID1. Subsequently, destination node returns Resv message to the previous node until source node receives the Resv message. If the capacity of existing service layer is not fully occupied, then fine granularity service that capacity is no more than remaining capacity of existing service layer can still be accepted in this service layer. If service layer does not exist, source node firstly establishes a service layer using signaling, then source node sends Path message to determine an available wavelength set until Path message arrives at destination node. If the available wavelength set exists, then destination node sends Resv message to source node to reserve available resources, and the switch reversing function of corresponding nodes are also finished simultaneously, otherwise, PathErr message is returned to the source node. In the circumstance of service layer exists, any node which Path message traverses detects the unavailable service layer, then PathErr message is also returned to the source node. If the capacity of new established service layer is not fully occupied, then fine granularity service that capacity is no more than remaining capacity of new established service layer can still be accepted in this service layer. The LSP setup delay from source node ID0 to destination node ID1 at T1 is dT means that source node ID0 sends the LSP setup message at time T1, and that source node receives the corresponding Resv message from destination node ID1 at time T1+dT.

The value of LSP establishment delay in multi-layer and multi-domain networks is a real number of milliseconds.

There is another case in which source node does not receive the optimal path computation result or the LSP confirm message of successful reservation within a reasonable period of time, and the value of LSP establishment delay in this case is marked undefined.

4.2.3. A Definition in Other Networks

There are two types of other networks: single layer and single domain networks and multi-layer and single domain networks. The definition in single layer and single domain networks is similar to the definition in single layer and multi-domain networks, and the difference is that inter-domain LSP establishment process in single layer and single domain networks is not considered. Accordingly, the definition in multi-layer and single domain networks is similar to the definition in multi-layer and multi-domain networks, and the difference is that inter-domain LSP establishment process in multi-layer and single domain networks is not considered.

The value of LSP establishment delay in single layer and single domain networks and multi-layer and single domain networks is a real

number of milliseconds.

There is another case in which source node does not receive the optimal path computation result or the LSP confirm message of successful reservation within a reasonable period of time, and the value of LSP establishment delay in this case is marked undefined.

4.3. Discussion

The reason that LSP establishment delay is set to undefined not only lies in source node never receives the corresponding reply message within a reasonable period of time, but also consists in that source node receives the PathErr message. There are many possible reasons for receiving the PathErr message: for example, network does not have enough resources to establish the service layer for the user requests or network element failure occurs.

5. LSP Establishment BER in Multi-Domain Networks

This section integrally defines a performance measurement metric named LSP establishment BER in multi-domain networks.

5.1. General Assumptions

- o Every node has a LID which stores the node physical information.
- o Destination node has a performance evaluation module which can evaluate the established LSP performance combining corresponding physical parameter information.

5.2. Definition

In the measurement process of LSP establishment BER, no matter that the network is single domain or multi-domain, the evaluation method is the same, and only the wavelength lightpath has physical parameters, so single layer network is considered.

In multi-domain networks, as physical parameters are collected and measured in the process of LSP setup, so only the LSP setup process is considered.

The LSP setup from source node to destination node mainly includes following process: source node firstly determine if service layers exist. If service layer exists, source node sends Path message to the next node to reserve resource and collects physical information of nodes and links, and the next node carries out the same function like source node until Path message arrives at destination node, then

destination node evaluates the LSP performance combining corresponding physical parameter information. If computed BER is within the tolerable range, then destination node returns Resv message to the previous node until source node receives the Resv message, otherwise, destination node returns PathErr message to the previous node until source node receives the PathErr message, and LSP setup fails. If service layer does not exist, source node firstly establishes a service layer using signaling, then source node sends Path message to determine an available wavelength set until Path message arrives at destination node. Meanwhile, signaling collects physical information of nodes and links. If the available wavelength set exists and BER that is computed by destination node is within the tolerable range, then destination node sends Resv message to source node to reserve available resources, otherwise, PathErr message is returned to the source node and LSP setup fails. In the circumstance of service layer exists, any node which Path message traverses detects the unavailable service layer, then PathErr message is also returned to the source node and LSP setup fails.

6. Methodologies

6.1. Definition

- o T0, a time when the path computation is attempted.
- o T1, a time when the LSP setup is attempted.
- o T2, a time when the optimal path computation result is returned.
- o T3, a time when the LSP confirm message of successful reservation is returned.

6.2. Methodologies

6.2.1. LSP Establishment Delay

- o Make sure that the PCE has enough computation ability to compute the path that conforms to user request.
- o Make sure that the network has enough resources to establish the requested path.
- o At the source node, form the path computation Req message. A timestamp (T0) may be stored locally on the source node when the path computation Req message is sent towards the PCE responsible for the source domain, and a timestamp (T1) may be stored locally on the source node when source node ID0 sends the LSP setup

message.

- o If the corresponding end-to-end path computation results and Resv message arrive at source node within a reasonable period of time, take the timestamp (T2) and timestamp (T3) upon receipt of the messages. By subtracting the two timestamps, an estimate of path computation delay (T2-T0) and LSP setup delay (T3-T1) can be computed.
- o If the corresponding end-to-end path computation results and Resv message fails to arrive at source node within a reasonable period of time, the path computation delay and LSP setup delay are considered to be undefined.
- o If the corresponding response is a PathErr message, then the path computation delay and LSP setup delay are considered to be undefined.

6.2.2. LSP Establishment BER

- o Make sure that the PCE has enough computation ability to compute the path that conforms to user request.
- o Make sure that the network has enough resources to establish the requested path.
- o In the process of path computation, BRPC is used as the computation method.
- o In the process of LSP setup, when Path message arrives at destination node, then the destination node computes the BER combining the corresponding physical parameter information which is collected from the traversing nodes and links. If the computed BER is within the tolerable range, then Resv message is returned to source node.
- o If the computed BER is outside the tolerable range, then PathErr message is returned to source node and LSP establishment fails.

7. Protocol Extension Requirements

- o In the measurement process of LSP establishment delay, the start time of LSP establishment and the stop time of LSP establishment need to be determined using corresponding protocol. In the process of path computation, a new object that includes timestamp needs to be added in routing protocol in order to record the start time of path computation and the stop time of path computation; In

the process of LSP setup, a new object that includes timestamp needs to be added in signaling protocol in order to record the start time of LSP setup and the stop time of LSP setup.

- o In the measurement process of LSP establishment BER, the physical information of nodes and links needs to be collected using signaling protocol, and BER is evaluated in the destination node combining corresponding physical parameter information, so a new object that includes network physical parameters needs to be added in signaling protocol in order to collect the physical information of nodes and links.

8. Security Considerations

This document involves some information collection about network physical parameters. Such information would need to be protected from intentional or unintentional disclosure.

9. Acknowledgments

We wish to thank Yongli Zhao, Linna Xia, Haoyuan Lin, Hongrui Han for their comments and help.

The RFC text was produced using Marshall Rose's xml2rfc tool.

10. References

10.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFC's to Indicate Requirement Levels", RFC 2119, March 1997.
- [RFC3473] Berger, L., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions", RFC 3473, January 2003.
- [RFC3945] Eric, M., "Generalized Multi-Protocol Label Switching (GMPLS) Architecture", RFC 3945, October 2004.
- [RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, August 2006.
- [RFC5440] Vasseur, J. and JL. Le Roux, "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, March 2009.

- [RFC5441] Vasseur, J., Zhang, R., Bitar, N., and JL. Le Roux, "A Backward-Recursive PCE-Based Computation (BRPC) Procedure to Compute Shortest Constrained Inter-Domain Traffic Engineering Label Switched Paths", RFC 5441, April 2009.
- [RFC5814] Sun, W. and G. Zhang, "Label Switched Path (LSP) Dynamic Provisioning Performance Metrics in Generalized MPLS Networks", RFC 5814, March 2010.

10.2. Informative References

- [I-D.ietf-ccamp-wson-impairments]
Lee, Y., Bernstein, G., Li, D., and G. Martinelli, "The Application of the Path Computation Element Architecture to the Determination of a Sequence of Domains in MPLS & GMPLS", July 2010.
- [Interdomain-LSP]
Aslam, F., Uzmi, ZA., and A. Farrel, "Interdomain Path Computation: Challenges and Solutions for Label Switched Networks", IEEE Communications Magazine, October 2007.
- [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.

Appendix A. Other Authors

1. Min Zhang

BUPT

No.10,Xitucheng Road,Haidian District

Beijing 100876

P.R.China

Phone: +8613910621756

Email: mzhang@bupt.edu.cn

URI: <http://www.bupt.edu.cn/>

2. Yunbin Xu

CATR

No.52 Hua Yuan Bei Lu,Haidian District

Beijing 100083

P.R.China

Phone: ++8613681485428

Email: xuyunbin@mail.ritt.com.cn

URI: <http://www.bupt.edu.cn/>

Authors' Addresses

Yuefeng Ji

BUPT

No.10,Xitucheng Road,Haidian District

Beijing 100876

P.R.China

Phone: +8613701131345

Email: jyf@bupt.edu.cn

URI: <http://www.bupt.edu.cn/>

Weiwei Bian

BUPT

No.10,Xitucheng Road,Haidian District

Beijing 100876

P.R.China

Phone: +8615210837998

Email: bianweiwei2008@163.com

URI: <http://www.bupt.edu.cn/>

Hongxiang Wang
BUPT
No.10,Xitucheng Road,Haidian District
Beijing 100876
P.R.China

Phone: +8613683683550
Email: wanghx@bupt.edu.cn
URI: <http://www.bupt.edu.cn/>

Shanguo Huang
BUPT
No.10,Xitucheng Road,Haidian District
Beijing 100876
P.R.China

Phone: +86 1062282048
Email: shghuang@bupt.edu.cn
URI: <http://www.bupt.edu.cn/>

Guoying Zhang
CATR
No.52 Hua Yuan Bei Lu,Haidian District
Beijing 100083
P.R.China

Phone: +86 1062300103
Email: zhangguoying@mail.ritt.com.cn
URI: <http://www.catr.cn/>

Internet Engineering Task Force
Internet-Draft
Intended status: Informational
Expires: April 17, 2011

M. Kattan, Ed.
G. Martinelli
D. Bianchi
Cisco
October 14, 2010

WSON Wavelength Property Information
draft-kattan-wson-property-00

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 17, 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 of 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 identifies 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.

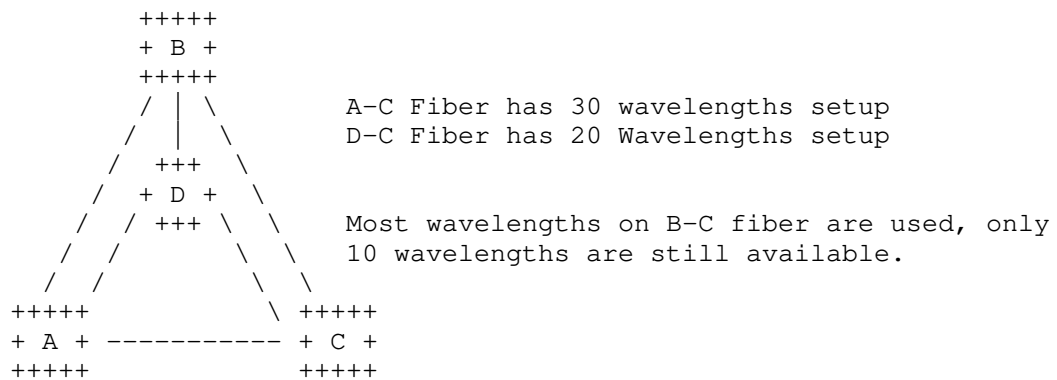


Figure 1

(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.

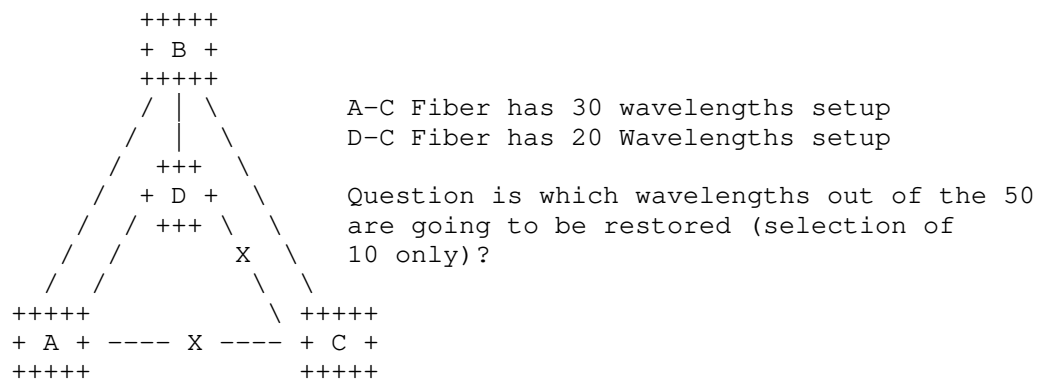


Figure 2

(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 other 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.

Wait-Time-to-Restore. This information will report a time elapsed before a wavelength go through a restoration process.

Hold-Off-Time.

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.

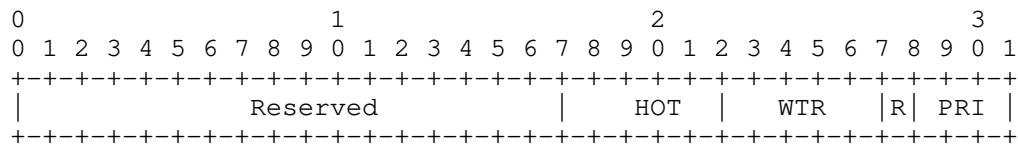


Figure 3

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 restoration

WRT is the wait time to restore. 5 bits with a granularity of 0.5 second will allow up to 16 seconds of delay on restoration.

Hold Off timer.

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

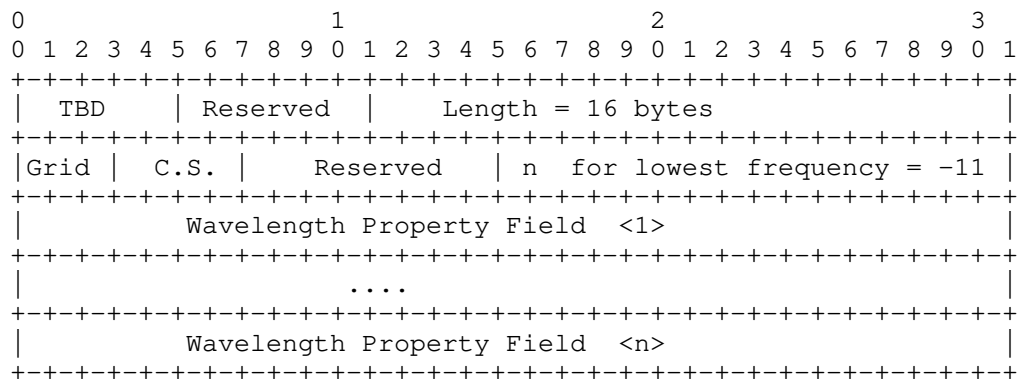


Figure 4

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

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: + 1 408 527 5101
Email: mkattan@cisco.com

Giovanni Martinelli
Cisco
Italy

Phone: +39 039 209 2044
Email: giomarti@cisco.com

David Bianchi
Cisco
Italy

Phone: +39 039 209
Email: davbianc@cisco.com

CCAMP Working Group
Internet-Draft
Intended status: Proposed Standard
Expires: April 27, 2011

Mohit Misra
Rajan Rao
Ashok Kunjidhapatham
Khuzema Pithewan
Infinera Corp
November 08, 2010

Signaling Extensions for Generalized MPLS (GMPLS) Control of
G.709 Optical Transport Networks
draft-khuzema-ccamp-gmpls-signaling-g709-00.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) 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

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	3
2. Conventions used in this document	4
3. Overview of GMPLS Signaling Extensions required for the Evolving OTN	4
4. Extensions to G.709 Traffic Parameters	5
4.1. Usage of Bit_Rate and Tolerance for ODUFlex Service	6
5. New Generalized Label Format	7
5.1. Label format for NVC or Multiplier > 1	9
6. Label Distribution Rules	9
7. Interoperability Considerations	10
7. Examples	10
8. Security Considerations	11
9. IANA Considerations	11
10. References	11
10.1. Normative References	11
10.2. Informative References	11
11. Acknowledgements	12
Author's Addresses	12

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) 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 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 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 multi-stage multiplexing

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.

ODUk-----ODUj-----ODUh
Label for Stage-1 Label for Stage-2

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.

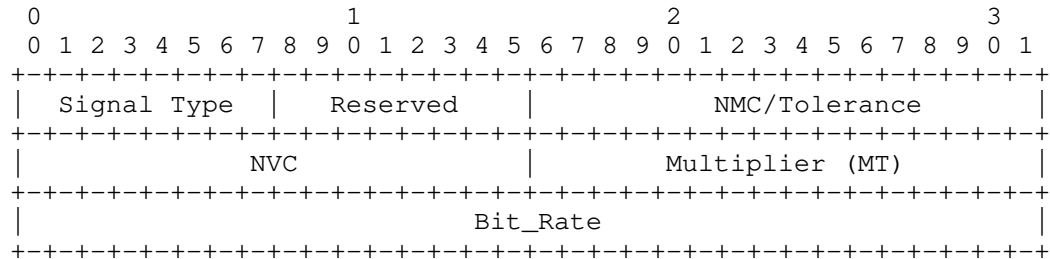
(e) Exchange of Tributary Port Number

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 added. This is captured in [OSPF-EXTN-FOR-OTN].

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.

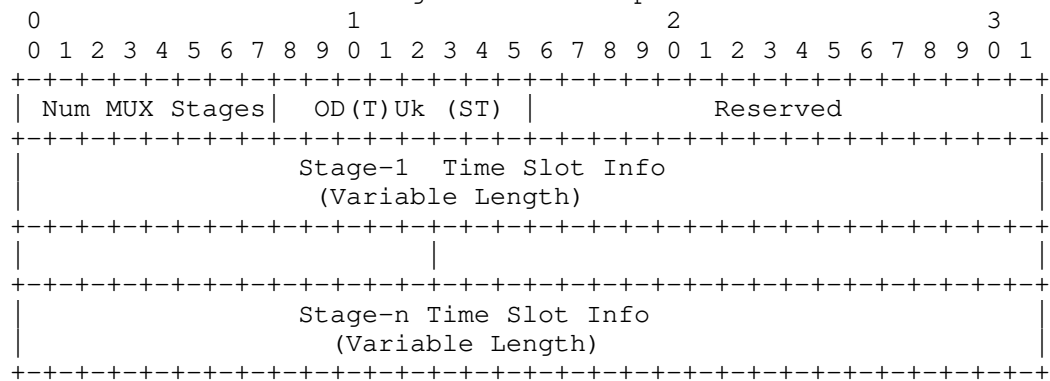
$$n = \frac{\text{Ceiling of Bit_Rate} * (1 + \text{Tolerance})}{\text{ODTUK.ts nominal bit rate} * (1 - \text{HO OPUk bit rate tolerance})}$$

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 more multi-stage label. A multi-stage label includes TS and TPN information for all stages of a multi-stage multiplexing hierarchy.

The format of a multi-stage label is explained below.



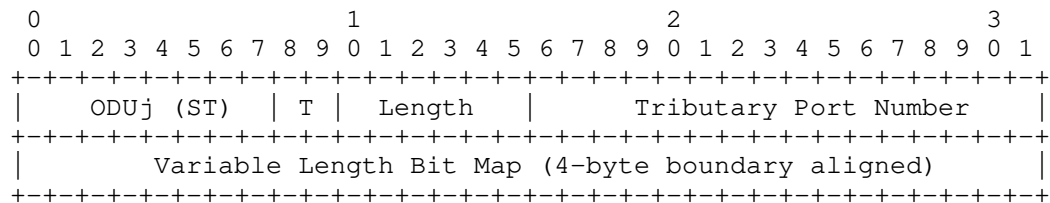
Num MUX Stages

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.

TS 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 OP Uk. 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 OP Uk type.

Bits Used	ODU Container
-----	-----
0-5	ODU1 to ODU3
0-6	ODU4
7-15	Reserved

The reserved bits should be coded as zeros for flexible assignment scheme.

Bit Map

This is a multi-byte bit map field. The length of this field varies depending on the number of TSs associated with the immediate HO-ODU

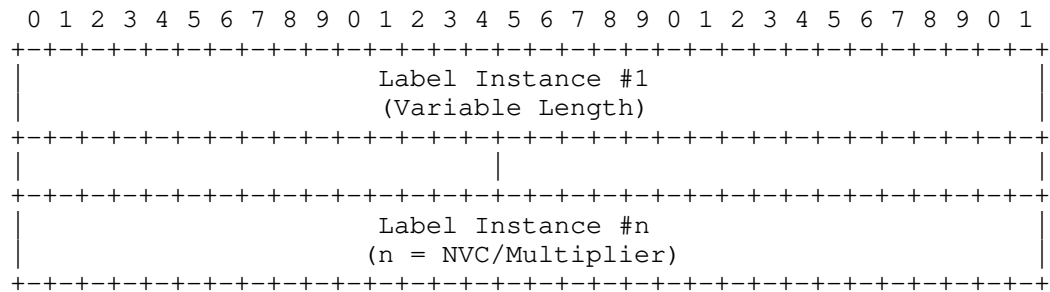
pertaining to the stage. Each bit represents one TS. Bit values are interpreted as follows

Bit Value	Meaning
0	Not Used
1	Used

This field must be 4 byte aligned using filler bytes.

5.1. Label format for NVC or Multiplier > 1

For NVC or Multiplier field value > 1, the label format defined in section 5 needs to be repeated NVC/multiplier times.



6. 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 a ODU LSP from its upstream neighbor, it reserves the resources required on an OTN interface and send the label back to upstream neighbor. Note that Label can also be explicitly specified by source node.

The encoding of Generalized Label is as follows:

For ODU_j to ODU_k multiplexing, the length field indicates the number of TS supported on ODU_k. TS reserved for ODU_j are marked as 1.

For ODU_k to OTU_k mapping, the length field is coded as 0 and BitMap field is not included.

B. Receiving Side

For ODUj to ODUK multiplexing, the node extracts the Bit Map field using the Length field. The position of Bit in the Bitmap interpreted as the Trib Slot Number. The value stored in the bit indicates if it is reserved for the ODUj.

For ODUK to OTUK mapping, the length is 0. Hence, bitmap field is not expected.

7. 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.

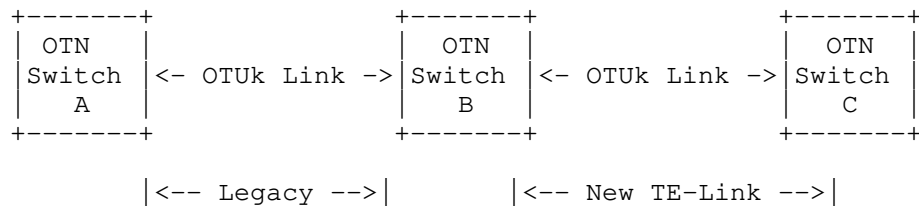


Figure-1: OTUk TE-Link

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.

7. Examples

<Will be added in the next revision of this draft>

8. Security Considerations

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

9. IANA Considerations

TBD

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.
- [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.
- [G.709-v3] ITU-T, "Interfaces for the Optical Transport Network (OTN)", G.709 Recommendation, December 2009.

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

Authors would like to thank Lou Berger and Biao Lu for review comments and suggestions.

Author's Addresses

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

Internet-Draft draft-khuzema-ccamp-gmpls-sig-g709-00.txtOctober 24, 2010

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

INTERNET-DRAFT
Intended Status: Proposed Standard
Expires: March 26, 2011

A. Malis, ed.
Verizon Communications
A. Lindem, ed.
Ericsson
September 22, 2010

Updates to ASON Routing for OSPFv2 Protocols (RFC 5787bis)
draft-malis-ccamp-rfc5787bis-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) 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 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	10
7.1 Import/Export Rules	11
7.2 Loop Prevention	11
7.2.1 Inter-RA Export Upward/Downward Sub-TLVs	11
7.2.2 Inter-RA Export Upward/Downward Sub-TLV Processing	12
8. OSPFv2 Scalability	13
9. Security Considerations	13
10. IANA Considerations	14
10.1. Sub-TLVs of the Link TLV	14

10.2. Sub-TLVs of the Node Attribute TLV	14
10.3. Sub-TLVs of the Router Address TLV	15
11. References	16
11.2. Informative References	16
12. Acknowledgements	17
Appendix A. ASON Terminology	18
Appendix B. ASON Routing Terminology	19
Authors' Addresses	20

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.

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.

A key ASON requirement is the support of multiple transport planes or layers. Each transport node has associated topology (links and reachability information) which is used for ASON routing.

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. 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].

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

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 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 [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. However, some link types may support several different signal types that are modeled as separate layers in the G.805 model [G.805] (e.g., SDH links may simultaneously support VC-3, VC-4, VC-4-4c, VC-4-16c, and VC-4-64c signals). Optimization refinements to reduce the overhead of advertising link characteristics separately for each signal type may be defined. However, further refinement of the ISCD sub-TLV for multi-layer networks is beyond the scope of this document.

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 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.

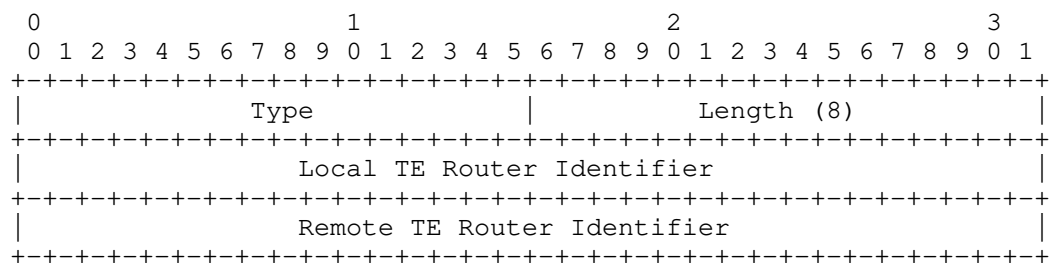
6.1. Link Advertisement (Local and Remote TE Router ID Sub-TLV)

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:



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

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 and link information will ever be mixed with global or local IP routing information. If there ever were 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. References

11.1. 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.
- [RFC5786] Aggarwal, R. and K. Kompella, "Advertising a Router's Local Addresses in OSPF TE Extensions", RFC 5786, March 2010.

11.2. 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)", November 2001 (and Revision, January 2003).

12. 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

Network Working Group
Internet Draft
Intended status: Informational
Expires: April 2011

E. Roch
Ciena
T. Marcot
France Telecom
L. Ong
Ciena
October 18, 2010

Extensions to Hierarchical LSPs for ASON identifiers support
draft-roch-ccamp-lsp-hier-ason-00.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 April 18, 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

A set of requirements and a proposed solution for the control of hierarchical Label Switched Paths (LSPs) is found in [HIER]. However, support of address separation as allowed by the Automatically Switched Optical Network (ASON) architecture [G.8080] is not covered by [HIER]. This internet draft describes additional requirements to consider for the use of LSP hierarchy in ASON networks and proposes extensions to address those requirements.

Table of Contents

1. Introduction and Problem Statement.....	2
1.1. Separate control plane instances at different layers.....	3
1.2. Address and identifier separation within a layer.....	3
2. Requirements.....	4
2.1. Routing Controller Identification.....	4
2.2. Signaling Controller Identification.....	4
3. Mechanisms and Protocol Extensions.....	4
3.1. LSP_TUNNEL_INTERFACE_ID sub-TLVs.....	4
3.1.1. Routing Controller Protocol Controller (RC PC) Identifier.....	5
3.1.2. Routing Controller Protocol Controller (RC PC) Reachable Address.....	5
3.1.3. Signaling Controller Protocol Controller (SC PC) Identifier.....	5
3.1.4. Signaling Controller Protocol Controller (SC PC) Reachable Address.....	5
4. Security Considerations.....	6
5. IANA Considerations.....	6
6. References.....	6
6.1. Normative References.....	6
6.2. Informative References.....	6
7. Acknowledgments.....	6

1. Introduction and Problem Statement

This problem statement applies to the operation of multilayer networks according to the ASON architecture.

[HIER] defines a set of extensions for the control of hierarchical Label Switched Paths (LSPs). This internet draft describes additional requirements for the use of LSP Hierarchy in ASON networks.

1.1. Separate control plane instances at different layers

In ASON architecture, the control plane instance in a client layer may be a separate instance than the control plane instance for the client layer. This requires that when a server layer link is created, sufficient information must be passed to allow a new control (signaling and optionally routing) association to be created between the client control instances at the ends of the new link. This includes identification and addressing information for both the signaling control instance and routing control instance at each end.

The ASON architecture [G.8080] allows for separate control plane instances for each controlled layer. In a real deployment, this can be seen in a few scenarios. For example, in networks mixing legacy equipment and emerging technologies, existing legacy control plane for some layers and new control plane for other layers may be based on different protocols, requiring different instances.

Additionally, some equipment may be entirely under management plane control whereas other is under control plane. There might also be business boundaries due to mergers and acquisitions or due to internal company organization. In these cases, the result is multiple instances of control plane.

Another scenario is that different instances may be used to solve scalability problems.

1.2. Address and identifier separation within a layer

Separate identification of routing controller instances, signaling controller instances and resource identifiers is required in order to support ASON signaling and routing. Separation of routing controller and resource identifier is already addressed as a requirement in [RFC4652], as referenced by the terms ''Li'' and ''Pi'' for the logical control plane entity and physical node identifiers, respectively. This allows 1:n relationships between the control entity and the physical resources being controlled, for example.

Separation of routing and signaling controller identifiers and their respective reachable addresses allows the routing and signaling controller identifiers to be independent of the specific network address by which they are reached. This allows the operator to modify the signaling communications network addressing scheme without impacting the control plane protocols. Routing controller addressing is further discussed in [RFC4258].

2. Requirements

2.1. Routing Controller Identification

In ASON architecture, a routing controller possesses two identifiers. The first is the Routing Controller Protocol Controller Identifier (RC PC ID). The second is the IPv4 address at which the routing controller can be reached, the Routing Controller Protocol Controller Signaling Control Network address (RC PC SCN address).

New requirement: It must be possible to exchange RC PC IDs and RC PC SCN addresses for the establishment of a routing adjacency in the client layer.

2.2. Signaling Controller Identification

In ASON architecture, signaling controller identifiers cannot be automatically derived from routing controller identifiers. In order to establish an RSVP-TE signaling adjacency between two client signaling controllers, a signaling mechanism is required in the server layer to identify the signaling controller. Each signaling controller requires two identifiers. The first is the Signaling Controller Protocol Controller Identifier (SC PC ID). The second is the IPv4 address at which the signaling controller can be reached, the Signaling Controller Protocol Controller Signaling Control Network address (SC PC SCN address).

New Requirement: It must be possible to exchange SC PC IDs and SC PC SCN addresses for the establishment of a signaling adjacency in the client layer.

3. Mechanisms and Protocol Extensions

This section defines protocol extensions to address the requirements described in the previous section.

3.1. LSP_TUNNEL_INTERFACE_ID sub-TLVs

The following sub-TLVs are optional sub-TLVs of the LSP_TUNNEL_INTERFACE_ID, in addition to already defined Target IGP Identifier and Component Link Identifier TLV. These sub-TLVs allow the client layer to use separate routing and signaling controller identifiers and reachable addresses.

3.1.1. Routing Controller Protocol Controller (RC PC) Identifier

The following sub-TLV is included to identify the RC PC associated with the client layer. The TLV is formatted as described in Section 3.1.2 of [HIER]. The Type field has the value 4 (TBD), and the Value field has the following content:

```

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
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
| Routing Controller Protocol Controller Identifier |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

```

3.1.2. Routing Controller Protocol Controller (RC PC) SCN Address

The following sub-TLV is included to provide the SCN reachable address for the RC PC associated with the client layer. The TLV is formatted as described in Section 3.1.2 of [HIER]. The Type field has the value 5 (TBD), and the Value field has the following content:

```

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
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
| Routing Controller Protocol Controller SCN IPv4 Address |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

```

3.1.3. Signaling Controller Protocol Controller (SC PC) Identifier

The following sub-TLV is included to identify the SC PC associated with the client layer. The TLV is formatted as described in Section 3.1.2 of [HIER]. The Type field has the value 6 (TBD), and the Value field has the following content:

```

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
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
| Signaling Controller Protocol Controller Identifier |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

```

3.1.4. Signaling Controller Protocol Controller (SC PC) SCN Address

The following sub-TLV is included to provide the reachable SCN address for the RC PC associated with the client layer. The TLV is

formatted as described in Section 3.1.2 of [HIER]. The Type field has the value 7 (TBD), and the Value field has the following content:

```

      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
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
| Routing Controller Protocol Controller SCN IPv4 Address          |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

```

4. Security Considerations

TBD

5. IANA Considerations

TBD

6. References

6.1. Normative References

[HIER] Shiomoto, K., and Farrel, A. (Editors), "Procedures for Dynamically Signaled Hierarchical Label Switched Paths", draft-ietf-ccamp-lsp-hierarchy-bis-08.txt, February 2010

[RFC4258] Brungard, D, Ed. "Requirements for Generalized Multi-Protocol Label Switching (GMPLS) Routing for the Automatically Switched Optical Network (ASON)", RFC4258, November 2005

[RFC4652] Papadimitriou, D., Ed. "Evaluation of Existing Routing Protocols against Automatic Switched Optical Network (ASON) Routing Requirements", RFC4652, October 2006

6.2. Informative References

[G.8080] ITU-T Rec G.8080/Y.1304 "Architecture for the Automatically Switched Optical Network (ASON)", June 2006

7. Acknowledgments

The authors would like to thank Vishnu Shukla (Verizon) for his contribution and comments to this document.

Authors' Addresses

Evelyne Roch
Ciena
Email: eroch@ciena.com

Thierry Marcot
France Telecom
Email: thierry.marcot@orange-ftgroup.com

Lyndon Ong
Ciena
Email: lyong@ciena.com

Network Working Group
Internet Draft
Intended status: Informational
Expires: April 2011

E. Roch
Ciena
T. Marcot
France Telecom
L. Ong
Ciena
October 18, 2010

Extensions to Hierarchical LSPs for Multi-Client Support
draft-roch-ccamp-lsp-hier-multi-client-00.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 April 18, 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

A set of requirements and a proposed solution for the control of hierarchical Label Switched Paths (LSPs) is found in [HIER]. However, support of multiple client layer networks are not covered by [HIER]. This internet draft describes additional requirements and proposes a solution to support multiple client layer networks.

Table of Contents

1. Introduction and Problem Statement.....	2
2. Requirements.....	2
3. Mechanisms and Protocol Extensions.....	2
4. Security Considerations.....	3
5. IANA Considerations.....	3
6. References.....	3
6.1. Normative References.....	3
6.2. Informative References.....	Error! Bookmark not defined.
7. Acknowledgments.....	4

1. Introduction and Problem Statement

[HIER] defines a set of extensions for the control of hierarchical Label Switched Paths (LSPs). It allows an LSP created in one network, i.e. server network, to be used in another network, i.e. client network. The LSP_TUNNEL_INTERFACE_ID is used during server network LSP establishment to exchange information about how to identify and optionally advertise the link in the client network.

2. Requirements

In order to support flexible adaptation where a server network LSP provides services to multiple client networks, it is necessary to identify to which layer the information carried in the LSP_TUNNEL_INTERFACE_ID applies to. For example, an OTN LSP is established and is available to carry Ethernet or SDH client traffic.

New Requirement: For each client network supported, it should be possible to exchange both the layer identification and a separate set of control plane identifiers associated with the client layer.

3. Mechanisms and Protocol Extensions

The extension continues to use the current format for LSP_TUNNEL_INTERFACE_ID object as defined in [HIER], and just adds a

new sub-TLV that identifies the specific client layer(s) that should use the information exchanged in the LSP_TUNNEL_INTERFACE_ID. This allows a single LSP_TUNNEL_INTERFACE_ID to carry information for multiple client layers if they share common control information. Several LSP_TUNNEL_INTERFACE_ID objects are required when there are client layers with different control plane identifier information.

A new CLIENT_LAYER_ID_SUB_TLV Object is defined to indicate to which client layer(s) the LSP_TUNNEL_INTERFACE_ID is applicable. The TLV is formatted as described in Section 3.1.2 of [HIER]. The Type field has the value 3 (TBD), and the Value field has the following content:

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
Encoding Type										Switching Type										Signal Type										Reserved									
...																		
Encoding Type										Switching Type										Signal Type										Reserved									

4. Security Considerations

TBD

5. IANA Considerations

TBD

6. References

6.1. Normative References

- [HIER] Shiomoto, K., and Farrel, A. (Editors), "Procedures for Dynamically Signaled Hierarchical Label Switched Paths", draft-ietf-ccamp-lsp-hierarchy-bis-08.txt, February 2010
- [RFC4258] Brungard, D, Ed. "Requirements for Generalized Multi-Protocol Label Switching (GMPLS) Routing for the Automatically Switched Optical Network (ASON)", RFC4258, November 2005
- [RFC4652] Papadimitriou, D., Ed. "Evaluation of Existing Routing Protocols against Automatic Switched Optical Network (ASON) Routing Requirements", RFC4652, October 2006

7. Acknowledgments

The authors would like to thank Vishnu Shukla (Verizon) and Fatai Zhang (Huawei) for their contribution and comments to this document.

Authors' Addresses

Evelyne Roch
Ciena
Email: eroch@ciena.com

Thierry Marcot
France Telecom
Email: thierry.marcot@orange-ftgroup.com

Lyndon Ong
Ciena
Email: lyong@ciena.com

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: April 21, 2011

L. Berger
LabN Consulting, L.L.C.
A. Takacs
D. Caviglia
Ericsson
D. Fedyk
Alcatel-Lucent
J. Meuric
France Telecom Orange
October 18, 2010

GMPLS Asymmetric Bandwidth Bidirectional Label Switched Paths (LSPs)
draft-takacs-ccamp-asymm-bw-bidir-lsps-bis-00.txt

Abstract

This document defines a method for the support of GMPLS asymmetric bandwidth bidirectional Label Switched Paths (LSPs). The presented approach is applicable to any switching technology and builds on the original Resource Reservation Protocol (RSVP) model for the transport of traffic-related parameters.

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 21, 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. Background	3
1.2. Approach Overview	4
1.3. Conventions Used in This Document	4
2. Generalized Asymmetric Bandwidth Bidirectional LSPs	5
2.1. UPSTREAM_FLOWSPEC Object	5
2.1.1. Procedures	5
2.2. UPSTREAM_TSPEC Object	5
2.2.1. Procedures	6
2.3. UPSTREAM_ADSPEC Object	6
2.3.1. Procedures	6
3. Packet Formats	7
4. Compatibility	9
5. IANA Considerations	10
5.1. UPSTREAM_FLOWSPEC Object	10
5.2. UPSTREAM_TSPEC Object	10
5.3. UPSTREAM_ADSPEC Object	10
6. Security Considerations	11
7. References	12
7.1. Normative References	12
7.2. Informative References	12
Authors' Addresses	14

1. Introduction

GMPLS [RFC3473] introduced explicit support for bidirectional Label Switched Paths (LSPs). The defined support matched the switching technologies covered by GMPLS, notably Time Division Multiplexing (TDM) and lambdas; specifically, it only supported bidirectional LSPs with symmetric bandwidth allocation. Symmetric bandwidth requirements are conveyed using the semantics objects defined in [RFC2205] and [RFC2210].

GMPLS asymmetric bandwidth bidirectional LSPs are bidirectional LSPs that have different bandwidth reservations in each direction. Support for bidirectional LSPs with asymmetric bandwidth, was previously discussed in the context of Ethernet, notably [GMPLS-PBBTE] and [RFC6003]. In that context, asymmetric bandwidth support was considered to be a capability that was unlikely to be deployed, and hence [RFC5467] was published as Experimental. The MPLS Transport Profile, MPLS-TP, requires that asymmetric bandwidth bidirectional LSPs be supported, see [RFC5654], and therefore this document is being published on the Standards Track. This document has no technical changes from the approach defined in [RFC5467]. This document removes an alternate approach that is not part of the Standards Track solution.

1.1. Background

Bandwidth parameters are transported within RSVP ([RFC2210], [RFC3209], and [RFC3473]) via several objects that are opaque to RSVP. While opaque to RSVP, these objects support a particular model for the communication of bandwidth information between an RSVP session sender (ingress) and receiver (egress). The original model of communication, defined in [RFC2205] and maintained in [RFC3209], used the SENDER_TSPEC and ADSPEC objects in Path messages and the FLOWSPEC object in Resv messages. The SENDER_TSPEC object was used to indicate a sender's data generation capabilities. The FLOWSPEC object was issued by the receiver and indicated the resources that should be allocated to the associated data traffic. The ADSPEC object was used to inform the receiver and intermediate hops of the actual resources allocated for the associated data traffic.

With the introduction of bidirectional LSPs in [RFC3473], the model of communication of bandwidth parameters was implicitly changed. In the context of [RFC3473] bidirectional LSPs, the SENDER_TSPEC object indicates the desired resources for both upstream and downstream directions. The FLOWSPEC object is simply confirmation of the allocated resources. The definition of the ADSPEC object is either unmodified and only has meaning for downstream traffic, or is implicitly or explicitly ([RFC4606] and [MEF-TRAFFIC]) irrelevant.

1.2. Approach Overview

The approach for supporting asymmetric bandwidth bidirectional LSPs defined in this document builds on the original RSVP model for the transport of traffic-related parameters and GMPLS's support for bidirectional LSPs.

The defined approach is generic and can be applied to any switching technology supported by GMPLS. With this approach, the existing SENDER_TSPEC, ADSPEC, and FLOWSPEC objects are complemented with the addition of new UPSTREAM_TSPEC, UPSTREAM_ADSPEC, and UPSTREAM_FLOWSPEC objects. The existing objects are used in the original fashion defined in [RFC2205] and [RFC2210], and refer only to traffic associated with the LSP flowing in the downstream direction. The new objects are used in exactly the same fashion as the old objects, but refer to the upstream traffic flow. Figure 1 shows the bandwidth-related objects used for asymmetric bandwidth bidirectional LSPs.

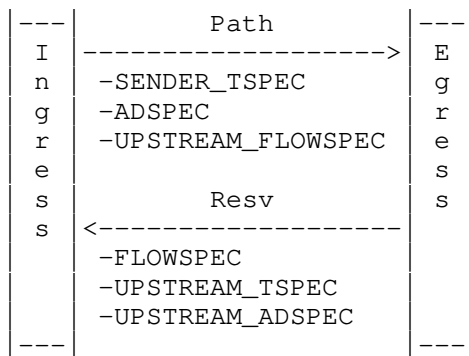


Figure 1: Generic Asymmetric Bandwidth Bidirectional LSPs

The extensions defined in this document are limited to Point-to-Point (P2P) LSPs. Support for Point-to-Multipoint (P2MP) bidirectional LSPs is not currently defined and, as such, not covered in this document.

1.3. 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. Generalized Asymmetric Bandwidth Bidirectional LSPs

The setup of an asymmetric bandwidth bidirectional LSP is signaled using the bidirectional procedures defined in [RFC3473] together with the inclusion of the new UPSTREAM_FLOWSPEC, UPSTREAM_TSPEC, and UPSTREAM_ADSPEC objects.

The new upstream objects carry the same information and are used in the same fashion as the existing downstream objects; they differ in that they relate to traffic flowing in the upstream direction while the existing objects relate to traffic flowing in the downstream direction. The new objects also differ in that they are carried on messages traveling in the opposite direction.

2.1. UPSTREAM_FLOWSPEC Object

The format of an UPSTREAM_FLOWSPEC object is the same as a FLOWSPEC object. This includes the definition of class types and their formats. The class number of the UPSTREAM_FLOWSPEC object is 120 (of the form 0bbbbbbb).

2.1.1. Procedures

The Path message of an asymmetric bandwidth bidirectional LSP MUST contain an UPSTREAM_FLOWSPEC object and MUST use the bidirectional LSP formats and procedures defined in [RFC3473]. The C-Type of the UPSTREAM_FLOWSPEC object MUST match the C-Type of the SENDER_TSPEC object used in the Path message. The contents of the UPSTREAM_FLOWSPEC object MUST be constructed using a format and procedures consistent with those used to construct the FLOWSPEC object that will be used for the LSP, e.g., [RFC2210] or [RFC4328].

Nodes processing a Path message containing an UPSTREAM_FLOWSPEC object MUST use the contents of the UPSTREAM_FLOWSPEC object in the upstream label and the resource allocation procedure defined in Section 3.1 of [RFC3473]. Consistent with [RFC3473], a node that is unable to allocate a label or internal resources based on the contents of the UPSTREAM_FLOWSPEC object MUST issue a PathErr message with a "Routing problem/MPLS label allocation failure" indication.

2.2. UPSTREAM_TSPEC Object

The format of an UPSTREAM_TSPEC object is the same as a SENDER_TSPEC object. This includes the definition of class types and their formats. The class number of the UPSTREAM_TSPEC object is 121 (of the form 0bbbbbbb).

2.2.1. Procedures

The UPSTREAM_TSPEC object describes the traffic flow that originates at the egress. The UPSTREAM_TSPEC object MUST be included in any Resv message that corresponds to a Path message containing an UPSTREAM_FLOWSPEC object. The C-Type of the UPSTREAM_TSPEC object MUST match the C-Type of the corresponding UPSTREAM_FLOWSPEC object. The contents of the UPSTREAM_TSPEC object MUST be constructed using a format and procedures consistent with those used to construct the FLOWSPEC object that will be used for the LSP, e.g., [RFC2210] or [RFC4328]. The contents of the UPSTREAM_TSPEC object MAY differ from contents of the UPSTREAM_FLOWSPEC object based on application data transmission requirements.

When an UPSTREAM_TSPEC object is received by an ingress, the ingress MAY determine that the original reservation is insufficient to satisfy the traffic flow. In this case, the ingress MAY issue a Path message with an updated UPSTREAM_FLOWSPEC object to modify the resources requested for the upstream traffic flow. This modification might require the LSP to be re-routed, and in extreme cases might result in the LSP being torn down when sufficient resources are not available.

2.3. UPSTREAM_ADSPEC Object

The format of an UPSTREAM_ADSPEC object is the same as an ADSPEC object. This includes the definition of class types and their formats. The class number of the UPSTREAM_ADSPEC object is 122 (of the form 0bbbbbbb).

2.3.1. Procedures

The UPSTREAM_ADSPEC object MAY be included in any Resv message that corresponds to a Path message containing an UPSTREAM_FLOWSPEC object. The C-Type of the UPSTREAM_TSPEC object MUST be consistent with the C-Type of the corresponding UPSTREAM_FLOWSPEC object. The contents of the UPSTREAM_ADSPEC object MUST be constructed using a format and procedures consistent with those used to construct the ADSPEC object that will be used for the LSP, e.g., [RFC2210] or [MEF-TRAFFIC]. The UPSTREAM_ADSPEC object is processed using the same procedures as the ADSPEC object and, as such, MAY be updated or added at transit nodes.

3. Packet Formats

This section presents the RSVP message-related formats as modified by this section. This document modifies formats defined in [RFC2205], [RFC3209], and [RFC3473]. See [RFC5511] for the syntax used by RSVP. Unmodified formats are not listed. Three new objects are defined in this section:

Object name	Applicable RSVP messages
-----	-----
UPSTREAM_FLOWSPEC	Path, PathTear, PathErr, and Notify (via sender descriptor)
UPSTREAM_TSPEC	Resv, ResvConf, ResvTear, ResvErr, and Notify (via flow descriptor list)
UPSTREAM_ADSPEC	Resv, ResvConf, ResvTear, ResvErr, and Notify (via flow descriptor list)

The format of the sender description for bidirectional asymmetric LSPs is:

```

<sender descriptor> ::= <SENDER_TEMPLATE> <SENDER_TSPEC>
                        [ <ADSPEC> ]
                        [ <RECORD_ROUTE> ]
                        [ <SUGGESTED_LABEL> ]
                        [ <RECOVERY_LABEL> ]
                        <UPSTREAM_LABEL>
                        <UPSTREAM_FLOWSPEC>

```

The format of the flow descriptor list for bidirectional asymmetric LSPs is:

```
<flow descriptor list> ::= <FF flow descriptor list>
                           | <SE flow descriptor>

<FF flow descriptor list> ::= <FLOWSPEC>
                              <UPSTREAM_TSPEC> [ <UPSTREAM_ADSPEC> ]
                              <FILTER_SPEC>
                              <LABEL> [ <RECORD_ROUTE> ]
                              | <FF flow descriptor list>
                              <FF flow descriptor>

<FF flow descriptor> ::= [ <FLOWSPEC> ]
                          [ <UPSTREAM_TSPEC> ] [ <UPSTREAM_ADSPEC> ]
                          <FILTER_SPEC> <LABEL>
                          [ <RECORD_ROUTE> ]

<SE flow descriptor> ::= <FLOWSPEC>
                        <UPSTREAM_TSPEC> [ <UPSTREAM_ADSPEC> ]
                        <SE filter spec list>

<SE filter spec list> is unmodified by this document.
```

4. Compatibility

This extension reuses and extends semantics and procedures defined in [RFC2205], [RFC3209], and [RFC3473] to support bidirectional LSPs with asymmetric bandwidth. To indicate the use of asymmetric bandwidth, three new objects are defined. Each of these objects is defined with class numbers in the form 0bbbbbbb. Per [RFC2205], nodes not supporting this extension will not recognize the new class numbers and should respond with an "Unknown Object Class" error. The error message will propagate to the ingress, which can then take action to avoid the path with the incompatible node or may simply terminate the session.

5. IANA Considerations

IANA has assigned new values for namespaces defined in this section and reviewed in this subsection.

The IANA has made the assignments described below in the "Class Names, Class Numbers, and Class Types" section of the "RSVP PARAMETERS" registry.

5.1. UPSTREAM_FLOWSPEC Object

A new class named UPSTREAM_FLOWSPEC has been created in the 0bbbbbbb range (120) with the following definition:

Class Types or C-types:

Same values as FLOWSPEC object (C-Num 9)

5.2. UPSTREAM_TSPEC Object

A new class named UPSTREAM_TSPEC has been created in the 0bbbbbbb range (121) with the following definition:

Class Types or C-types:

Same values as SENDER_TSPEC object (C-Num 12)

5.3. UPSTREAM_ADSPEC Object

A new class named UPSTREAM_ADSPEC has been created in the 0bbbbbbb range (122) with the following definition:

Class Types or C-types:

Same values as ADSPEC object (C-Num 13)

6. Security Considerations

This document introduces new message objects for use in GMPLS signaling [RFC3473] -- specifically the UPSTREAM_TSPEC, UPSTREAM_ADSPEC, and UPSTREAM_FLOWSPEC objects. These objects parallel the existing SENDER_TSPEC, ADSPEC, and FLOWSPEC objects but are used in the opposite direction. As such, any vulnerabilities that are due to the use of the old objects now apply to messages flowing in the reverse direction.

From a message standpoint, this document does not introduce any new signaling messages or change the relationship between LSRs that are adjacent in the control plane. As such, this document introduces no additional message- or neighbor-related security considerations.

See [RFC3473] for relevant security considerations, and [SEC-FRAMEWORK] for a more general discussion on RSVP-TE security discussions.

7. References

7.1. Normative References

- [RFC2205] Braden, R., Ed., Zhang, L., Berson, S., Herzog, S.,
and S. Jamin, "Resource ReSerVation Protocol (RSVP)
-- Version 1 Functional Specification", RFC 2205,
September 1997.
- [RFC2210] Wroclawski, J., "The Use of RSVP with IETF Integrated
Services", RFC 2210, September 1997.
- [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., Ed., "Generalized Multi-Protocol Label
Switching (GMPLS) Signaling Resource ReserVation
Protocol-Traffic Engineering (RSVP-TE) Extensions",
RFC 3473, January 2003.

7.2. Informative References

- [GMPLS-PBBTE] Fedyk, D., et al "GMPLS Control of Ethernet", Work in
Progress, July 2008.
- [RFC6003] Papadimitriou, D., "MEF Ethernet Traffic Parameters,"
RFC 6003, October 2008.
- [RFC5654] B. Niven-Jenkins, Ed., D. Brungard, Ed. and
M. Betts, Ed., "Requirements of an MPLS Transport
Profile," RFC 5654, September 2009.
- [RFC4606] Mannie, E. and D. Papadimitriou, "Generalized Multi-
Protocol Label Switching (GMPLS) Extensions for
Synchronous Optical Network (SONET) and Synchronous
Digital Hierarchy (SDH) Control", RFC 4606, August
2006.

- [RFC4328] Papadimitriou, D., Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Extensions for G.709 Optical Transport Networks Control", RFC 4328, January 2006.
- [RFC5511] Farrel, A. "Reduced Backus-Naur Form (RBNF) A Syntax Used in Various Protocol Specifications", RFC 5511, April 2009.
- [RFC5920] Fang, L., Ed., "Security Framework for MPLS and GMPLS Networks", RFC 5920, July 2010.
- [RFC5467] L. Berger, A. Takacs, D. Caviglia, D. Fedyk and J. Meuric, "GMPLS Asymmetric Bandwidth Bidirectional Label Switched Paths (LSPs)", RFC 5467, March 2009.

Authors' Addresses

Lou Berger
LabN Consulting, L.L.C.

Email: lberger@labn.net

Attila Takacs
Ericsson
Laborc u. 1.
Budapest, 1037
Hungary

Email: attila.takacs@ericsson.com

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

Phone: +390106003738

Fax:

Email: diego.caviglia@ericsson.com

Don Fedyk
Alcatel-Lucent
Groton, MA
USA

Email: donald.fedyk@alcatel-lucent.com

Julien Meuric
France Telecom Orange
2, avenue Pierre Marzin
Lannion Cedex, 22307
France

Email: julien.meuric@orange-ftgroup.com

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

Q. Wang
X. Fu
ZTE Corporation
Oct 25, 2010

GMPLS extensions to communicate latency as a TE performance metric
draft-wang-ccamp-latency-te-metric-01

Abstract

Latency is such requirement that must be achieved according to the SLA signed between customers and service providers, so mechanism is needed to collect, compute and identify the latency by signaling and routing protocol.

This document describes the requirement and method to compute and identify the latency by control plane in today's network which is consisted of packet transport network and optical transport network in order to meet the latency SLA of the customer. This document also describes RSVP-TE signaling and OSPF routing extensions needed to support the computation and identification of latency. These extensions are intended to advertise and convey the information of node latency and link latency as TE 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 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	4
2. Terminology	4
2.1. List of Acronyms	5
3. Analysis of the Latency Measurement Mechanism	5
3.1. Support of SLA	6
3.2. Latency Value	6
3.3. Latency of Server Layer Network	7
3.4. Role of the Control Plane	7
3.5. Impact of the Change of Link Latency	8
4. A New Latency Measurement Mechanism	8
4.1. Advertisement of the Latency Value	8
4.2. Latency Collection and Verification	9
5. Signaling and Routing Extensions to Support Latency Measurement	9
5.1. Routing Extensions to Support the Advertisement of Latency	10
5.2. Signaling Extensions to Support the Latency Measurement	11
6. Security Considerations	11
7. IANA Considerations	11
8. References	12
8.1. Normative References	12
8.2. Informative References	12
Authors' Addresses	12

1. Introduction

In a network, latency, a synonym for delay, is an expression of how much time it takes for a packet of data to get from one designated point to another. In some usages, latency is measured by sending a packet that is returned to the sender and the round-trip time is considered the latency. In this document, we refer to the former expression.

In many cases, latency is a sensitive topic. For example, two stock exchanges, one in Beijing, which is a city of north China and another in Shenzhen, which is a city of south China. Both of them need to synchronize with each other. A little change may result in large loss. So something SHOULD be assured that the network path latency MUST be limited to a value lower than the upper limit. SLA contract which includes the requirement of latency is signed between service providers and customers. In the future, latency demand will be needed by more and more customers.

Measurement mechanism of link latency has been defined in many technologies. For example, the measurement mechanism of 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. The detailed measurement mechanism of link latency is out of scope of this document. You can refer to ITU-T G.709 for more messages. Technologies that do not support the measurement of latency SHOULD be developed to allow the measurement of link latency in scenario similar to the above. This is out of scope of this document. Node latency can also be recorded at each node by recording the process time at the beginning and at the end. More detail of the node latency is described in section 3.2.

Current operation and maintenance mode of latency measurement is high in cost and low in efficiency. Only after the path needed by the customers' business is determined, signal can be sent to detect whether the latency of the path fit the requirement of the customers. If not, another path SHOULD be determined by the ingress node until one can. So a low cost and high efficiency latency measurement method SHOULD be provided in order to support the SLA. However, the control plane does not provide latency measure mechanism. A new method is provided that the node latency, link latency and latency variation can be collected by control plane from the transport plane. Then node latency, link latency values and latency variation can be

used by service provider through control plane to provide a path correspond with the customers' requirement. As there is demand from the customer, this method can be used to select a path correspond with customers' latency demand. In this document, link latency refers to the latency of the link between two neighbor nodes or a FA-LSP.

This document describes the requirement and method to compute and identify the latency by control plane in today's network which is consisted of packet transport network and optical transport network in order to meet the latency SLA of the customer. This document also describes RSVP-TE signaling and OSPF routing extensions needed to support the computation and identification of latency. Latency can be divided into two types as described above: node latency which is provided by the node as a result of process time at each node and link latency as a result of packet traverse between two neighbor nodes or a FA-LSP. Latency variation is also a parameter that is used to indicate the variation range of the latency value. Extensions are also intended to advertise and convey the information of node latency, link latency and latency variation as TE performance metric.

[RFC4203] details the OSPF extensions in support of Generalized Multi-Protocol Label Switching (GMPLS). In order to support the advertisement of the attributes of the node latency, link latency and latency variation by routing, extensions SHOULD be made to [RFC4203] in this document. Thus ingress node that is responsible for the creation of the path will have a good knowledge of the latency of the path.

[RFC3473] details the Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions. Extensions SHOULD be made to [RFC3473] to collect the node, link latency and latency variation along the path, so egress node can determine whether such a path is adaptive. This extensions is not necessary unless there is a need.

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. Terminology

The reader is assumed to be familiar with the terminology in [RFC3473] and [RFC4203].

Frame Delay:

The definition of Frame Delay in ITU-T Y.1731 can be seen below. Frame Delay can be specified as round-trip delay for a frame, where Frame Delay is defined as the time elapsed since the start of transmission of the first bit of the frame by a source node until the reception of the last bit of the loop backed frame by the same source node, when the loop back is performed at the frame's destination node.

Frame Delay Variation:

The definition of Frame Delay in ITU-T [Y.1731] can be seen below. Frame Delay Variation is a measure of the variations in the Frame Delay between a pair of service frames.

Path Monitoring & Tandem Connection Monitoring:

Path Monitoring & Tandem Connection Monitoring is a field contained in [G.709] OTN ODUk overhead, which can be used to support the measurement of latency between two OTN nodes.

Service Level Agreement:

A service level agreement is a part of a service contract where the level of service is formally defined between service providers and customers.

2.1. List of Acronyms

FD: Frame Delay

FDV: Frame Delay Variation

PM&TCM: Path Monitoring & Tandem Connection Monitoring

SLA: Service Level Agreement

3. Analysis of the Latency Measurement Mechanism

As described in the Introduction section, latency is sensitive in many cases like finance, storage. A little frame delay may result in large loss. So network latency values MUST be strictly limited to a value lower than the upper limit described in the SLA. Latency measurement mechanism is important to certain customers. However, the control plane does not provide latency measure mechanism. A method is provided that the node latency, link latency and latency variation can be collected by control plane from the latency measurement of the transport plane. Then node latency, link latency values and latency variation can be used by service provider through control plane to provide a path correspond with the customers' demand. In this document, link latency refers to the latency of the link between two neighbor nodes or a FA-LSP. This section analyzes latency support for SLA contract signed between customers and

providers, analysis of the mechanism of latency measurement, latency of the server layer network and role of the control plane in this new latency measurement mechanism.

3.1. Support of SLA

In today's network (e.g., DWDM), latency measurement is required by many service providers because of the demand from the customers. Latency is especially important for the customers who provide service like finance, storage. As a result of the demand, SLA contract which includes the demand of latency is signed between service providers and customers. According to the definition in section 2, SLA (i.e., Service Level Agreement) is a part of a service contract where the level of service is formally defined between service providers and customers. Service providers MUST provide accurate latency measurement result to the customers per SLA levels. Latency to different customers can be different per SLA levels.

However, current operation and maintenance mode of latency measurement through transport plane is high in cost and low in efficiency. Only after the path needed by the customers' business is determined, signal can be sent to detect whether the latency of the path fit the requirement of the customers. A new method described in this document is provided to support a low cost and high efficiency latency measurement mechanism in order to support the SLA. This can be seen in the 4th section and 5th section.

3.2. Latency Value

The mechanism of latency measurement can be sorted into two types. In order to monitor the performance, pro-active latency measurement is required. Generally, every 15 minutes or 24 hours, the value of FD and FDV 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).

As described in [CL-REQ], when a traffic flow moves from one component link to another in the same composite link between a set of nodes (or sites), it MUST be processed in a minimally disruptive manner. When a traffic flow moves from a current link to a target link with different latency, reordering can occur if the target link latency is less than that of the current and clumping can occur if target link latency is more than that of the current. Therefore, the solution SHALL provide a means to indicate that a traffic flow shall select a component link with the minimum latency value and a maximum acceptable latency value.

Similarly, the value of latency is not fixed because of different

signal process technology (The packet transport network use statistical multiplexing and the optical transport network use time division multiplex). For example, in statistical multiplexing business, latency for every business may be different because of the existence of buffering and priority. At this time, average latency value is needed when refer to node latency. Average latency value of node can be derived through the computation of the node or management plane configuration.

latency variation is also needed in the case the latency value of, for example, average latency value's variation range.

Measurement mechanism of link latency has been defined in many technologies like Ethernet, OTN. You can refer to ITU-T [G.8021], [Y.1731] and [G.709] for more information.

3.3. Latency of Server Layer Network

When a LSP traverses a server layer FA-LSP, the latency information of the FA-LSP SHOULD be provided by signaling protocol message if needed. Extension to the current signaling protocol is done to carry the latency information of the server layer FA-LSP. This is described in section 4 and section 5.

The boundary nodes of the FA-LSP SHOULD be aware of the latency information of this FA-LSP (i.e., minimum latency, maximum latency, average latency). If the latency information of the FA-LSP changes, the ingress node of the FA-LSP will receive the TE link information advertisement including the latency value which is already changed, then it will compute the total latency value of the FA-LSP again. If this value changes, the client layer of the FA-LSP MUST also be notified about the total value of the latency.

The ingress node or egress node of the FA-LSP can advertise the total value of the latency to the client layer nodes connecting to the ingress node or egress node through signaling protocol message (e.g., notify message or refresh message). If the FA-LSP is able to form a routing adjacency and/or as a TE link in the client network, the value of the FA-LSP can be used as TE link metric and advertised into the client layer routing instances or PCE.

3.4. Role of the Control Plane

Current mechanism of latency measurement is provided by transport plane instead of control plane. The latency information between two specified nodes will be detected if there is latency demand of the path between the two nodes. This is low in efficiency and high in cost if the latency information does not correspond with the

customers' demand.

A new method of latency measurement mechanism is provided by collecting the node latency value, link latency value between two neighbor nodes or a FA-LSP and latency variation, then these values is provided to the control plane. Control plane can compute a path correspond with customers' demand with these latency values.

3.5. Impact of the Change of Link Latency

If the link latency of a LSP which have a latency value corresponds with customers' demand changes, the ingress node or PCE will be aware of the latency value change in the network. Total latency value of the LSP affected by the latency value change will be re-computed through the ingress node or PCE. Client service SHOULD be switched to a new LSP which have a latency value corresponds with customers' demand if current changed latency value is invalid. This is much like the recovery, but not recovery. All the LSPs affected by this latency change may not be rerouted to find appropriate LSPs if they still have appropriate latency values. All the LSPs affected will be rerouted to find a recovery path if there is a link failure.

As a result of the change of link latency in the LSP, current LSP may be frequently switched to a new LSP with a appropriate latency value. In order to avoid this, solution SHOULD indicate the switchover of the LSP according to maximum acceptable value of the customers.

4. A New Latency Measurement Mechanism

This new latency measurement can be divided into two phases. The first phase is the advertisement of the latency information by routing protocol, including node latency, link latency between two neighbor nodes or a FA-LSP and latency variation, so every node in the network can be aware of the latency of every node and link. The second phase is the latency collection and verification along the path from the ingress node to the egress node by signaling protocol, so an adaptive LSP can be found out and verified.

4.1. Advertisement of the Latency Value

As described in the introduction section, a node in the packet transport network or optical transport network can detect link latency value which has connection with it. Also the node latency can be recorded at every node. Then these link latency values of the neighbor nodes, node latency and latency variation is notified to the control plane. The control plane instances then advertise these link latency values, node latency values and latency variation as

attributes of the TE link to the other nodes in the routing domain or PCE by routing protocol. If any latency values change, then the change MUST be notified to the control plane instances, then advertise by routing protocol in the routing domain or to the PCE. As a result, control plane instances and PCE can have every node latency values, link latency values and latency variation in the network.

4.2. Latency Collection and Verification

When the PCE receives the request which indicates the demand of latency, PCE can compute a path which satisfies customers' latency demand with the node latency values, link latency values and latency variation in the network. The ingress node initializes the creation of the LSP with path signaling message which includes the latency demand parameter. The path signaling message collects the node latency value, link latency value and latency variation along the path. When the path signaling message reaches the egress node, the egress node can verify whether the value of the latency is applicable by comparing the LSP latency with the latency demand parameter carried in the path message. Similarly, when egress node returns recv signaling message to ingress node, node latency values, link latency values and latency variation will also be gathered in the reverse direction. The ingress node verifies whether the latency values from the egress node to the ingress node is applicable. This extensions is not necessary unless there is a need.

When a LSP traverses a server layer FA-LSP, the latency information of the FA-LSP is advertised by routing protocol and carried in the signaling message. The latency information of the server layer FA-LSP can be carried in the ERBO object which is defined in [draft-fuxh-ccamp-boundary-explicit-control-ext]. Region boundaries carried in ERBO contain one pair or multiple pair of nodes. One pair of boundary nodes indicates the head node and the end node of the FA-LSP (i.e., the region boundary). The latency values information of the FA-LSP between two boundary nodes is carried in the signaling message directly behind a pair of boundary nodes in the ERBO. Ingress node will re-compute the total latency value of the FA-LSP if the total latency value of the FA-LSP changes. The latency value of the FA-LSP SHOULD be announced to the client layer of the FA-LSP, also advertised in the routing domain.

5. Signaling and Routing Extensions to Support Latency Measurement

Extensions SHOULD be done to existing OSPF-TE routing protocol and RSVP-TE routing protocol, in order to support the advertisement, the collection and the verification of the latency values. In this

section, routing extensions and signaling extensions will be described.

5.1. Routing Extensions to Support the Advertisement of Latency

Some extensions to the existing OSPF-TE routing protocol to support the advertisement of the node latency value, link latency and latency variation value in the routing domain or to the PCE as TE metric. OSPF-TE routing protocol can be used to carry latency information by adding a sub-TLV to the TE link which is defined in [RFC4203]. The latency value can be used as constraint for routing computation and as a factor impacting the node and link performance.

As defined in [RFC3630] and [RFC4203], the top-level TLV can take one of two values (1) Router address or (2) Link. Node latency sub-TLV and link latency sub-TLV can be added behind the top-level TLV. The link latency sub-TLV has the same format as node latency TLV. They both include these parameters like minimum latency value, minimum latency variation value, maximum latency value, maximum latency variation value, average latency value, average latency variation value. The format of the sub-TLV can be seen below.

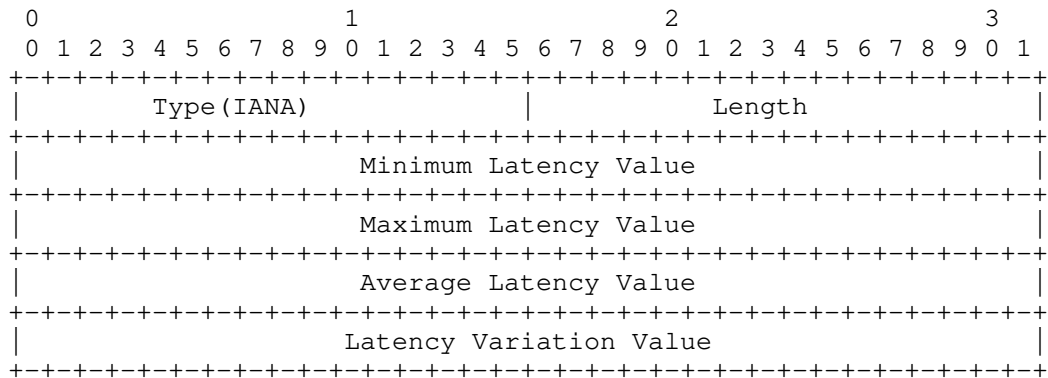


Figure 1: Format of the sub-TLV

- Minimum Latency Value: a value indicates the boundary of the node latency or link latency along with maximum latency value.
- Maximum Latency Value: a value indicates the boundary of the node latency or link latency along with maximum latency value.

- Average Latency Value: a value indicates the average of the node latency or link latency.
- Latency Variation Value: a value indicates the variation range of the minimum latency value, maximum latency value or average latency value.

5.2. Signaling Extensions to Support the Latency Measurement

Extensions SHOULD also be done to the RSVP-TE signaling protocol to support the collection and verification of the latency measurement. This can be achieved base on the extension to the RRO which is defined in [RFC3209] by adding an interface ID (i.e., IP Address) or interface identifier defined in [RFC3477], then adding the sub-TLV which has the same format with that described above. When a node receives the path message, node latency value, link latency value and latency variation along the path which has correlation to the node will be added behind the interface identifier and node ID sub-object. At the same time, the latency values requirement from the ingress node to the egress node have been added into the TE metric TLV. When the egress node receives the path message, the latency value of the LSP can be compute by the node latency value, link latency value and latency variation carried behind RRO. If the total latency value does not meet the requirement of the customer, patherr message SHOULD be created and return to the ingress node. Recv message can be used to collect and verify the latency information in the reverse direction in the same way.

The signaling format of the sub-TLV has the same format as that described in the section 5.1. This format can also been used behind a pair of boundary nodes which are carried in ERBO to indicate the latency information of the FA-LSP if there are requirement of the server layer.

6. Security Considerations

TBD

7. IANA Considerations

TBD

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.
- [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.

8.2. 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

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

Email: wang.qilei@zte.com.cn
URI: <http://wwwen.zte.com.cn/>

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://wwwen.zte.com.cn/>

Network work group
Internet Draft
Intended status: Standards Track

Fatai Zhang
Young Lee
Jianrui Han
Huawei
G. Bernstein
Grotto Networking
Yunbin Xu
CATR
September 10, 2010

Expires: March 10, 2011

OSPF-TE Extensions for General Network Element Constraints

draft-zhang-ccamp-general-constraints-ospf-ext-00.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 March 10, 2011.

Abstract

Generalized Multiprotocol Label Switching can be used to control a wide variety of technologies including packet switching (e.g., MPLS), time-division (e.g., SONET/SDH, OTN), wavelength (lambdas), and

spatial switching (e.g., incoming port or fiber to outgoing port or fiber). 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 describes OSPF routing protocol extensions to support these kinds of constraints under the control of Generalized MPLS (GMPLS).

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.....	2
2. Node Information.....	3
2.1. Connectivity Matrix.....	4
3. Link Information.....	4
3.1. Port Label Restrictions.....	5
3.2. Available Labels.....	5
3.3. Shared Backup Labels.....	6
4. Routing Procedures.....	6
5. Security Considerations.....	7
6. IANA Considerations.....	7
6.1. Node Information.....	7
6.2. Link Information.....	7
7. References.....	8
8. Authors' Addresses.....	9
Acknowledgment.....	11

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 should be capable of performing 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.

[GEN-Encode] 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.

This document defines extensions to the OSPF routing protocol based on [GEN-Encode] to enhance the Traffic Engineering (TE) properties of GMPLS TE which are defined in [RFC3630], [RFC4202], and [RFC4203]. The enhancements to the Traffic Engineering (TE) properties of GMPLS TE links can be announced in OSPF TE LSAs. The TE LSA, which is an opaque LSA with area flooding scope [RFC3630], has only one top-level Type/Length/Value (TLV) triplet and has one or more nested sub-TLVs for extensibility. The top-level TLV can take one of three values (1) Router Address [RFC3630], (2) Link [RFC3630], (3) Node Attribute [RFC5786]. In this document, we enhance the sub-TLVs for the Link TLV and Node Attribute TLV in support of the general network element constraints under the control of GMPLS.

The detailed encoding of OSPF extensions are not defined in this document. [GEN-Encode] provides encoding detail.

2. Node Information

According to [GEN-Encode], the additional node information representing node switching asymmetry constraints includes Node ID, connectivity matrix. Except for the Node ID which should comply with Routing Address described in [RFC3630], the other pieces of information are defined in this document.

[RFC5786] defines a new top TLV named the Node Attribute TLV which carries attributes related to a router/node. This Node Attribute TLV contains one or more sub-TLVs.

Per [GEN-Encode], we have identified the following new Sub-TLVs to the Node Attribute TLV. Detail description for each newly defined Sub-TLV is provided in subsequent sections:

Sub-TLV Type	Length	Name
TBD	variable	Connectivity Matrix

In some specific technologies, e.g., WSON networks, Connectivity Matrix sub-TLV may be optional, which depends on the control plane implementations. Usually, for example, in WSON networks, Connectivity Matrix sub-TLV may appear in the LSAs because WSON switches are asymmetric at present. It is assumed that the switches are symmetric switching, if there is no Connectivity Matrix sub-TLV in the LSAs.

2.1. Connectivity Matrix

It is necessary to identify which ingress ports and labels can be switched to some specific labels on a specific egress port, if the switching devices in some technology are highly asymmetric.

The Connectivity Matrix is used to identify these restrictions, which can represent either the potential connectivity matrix for asymmetric switches (e.g. ROADMs and such) or fixed connectivity for an asymmetric device such as a multiplexer as defined in [WSON-Info].

The Connectivity Matrix is a sub-TLV (the type is TBD by IANA) of the Node Attribute TLV. The length is the length of value field in octets. The meaning and format of this sub-TLV are defined in Section 5.3 of [GEN-Encode]. One sub-TLV contains one matrix. The Connectivity Matrix sub-TLV may occur more than once to contain multi-matrices within the Node Attribute TLV.

3. Link Information

The most common link sub-TLVs nested to link top-level TLV are already defined in [RFC3630], [RFC4203]. For example, Link ID, Administrative Group, Interface Switching Capability Descriptor (ISCD), Link Protection Type, Shared Risk Link Group Information (SRLG), and Traffic Engineering Metric are among the typical link sub-TLVs.

Per [GEN-Encode], we add the following additional link sub-TLVs to the link-TLV in this document.

Sub-TLV Type	Length	Name
TBD	variable	Port Label Restrictions
TBD	variable	Available Labels
TBD	variable	Shared Backup Labels

Generally all the sub-TLVs above are optional, which depends on the control plane implementations. If it is default no restrictions on labels, Port Label Restrictions sub-TLV may not appear in the LSAs. In order to be able to compute label assignment, Available Labels sub-TLV may appear in the LSAs. For example, in WSON networks, without available wavelength information, path computation need guess

what lambdas may be available (high blocking probability or distributed wavelength assignment may be used).

3.1. Port Label Restrictions

Port label restrictions describe the label restrictions that the network element (node) and link may impose on a port. These restrictions represent 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.

For example, Port Label Restrictions describes the wavelength restrictions that the link and various optical devices such as OXCs, ROADMs, and waveband multiplexers may impose on a port in WSON. These restrictions represent what wavelength may or may not be used on a link and are relatively static. The detailed information about Port label restrictions is described in [WSON-Info].

The Port Label Restrictions is a sub-TLV (the type is TBD by IANA) of the Link TLV. The length is the length of value field in octets. The meaning and format of this sub-TLV are defined in Section 5.4 of [GEN-Encode]. The Port Label Restrictions sub-TLV may occur more than once to specify a complex port constraint within the link TLV.

3.2. Available Labels

Available Labels indicates the labels available for use on a link as described in [GEN-Encode]. The Available Labels is a sub-TLV (the type is TBD by IANA) of the Link TLV. The length is the length of value field in octets. The meaning and format of this sub-TLV are defined in Section 5.1 of [GEN-Encode]. The Available Labels sub-TLV may occur at most once within the link TLV.

Note that there are five approaches for Label Set which is used to represent the Available Labels described in [GEN-Encode]. Usually, it depends on the implementation to one of the approaches. In WSON networks, considering that the continuity of the available or unavailable wavelength set can be scattered for the dynamic wavelength availability, so it may burden the routing to reorganize the wavelength set information when the Inclusive (/Exclusive) List (/Range) approaches are used to represent Available Wavelengths information. Therefore, it is RECOMMENDED that only the Bitmap Set be used for representation Available Wavelengths information.

The "Base Label" and "Last Label" in label set defined in [GEN-Encode] corresponds to base wavelength label and last wavelength label in WSON, the format of which is described 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	2	3	4	5	6	7	8	9
Grid										C.S.										Reserved										n									

The detailed information related to wavelength label can be referred to [Lambda-Labels].

3.3. Shared Backup Labels

Shared Backup Labels indicates the labels available for shared backup use on a link as described in [GEN-Encode].

The Shared Backup Labels is a sub-TLV (the type is TBD by IANA) of the Link TLV. The length is the length of value field in octets. The meaning and format of this sub-TLV are defined in Section 5.2 of [GEN-Encode]. The Shared Backup Labels sub-TLV may occur at most once within the link TLV.

4. Routing Procedures

All the sub-TLVs are nested to top-level TLV(s) and contained in Opaque LSAs. The flooding of Opaque LSAs must follow the rules specified in [RFC2328], [RFC2370], [RFC3630], [RFC4203] and [RFC5786].

Considering the routing scalability issues in some cases, the routing protocol should be capable of supporting the separation of dynamic information from relatively static information.

In the WSON networks, the node information and link information can be classified as two kinds: one is relatively static information such as Node ID, Connectivity Matrix information; the other is dynamic information such as Available Wavelengths information. [GEN-Encode] give recommendations of typical usage of previously defined sub-TLVs which contain relatively static information and dynamic information. An implementation SHOULD take measures to avoid frequent updates of relatively static information when the relatively static information is not changed.

For node information, since the Connectivity Matrix information is static, the LSA containing the Node Attribute TLV can be updated with a lower frequency to avoid unnecessary updates.

For link information, a mechanism MAY be applied such that static information and dynamic information of one TE link are contained in separate Opaque LSAs, which are updated with different frequencies, to avoid unnecessary updates of static information when dynamic information is changed.

Note that as with other TE information, an implementation SHOULD take measures to avoid rapid and frequent updates of routing information that could cause the routing network to become swamped. A threshold mechanism MAY be applied such that updates are only flooded when a number of changes have been made to the label availability (e.g., wavelength availability) information within a specific time. Such mechanisms MUST be configurable if they are implemented.

5. Security Considerations

This document does not introduce any further security issues other than those discussed in [RFC 3630], [RFC 4203].

6. IANA Considerations

[RFC3630] says that the top level Types in a TE LSA and Types for sub-TLVs for each top level Types must be assigned by Expert Review, and must be registered with IANA.

IANA is requested to allocate new Types for the sub-TLVs as defined in Sections 2.1, 3.1, 3.2 and 3.3 as follows:

6.1. Node Information

This document introduces the following sub-TLVs of Node Attribute TLV (Value TBD, see [RFC5786]):

Type	sub-TLV
TBD	Connectivity Matrix

6.2. Link Information

This document introduces the following sub-TLVs of TE Link TLV (Value 2):

Type	sub-TLV
TBD	Port Label Restrictions
TBD	Available Labels
TBD	Shared Backup Labels

7. 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.
- [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.
- [RFC3945] E. Mannie, Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Architecture", RFC 3945, October 2004.
- [RFC4655] Farrel, A., Vasseur, J.-P., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, August 2006.
- [RFC2328] Moy, J., "OSPF Version 2", STD 54, RFC 2328, April 1998.
- [RFC2370] Coltun, R., "The OSPF Opaque LSA Option", RFC 2370, July 1998.
- [RFC5786] R. Aggarwal and K. Kompella, "Advertising a Router's Local Addresses in OSPF Traffic Engineering (TE) Extensions", RFC 5786, March 2010.

- [Lambda-Labels] T. Otani, H. Guo, K. Miyazaki, D. Caviglia, "Generalized Labels for Lambda-Switching Capable Label Switching Routers", work in progress: draft-ietf-ccamp-gmpls-g-694-lambda-labels-07.txt, April 2010.
- [WSON-Frame] Y. Lee, G. Bernstein, W. Imajuku, "Framework for GMPLS and PCE Control of Wavelength Switched Optical Networks (WSON)", work in progress: draft-ietf-ccamp-rwa-WSON-Framework-06.txt, April 2010.
- [WSON-Info] Y. Lee, G. Bernstein, D. Li, W. Imajuku, "Routing and Wavelength Assignment Information Model for Wavelength Switched Optical Networks", work in progress: draft-ietf-ccamp-rwa-info-09.txt, September 2010.
- [RWA-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-05.txt, July 2010.
- [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-02.txt, June 2010.

8. 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

Young Lee
Huawei Technologies
1700 Alma Drive, Suite 100
Plano, TX 75075
USA

Phone: (972) 509-5599 (x2240)
Email: ylee@huawei.com

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

Greg Bernstein
Grotto Networking
Fremont CA, USA

Phone: (510) 573-2237
Email: gregb@grotto-networking.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

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

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

Ming Chen
European Research Center
Huawei Technologies
Riesstr. 25, 80992 Munchen, Germany

Phone: 0049-89158834072
Email: minc@huawei.com

Yabin Ye
European Research Center
Huawei Technologies
Riesstr. 25, 80992 Munchen, Germany

Phone: 0049-89158834074
Email: yabin.ye@huawei.com

Acknowledgment

We thank Ming Chen and Yabin Ye from DICONNET Project who provided valuable information for this document.

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.

Full Copyright Statement

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
Huawei
Guoying Zhang
CATR
Sergio Belotti
Alcatel-Lucent
D. Ceccarelli
Ericsson
October 21, 2010

Expires: April 21 2011

Generalized Multi-Protocol Label Switching (GMPLS) Signaling
Extensions for the evolving G.709 Optical Transport Networks Control

draft-zhang-ccamp-gmpls-evolving-g709-06.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 April 21, 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.....	4
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 Traffic Parameter.....	6
4.2. Example of ODUFlex Traffic Parameter.....	7
5. Generalized Label.....	8
5.1. New definition of ODUk Label.....	8
5.2. Examples.....	10
5.3. Label Distribution Procedure.....	11
5.4. Control Plane Backward Compatibility Considerations.....	12
6. Tributary Port Number Assignment.....	13
6.1. TPN Object.....	13
6.2. Procedure of TPN Assignment.....	14
6.2.1. Downstream Node Assignment by Control Plane.....	14
6.2.2. Upstream Node Assignment by Control Plane.....	15
6.3. Collision Management.....	15
7. Security Considerations.....	15
8. IANA Considerations.....	16
9. References.....	16
9.1. Normative References.....	16
9.2. Informative References.....	17
10. Authors' Addresses.....	18
Acknowledgment.....	19

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 superseded 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].

[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:

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								
+-----																																							

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 (i.e., 1.25*N Gbps)
33~255	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.

For example, for an ODUflex(CBR) service with Bit_Rate = 2.5Gbps and Tolerance = 50ppm, the actual bandwidth of the ODUflex is:

$$2.5\text{Gbps} * (1 - 50\text{ppm}) \sim 2.5\text{Gbps} * (1 + 50\text{ppm})$$

In case of other ODUk signal types, the Bit_Rate and Tolerance fields are not necessary and MUST be filled with 0.

4.1. Usage of ODUflex 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:

$$N = \text{Ceiling of}$$

$$\frac{\text{ODUflex(CBR) nominal bit rate} * (1 + \text{ODUflex(CBR) bit rate tolerance})}{\text{ODUk.ts nominal bit rate} * (1 - \text{HO OPUk bit rate tolerance})}$$

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:

$$\text{Minimum bandwidth of ODUk.ts} = \text{ODUk.ts nominal bit rate} * (1 - \text{HO OPUk bit rate tolerance})$$

$$\text{Maximum bandwidth of ODTUk.ts} = \text{ODTUk.ts nominal bit rate} * (1 + \text{HO OPUk bit rate tolerance})$$

Where: HO OPUk bit rate tolerance = 20ppm

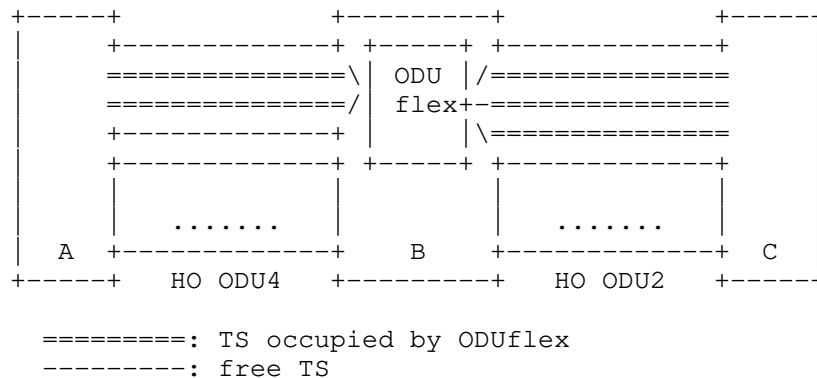
For different ODUk, the bandwidths of the tributary slot are different, and so the total number of tributary slots to be reserved for the ODUflex(CBR) may not be the same on different HO ODUk links. This is why the traffic parameter should bring the actual bandwidth information other than the NMC field.

[Editors note] In case of ODUflex(GFP), the calculation of the total number of tributary slots to be reserved along the path is now under discussion in ITU-T. Therefore, the traffic parameters for ODUflex(GFP) is for further study.

4.2. Example of ODUflex Traffic Parameter

This section gives an example to illustrate the usage of `ODUflex(CBR)` traffic parameter.

Assume there is an ODUflex(CBR) service requesting a bandwidth of (2.5Gbps, +/-20ppm) from node A to node C. In other words, the ODUflex traffic parameter indicates that Signal Type is 32 (ODUflex), Bit_Rate is 2.5Gbps and Tolerance is 20ppm.



- On the HO ODU4 link between node A and B:

The maximum bandwidth of the ODUflex equals $2.5\text{Gbps} * (1 + 20\text{ppm})$, and the minimum bandwidth of the tributary slot of ODU4 equals

1.301 683 217Gbps, so the total number of tributary slots N1 to be reserved on this link is:

$$N1 = \text{ceiling} (2.5\text{Gbps} * (1 + 20\text{ppm}) / 1.301\ 683\ 217) = 2$$

- On the HO ODU2 link between node B and C:

The maximum bandwidth of the ODUflex equals $2.5\text{Gbps} * (1 + 20\text{ppm})$, and the minimum bandwidth of the tributary slot of ODU2 equals 1.249 384 632Gbps, so the total number of tributary slots N2 to be reserved on this link is:

$$N2 = \text{ceiling} (2.5\text{Gbps} * (1 + 20\text{ppm}) / 1.249\ 384\ 632) = 3$$

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:

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										
ODUj										OD(T)Uk										T										Reserved										Bit Map									
																																																

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. 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)

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.

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 1| Reserved |          Padded Bits (0)          |
+-----+-----+-----+-----+-----+-----+-----+-----+

```

(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 |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.

- 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 |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.

- 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 |0 1 1 0 1 0 1 0 0 0 0 0 0 0 0 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.

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. Note that these labels can also be specified by the source node of the connection.

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.

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.

In case of ODUk to OTUk mapping, the size of Bit Map field is 0 and no additional procedure is needed.

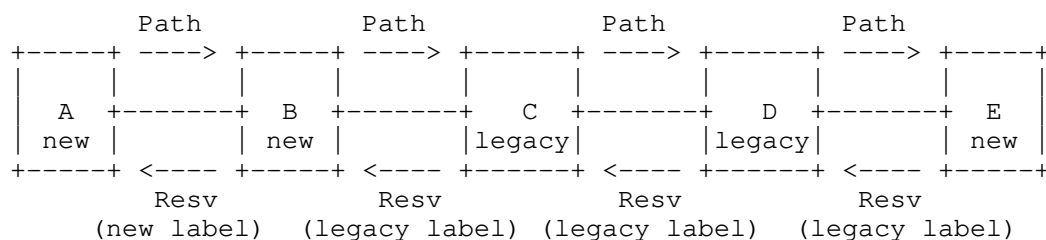
5.4. Control Plane Backward Compatibility Considerations

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 below:



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. Tributary Port Number Assignment

When an LO ODU_j is multiplexed into HO ODU_k occupying one or more TSs, a Tributary Port Number (TPN) 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.

For HO ODU2 or ODU3 link, the TPN value (6 bits) MUST be different from each other for one type of LO ODU. For HO ODU4 link, the TPN value (7 bits) MUST be different from each other for all types of LO ODU_j.

Referring to Section 4.2 of [OTN-info], the RSVP-TE signaling is necessary to be extended to support the TPN assignment function.

6.1. TPN Object

A new TPN object is introduced in the PATH and RESV message to support TPN assignment. The TPN object is optional and has the following format:

TPN Class-Num = xx (TBD), C_Type = 1

```

      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
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|D|           Reserved           |           TPN           |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

```

D (Downstream Assignment) (1 bit): indicates which node to assign the TPN. When set, the TPN is assigned by the downstream node; when cleared, the TPN is assigned by the upstream node.

TPN (16 bits): indicates the Tributary Port Number 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 object SHOULD not appear in the RSVP-TE message.

6.2. Procedure of TPN Assignment

Since the TPN is not needed in case of ODU mapping, the following sub-sessions are only applicable for the ODU multiplexing cases.

6.2.1. Downstream Node Assignment by Control Plane

In this case, the upstream node sends a PATH message, which contains a TPN Object with the D bit set to 1, to its downstream neighbor node to request creation of LO ODU_j. The TPN field in this object is set to 0 and MUST be ignored.

On receiving the PATH message, the downstream neighbor node performs a normal tributary slot selection and reservation in the selected HO ODU_k link. After that, the downstream node assigns a valid TPN, which does not collided with other TPN value used by existing LO ODU connections in the selected HO ODU link and configures the expected multiplex structure identifier (ExMSI) using this TPN. Then, the

assigned TPN is filled into the TPN Object and sent to the upstream neighbor node via the RESV message.

The upstream node, when receiving the RESV message, gets the TPN assigned by its downstream neighbor node and fills 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.

6.2.2. Upstream Node Assignment by Control Plane

In this case, the upstream node performs a normal tributary slot selection and reservation in the selected HO ODUk link for LO ODUj, and then assigns a valid TPN, which does not collided with other TPN value used by existing LO ODU connections in the selected HO ODU link, for the reserved tributary slot(s).

Then, the upstream node sends a PATH message, which contains the assigned TPN value in the TPN Object (D = 0) and contains the selected tributary slots information (e.g., via the existing LABEL_SET Object), to its downstream neighbor node to request creation of LO ODUj.

The downstream neighbor node, based on the received tributary slots information and the TPN value, configures the ExMSI in the data plane, so that the data plane MSI procedure can be performed, as described in the previous sub-session.

6.3. Collision Management

[Editors note] This chapter should indicate the procedure in case of collision between Tributary Port Numbers and/or Tributary Slots e.g. two different LSP setups may choose a disjoint set of Tributary Slots but they may request the same Tributary Port Number value (same MSI in G.709 OPUK field).

In this case the first signaling should be successful and the second one must fail.

7. 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.

8. IANA Considerations

- TPN Object:

A new value is needed to be defined by IANA for this document:

- o TPN Object (Session 6): Class-Num = xx (TBD), C-Type = 1

- 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.

9. References

9.1. 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.
- [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.
- [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.

9.2. 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.
- [G709-V3] ITU-T, "Interfaces for the Optical Transport Network (OTN)", G.709/Y.1331, December 2009.
- [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.

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

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: linyi_hw@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 I+D
October 21, 2010

Expires: April 21, 2011

GMPLS-based Hierarchy LSP creation
in Multi-Region and Multi-Layer Networks

draft-zhang-ccamp-gmpls-h-lsp-mln-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 April 21, 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], [RFC4206bis] and [MLN-EXT] 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
4. Requirements of SC and Switching Granularity Selection.....	6
4.1. Model 1: Pre-provisioning of FA-LSP.....	7
4.2. Model 2: Signaling trigger server layer path computation.....	8
4.3. Model 3: Full path computation at source node.....	8
5. ERO Sub-Object.....	9
5.1. Application of SERVER_LAYER_INFO sub-object.....	10
6. Security Considerations.....	11
7. IANA Considerations.....	11
8. Acknowledgments.....	11
9. References.....	11
10. Authors' Addresses.....	13

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 [MLN-EXT].

[RFC4206] and [RFC4206bis] 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 supporting multiple interface switching capabilities and multiple switching granularities, and then provides the extensions to the RSVP-TE protocol in order to

2. Terminology

3. Provisioning of FA-LSP in Server Layer Network

As described in [RFC5212], the edge node of a region always has multiple Interface Switching Capabilities (ISCs), i.e., it contains multiple matrices which may be connected to each other by internal links. Nodes with multiple Interface Switching Capabilities are further classified as "simplex" or "hybrid" nodes by [RFC5212] and [RFC5339], where the simplex node advertises several TE links each with a single ISC value carried in its ISCD sub-TLV, while the hybrid node advertises a single TE link containing more than one ISCD each with a different ISC value. An example hybrid node with a link having multiple ISCs is shown in Figure 1, copied from [RFC5339].



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.

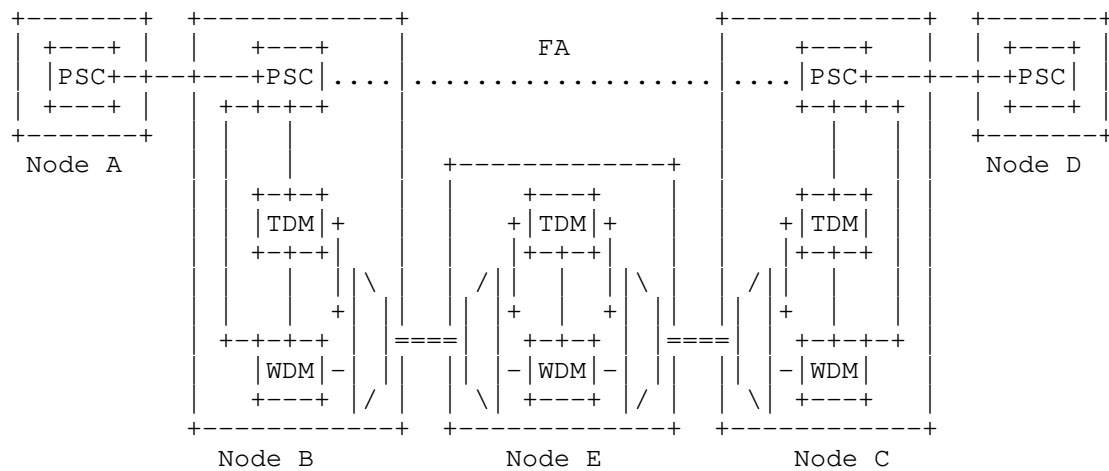


Figure 2 - MLN with multiple ISCs at edge node

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.

[MLN-EXT] 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.

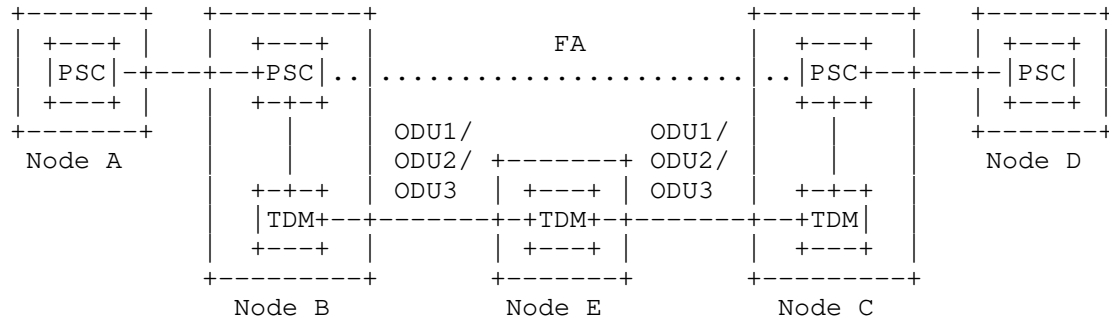


Figure 3a - Multiple switching granularities in server layer

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.

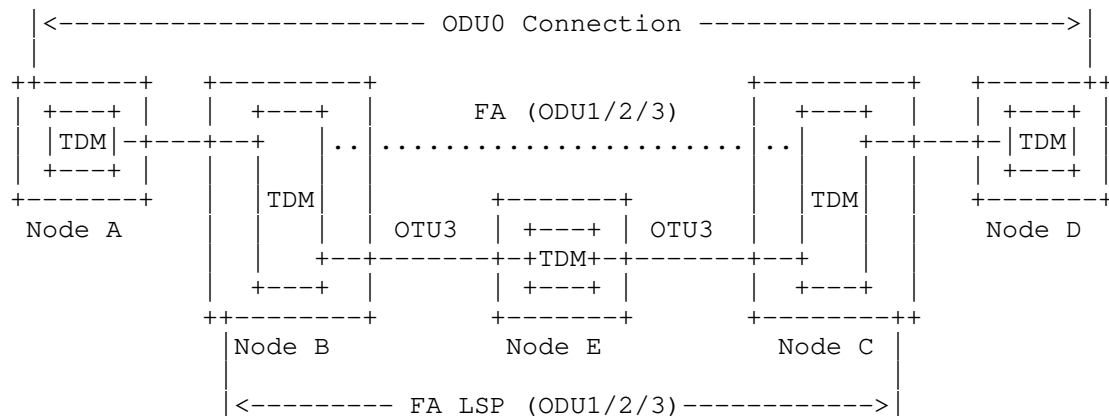


Figure 3b - TDM nested LSP provisioning

Figure 3b is another example multi-layer network within the same region. When there is a need to set up an FA between node B and C for the client layer ODU0 connection, the server layer has multiple choices, e.g., ODU1 or ODU2 or ODU3, for the FA-LSP if the multi-stage multiplexing is supported at node B and C.

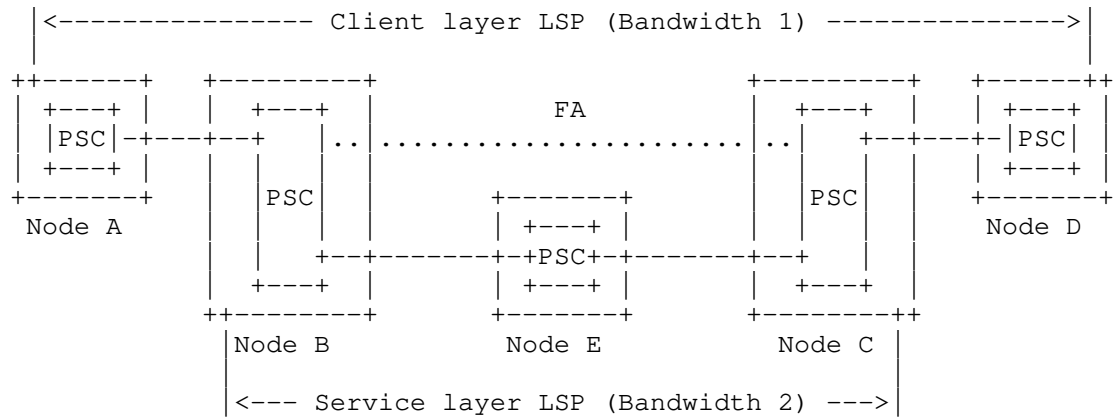


Figure 3c - PSC nested LSP provisioning

Figure 3c is a third example showing an LSP nesting scenario in a PSC signal-layer network (e.g., an MPLS-TP network). A PSC tunnel passing through node B, E and C is requested to carry the client layer LSP. There are multiple choices of the bandwidth of the tunnel, on the premise that the bandwidth of the FA-LSP is equal to or larger than the client layer LSP.

The selection of server layer switching matrix and switching granularity is based on both policy and bandwidth resources. The selection can be performed by planning tool and/or NMS/PCE/VNTM (Virtual Network Topology Manager, see [RFC5623]) and/or the network node.

4. Requirements of SC and Switching Granularity 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 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 a 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 [RFC4206bis].

It's possible that the source node of the client layer LSP selects the server layer SC and/or granularity 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 ([MLN-EXT]) 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 server layer granularity selection. Therefore, in this case, new extensions for server layer SC and switching granularity selection 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 PATH message of the client layer LSP. 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 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 server layer SC and switching granularity selection 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:

```

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
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|L|   Type   |   Length   |M|   Reserved   |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
| LSP Enc. Type |Switching Type |   Reserved   |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|                                     Traffic Parameters                                     |
~                                                                 ~
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

```

- 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 and Switching Type: These two fields are used to point out which switching layer is requested to set up the FA-LSP. The values of these two fields are inherited from the Generalized Label Request in GMPLS signaling, referring to [RFC3471], [RFC3473] and other related standards and drafts. Note that the G-PID of the Server layer FA-LSP can be deduced from the type of client layer LSP by these two fields.
- 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 [ETH-TP], [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.

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 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. Shiimoto 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.
- [RFC4206bis] K. Shiimoto, A. Farrel, "Procedures for Dynamically Signaled Hierarchical Label Switched Paths", draft-ietf-ccamp-lsp-hierarchy-bis-08.txt, February 2010.
- [MLN-EXT] Dimitri Papadimitriou et al, "Generalized Multi-Protocol Label Switching (GMPLS) Protocol Extensions for Multi-Layer and Multi-Region Networks (MLN/MRN)", draft-ietf-ccamp-gmpls-mln-extensions-12.txt, February 21, 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.
- [ETH-TP] D. Papadimitriou, "Ethernet Traffic Parameters", draft-ietf-ccamp-ethernet-traffic-parameters-10.txt, January 20, 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: 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: linyi_hw@huawei.com

Francisco Javier Jimenez Chico
Telefonica I+D
Emilio Vargas 6
Madrid, 28043 Spain

Phone: +34 913379037
Email: fjjc@tid.es

Oscar Gonzalez de Dios
Telefonica I+D
Emilio Vargas 6
Madrid, 28045 Spain

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

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
October 20, 2010

Expires: April 20, 2011

RSVP-TE Extensions for Configuration SRLG of an FA
draft-zhang-ccamp-srlg-fa-configuration-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 April 20, 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 Indication.....	4
3.2. SRLG Information Object.....	4
3.3. Signaling Procedures.....	5
4. Manageability Considerations.....	5
5. IANA Considerations.....	6
6. Security Considerations.....	6
7. References.....	6

1. Introduction

As described in [RFC4206], H-LSP (Hierarchical LSP) can be used for carrying one or more other LSPs. [LSP-Hierarchy-bis] 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 [LSP-Hierarchy-bis], 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 [LSP-

Hierarchy-bis], 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 [LSP-Hierarchy-bis].

In addition, Section 1.2 of [LSP-Hierarchy-bis] 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 Indication

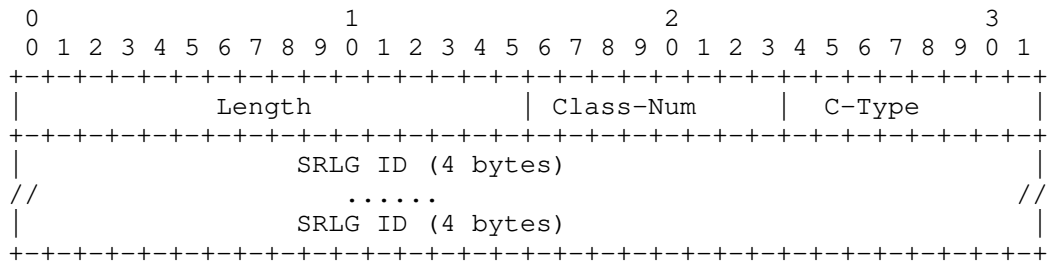
In order to indicate nodes that SRLG collection is desired, a new flag in the SESSION_ATTRIBUTE is needed:

0x08 SRLG recording desired

This flag indicate that SRLG information should be recorded along the LSP.

3.2. SRLG Information Object

An SRLG information object is defined to carry the SRLG information of the LSP. The Class-Num and the C-Type of the SRLG Information Object need to be assigned by the IANA.



The SRLG ID list carries the SRLG information of the LSP. This object can be carried in a Path/Resv message.

The SRLG ID can be added to the SRLG Information Object in a Path/Resv message hop by hop. Then the endpoints of the LSP can get the SRLG information of the path.

3.3. Signaling Procedures

When an LSP head node determines that it needs to get the SRLG information of the LSP, it sets the "SRLG recording desired" in the SESSION_ATTRIBUTE when it sends the Path message to the downstream node. The downstream nodes record the SRLG information in the SRLG Information Object hop by hop. Then the tail node of the LSP can get the SRLG information from the SRLG Information Object.

When the tail node of the LSP receives the Path message and the "SRLG recording desired" is set in the SESSION_ATTRIBUTE object, it can get the SRLG information from the SRLG Information Object of the Path message. Hence it can add the collected SRLG information into the SRLG Information Object of a Resv message which will be forwarded hop by hop in the upstream direction until it arrives the head node. Then the head node can also get the SRLG information of the LSP from the SRLG Information Object in the Resv message.

Based on the above procedure, the endpoints can get the SRLG information automatically. Then the endpoints can for instance configure the SRLG information and advertise it as a TE link to the routing instance based on the procedure described in [LSP-Hierarchy-bis].

4. Manageability Considerations

TBD.

5. IANA Considerations

TBD.

6. Security Considerations

TBD.

7. References

- [LSP-Hierarchy-bis] K. Shiimoto, A. Farrel, " Procedures for Dynamically Signaled Hierarchical Label Switched Paths ", draft-ietf-ccamp-lsp-hierarchy-bis-08, August 2010.
- [RFC 4206] K. Kompella, Y. Rekhter, " Label Switched Paths (LSP) Hierarchy with Generalized Multi-Protocol Label Switching (GMPLS) Traffic Engineering (TE) ", rfc4206, October 2005.
- [RFC 4874] CY. Lee, A. Farrel, S. De Cnodder, " Exclude Routes - Extension to Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) ", rfc4874, April 2007.
- [RFC 3477] K. Kompella, Y. Rekhter, " Signalling Unnumbered Links in Resource ReSerVation Protocol - Traffic Engineering (RSVP-TE) ", rfc3477, January 2003.
- [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.

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

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: April 28, 2011

F. Zhang, Ed.
F. Yang
L. Jin
W. Jiang
ZTE Corporation
October 25, 2010

RSVP-TE Extensions to Establish Associated Bidirectional LSP
draft-zhang-mps-tp-rsvp-tp-ext-associated-lsp-01

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.berger-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 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
2. Conventions used in this document	4
3. Extensions to the Extended ASSOCIATION object.	4
4. Association of Two Reverse Unidirectional LSPs	6
4.1. Asymmetric Bandwidth LSPs	7
5. IANA Considerations	8
6. Security Considerations	8
7. Acknowledgement	8
8. Normative references	8
Authors' Addresses	9

1. Introduction

The associated bidirectional path, defined in[RFC5654], is constructed from a pair of unidirectional paths that are associated with one another at the path's ingress/egress points. The forward and backward directions are setup, monitored, and protected independently.

[RFC5654] specifies the requirements about associated bidirectional paths in requirement 7, 11, 12, 50:

7 MPLS-TP MUST support associated bidirectional point-to-point transport paths.

11 The end points of an associated bidirectional transport path MUST be aware of the pairing relationship of the forward and reverse paths used to support the bidirectional service.

12 Nodes on the path of an associated bidirectional transport path where both the forward and backward directions transit the same node in the same (sub)layer as the path SHOULD be aware of the pairing relationship of the forward and the backward directions of the transport path.

50 The MPLS-TP control plane MUST support stabling associated bidirectional P2P path 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 associated bidirectional LSP is useful for protection switching, for OAM that requires a reply (from MIP or MEP), and for defect correlation. The binding can happen when the two unidirectional LSPs are being established or have been established.

The notion of association as well as the corresponding RSVP ASSOCIATION object are defined in [RFC4872] and [RFC4873]. 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 [I-D.berger-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. Extensions to the Extended ASSOCIATION object.

The Extended ASSOCIATION object is defined in [I-D.berger-ccamp-assoc-info].

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

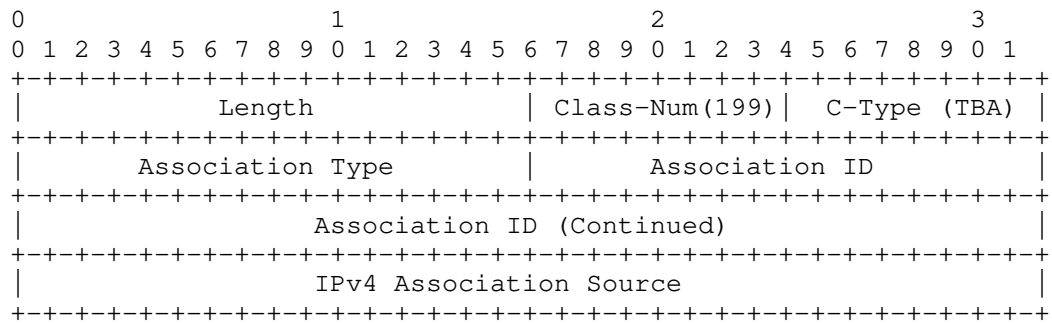


Figure 1: Extended IPv4 ASSOCIATION object

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

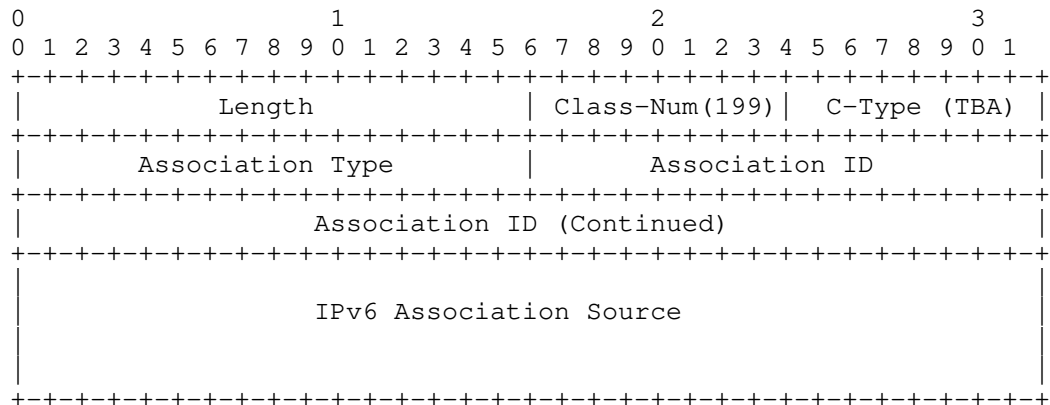


Figure 2: Extended IPv6 ASSOCIATION object

- 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 LSP (A)

If the midpoints do not know this Association Type, just ignore it and forward it to the next node, but the egress nodes MUST return a PathErr message with error code/sub-code "LSP Admission Failure/Bad Association Type" if they do not know this 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 sapce) to identified association, see [RFC4872] and [I-D.berger-ccamp-assoc-info]. This document adds specific rules: the Association source MUST 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. This document adds specific rules: the first 16 bits MUST set to the tunnel ID of the initiating node and the following 16 bits MUST 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.berger-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.

[I-D.berger-ccamp-assoc-info] discussed the rules associated with the processing of the Extended ASSOCIATION objects in RSVP message. 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 be operation based on the same Extended ASSOCIATION objects.

4. Association of Two Reverse Unidirectional LSPs

Consider the following example:

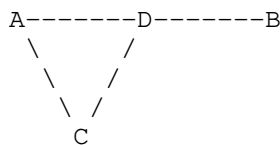


Figure 3: An example of associated bidirectional LSP

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].

Just like the end-to-end recovery LSP association [I-D.berger-ccamp-assoc-info], the following combination may occur when binding LSP1 and LSP2 together to be an associated bidirectional LSP

Case 1. 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.

Case 2. 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.

Case 3. The Extended ASSOCIATION object of both the LSPs 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.berger-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.

4.1. Asymmetric Bandwidth LSPs

There are some kind of services which have different bandwidth requirements for each direction, and associated bidirectional LSPs SHOULD support them. [RFC5467] defined three new objects named UPSTREAM_FLOWSPEC object, UPSTREAM_TSPEC object and UPSTREAM_ADSPEC object, which refer to the upstream traffic flow. In this document, UPSTREAM_TSPEC MUST be carried in the initializing LSP's Path message to trigger the egress node to setup the reverse LSP with corresponding asymmetric bandwidth. If the midpoints do not know this object, just ignore it and forward it to the next node, but the egress nodes MUST return a PathErr message with error code/sub-code "Admission Control failure/Unknown object" if they do not know this

object.

5. IANA Considerations

Within the current document, a new Association Type is defined in Extended ASSOCIATION object.

Value	Type
-----	----
4	Association of two reverse unidirectional LSPs (A)

There are no other IANA considerations introduced by this document.

6. Security Considerations

No new security considerations are introduced in this document.

7. Acknowledgement

The author would like to thank Xihua Fu and Bo Wu for their useful discussion; thank Lou Berger for his suggestion on the organization of this document; thank Lamberto Sterling for his valuable comments on the section of asymmetric bandwidths.

8. Normative references

- [I-D.berger-ccamp-assoc-info]
Berger, L., Faucheur, F., and A. Narayanan, "Usage of The RSVP Association Object", draft-berger-ccamp-assoc-info-01 (work in progress), March 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-03 (work in progress), October 2010.
- [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.
- [RFC5467] Berger, L., Takacs, A., Caviglia, D., Fedyk, D., and J. Meuric, "GMPLS Asymmetric Bandwidth Bidirectional Label Switched Paths (LSPs)", RFC 5467, March 2009.
- [RFC5654] Niven-Jenkins, B., Brungard, D., Betts, M., Sprecher, N., and S. Ueno, "Requirements of an MPLS Transport Profile", RFC 5654, September 2009.

Authors' Addresses

Fei Zhang (editor)
ZTE Corporation

Email: zhang.fei3@zte.com.cn

Fan Yang
ZTE Corporation

Email: yang.fan5@zte.com.cn

LZ Jin
ZTE Corporation

Email: lizhong.jin@zte.com.cn

WL jiang
ZTE Corporation

Email: jiang.weilian@zte.com.cn

Network Working Group
Internet-Draft
Intended status: Informational
Expires: April 21, 2011

M. Zhang
H. Ding
YL. Zhao
BUPT
HY. Zhang
YB. Xu
CATR
October 18, 2010

Performance Metric of Convergence Time of Information Flooding in Multi-
Domain GMPLS Networks
draft-zhangm-ccamp-metric-00

Abstract

To keep the information of topology and links resource synchronized at each control node, massive messages are necessary to be flooded in the control plane of General Multi-Protocol Label Switching (GMPLS) based multi-domain networks. The convergence time of information flooding will have a significant impact on the performance of the networks. So measuring and analyzing the convergence time of information flooding in multi-domains becomes very important. A performance metric of convergence time of information flooding is proposed to characterize the ability of information synchronization in multi-domain networks.

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 21, 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. Motivations	3
2. Conventions Used in this Document	4
3. Overview of the Performance metric	4
4. convergence time of information flooding in single domain . .	4
4.1. Initial convergence time of information flooding	4
4.1.1. Definition	4
4.1.2. Methodology	4
4.1.3. Sample	5
4.2. convergence time of information flooding with LSPs	6
4.2.1. Definition	6
4.2.2. Methodology	6
4.2.3. Sample	7
5. Convergence time of information flooding in multi-domain networks	8
5.1. Initial convergence time of information flooding	8
5.1.1. Definition	8
5.1.2. Methodology	8
5.1.3. Sample	9
5.2. Convergence time of information flooding with LSPs	10
5.2.1. Definition	10
5.2.2. Methodology	11
5.2.3. Sample	12
6. Protocol Extension Requirements	14
6.1. OSPF-TE Extension Requirements	14
6.2. RSVP-TE Extension Requirements	14
7. Discussion	14
8. Security Considerations	15
9. Acknowledgement	15
10. Normative References	15
Appendix A. author	15
Authors' Addresses	16

1. Introduction

Generalized Multi-Protocol Label Switching (GMPLS) [RFC3945], which can handle multiple switching technologies: packet switching (PSC), Layer 2 switching (L2SC), Time-Division Multiplexing (TDM) Switching, wavelength switching (LSC) and fiber switching (FSC), is a key technology for transport and service network. As the network varies from time to time, division of the network is a natural solution to cope with large scale network control and management. In order to keep the information of topology and links resource synchronized, which is necessary during the process of path computation, massive messages are necessary to be flooded in the control plane of multi-domain networks. So measuring and analyzing the convergence time of information flooding in multi-domain networks becomes very important. RFC 5814 has defined three metric to characterize the dynamic Label Switching Path (LSP) provisioning performance of signaling protocol. However, performance metric that concerns routing protocols, especially in multi-domain network are not specified in previous documents.

In this document, we define a performance metric of convergence time of information flooding from the routing aspect to characterize the ability of information synchronization in multi-domain networks. The metric can be used to measure convergence time of information flooding in single domain, multi-domains, initial and with LSPs. Methodologies and samples of the testing procedure for different scenarios are also included in the following sections.

1.1. Motivations

Convergence time of information flooding is useful for several reasons.

- o When a large scale network is deployed, a series of tests are to be conducted to evaluate the network performance, such as adding or deleting the clients, establishing or tearing the connections, and so on. The convergence time, which indicates the ability to synchronize the topology and link resource states, is also worth measuring and analyzing meantime, since it can further illustrate the reasonability of domain division.
- o During the operation, nodes or links failures MAY cause congestion due to both resource unavailability and a large amount of information flooding. Measuring and monitoring convergence time of information flooding is helpful for network failure detection.
- o After network updating and large scale reconfiguration, convergence time of information flooding should be measured,

because it MAY reflect the reasonability of this updating by comparing the convergence time of information flooding with the LSP setup delay.

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. Overview of the Performance metric

To evaluate the convergence ability of a GMPLS-based network from routing aspect, we define a performance metric of convergence time of information flooding. The following sections specify the metric in different situations, initial or with LSPs, in single domain or multi-domains. The initial convergence time of information flooding measures the time that one network spends from its power-on to its full operation. The convergence time with LSPs measures the time needed to synchronize the information after several changes in link resource states. The convergence time in single domain is intra-domain time whereas the convergence time in multi-domains comprises intra-domain time and inter-domain time and the two kinds of convergence time need to compute separately.

The convergence time of information flooding is either a real number of milliseconds or undefined. And in methodology "undefined" convergence time is defined.

4. convergence time of information flooding in single domain

4.1. Initial convergence time of information flooding

4.1.1. Definition

The initial convergence time of information flooding describes a period of time, which starts at the moment when the first bit of the first Hello packet is sent and ends at the moment when all nodes database synchronized in this domain.

4.1.2. Methodology

Generally, the convergence time of initial information flooding in single domain proceeds as follows,

- o All control nodes in this domain switch on normally;
- o Store a time structure, which records the first sending moment of Hello packets, in every control node. And the starting point (T1) of the convergence time is the earliest sending moment among all control nodes in this domain;
- o After the neighbor relationships are established, Database Description (DD) packets are sent and received to synchronize the information topology and links resource according to RFC 2328. Then all nodes in the domain are marked as full adjacency. Store a time structure, which records the latest receiving moment of DD packets, in every node. Choose the latest moment among all control nodes as end point (T2);
- o Initial convergence time of information flooding in single domain can be computed by subtracting the starting point from end point (T2-T1).

4.1.3. Sample

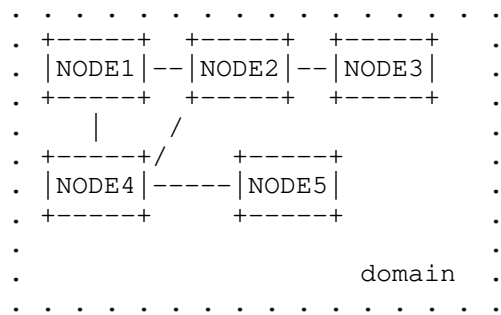


Figure 1: Initial information synchronization in single domain

Control nodes are represented by vertices and physical links are represented by dashed lines. When all these control nodes switch on normally and their interfaces first become operational, Hello packets are flooded in this domain to establish neighbor relationships as described in Section 7, RFC1583.

For a domain as shown in Figure 1, after the power-on, all five nodes begin to send and receive Hello packets. In most occasions, these nodes receive the message in a very short interval which is normally a few microseconds. So practically a structure of time is stored in every node to record the moment, which is the time when the node sends a Hello packet. The starting point of the initial convergence

time of information flooding is the earliest sending moment of all the Hello packets.

After the establishment of neighbor relationships, DD packets are sent and received to compare the nodes database so as to establish adjacency relationships. And the receiving moments of DD packets SHOULD be stored in another time structure in every node. Then choose the latest moment among all the receiving moments relating to DD packets as the end point of the convergence time of information flooding in this scenario.

So the initial convergence time of information flooding in single domain can be computed by subtracting the starting point from the end point.

4.2. convergence time of information flooding with LSPs

4.2.1. Definition

The convergence time of information flooding with LSPs describes a time period, that starts at the moment when the ingress node sends the first bit of a Link State Update (LSU) packet and ends at the moment when the last node in this domain receives the LSU packet.

The undefined convergence time of information flooding with LSPs means ingress node fails to receive corresponding ResvConf message, which MAY indicates resource reservation failure or nodes breaking down along the LSP.

4.2.2. Methodology

Generally, the methodology proceeds as follows,

- o All control nodes in this domain switch on normally;
- o Initial information synchronization is complete, which means all node Link State Database (LSDBs) are up-to-date;
- o Select an ingress node ID0 and an egress node ID1, and create a LSP path from ID0 to ID1.
- o Wait until ID0 receives the corresponding ResvConf message and updates its LSDB and forms a LSU packet LSU0. Store a timestamp (T1) at ID0 as soon as possible;
- o Another timestamp (T2) SHOULD be stored when the last node within this domain receives LSU0 at the exact last node;

- o The convergence time of information flooding with LSPs in single domain can be computed by subtracting the two timestamps (T2-T1);
- o If ID0 fails to receive the corresponding ResvConf message in a reasonable period of time, the convergence time is set to undefined.

4.2.3. Sample

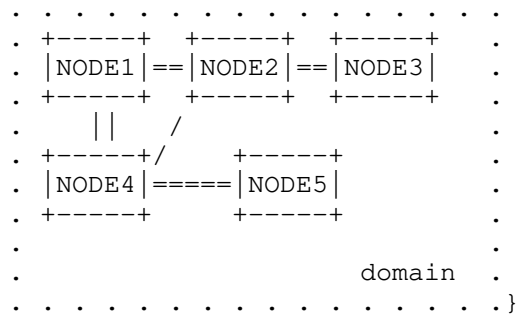


Figure 2: Convergence time of information flooding with LSPs in single domain

Control nodes are represented by vertices and physical links are represented by dashed lines, LSPs are represented by equal signs. For a single domain as shown in Figure 2, NODE1->NODE2->NODE3 constitute LSP1 which already exists, and NODE1->NODE4->NODE5 constitute LSP2 which need to be measured.

When NODE1, the ingress node of LSP2, receives a ResvConf message correspondingly with LSP2, it forms an LSU packet including changed LSAs and floods it within the domain. NODE2 and NODE4 receive the LSU and then NODE3 and NODE5 receive the message. A time structure is stored in every control node to record the earliest sending moment and latest receiving moment of LSU packets. The ingress node NODE1's sending moment can be the starting point and choose the latest moment among NODE2's, NODE3's, NODE4's, NODE5's receiving moments as the end point of the convergence time.

So the convergence time of information flooding with LSPs in single domain as depicted in Figure 2 can be computed by subtracting the starting point from the end point.

5. Convergence time of information flooding in multi-domain networks

5.1. Initial convergence time of information flooding

5.1.1. Definition

Initial convergence time of information flooding in multi-domain network defines a time period, which starts at the moment when the first Hello packet is sent and ends at the moment when all nodes database synchronized in the network.

5.1.2. Methodology

Generally, convergence time of information flooding in multi-domain network proceeds as follows,

- o All control nodes in multi-domains switch on normally;
- o Store a time structure, which records the first sending moment of Hello packets in every control node. And the starting point (T1) of the convergence time is the earliest sending moment among all control nodes in the multi-domain network;
- o After the neighbor relationships are established, Database Description (DD) packets are sent and received to synchronize the information topology and links resource according to RFC 2328. Then all nodes in the domain are marked as full adjacency. Store a time structure, which records latest receiving moment of DD packets, in every node. Choose the latest moment among all these structures as another timestamp (T2);
- o Initial convergence time of information flooding in the multi-domain network can be computed by subtracting the two timestamps (T2-T1).

5.1.3. Sample

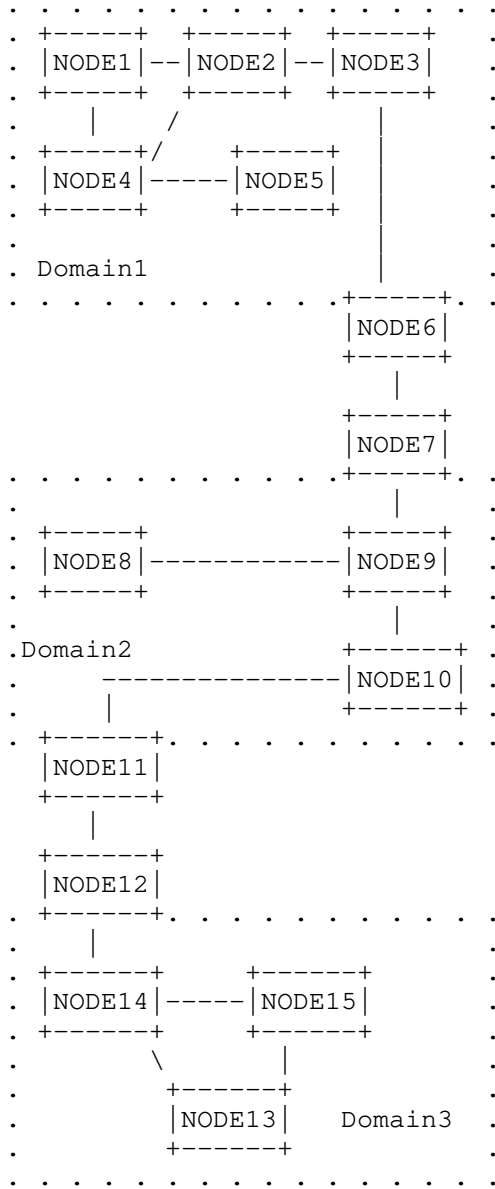


Figure 3: Initial convergence time of information flooding in multi-domain networks

Control nodes are represented by vertices and physical links are represented by dashed lines. Border nodes NODE6, NODE7, NODE11 and NODE12 connect Domain1, Domain2 and Domain3.

When all these nodes switch on normally and their interfaces first become operational, Hello packets are flooded in the multi-domain network to establish neighbor relationships.

For the three domains in the network as depicted in Figure 3, all these nodes are switched on more or less simultaneously, so store the same time structure in every node is necessary to record the first sending moment of Hello packets. The earliest sending moment is the starting point of the convergence time.

After the establishment of neighbor relationships, DD packets are sent and received to compare the nodes database so as to establish adjacency relationships. And the receiving moments of DD packets SHOULD be stored in another time structure in every node. Then choose the latest moment among all receiving moments relating to DD packets as the end point of the convergence time of information flooding in this scenario.

So the initial convergence time of information flooding in the multi-domain network can be computed by subtracting the starting point from the end point.

5.2. Convergence time of information flooding with LSPs

5.2.1. Definition

Convergence time of information flooding with LSPs in multi-domain networks comprises two parts: intra-domain convergence time of information flooding and inter-domain convergence time of information flooding. While intra-domain convergence time of information flooding is further divided into several single domain convergence time of information flooding according to the domains the cross-domain-LSP traversed through. And inter-domain convergence time of information flooding refers to a time period during which the domain border nodes synchronize their information according to RFC1583.

Every single domain convergence time of information flooding can refer to section 4.2. Note that for source domain the convergence time starts at the moment when the LSP ingress node updates its information. Otherwise, for intermediate domains, as well as destination domain, the convergence time starts at the moment when the ingress node of that domain updates its information with respect to the LSP. All the convergence time end with the last node!_s receipt of the updating information.

Setting the convergence time of information flooding with LSPs as undefined means the ingress node fails to receive the corresponding ResvConf message, which MAY indicate resource reservation failures or nodes breaking down along the cross-domain LSP.

5.2.2. Methodology

The procedure of convergence time of information flooding in multi-domain network, as mentioned above, comprises two parts: intra-domain convergence time of information flooding and inter-domain convergence time of information flooding. Methodology for convergence time of information flooding in single domain has been specified in Section 4.2.2; and the methodology for inter-domain convergence time of information flooding proceeds as follows,

- o All control nodes in all domains switch on normally;
- o Initial information synchronization is complete, which means all nodes!_ LSDB are up-to-date;
- o Select an ingress node ID0 and an egress node ID1 in a different domain, and create a cross-domain LSP from ID0 to ID1;
- o Wait until ID0 receives all the corresponding ResvConf messages that confirm the completion of resource reservation, ID0 updates its LSDB and forms a LSU packet LSU0. Store a timestamp (T1) at ID0 as soon as the LSU packet is sent;
- o Then the source domain border node receives the LSU packet and summarizes the source domain information into an advertisement. Then the border node distributes the advertisement to backbone area. Store a timestamp (T2) at the border node as soon as the advertisement is distributed;
- o Another timestamp (T3) SHOULD be stored locally as soon as the border node in destination domain receives an advertisement from backbone area;
- o Inter-domain convergence time of information flooding can be computed by subtracting two timestamps (T3-T2);
- o The convergence time of information flooding of the network can also be computed by subtracting two timestamps (T3-T1);
- o If ID0 fails to receive the corresponding ResvConf message in a reasonable period of time, the inter-domain convergence time of information flooding is set to undefined.

5.2.3. Sample

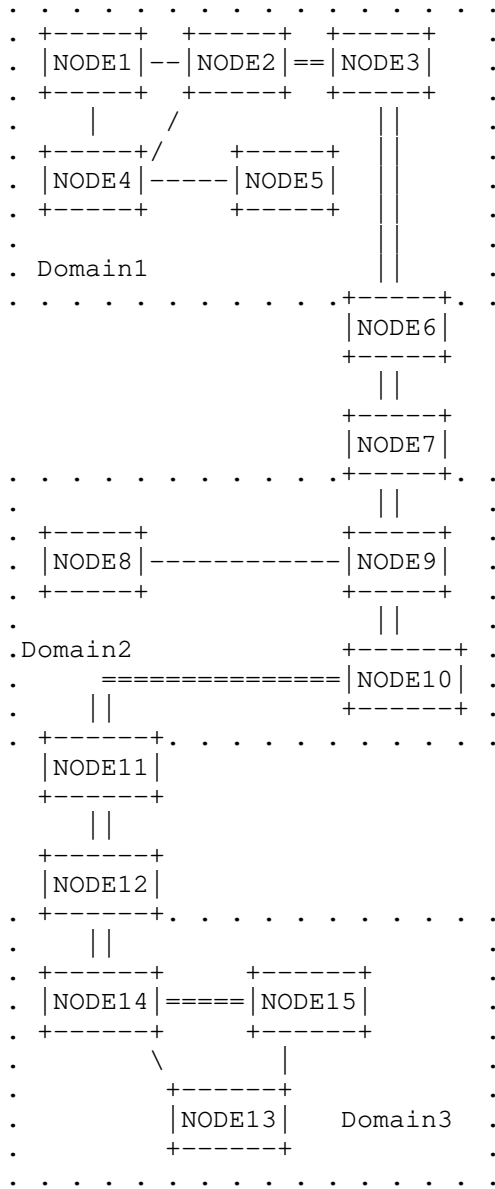


Figure 4: Convergence time of information flooding with LSPs in multi-domain networks

Control nodes are represented by vertices and physical links are represented by dashed lines and LSPs are represented by equal signs. Border nodes are NODE6, NODE7, NODE11 and NODE12 which connect Domain1, Domain2 and Domain3, as shown in Figure 4.

The cross-domain LSP is NODE2->NODE3->NODE6->NODE7->NODE9->NODE10->NODE11->NODE12->NODE14->NODE15. When ResvConf message is received from every node along this LSP, meaning the resource reservation is complete, LSAs flooding begins from the source domain, Domain1.

NODE2 sends out a LSU packet LSU1 which contains link states changes in Domain1, and LSU1 is then flooded in Domain1. When Domain1's border node NODE6 receives LSU1, it not only updates its own database but also forms an inter-domain LSU packet LSU12 summarizing Domain1's states changes and sends it to Domain2's border node NODE7. As ingress node of partial LSP in Domain2, NODE7 also sends out a LSU packet LSU2 which includes link states changes in Domain2 and then LSU2 is flooded in Domain2. Similarly, NODE11 forms an inter-domain LSU packet LSU23 and sends it to NODE12. NODE12, as ingress node of partial LSP in Domain3, forms LSU3 which is then flooded in Domain3.

LSU1, LSU2 and LSU3 are intra-domain LSU packets while LSU12, LSU23 are inter-domain LSU packets.

Time structures are stored in every node along the sending and receiving of LSU packets. Moments related to different LSU packets are recorded in different time structures.

Intra-domain convergence time of information flooding in Domain1 can be computed by subtracting the end point, which can be obtain by choosing the latest receiving moment in Domain1, from NODE2s sending moment. Similarly, the starting point and end point of the intra-domain convergence time of information flooding in Domain2 are NODE7s LSU2 sending moment and the latest receiving moment among NODE8, NODE9, NODE10 and NODE11. For Domain3 the starting point and end point are NODE11s LSU23 sending moment and the latest receiving moment among NODE12, NODE13, NODE14 and NODE15.

Inter-domain convergence time of information flooding reflects the time required to synchronize information among border nodes: NODE6, NODE7, NODE11 and NODE12. So the starting point is NODE6s LSU12 sending moment while the endpoint is the latest inter-domain LSU packets receiving moment. By subtracting the two moments, inter-domain convergence time of information flooding for Domain1, Domain2 and Domain3 is computed. Note that in Figure 4, there is only one entrance border node and one exit border node between two domains. As for Domain1, NODE6 is the only exit node, so the starting point of inter-domain convergence time of information flooding is the moment

when NODE6 sends LSU12. In the topology where there are more than one exit nodes in the source domain, the starting moment will be the earliest LSU12 sending moments among the exit border nodes.

The convergence time of information flooding in the network can be computed by subtracting the following two moments, one is the NODE2s sending moment in Domain1, the other is the latest LSU23 receiving moment.

6. Protocol Extension Requirements

6.1. OSPF-TE Extension Requirements

The measurement procedure of the initial convergence time of information flooding requires the extensions in OSPF-TE protocol. During the procedure, sending and receiving moments of Hello packets and DD packets need to be recorded. Corresponding timestamps are needed to symbolize the sending and receiving of Hello packets and DD packets in every node.

6.2. RSVP-TE Extension Requirements

7. Discussion

The following issues are likely to come up in practice.

- o The accuracy of convergence time of information flooding depends largely on the clock resolution in every node, where time structures are stored; so synchronization among all nodes in the network is crucial.
- o Whether a convergence time of information flooding is a real number or undefined largely depends on the choosing of the reasonable waiting time before the ResvConf is received. However, choosing the waiting time is complicated. If the time is set too short, there will be too much "undefined" convergence time and the result does not reflect the network performance properly. However, if the time is set too long, time is wasted waiting when there are resource reservation failures or breaking down nodes. Choose the appropriate waiting time is also depending on the network status, if the network is light loaded, the waiting time can be set shorter than it is set when the network is heavy loaded.

8. Security Considerations

9. Acknowledgement

We wish to thank Shengwei Meng, and Koubo Wu in the Key Laboratory of Information Photonics and Optical Communications (BUPT), Ministry of Education, for their valuable comments. We also wish to thank the support from National 863 program.

10. Normative References

- [RFC1583] Moy, J., "OSPF Version 2", March 1994.
- [RFC2119] Bradner, S., "Key words for use in RFC's to Indicate Requirement Levels", RFC 2119, March 1997.
- [RFC2205] Braden, R., Zhang, L., Berson, S., Herzog, S., and S. Jamin, "Resource ReSerVation Protocol (RSVP) -- Version 1 Functional Specification", RFC 2205, September 1997.
- [RFC3630] Katz, D., Kompella, K., and D. Yeung, "Traffic Engineering (TE) Extensions to OSPF Version 2", September 2003.
- [RFC3945] Mannie, E., "Generalized Multi-Protocol Label Switching (GMPLS) Architecture", October 2004.
- [RFC5151] Farrel, A., Ayyangar, A., and JP. Vasseur, "Inter-Domain MPLS and GMPLS Traffic Engineering -- Resource Reservation Protocol-Traffic Engineering (RSVP-TE) Extensions", February 2008.
- [RFC5152] Vasseur, JP., Ayyangar, A., and R. Zhang, "A Per-Domain Path Computation Method for Establishing Inter-Domain Traffic Engineering (TE) Label Switched Paths (LSPs)", February 2008.

Appendix A. author

Jie Zhang

Beijing University of Post and Telecommunication

No.10,Xitucheng Road,Haidian District

Beijing 100876

China

Phone: +8613911060930

Email: lgr24@bupt.edu.cn

Authors' Addresses

Min Zhang
Beijing University of Post and Telecommunication
No.10,Xitucheng Road,Haidian District
Beijing 100876
P.R.China

Phone: +8613910621756
Email: mzhang@bupt.edu.cn

Hui Ding
Beijing University of Post and Telecommunication
No.10,Xitucheng Road,Haidian District
Beijing 100876
P.R.China

Phone: +8613426082796
Email: dinghui.ei@gmail.com

Yongli Zhao
Beijing University of Post and Telecommunication
No.10,Xitucheng Road,Haidian District
Beijing 100876
P.R.China

Phone: +8613811761857
Email: yufengx386@gmail.com

Haiyi Zhang
China Academy of Telecommunication Research, MIIT, China.
No.52 Hua Yuan Bei Lu,Haidian District
Beijing 100083
P.R.China

Phone: +861062300100
Email: zhanghaiyi@mail.ritt.com.cn

Yunbin Xu
China Academy of Telecommunication Research, MIIT, China.
No.52 Hua Yuan Bei Lu, Haidian District
Beijing 100083
P.R.China

Phone: +8613681485428
Email: xuyunbin@mail.ritt.com.cn

Network Working Group
Internet-Draft
Intended status: Informational
Expires: April 21, 2011

M. Zhang
LF. Zhang
YF. Ji
BUPT
YB. Xu
Y. Wang
MIIT
October 18, 2010

Network Survivability Evaluation Metrics in Multi-domain Generalized
MPLS Networks
draft-zhangm-ccamp-reroute-00

Abstract

The ubiquitous presence of the internet coupled with the increasing demand for high bandwidth dedicated large scale network has made it imperative that the multi-domain networks are facilitated by the development of GMPLS. In such large scale network, the high performance network survivability is a significant factor to resist the fault service discontinue and interruption even to decrease economic loss and the society impact. This document proposes a series of network survivability evaluation metrics and methodologies that can be used to demonstrate the network survivability performance in single and multi-domain GMPLS networks, more specifically, the network fault restoration performance.

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 21, 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. motivation	3
1.2. Terminology	3
2. Conventions Used in This Document	4
3. Overview of Network Survivability Evaluation Metrics	4
4. Network Survivability Evaluation Metrics	4
4.1. Fault Restoration Time Phases	4
4.2. Restoration Schemes and Scenarios	6
4.2.1. Faults in single domain	7
4.2.2. Faults in multi-domain	7
4.2.2.1. Faults within a domain	8
4.2.2.2. Inter-domain faults	8
5. Methodologies	9
5.1. Fault restoration in single domain network	9
5.1.1. Reroute	9
5.1.2. Fast Reroute	10
5.2. Fault restoration within a domain in multi-domain network	12
5.2.1. Reroute	12
5.2.2. Fast Reroute	13
5.3. Inter-domain fault restoration in multi-domain network	15
6. Protocol Extension Requirements	17
7. Security Considerations	17
8. Acknowledgments	17
9. Normative References	17
Appendix A. Other author	18
Authors' Addresses	18

1. Introduction

1.1. motivation

Generalized Multi-Protocol Label Switching (GMPLS) network is a promising choice with the use of optical technology in core networks combined with IP/Multi-Protocol Label Switching (MPLS) solution for the next generation Internet architecture. The ubiquitous presence of the internet coupled with the increasing demand for high bandwidth and dedicated large scale network has made it imperative that the multi-domain networks are facilitated by the development of GMPLS.

Survivability is the capability of the network to maintain service continuity in the presence of faults within the network, at the same time, service influenced could be switched over to free resource. All kinds of intra-domain and inter-domain faults occurs in multi-domain GMPLS Networks, therefore, in such large scale network, the high performance network survivability is a significant factor to resist the fault service interruption even to decrease economic loss and the society impact due to faults. Recovery time is a key factor to measure network survivability performance which has an impact on the link and service evaluation. The long recovery time could increase the traffic delay, packet losses, the resource collision, preemption and service discontinue even the whole network can not reach the level of reliability required by traffic service. The time of every recovery phrase is required to be known by a series of measurement methodologies in order to reduce the fault restoration time. Certain method could be adopted to reduce the every phrase time to achieve the aim of reducing the whole recovery time. Therefore, network survivability evaluation metrics is necessary in multi-domain Generalized MPLS Networks.

This document proposes a series of network survivability evaluation metrics and methodologies that can be used to demonstrate the network survivability performance in single and multi-domain GMPLS networks, more specifically, the network fault restoration performance. The time of every fault restoration phase is measured precisely to evaluate the whole network performance by proposed evaluation metrics.

1.2. Terminology

LSP: Label Switched Path.

LSR: Label Switched Router.

QoS: Quality of Service.

PSL: Path Switch LSR.

ML: Merge LSR.

LMP: Link Management Protocol.

NMS: Network Management System.

RSVP: Resource Reserve Protocol.

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]. In addition, the reader is assumed to be familiar with the terminology used in [RFC3945], [RFC3471], [RFC3473] and referenced as well as in [RFC4427] and [RFC4426].

3. Overview of Network Survivability Evaluation Metrics

There are two recovery mechanisms (eg. protection and restoration) and the former is outside the scope of this document currently. Network survivability evaluation metric is used to measure precise recovery time which is a key factor during the whole fault recovery process (eg. fault detection, fault location, fault notification, fault recovery and reversion). These phases define the sequence of generic operations that need to be performed when a failure occurs. The evaluation metrics take the time of every phase into account and give the specific measurement steps and methodologies.

4. Network Survivability Evaluation Metrics

High performance of network survivability has become a key issue to improve and satisfy the increasing requirements of reliability and Quality of Service (QoS) of the whole network. This section defines a network survivability evaluation metric in single and multi-domain Generalized MPLS networks.

4.1. Fault Restoration Time Phases

This section gives several typical definitions of restoration times and durations as shown in figure 1.

Phase 1: Fault detection.

Phase 2: Fault localization and isolation.

Phase 3: Fault notification.

Phase 4: Recovery.

Phase 5: Reversion.

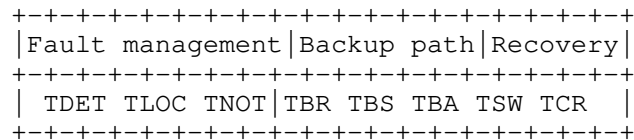


Figure 1: Failure Restoration Time Phases

A detailed analysis and specific definition is provided for each of the restoration phases as identified in [RFC4427] and [RFC4428].

o Fault detection time TDET

Fault detection time is defined as the time between occurrence of fault and detecting the fault and degradation.

TDET depends on several factors pertaining to the link propagation time, link transmission time, node processing time and node queuing time.

o Fault Localization and isolation time TLOC

Fault Localization and isolation time is defined as the time the signal indication information is delivered from fault node to PSL.

o Fault notification time TNOT

Fault notification time is defined as the time to inform the node responsible of the switchover that a failure has occurred.

TNOT depends on failure notification delay and the notification method used.

o Backup routing time TBR

Backup routing time is defined as the time for new backup creation, routing (TBR) and signaling (TBS).

TBR depends on the routing method applied.

- o Backup signaling time TBS

Backup signaling time is defined as the time that is required to activate the backup path before the switchover.

TBS depends on the signaling method applied.

- o Backup activation time TBA

Backup activation time is defined as the time between the settlement of backup path and the switching over the traffic.

TBA depends on the backup path distance and signaling process.

- o Switchover time TSW

Switchover time is defined as the time of switching the traffic from the working path through which the traffic is flowing, to the alternative/backup path.

TSW depends on the node technology.

- o Restoration completion time TCR

Restoration completion time is defined as the time to complete the fault recovery, i.e. the time it takes the first packet to arrive from the backup path to the ML.

TCR depends on the backup distance.

- o The total restoration time

The total restoration time is defined as the sum of TDET, TNOT, TBR, TBS, TBA , TSW and TCR.

4.2. Restoration Schemes and Scenarios

Link restoration could effectively take use of network bandwidth to eliminate faults. The restoration technique is also referred as reroute and fast reroute, for instance, no backup path is established prior to the failure to protect the working path. Therefore, restoration requires dynamic routing algorithms and bandwidth allocation to establish a backup path on demand upon network failure. Once the backup path has been set up, traffic is then switched from the working path.

4.2.1. Faults in single domain

There are two restoration methods in allusion to fault in single domain. Fast reroute mainly provides the local repair function such as span restoration and segment restoration. The start node of span and segment restoration is responsible for backup path computation and traffic switching as the PSL(Path Switch LSR) instead of the source node in reroute scheme.

o Reroute

In the scenario of single domain, detecting entities in transport plane detect related fault information when node or link failure occurs. Failure localization/isolation is triggered immediately after the failure detection. And then the fault indication signaling is sent to the source node through the GMPLS-based signaling or flooding method by the detecting node. In the case of flooding method, intermediate nodes pertaining to the fault end-to-end LSP are informed the fault indication signaling between the upstream node and source node through a notification mechanism. In the signaling-based technique, detecting node sends fault indication signaling such as RSVP-TE to each LSP affected by the failure through different notification mechanism.

After receiving the fault indication signaling, the source node computes a backup path by a series of routing algorithms or route pre-computation scheme and then allocates the bandwidth. Path and RESV signaling are responsible for path establishment and resource reservation respectively for the new backup path. After that, the traffic is switched to the backup path from the working path.

o Fast reroute

In the scheme of fast reroute, failure localization/isolation is triggered immediately after the failure detection. And then the fault indication signaling is sent to the span or segment PSL from the upstream node of failure link through the GMPLS-based signaling or flooding method. These two notification methods are described in reroute part of section 4.2.1. On receiving the fault indication signaling, PSL is computes a new path by a series of routing algorithms and allocates the bandwidth to the backup path bypass the fault LSP. After that, the traffic is switched to the backup path from the working path.

4.2.2. Faults in multi-domain

There are three types of faults in multi-domain network, such as link or node failure within the domain, failure of a link at a domain

border and failure of domain border node. Inter-domain and Intra-domain restoration mechanisms are independent with each other.

4.2.2.1. Faults within a domain

When an intra-domain failure occurs, intra-domain restoration mechanism is set up first within a domain and the restoration scheme is similar to that of single domain in the scenario of multi-domain. Inter-domain restoration mechanism would be triggered only if the previous restoration mechanism fails.

o Reroute

Detecting entities in transport plane detect related fault information when node or link failure occurs within a domain. Failure localization/isolation is triggered immediately after the failure detection. And then the fault indication signaling is sent to the source node across intermediate domains through the GMPLS-based signaling or flooding method. After receiving the fault indication signaling, the source node computes a new path by a series of routing algorithms and allocates the bandwidth. Path and RESV signaling are responsible for path establishment and resource reservation respectively for the backup path. After that, the traffic is switched to the backup path from the working path.

o Fast reroute

In the same scenario above, detecting entities in transport plane detect related fault information when node or link failure occurs within a domain. Failure localization/isolation is triggered immediately after the failure detection. And then the fault indication signaling is sent to the local or segment PSL from the upstream node of failure link through the GMPLS-based signaling or flooding method. These two notification methods are described in section 4.2.1.1. After receiving the fault indication signaling, Path Switch LSR (PSL) is responsible for computing a new path by a series of routing algorithms and allocates the bandwidth to establish the backup path bypass the fault LSP. After that, the traffic is switched to the backup path from the working path.

4.2.2.2. Inter-domain faults

Inter-domain faults comprise inter-domain link fault and border node fault of the domain. Each domain has its own domain border node, and these two border nodes are connected by a TE link. TE link is invalid once the border node fails.

When the LSP traverses multiple domains and inter-domain failure

occurs, the process of failure detection and localization/isolation is the same to that of single domain whose detail is described in section 4.2.1. If the fault TE link is the only one between two domains, the restoration mechanism adopts the end-to-end reroute restoration scheme. The fault indication signal is sent to source node by the upstream node along the LSP, and then the source node computes another path and allocates the resource avoiding the domain relative to the fault node and link. Otherwise, the restoration mechanism could adopt either the reroute or the fast reroute scheme if there is more than one link between two domains. Path and RESV signaling are responsible for path establishment and resource reservation respectively between PSL and ML. After that, the traffic is switched to the backup path from the working path.

5. Methodologies

It is difficult to measure Detection time TDEF which depends on the monitoring technique and reversion is a normalization process. Therefore, the methodology of detection and reversion time are outside the scope of this document.

5.1. Fault restoration in single domain network

This section gives two measurement methods of fault restoration which are end-to-end reroute and fast reroute respectively in single domain network. It is assumed that there exists an LSP (1-2-3-4) where data flow is from node 1 to node 4 as an example shown in figure 2 and 3. The link fault occurs between node 2 and node 3.

5.1.1. Reroute

Generally, when the failure occurs the methodology would proceed as follows:

- o The node 3 sends Channelstatus Message to node 2 indicating the failure to the corresponding the upstream node.
- o Record the timestamp (T1) when the first bit of Channelstatus Message is sent to the node 2 along the LSP.
- o When node 2 receives the ChannelStatus message from node 3, it returns a ChannelStatusAck message back to node 3 and correlates the failure locally. When Node 2 correlates the failure and verifies that the failure is clear, it has localized the failure to the data link between node 3 and node 2. At that time, node 2 sends a ChannelStatus message to node 3 indicating that the failure has been localized.

- o Then record the timestamp (T2) when the last bit of ChannelStatus message from node 2 is received by node 3.
- o The fault localization delay is T2-T1.
- o The node 2 sends the notification information to the source node(node 1) of the LSP traversing intermediate nodes. Then record the timestamp (T3) when the first bit of PathErr information is sent out.
- o Record the timestamp (T4) when the node 1 receives the last bit of the PathErr Message.
- o Notification delay is T4-T3.
- o Record the timestamp (T5) after node receives the notification information. Node 1 as the PSL computes a new path through either a series of route algorithms or pre-computed schemes.
- o PATH and RESV signaling are responsible for path establishment request and resource reservation respectively for a new backup path. Then the traffic is switched from a working path to the backup path. Record the timestamp (T6) when the first packet of traffic arrives at the ML(node 4) through the backup path.
- o Recovery time is T6-T5.
- o The total fault restoration time is T2+T4+T6-T1-T3-T5.

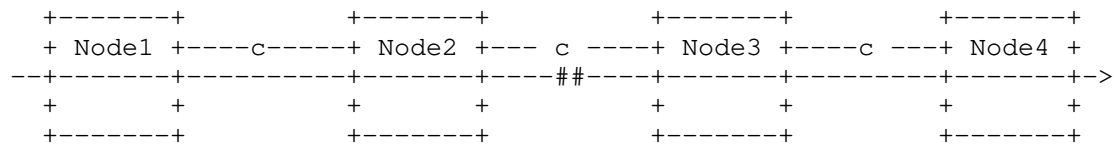


Figure 2: Reroute of fault in single domain (indicated by ## in the figure)

5.1.2. Fast Reroute

Generally, when the failure occurs, the methodology would proceed as follows:

- o The process of fault localization is similar to that of reroute restoration in single domain network which is described in section 5.1.1.
- o The fault localization delay is also $T_2 - T_1$.
- o PathErr information is sent to different PSL that differs from fast reroute restoration scheme. Node 2 is the PSL as the ingress node of backup path if the span recovery scheme is adopted. Otherwise, consider other PSL as the ingress node of backup path if segment restoration scheme is implemented.
- o Then record the timestamp (T_3) when the first bit of notification information is sent out by the node 2 to the PSL which is responsible for switching over the traffic.
- o Record the timestamp (T_4) when the PSL receives the last bit of the notification information.
- o Notification delay is $T_4 - T_3$.
- o Record the timestamp (T_5) after PSL receives the PathErr Message. The PSL computes a new path through either a series of route algorithms or pre-computed scheme (eg. 1-2-5-3-4).
- o PATH and RESV signaling are responsible for path establishment request and resource reservation respectively for the backup path. Then the traffic is switched from a working path to the backup path. Record the timestamp (T_6) when the first packet of traffic arrives at the ML(node 3) through the backup path.
- o Recovery time is $T_6 - T_5$.
- o The total fault restoration time is $T_2 + T_4 + T_6 - T_1 - T_3 - T_5$.

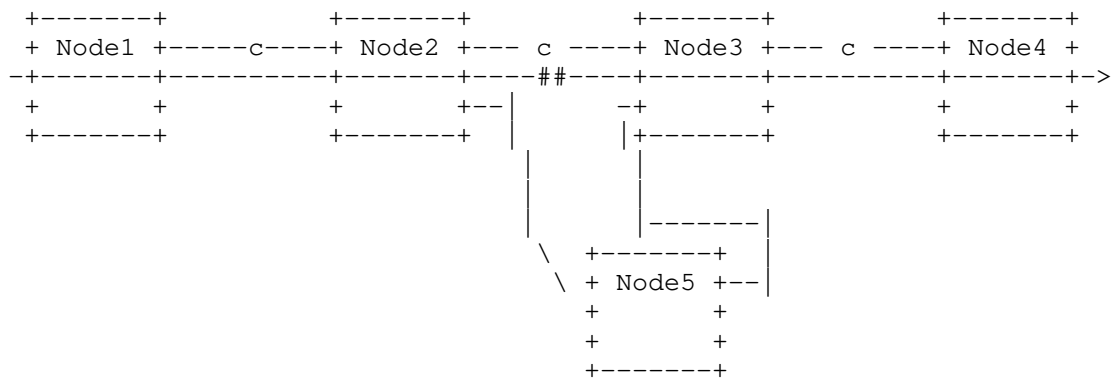


Figure 3: Fast reroute of fault in single domain (indicated by ## in the figure)

5.2. Fault restoration within a domain in multi-domain network

5.2.1. Reroute

Figure 4 describes the node connection situation. As illustrated node 1 and node 4 are in domain A and B respectively and node 2 and 3 are all in domain B. Generally, when the failure occurs, the methodology would proceed as follows:

- o The node 3 sends Channelstatus Message to node 2 indicating the failure to the corresponding upstream node.
- o Record the timestamp (T1) when the first bit of Channelstatus Message is sent to the node 2 along the LSP.
- o When node 2 receives the ChannelStatus message from node 3, it returns a ChannelStatusAck message back to node 3 and correlates the failure locally. When Node 2 correlates the failure and verifies that the failure is clear, it has localized the failure to the data link between Node 3 and node 2. At that time, Node 2 sends a ChannelStatus message to Node 3 indicating that the failure has been localized.
- o Then record the timestamp (T2) when the last bit of ChannelStatus message from node 2 is received by node 3.
- o The fault localization delay is $T2 - T1$.
- o Node 2 sends the notification information to the source node of the LSP (node 1) traversing intra-domain nodes and border nodes.

- o Record the timestamp (T_4) when the node 1 receives the last bit of the notification information.
- o Notification delay is $T_4 - T_3$.
- o Record the timestamp (T_5) after node 1 receives the PathErr Message. As the PSL, node 1 finds a new path through either a series of route algorithms or pre-computation scheme.
- o PATH and RESV signaling are responsible for path establishment request and resource reservation respectively for a new backup path. Then the traffic is switched from a working path to the backup path. Record the timestamp (T_6) when the first packet of traffic arrives at the destination node (node 4) through the backup path.
- o Recovery time is $T_6 - T_5$.
- o The total fault restoration time is $T_2 + T_4 + T_6 - T_1 - T_3 - T_5$.

- o The process of fault localization is similar to that of reroute restoration in single domain network which is described in section 5.1.2. The fault localization delay is also $T_2 - T_1$.
- o Notification information is sent to different PSL that differs from fast reroute restoration scheme. Node 2 is the PSL as the ingress node of restoration path if the span recovery scheme is adopted. Otherwise, consider other PSL as the ingress node of restoration path if segment recovery scheme is implemented.
- o Then record the timestamp (T_3) when the first bit of notification information is sent out by the node 2 to the PSL which is responsible for switching over the traffic.
- o Record the timestamp (T_4) when the PSL receives the last bit of the PathERR message.
- o Notification delay is $T_4 - T_3$.
- o Record the timestamp (T_5) after PSL receives the PathErr Message. The PSL finds a new path through either a series of route algorithms or pre-computed schemes.
- o PATH and RESV signaling are responsible for path establishment request and resource reservation respectively for a new backup path. Then the traffic is switched from a working path(2-3) to the backup path(2-5-3). Record the timestamp (T_6) when the first packet of traffic arrives at the ML(node 3) through the backup path.
- o Recovery time is $T_6 - T_5$.
- o The total fault restoration time is $T_2 + T_4 + T_6 - T_1 - T_3 - T_5$.
- o If the intra-domain fast reroute mechanism fails, reroute restoration is triggered whose methodology is illustrated in section 5.2.1.

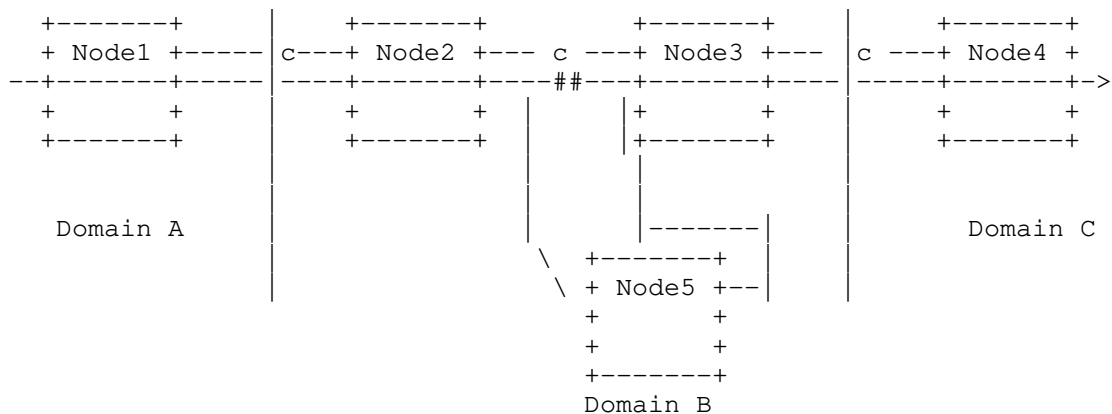


Figure 5: Fast reroute of fault within a domain in multi-domain network(indicated by ## in the figure)

5.3. Inter-domain fault restoration in multi-domain network

Figure 6 describes the node connection situation. As illustrated node 1 and node 4 are in domain A and C respectively and node 2,3 and 5 are all in domain B.

Generally, when the failure between domain A and B occurs, the methodology would proceed as follows:

- o The node 4 sends Channelstatus Message to node 3 indicating the failure to the corresponding upstream node.
- o Record the timestamp (T1) when the first bit of Channelstatus Message is sent to the node 3 along the LSP.
- o When node 3 receives the ChannelStatus message from node 4, it returns a ChannelStatusAck message back to node 4 and correlates the failure locally. When Node 3 correlates the failure and verifies that the failure is clear, it has localized the failure to the data link between Node 3 and node 4. At that time, Node 3 sends a ChannelStatus message to Node 4 indicating that the failure has been localized.
- o Record the timestamp (T2) when the last bit of ChannelStatus message from node 3 is received by node 4.
- o The fault localization delay is T2-T1.

- o Measurement method of notification delay is the same to that of fault reroute restoration within a domain in multi-domain network as described in section 5.2.1.
- o Notification delay is $T_4 - T_3$.
- o Record the timestamp (T_5) after node 1 receives the PathErr Message. Node 1 as the PSL computes a new path through either a series of route algorithms or pre-computed scheme. Consider to choose a backup path bypass the upstream domain of fault link if the fault link is the only link between domain B and domain C.
- o PATH and RESV signaling are responsible for path establishment request and resource reservation respectively for a new backup path. Then the traffic is switched from a working path to the backup path. Record the timestamp (T_6) when the first packet of traffic arrives at the destination node (node 4) through the backup path.
- o Recovery time is $T_6 - T_5$.
- o The total fault restoration time is $T_2 + T_4 + T_6 - T_1 - T_3 - T_5$.

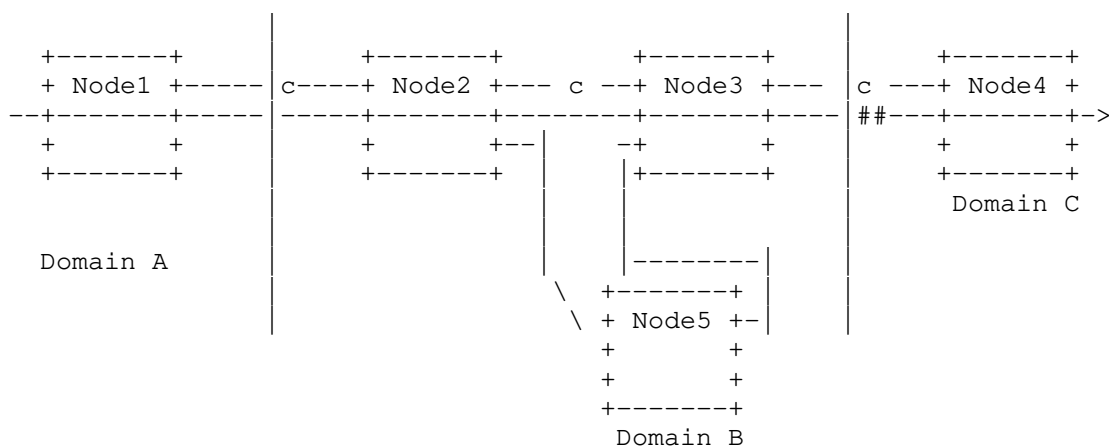


Figure 6: Inter-domain fault in multi-domain network (indicated by ## in the figure)

6. Protocol Extension Requirements

It is assumed that clock of every control node is synchronous during the process of measurement. Control plane reports different time to NMS(Network Management System) which is responsible for computing the sum of different fault restoration duration time. LMP and RSVP extensions are required in order to record precise the start and end time in every restoration phrase.

In the process of fault location measurement, detection entities send alarm information to upstream neighbor node through signaling of LMP when it detects the fault in control plane. It is necessary to extend LMP by adding a FAULT_TIMESTAMP object as a timestamp in the ChannelStatus Message. The FAULT_TIMESTAMP Object could be used to record the time when the signaling is sent and received to measure the precise fault location notification time. Then when the fault notification is implemented, the fault indicating signal is delivered to the PSL through the PathErr signal of RSVP. SEND_ERR_TIMESTAMP and RECEIVE_ERR_TIMESTAMP Objects are added in PathErr signal and defined to record the time of notification signal sent and received by upstream node next to the fault and PSL respectively.

7. Security Considerations

As this document is solely for the purpose of providing metric methodology and describes neither a protocol nor a protocol implementation, there is no security considerations associated with this document.

8. Acknowledgments

We wish to thank Jiuyu Xie, Yongli Zhao and Shengwei Meng for their comments and help.

The RFC text was produced using Marshall Rose's xml2rfc tool.

9. Normative References

- [RFC3473] Berger, L., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource Reservation Protocol-Traffic Engineering (RSVP-TE) Extensions", RFC 3473, January 2003.
- [RFC4090] Pan, P., Swallow, G., and A. Atlas, "Fast Reroute Extensions to RSVP-TE for LSP Tunnels", RFC 4090, May 2005.

- [RFC4204] Lang, J., "Link Management Protocol (LMP)", RFC 4204, October 2005.
- [RFC4426] Lang, J., Rajagopalan, B., and D. Papadimitriou, "Generalized Multiprotocol Label Switching (GMPLS) Recovery Functional Specification", RFC 4426, March 2006.
- [RFC4427] Mannie, E. and D. Papadimitriou, "Recovery (Protection and Restoration) Terminology for Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4427, March 2006.
- [RFC4428] Papadimitriou, D. and E. Mannie, "Recovery (Protection and Restoration) Terminology for Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4428, March 2006.

Appendix A. Other author

Haiyi Zhang

MIIT

No.52 Hua Yuan Bei Lu, Haidian District

Bei Jing 100083

P.R.China

Phone:+861062300100

Email:Zhanghaiyi@mail.ritt.com.cn

Authors' Addresses

Min Zhang

BUPT

No.10,Xitucheng Road,Haidian District

Bei Jing 100876

P.R.China

Phone: +8613910621756

Email: mzhang@bupt.edu.cn

URI: <http://www.bupt.edu.cn/>

Lifang Zhang
BUPT
No.10,Xitucheng Road,Haidian District
Bei Jing 100876
P.R.China

Phone: +8615210889041
Email: capricorn7111@hotmail.com
URI: <http://www.bupt.edu.cn/>

Yuefeng Ji
BUPT
No.10,Xitucheng Road,Haidian District
Bei Jing 100876
P.R.China

Phone: +8613701131345
Email: jyf@bupt.edu.cn
URI: <http://www.bupt.edu.cn/>

Yunbin Xu
MIIT
No.52 Hua Yuan Bei Lu,Haidian District
Bei Jing 100083
P.R.China

Phone: +8613681485428
Email: xuyunbin@mail.ritt.com.cn

Yu Wang
MIIT
No.52 Hua Yuan Bei Lu,Haidian District
Bei Jing 100083
P.R.China

Phone: +8613651161646
Email: wangyu@mail.ritt.com.cn

