

PCE Working Group
Internet Draft
Intended status: Informational

Zafar Ali
Antonello Bonfanti
Cisco Systems
F. Zhang
Huawei Technologies
August 29, 2014

Expires: February 28, 2015

Resource ReserVation Protocol-Traffic Engineering (RSVP-TE)
Extension for Additional Signal Types in G.709 OTN
draft-ali-ccamp-additional-signal-type-g709v3-05.txt

Status of this Memo

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

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). 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 February 28, 2015.

Copyright Notice

Copyright (c) 2014 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.

Abstract

[RFC4328] and [RFC7139] provide the extensions to the Generalized Multi-Protocol Label Switching (GMPLS) signaling to control the full set of OTN features including ODU0, ODU1, ODU2, ODU3, ODU4, ODU2e and ODUflex. However, these specifications do not cover additional signal types ODU1e, ODU3e1, and ODU3e2 mentioned in [G.Sup43]. This draft provides GMPLS signaling extension for these additional signal types.

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. RSVP-TE extension for Additional Signal Types	2
3. Security Considerations	3
4. IANA Considerations	3
5. Acknowledgments	3
6. References	3
6.1. Normative References	3
6.2. Informative References	4

1. Introduction

[RFC7139] updates the ODU-related portions of [RFC4328] to provide Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) extensions to support control for [G.709-v3]. However, it does not cover additional signal types mentioned in [G.Sup43] (ODU1e, ODU3e1, and ODU3e2). This draft provides GMPLS signaling extension to support these additional signal types mentioned in [G.Sup43].

2. RSVP-TE extension for Additional Signal Types

[RFC7139] defines the format of Traffic Parameters in OTN-TDM SENDER_TSPEC and OTN-TDM FLOWSPEC objects. The said traffic parameters have a signal type field. This document defines the

Internet-Draft draft-ali-ccamp-additional-signal-type-g709v3-05.txt

signal type for ODU1e, ODU3e1 and ODU3e2 as defined in the IANA consideration section.

3. Security Considerations

This document does not introduce any additional security issues above those identified in [RFC7139].

4. IANA Considerations

This document defines signal type for ODU1e, ODU3e1 and ODU3e2, as follows:

Value	Type
-----	----
TBD	ODU1e (10Gbps Ethernet [GSUP.43])
TBD	ODU3e1 (40Gbps Ethernet [GSUP.43])
TBD	ODU3e2 (40Gbps Ethernet [GSUP.43])

These signaled types are carried in Traffic Parameters in OTN-TDM SENDER_TSPEC and OTN-TDM FLOWSPEC objects [RFC7139].

5. Acknowledgments

The authors would like to thank Lou Berger, Adrian Farrel and Sudip Shukla for comments.

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.
- [RFC4328] Papadimitriou, D., Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Extensions for G.709 Optical Transport Networks Control", RFC 4328, January 2006.
- [RFC7139] Zhang, F., Ed., Zhang, G., Belotti, S., Ceccarelli, D., and K. Pithewan, "GMPLS Signaling Extensions for Control of Evolving G.709 Optical Transport Networks", RFC 7139, March 2014.
- [RFC7139] F.Zhang, G.Zhang, S.Belotti, D.Ceccarelli, K.Pithewan, "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Extensions for the evolving G.709 Optical Transport Networks Control, draft-ietf-ccamp-gmpls-signaling-g709v3, work in progress.

Internet-Draft draft-ali-ccamp-additional-signal-type-g709v3-05.txt

6.2. Informative References

[G.709-v3] ITU-T, "Interface for the Optical Transport Network (OTN)", G.709/Y.1331 Recommendation, February, 2012.

[GSUP.43] ITU-T, "Proposed revision of G.sup43 (for agreement)", February, 2011.

Authors' Addresses

Zafar Ali
Cisco Systems
Email: zali@cisco.com

Antonello Bonfanti
Cisco Systems
abonfant@cisco.com

Fatai Zhang
Huawei Technologies
Email: zhangfatai@huawei.com

PCE Working Group
Internet Draft
Intended status: Standards Track

Zafar Ali
Antonello Bonfanti
Cisco Systems
F. Zhang
Huawei Technologies
August 29, 2014

Expires: February 28, 2015

IANA Allocation Procedures for OTN Signal Type Subregistry to
the GMPLS Signaling Parameters Registry
draft-ali-ccamp-otn-signal-type-subregistry-02.txt

Status of this Memo

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

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). 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 February 28, 2015.

Copyright Notice

Copyright (c) 2013 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.

Abstract

IANA has defined an "OTN Signal Type" subregistry to the "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Parameters" registry. This draft proposes changes to OTN Signal Type subregistry to include Specification Required policies, as defined in [RFC5226].

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. IANA Considerations	3
3. References	3
3.1. Normative References	3
3.2. Informative References	3

1. Introduction

[RFC4328] and [RFC7139] provide the extensions to the Generalized Multi-Protocol Label Switching (GMPLS) signaling to control the full set of OTN features including ODU0, ODU1, ODU2, ODU3, ODU4, ODU2e and ODUFlex. However, it does not cover additional signal types mentioned in [G.Sup43] (ODU1e, ODU3e1, and ODU3e2). As ODU1e, ODU3e1, and ODU3e2 signal types are only defined in an ITU-T supplementary document, IANA cannot allocate values from the Standards Action registration policy defined in [RFC5226].

IANA maintains "OTN Signal Type" subregistry to the "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Parameters" registry for the OTN signal defined in [RFC4328] and [RFC7139]. However, this subregistry currently is defined to only use the Standards Action registration policy as defined by [RFC5226]. This document extends "OTN Signal Type" subregistry to also support Specification Required policies, as defined in [RFC5226].

2. IANA Considerations

IANA maintains the an "OTN Signal Type" subregistry to the "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Parameters" registry. The registry currently is defined to use the Standards Action registration policy as defined by [RFC5226]. This document directs that both Standards Action and Specification Required policies, as defined in [RFC5226], be applied to this subregistry. When needed, the Designated Expert shall be identified by a CCAMP WG chair or, in the case the group is no longer active, by the IESG.

3. Acknowledgments

The authors would like to thank Lou Berger and Adrian Farrel for comments.

4. References

4.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC4328] Papadimitriou, D., Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Extensions for G.709 Optical Transport Networks Control", RFC 4328, January 2006.
- [RFC7139] Zhang, F., Ed., Zhang, G., Belotti, S., Ceccarelli, D., and K. Pithewan, "GMPLS Signaling Extensions for Control of Evolving G.709 Optical Transport Networks", RFC 7139, March 2014.
- [RFC5226] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 5226, May 2008.

4.2. Informative References

- [GSUP.43] ITU-T, "Proposed revision of G.sup43 (for agreement)", February, 2011.

Authors' Addresses

draft-ali-ccamp-otn-signal-type-subregistry-02.txt

Zafar Ali
Cisco Systems
Email: zali@cisco.com

Antonello Bonfanti
Cisco Systems
abonfant@cisco.com

Fatai Zhang
Huawei Technologies
Email: zhangfatai@huawei.com

CCAMP Working Group
Internet Draft
Intended status: Standards Track

Vishnu Pavan Beeram (Ed)
Juniper Networks
Igor Bryskin (Ed)
ADVA Optical Networking

Expires: January 03, 2015

July 03, 2014

Network Assigned Upstream-Label
draft-beeram-ccamp-network-assigned-upstream-label-03

Status of this Memo

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

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

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

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

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

This Internet-Draft will expire on January 03, 2015.

Copyright Notice

Copyright (c) 2014 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 discusses a GMPLS RSVP-TE protocol mechanism that enables the network to assign an upstream-label for a given LSP. This is useful in scenarios where a given node does not have sufficient information to assign the correct upstream-label on its own and needs to rely on the network to pick an appropriate label.

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. Use-Case: Alien Wavelength Setup.....	3
3. The "crank-back" approach.....	3
4. Symmetric Labels.....	5
5. Unassigned Upstream Label.....	5
5.1. Processing Rules.....	5
5.2. Backwards Compatibility.....	6
6. Applicability.....	6
6.1. Initial Setup.....	7
6.2. Wavelength Change.....	8
7. Security Considerations.....	8
8. IANA Considerations.....	8
9. Normative References.....	8
10. Acknowledgments.....	8

1. Introduction

The GMPLS RSVP-TE extensions for setting up a Bidirectional LSP are discussed in [RFC3473]. The Bidirectional LSP setup is indicated by the presence of an UPSTREAM_LABEL Object in the PATH message. As per the existing setup procedure outlined for a Bidirectional LSP, each upstream-node must allocate a valid upstream-label on the outgoing interface before sending the initial PATH message downstream. However, there are certain scenarios where it is not desirable or possible for a given node to pick the upstream-label on its own. This document defines the protocol mechanism to be used in such

scenarios. This mechanism enables a given node to offload the task of assigning the upstream-label for a given LSP onto the network.

2. Use-Case: Alien Wavelength Setup

Consider the network topology depicted in Figure 1. Nodes A and B are client IP routers that are connected to an optical WDM transport network. F, H and I represent WDM nodes. The transponder sits on the router and is directly connected to the add-drop port on a WDM node.

The optical signal originating on "Router A" is tuned to a particular wavelength. On "WDM-Node F", it gets multiplexed with optical signals at other wavelengths. Depending on the implementation of this multiplexing function, it may not be acceptable to have the router send signal into the optical network unless it is at the appropriate wavelength. In other words, having the router send signal with a wrong wavelength may adversely impact existing optical trails. If the clients do not have full visibility into the optical network, they are not in a position to pick the correct wavelength up-front.

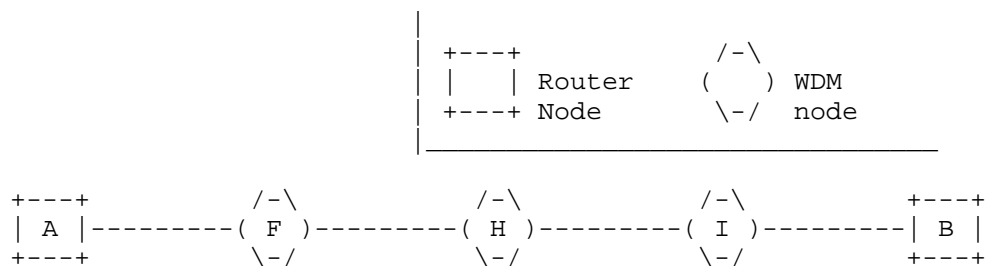


Figure 1: Sample topology

3. The "crank-back" approach

There are currently no GMPLS RSVP-TE protocol mechanisms that an upstream-node can use for indicating that it does not know what upstream-label to use and that it needs the downstream-node to pick the label on its behalf.

The following setup sequence is an attempt to address the above use-case using existing protocol mechanisms:

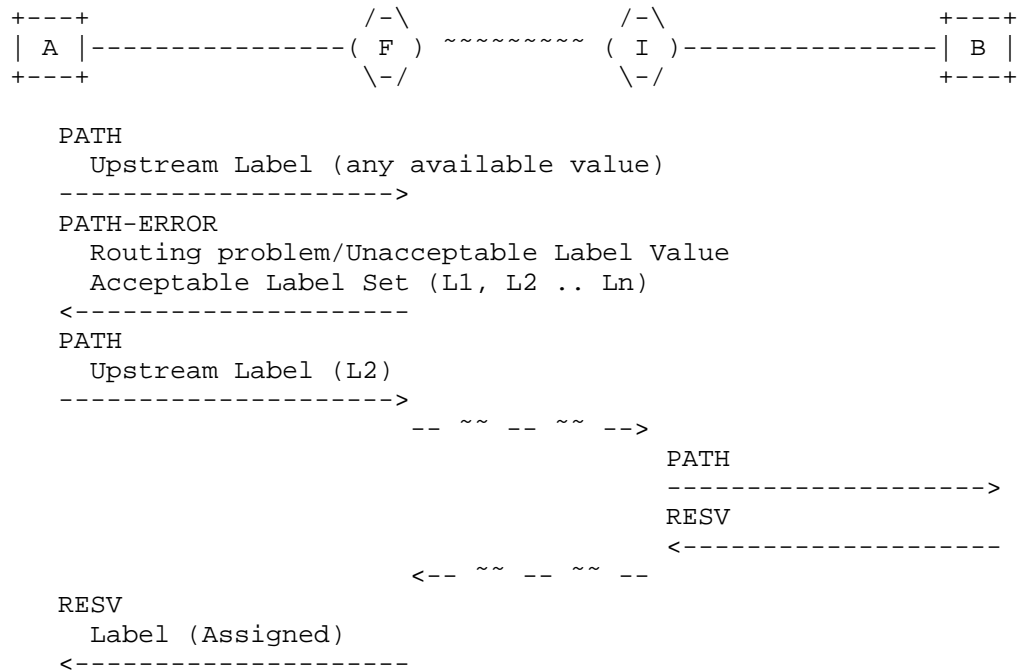


Figure 2: Setup Sequence - Crank-back Approach

The above approach does sort of work, but there are a few obvious concerns:

- Since "Router-A" does not know which upstream-label to use, it picks some random label and signals it without programming its data-plane. As a result, the outgoing PATH message has no indication of whether the upstream-label has been installed along the data-path or not.
- If "Router-A" somehow correctly guesses (by sheer luck) an acceptable upstream label upfront, the network may end up finding a path which is suboptimal (there could be a different acceptable upstream label which corresponds to a better path in the network)
- The "Path-Error with Acceptable Label Set" retry approach is usually used for exception handling. The above solution uses it for almost every single setup request (except in the rare scenario where the appropriate upstream-label is guessed correctly).
- There is an awkward window between the time the network sends out the Path-Error (with the `ACCEPTABLE_LABEL_SET`) and receives the corresponding Path (with the selected `UPSTREAM_LABEL`); this window

- opens up the possibility of the selected UPSTREAM_LABEL to be stale by the time the network receives the retry PATH.
- The above solution assumes the use of "symmetric labels" by default.

The rest of the sections in this draft discuss a solution proposal that is devoid of any of the above concerns.

4. Symmetric Labels

As per [RFC3471], the upstream-label and the downstream-label for an LSP at a given hop need not be the same. The use-case discussed in this document pertains to Lambda Switch Capable (LSC) LSPs and it is an undocumented fact that in practice, LSC LSPs always have symmetric labels at each hop along the path of the LSP.

The use of the protocol mechanism discussed in this document mandates "Label Symmetry". This mechanism is meant to be used only for Bidirectional LSPs that assign Symmetric Labels at each hop along the path of the LSP.

5. Unassigned Upstream Label

This document proposes the use of a special label value - "0xFFFFFFFF" - to indicate an Unassigned Label. The presence of this value in the UPSTREAM_LABEL object of a PATH message indicates that the upstream-node has not assigned an upstream label on its own and has requested the downstream-node to provide a label that it can use in both forward and reverse directions. The presence of this value in the UPSTREAM_LABEL object of a PATH message can also be interpreted as a request to mandate "symmetric labels" for the LSP at the given hop.

5.1. Processing Rules

The Unassigned Upstream Label is used by an upstream-node when it is not in a position to pick the upstream label on its own. In such a scenario, the upstream-node sends a PATH message downstream with an Unassigned Upstream Label and requests the downstream-node to provide a symmetric label. If the upstream-node desires to make the downstream-node aware of its limitations with respect to label selection, it has the option to specify a list of valid labels via the LABEL_SET object.

In response, the downstream-node picks an appropriate symmetric label and sends it via the LABEL object in the RESV message. The

upstream-node would then start using this symmetric label for both directions of the LSP. If the downstream-node cannot pick the symmetric label, it MUST issue a PATH-ERR message with a "Routing Problem/Unacceptable Label Value" indication.

The upstream-node will continue to signal the Unassigned Upstream Label in the PATH message even after it receives an appropriate symmetric label in the RESV message. This is done to make sure that the downstream-node would pick a symmetric label if and when it needs to change the RESV label at a later point in time.

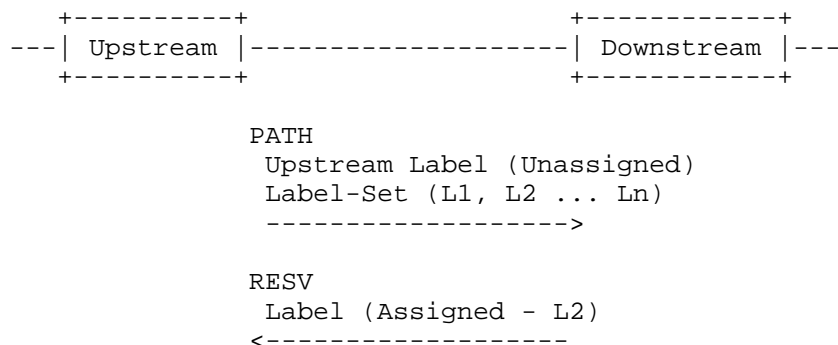


Figure 3: Unassigned UPSTREAM_LABEL

5.2. Backwards Compatibility

If the downstream-node is running an older implementation (which may be using the "crank-back" approach discussed in Section 3) and doesn't understand the semantics of an Unassigned UPSTREAM_LABEL, it will either (a) reject the special label value and generate an error or (b) accept it and treat it as a valid label.

If the behavior that is exhibited is (a), then there are obviously no backwards compatibility concerns. Ingress implementations may even choose to adopt the "crank-back" approach in such cases. If there is some existing implementation that exhibits the behavior in (b), then there could be some potential issues. The use-case discussed in this draft pertains to LSC LSPs and it is safe to assume that the behavior in (b) will not be exhibited for such LSPs.

6. Applicability

Let us revisit the "alien wavelength" use-case discussed in Section 2 and examine how the mechanism proposed in this document allows the

optical network to select and communicate the correct wavelength to its clients.

6.1. Initial Setup

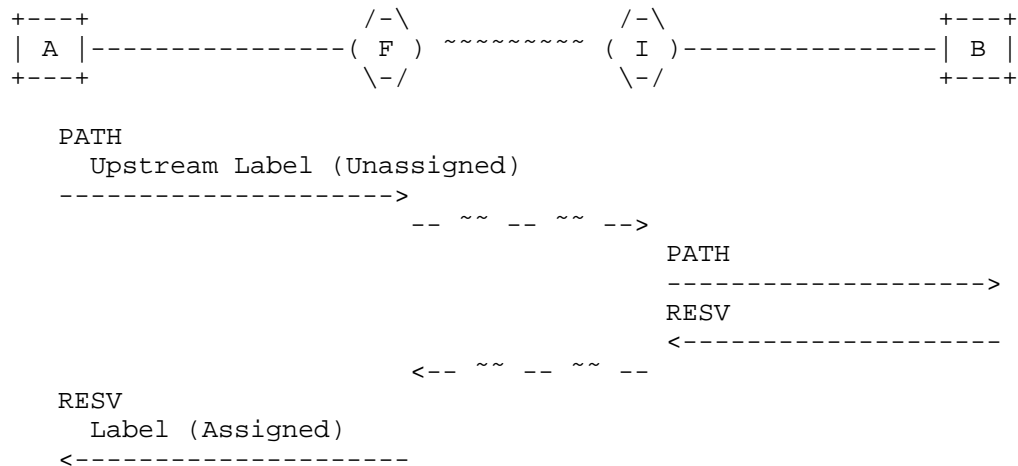


Figure 4: Alien Wavelength - Initial Setup

Steps:

- "Router A" does not have enough information to pick an appropriate client wavelength. It sends a PATH downstream requesting the network to assign an appropriate symmetric label for it to use. Since the client wavelength is unknown, the laser is off at the ingress client.
- The network receives the PATH, chooses the appropriate wavelength values and forwards them in appropriate label fields to the egress client ("Router B")
- "Router B" receives the PATH, turns the laser ON and tunes it to the appropriate wavelength (received in the UPSTREAM_LABEL/LABEL_SET of the PATH) and sends out a RESV upstream.
- The RESV received by the ingress client carries a valid symmetric label in the LABEL object. "Router A" turns on the laser and tunes it to the wavelength specified in the network assigned symmetric LABEL.

For cases where the egress-node relies on RSVP signaling to determine exactly when to start using the LSP, this draft recommends

integrating the above sequence with any of the existing graceful setup procedures:

- "RESV-CONF" setup procedure (or)
- 2-step "ADMIN STATUS" based setup procedure ("A" bit set in the first step; "A" bit cleared when the LSP is ready for use).

6.2. Wavelength Change

After the LSP is set up, the network MAY decide to change the wavelength for the given LSP. This could be for a variety of reasons - policy reasons, restoration within the core, preemption etc.

In such a scenario, if the ingress client receives a changed label via the LABEL object in a RESV modify, it MUST retune the laser at the ingress to the new wavelength. Similarly if the egress client receives a changed label via UPSTREAM_LABEL/LABEL_SET in a PATH modify, it MUST retune the laser at the egress to the new wavelength.

7. Security Considerations

TBD

8. IANA Considerations

TBD

9. 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 Signaling Functional Description", RFC 3471, January 2003
- [RFC3473] Berger, L., "Generalized Multi-Protocol Label Switching Signaling Resource Reservation Protocol-Traffic Engineering Extensions", RFC 3473, January 2003.

10. Acknowledgments

TBD

Authors' Addresses

Vishnu Pavan Beeram
Juniper Networks
Email: vbeeram@juniper.net

John Drake
Juniper Networks
Email: jdrake@juniper.net

Gert Grammel
Juniper Networks
Email: ggrammel@juniper.net

Igor Bryskin
ADVA Optical Networking
Email: ibryskin@advaoptical.com

Pawel Brzozowski
ADVA Optical Networking
Email: pbrzozowski@advaoptical.com

Daniele Ceccarelli
Ericsson
Email: daniele.ceccarelli@ericsson.com

Oscar Gonzalez de Dios
Telefonica
Email: ogondio@tid.es

Internet Engineering Task Force
Internet-Draft
Intended status: Standards Track
Expires: January 7, 2016

D. Hiremagalur, Ed.
G. Grammel, Ed.
Juniper
G. Galimberti, Ed.
Z. Ali, Ed.
Cisco
R. Kunze, Ed.
Deutsche Telekom
D. Beller, Ed.
ALU
July 6, 2015

Extension to the Link Management Protocol (LMP/DWDM -rfc4209) for Dense
Wavelength Division Multiplexing (DWDM) Optical Line Systems to manage
the application code of optical interface parameters in DWDM application
draft-dharinigert-ccamp-g-698-2-lmp-10

Abstract

This memo defines extensions to LMP(rfc4209) for managing Optical parameters associated with Wavelength Division Multiplexing (WDM) systems or characterized by the Optical Transport Network (OTN) in accordance with the Interface Application Code approach defined in ITU-T Recommendation G.698.2.[ITU.G698.2], G.694.1.[ITU.G694.1] and its extensions.

Copyright Notice

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

Status of This Memo

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

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

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

This Internet-Draft will expire on January 7, 2016.

Copyright Notice

Copyright (c) 2015 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

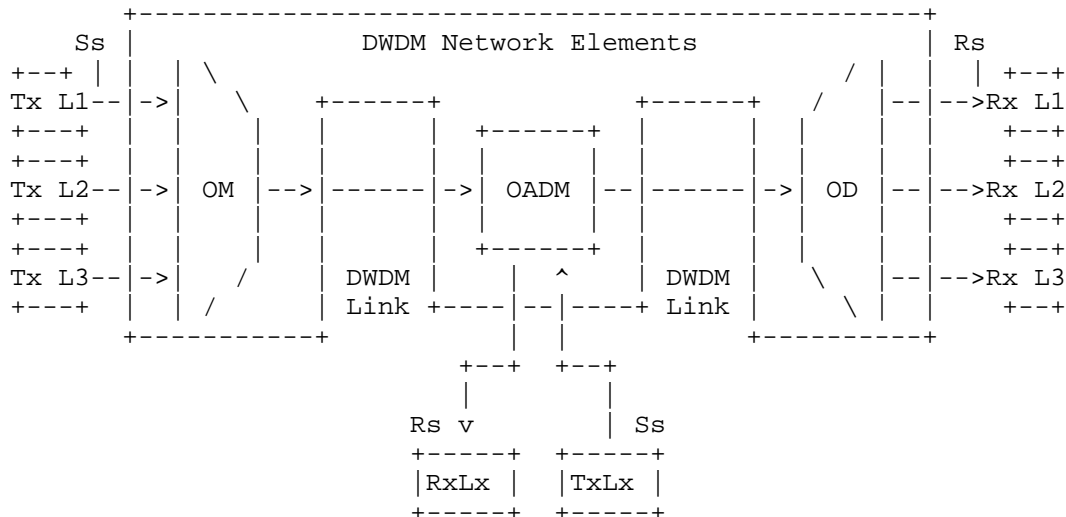
1. Introduction	2
2. Use Cases	4
3. Extensions to LMP-WDM Protocol	11
4. General Parameters - OCh_General	11
5. ApplicationIdentifier - OCh_ApplicationIdentifier	13
6. OCh_Ss - OCh transmit parameters	15
7. OCh_Rs - receive parameters	15
8. Security Considerations	16
9. IANA Considerations	16
10. Contributors	17
11. References	17
11.1. Normative References	17
11.2. Informative References	18
Authors' Addresses	18

1. Introduction

This extension is based on "draft-galikunze-ccamp-g-698-2-snmp-mib-10", for the relevant interface optical parameters described in recommendations like ITU-T G.698.2 [ITU.G698.2] and G.694.1.[ITU.G694.1]. The LMP Model from RFC4902 provides link property correlation between a client and an OLS device. LMP link property correlation, exchanges the capabilities of either end of the link where the term 'link' refers to the attachment link between OXC and OLS (see Figure 1). By performing link property correlation, both ends of the link exchange link properties, such as application identifiers. This allows either end to operate within a commonly understood parameter window. Based on known parameter limits, each device can supervise the received signal for conformance using mechanisms defined in RFC3591. For example if the Client transmitter power (OXC1) has a value of 0dBm and the ROADM interface measured

power (at OLS1) is -6dBm the fiber patch cord connecting the two nodes may be pinched or the connectors are dirty. More, the interface characteristics can be used by the OLS network Control Plane in order to check the Optical Channels feasibility. Finally the OXC1 transceivers parameters (Application Code) can be shared with OXC2 using the LMP protocol to verify the Transceivers compatibility. The actual route selection of a specific wavelength within the allowed set is outside the scope of LMP. In GMPLS, the parameter selection (e.g. central frequency) is performed by RSVP-TE.

Figure 1 shows a set of reference points, for the linear "black link" approach, for single-channel connection (Ss and Rs) between transmitters (Tx) and receivers (Rx). Here the DWDM network elements include an OM and an OD (which are used as a pair with the opposing element), one or more optical amplifiers and may also include one or more OADMs.

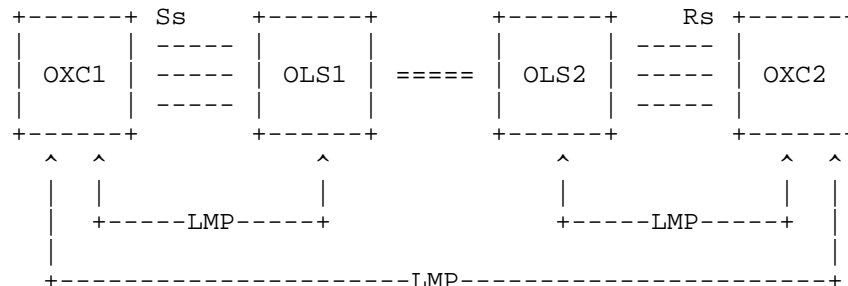


Ss = reference point at the DWDM network element tributary output
Rs = reference point at the DWDM network element tributary input
Lx = Lambda x
OM = Optical Mux
OD = Optical Demux
OADM = Optical Add Drop Mux

from Fig. 5.1/G.698.2

Figure 1: Linear Black Link approach

Figure 2 Extended LMP Model (from [RFC4209])



OXC : is an entity that contains transponders
 OLS : generic optical system, it can be -
 Optical Mux, Optical Demux, Optical Add
 Drop Mux, etc.
 OLS to OLS : represents the black-Link itself
 Rs/Ss : in between the OXC and the OLS

Figure 2: Extended LMP Model

2. Use Cases

The use cases described below are assuming that power monitoring functions are available in the ingress and egress network element of the DWDM network, respectively. By performing link property correlation it would be beneficial to include the current transmit power value at reference point Ss and the current received power value at reference point Rs. For example if the Client transmitter power (OXC1) has a value of 0dBm and the ROADM interface measured power (at OLS1) is -6dBm the fiber patch cord connecting the two nodes may be pinched or the connectors are dirty. More, the interface characteristics can be used by the OLS network Control Plane in order to check the Optical Channels feasibility. Finally the OXC1 transceivers parameters (Application Code) can be shared with OXC2 using the LMP protocol to verify the Transceivers compatibility. The actual route selection of a specific wavelength within the allowed set is outside the scope of LMP. In GMPLS, the parameter selection (e.g. central frequency) is performed by RSVP-TE.

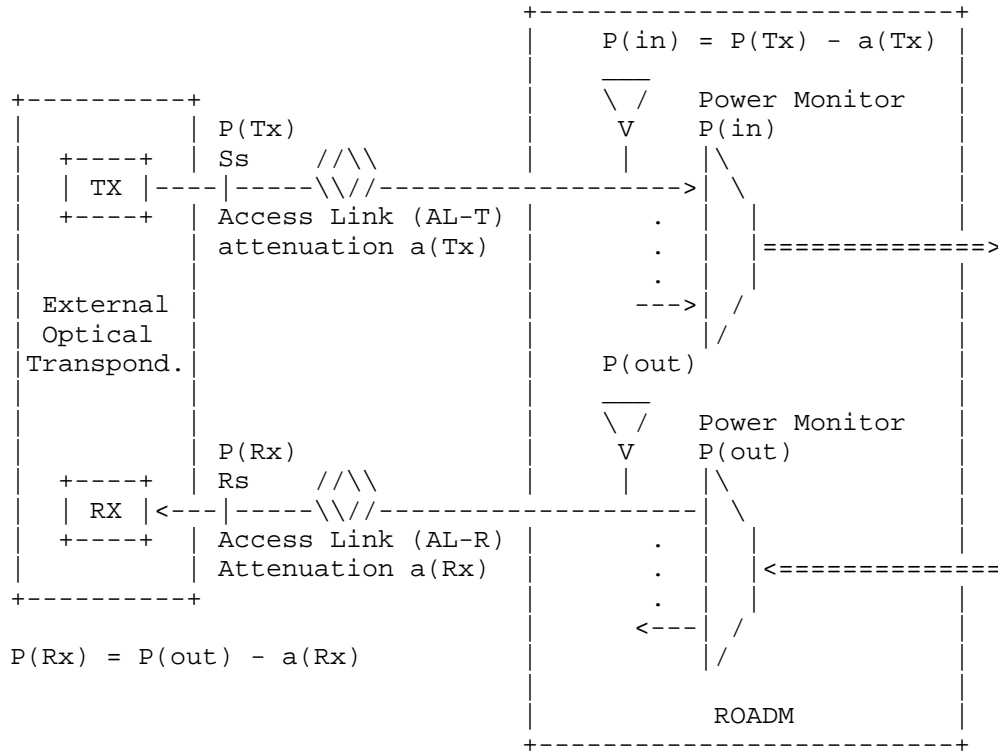
G.698.2 defines a single channel optical interface for DWDM systems that allows interconnecting network-external optical transponders across a DWDM network. The optical transponders are considered to be external to the DWDM network. This so-called 'black link' approach

illustrated in Figure 5-1 of G.698.2 and a copy of this figure is provided below. The single channel fiber link between the Ss/Rs reference points and the ingress/egress port of the network element on the domain boundary of the DWDM network (DWDM border NE) is called access link in this contribution. Based on the definition in G.698.2 it is considered to be part of the DWDM network. The access link typically is realized as a passive fiber link that has a specific optical attenuation (insertion loss). As the access link is an integral part of the DWDM network, it is desirable to monitor its attenuation. Therefore, it is useful to detect an increase of the access link attenuation, for example, when the access link fiber has been disconnected and reconnected (maintenance) and a bad patch panel connection (connector) resulted in a significantly higher access link attenuation (loss of signal in the extreme case of an open connector or a fiber cut). In the following section, two use cases are presented and discussed:

- 1) pure access link monitoring
- 2) access link monitoring with a power control loop

These use cases require a power monitor as described in G.697 (see section 6.1.2), that is capable to measure the optical power of the incoming or outgoing single channel signal. The use case where a power control loop is in place could even be used to compensate an increased attenuation as long as the optical transmitter can still be operated within its output power range defined by its application code.

Figure 3 Access Link Power Monitoring



- For AL-T monitoring: $P(Tx)$ and $a(Tx)$ must be known
- For AL-R monitoring: $P(Rx)$ and $a(Rx)$ must be known

An alarm shall be raised if $P(in)$ or $P(Rx)$ drops below a configured threshold (t [dB]):

- $P(in) < P(Tx) - a(Tx) - t$ (Tx direction)
- $P(Rx) < P(out) - a(Rx) - t$ (Rx direction)
- $a(Tx) = | a(Rx)$

Figure 3: Extended LMP Model

Pure Access Link (AL) Monitoring Use Case

Figure 4 illustrates the access link monitoring use case and the different physical properties involved that are defined below:

- S_s, R_s : G.698.2 reference points
- $P(Tx)$: current optical output power of transmitter Tx
- $a(Tx)$: access link attenuation in Tx direction (external transponder point of view)
- $P(in)$: measured current optical input power at the input port of border DWDM NE
- t : user defined threshold (tolerance)
- $P(out)$: measured current optical output power at the output port of border DWDM NE
- $a(Rx)$: access link attenuation in Rx direction (external transponder point of view)
- $P(Rx)$: current optical input power of receiver Rx

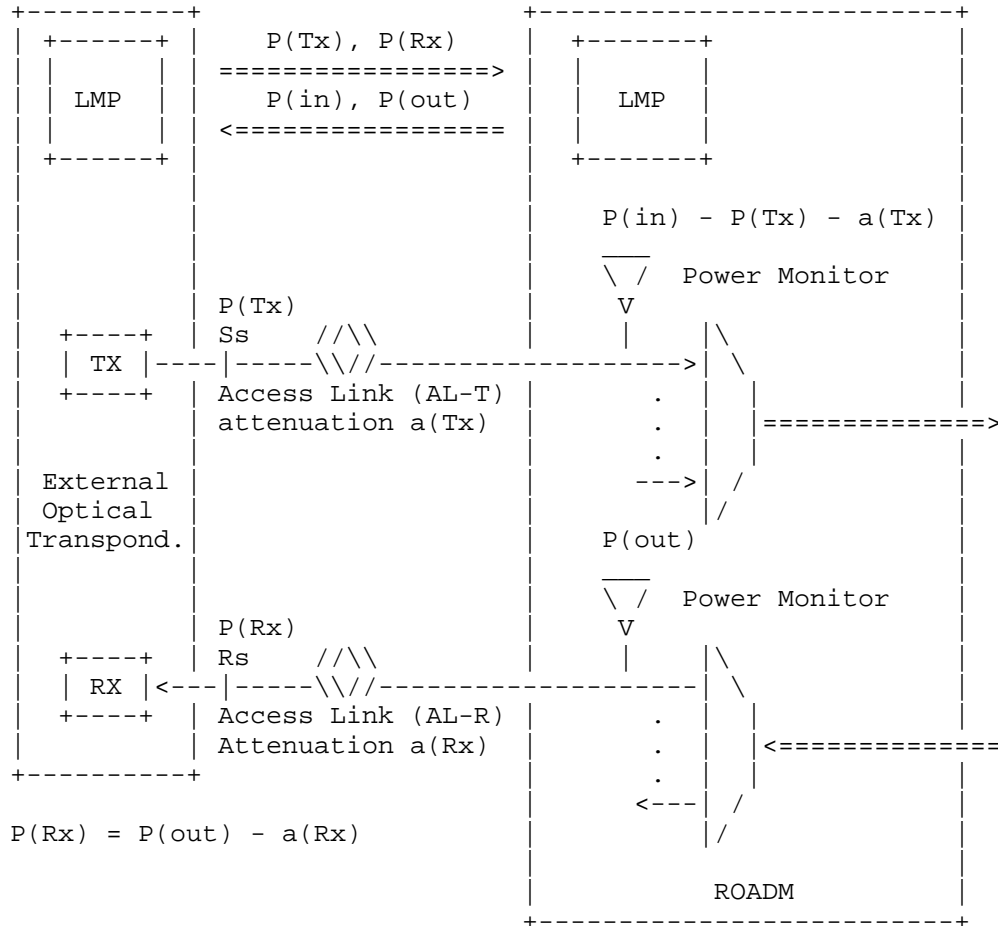
Assumptions:

- The access link attenuation in both directions ($a(Tx)$, $a(Rx)$) is known or can be determined as part of the commissioning process. Typically, both values are the same.
- A threshold value t has been configured by the operator. This should also be done during commissioning.
- A control plane protocol (e.g. this draft) is in place that allows to periodically send the optical power values $P(Tx)$ and $P(Rx)$ to the control plane protocol instance on the DWDM border NE. This is illustrated in Figure 3.
- The DWDM border NE is capable to periodically measure the optical power P_{in} and P_{out} as defined in G.697 by power monitoring points depicted as yellow triangles in the figures below.

AL monitoring process:

- Tx direction: the measured optical input power P_{in} is compared with the expected optical input power $P(Tx) - a(Tx)$. If the measured optical input power P_{in} drops below the value $(P(Tx) - a(Tx) - t)$ a low power alarm shall be raised indicating that the access link attenuation has exceeded $a(Tx) + t$.
- Rx direction: the measured optical input power $P(Rx)$ is compared with the expected optical input power $P(out) - a(Rx)$. If the measured optical input power $P(Rx)$ drops below the value $(P(out) - a(Rx) - t)$ a low power alarm shall be raised indicating that the access link attenuation has exceeded $a(Rx) + t$.

Figure 4 Use case 1: Access Link power monitoring



- For AL-T monitoring: $P(Tx)$ and $a(Tx)$ must be known
 - For AL-R monitoring: $P(Rx)$ and $a(Rx)$ must be known
- An alarm shall be raised if $P(in)$ or $P(Rx)$ drops below a configured threshold (t [dB]):
- $P(in) < P(Tx) - a(Tx) - t$ (Tx direction)
 - $P(Rx) < P(out) - a(Rx) - t$ (Rx direction)
 - $a(Tx) = a(Rx)$

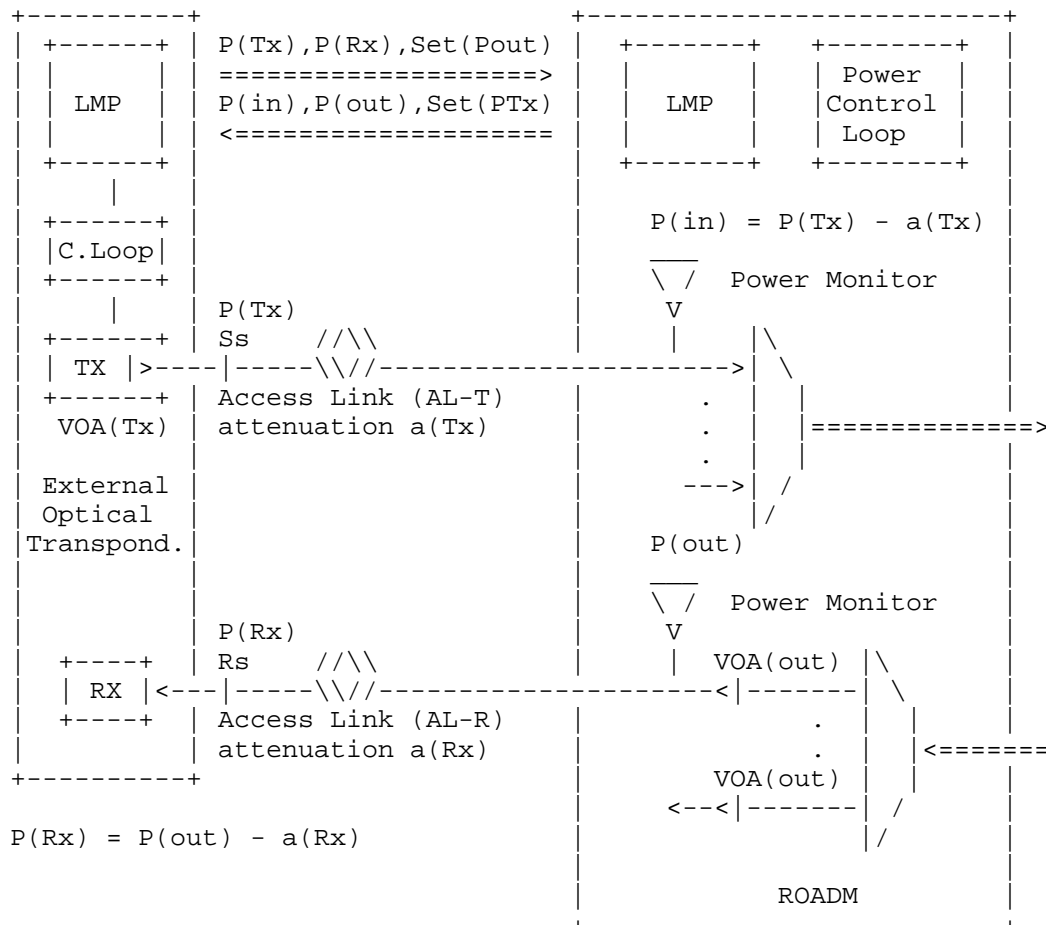
Figure 4: Extended LMP Model

Power Control Loop Use Case

This use case is based on the access link monitoring use case as described above. In addition, the border NE is running a power control application that is capable to control the optical output power of the single channel tributary signal at the output port of the border DWDM NE (towards the external receiver Rx) and the optical output power of the single channel tributary signal at the external transmitter Tx within their known operating range. The time scale of this control loop is typically relatively slow (e.g. some 10s or minutes) because the access link attenuation is not expected to vary much over time (the attenuation only changes when re-cabling occurs).

From a data plane perspective, this use case does not require additional data plane extensions. It does only require a protocol extension in the control plane (e.g. this LMP draft) that allows the power control application residing in the DWDM border NE to modify the optical output power of the DWDM domain-external transmitter Tx within the range of the currently used application code. Figure 5 below illustrates this usecase utilizing the LMP protocol with extensions defined in this draft.

Figure 5 Use case 2: Power Control Loop



- The Power Control Loops in Transponder and ROADM regulate the Variable Optical Attenuators (VOA) to adjust the proper power in base of the ROADM and Receiver characteristics and the Access Link attenuation

Figure 5: Extended LMP Model

3. Extensions to LMP-WDM Protocol

This document defines extensions to [RFC4209] to allow the Black Link (BL) parameters of G.698.2, to be exchanged between a router or optical switch and the optical line system to which it is attached. In particular, this document defines additional Data Link sub-objects to be carried in the LinkSummary message defined in [RFC4204] and [RFC6205]. The OXC and OLS systems may be managed by different Network management systems and hence may not know the capability and status of their peer. The intent of this draft is to enable the OXC and OLS systems to exchange this information. These messages and their usage are defined in subsequent sections of this document.

The following new messages are defined for the WDM extension for ITU-T G.698.2 [ITU.G698.2]/ITU-T G.698.1 [ITU.G698.1]/ITU-T G.959.1 [ITU.G959.1]

- OCh_General (sub-object Type = TBA)
- OCh_ApplicationIdentifier (sub-object Type = TBA)
- OCh_Ss (sub-object Type = TBA)
- OCh_Rs (sub-object Type = TBA)

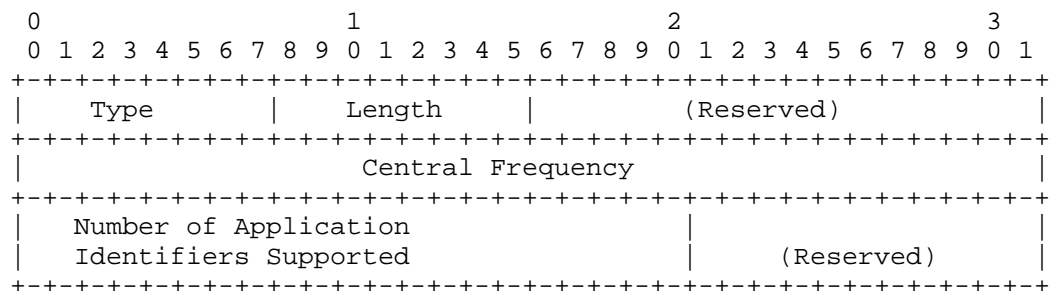
4. General Parameters - OCh_General

These are the general parameters as described in [G698.2] and [G.694.1]. Please refer to the "draft-galikunze-ccamp-g-698-2-snmp-mib-12" for more details about these parameters and the [RFC6205] for the wavelength definition.

The general parameters are

1. Central Frequency - (Tera Hz) 4 bytes (see RFC6205 sec.3.2)
2. Number of Application Identifiers (A.I.) Supported
3. Single-channel Application Identifier in use
4. Application Identifier Type in use
5. Application Identifier in use

Figure 6: The format of the this sub-object (Type = TBA, Length = TBA) is as follows:



Single-channel Application Identifier Number in use	A.I. Type in use	A.I. length
Single-channel Application Identifier in use		
Single-channel Application Identifier in use		
Single-channel Application Identifier in use		

A.I. Type in use: STANDARD, PROPRIETARY

A.I. Type in use: STANDARD

Refer to G.698.2 recommendation : B-DScW-ytz(v)

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			
Single-channel Application Code			
Single-channel Application Code			
Single-channel Application Code			

A.I. Type in use: PROPRIETARY

Note: if the A.I. type = PROPRIETARY, the first 6 Octets of the Application Identifier in use are six characters of the PrintableString must contain the Hexadecimal representation of an OUI (Organizationally Unique Identifier) assigned to the vendor whose implementation generated the Application Identifier; the remaining octets of the PrintableString are unspecified.

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			
OUI			
OUI cont.		Vendor value	
Vendor Value			

```

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

```

Figure 6: OCh_General

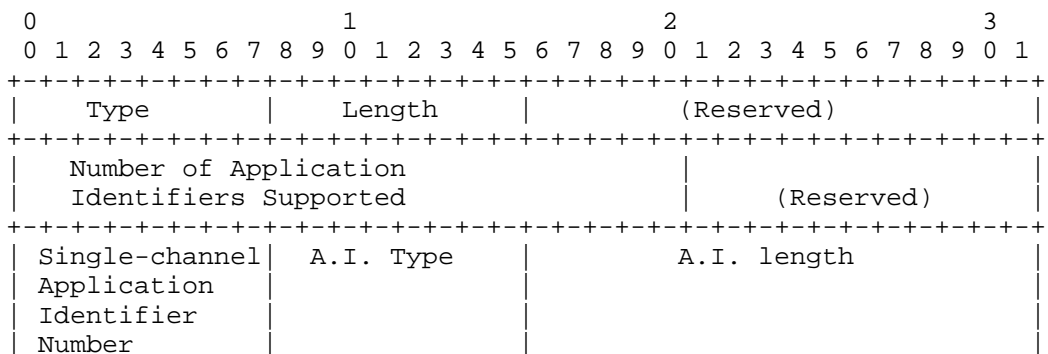
5. ApplicationIdentifier - OCh_ApplicationIdentifier

This message is to exchange the application identifiers supported as described in [G698.2]. Please refer to the "draft-galikunze-ccamp-g-698-2-snmp-mib-10". For more details about these parameters. There can be more than one Application Identifier supported by the OXC/OLS. The number of application identifiers supported is exchanged in the "OCh_General" message. (from [G698.1]/[G698.2]/[G959.1] and G.874.1)

The parameters are

1. Number of Application Identifiers (A.I.) Supported
 2. Single-channel application identifier Number uniquely identifies this entry - 8 bits
 3. Application Identifier Type (A.I.) (STANDARD/PROPRIETARY)
 4. Single-channel application identifier -- 96 bits (from [G698.1]/[G698.2]/[G959.1])
- this parameter can have multiple instances as the transceiver can support multiple application identifiers.

Figure 7: The format of the this sub-object (Type = TBA, Length = TBA) is as follows:



```

+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|               Single-channel Application Identifier               |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|               Single-channel Application Identifier               |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|               Single-channel Application Identifier               |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
//               ....               //
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| Single-channel |           A.I. Type           |           A.I. length           | |
| Application   |           |           |           |
| Identifier    |           |           |           |
| Number       |           |           |           |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|               Single-channel Application Identifier               |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|               Single-channel Application Identifier               |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|               Single-channel Application Identifier               |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+

```

A.I. Type in use: STANDARD, PROPRIETARY

A.I. Type in use: STANDARD

Refer to G.698.2 recommendation : B-DScW-ytz(v)

```

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
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|               Single-channel Application Code                     |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|               Single-channel Application Code                     |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|               Single-channel Application Code                     |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+

```

A.I. Type in use: PROPRIETARY

Note: if the A.I. type = PROPRIETARY, the first 6 Octets of the Application Identifier in use are six characters of the PrintableString must contain the Hexadecimal representation of an OUI (Organizationally Unique Identifier) assigned to the vendor whose implementation generated the Application Identifier; the remaining octets of the PrintableString are unspecified.

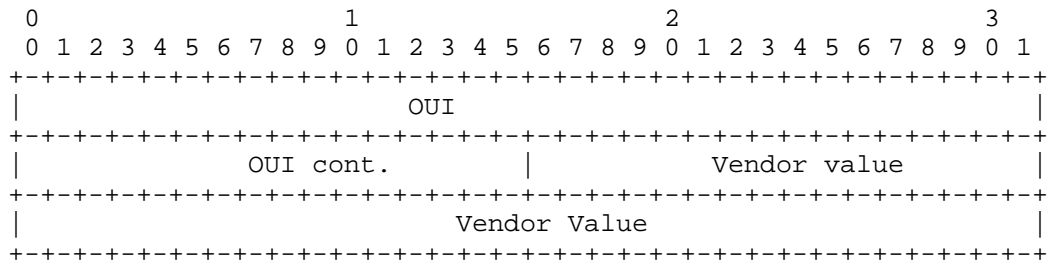


Figure 7: OCh_ApplicationIdentifier

6. OCh_Ss - OCh transmit parameters

These are the G.698.2 parameters at the Source(Ss reference points). Please refer to "draft-galikunze-ccamp-g-698-2-snmp-mib-10" for more details about these parameters.

1. Output power

Figure 8: The format of the OCh sub-object (Type = TBA, Length = TBA) is as follows:

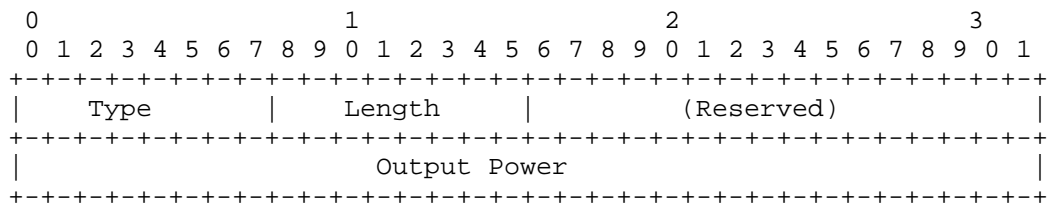


Figure 8: OCh_Ss transmit parameters

7. OCh_Rs - receive parameters

These are the G.698.2 parameters at the Sink (Rs reference points). Please refer to the "draft-galikunze-ccamp-g-698-2-snmp-mib-10" for more details about these parameters.

1. Current Input Power - (0.1dbm) 4bytes

Figure 9: The format of the OCh receive sub-object (Type = TBA, Length = TBA) is as follows:

The format of the OCh receive/OLS Sink sub-object (Type = TBA, Length = TBA) is as follows:

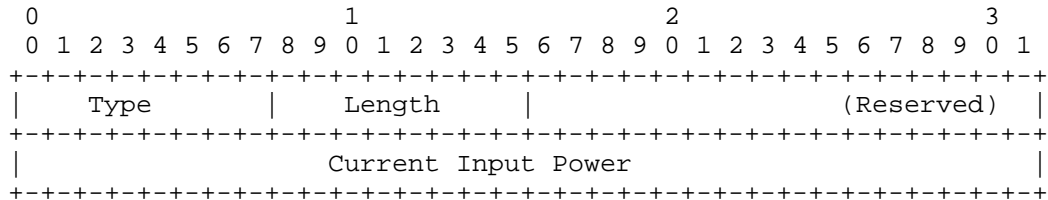


Figure 9: OCh_Rs receive parameters

8. Security Considerations

LMP message security uses IPsec, as described in [RFC4204]. This document only defines new LMP objects that are carried in existing LMP messages, similar to the LMP objects in [RFC:4209]. This document does not introduce new security considerations.

9. IANA Considerations

LMP <xref target="RFC4204"/> defines the following name spaces and the ways in which IANA can make assignments to these namespaces:

- LMP Message Type
 - LMP Object Class
 - LMP Object Class type (C-Type) unique within the Object Class
 - LMP Sub-object Class type (Type) unique within the Object Class
- This memo introduces the following new assignments:

LMP Sub-Object Class names:

under DATA_LINK Class name (as defined in <xref target="RFC4204"/>)

- OCh_General (sub-object Type = TBA)
- OCh_ApplicationIdentifier (sub-object Type = TBA)
- OCh_Ss (sub-object Type = TBA)
- OCh_Rs (sub-object Type = TBA)

10. Contributors

Arnold Mattheus
Deutsche Telekom
Darmstadt
Germany
email a.mattheus@telekom.de

John E. Drake
Juniper
1194 N Mathilda Avenue
HW-US, Pennsylvania
USA
jdrake@juniper.net

11. References

11.1. Normative References

- [RFC4204] Lang, J., "Link Management Protocol (LMP)", RFC 4204, October 2005.
- [RFC4209] Fredette, A. and J. Lang, "Link Management Protocol (LMP) for Dense Wavelength Division Multiplexing (DWDM) Optical Line Systems", RFC 4209, October 2005.
- [RFC6205] Otani, T. and D. Li, "Generalized Labels for Lambda-Switch-Capable (LSC) Label Switching Routers", RFC 6205, March 2011.
- [RFC4054] Strand, J. and A. Chiu, "Impairments and Other Constraints on Optical Layer Routing", RFC 4054, May 2005.
- [ITU.G698.2]
International Telecommunications Union, "Amplified multichannel dense wavelength division multiplexing applications with single channel optical interfaces", ITU-T Recommendation G.698.2, November 2009.
- [ITU.G694.1]
International Telecommunications Union, "Spectral grids for WDM applications: DWDM frequency grid", ITU-T Recommendation G.698.2, February 2012.

- [ITU.G709] International Telecommunications Union, "Interface for the Optical Transport Network (OTN)", ITU-T Recommendation G.709, February 2012.
- [ITU.G872] International Telecommunications Union, "Architecture of optical transport networks", ITU-T Recommendation G.872, October 2012.
- [ITU.G874.1] International Telecommunications Union, "Optical transport network (OTN): Protocol-neutral management information model for the network element view", ITU-T Recommendation G.874.1, October 2012.

11.2. Informative References

- [RFC3410] Case, J., Mundy, R., Partain, D., and B. Stewart, "Introduction and Applicability Statements for Internet-Standard Management Framework", RFC 3410, December 2002.
- [RFC2629] Rose, M., "Writing I-Ds and RFCs using XML", RFC 2629, June 1999.
- [RFC4181] Heard, C., "Guidelines for Authors and Reviewers of MIB Documents", BCP 111, RFC 4181, September 2005.
- [I-D.kunze-g-698-2-management-control-framework] Kunze, R., "A framework for Management and Control of optical interfaces supporting G.698.2", draft-kunze-g-698-2-management-control-framework-00 (work in progress), July 2011.

Authors' Addresses

Dharini Hiremagalur (editor)
Juniper
1194 N Mathilda Avenue
Sunnyvale - 94089 California
USA

Phone: +1408
Email: dharinih@juniper.net

Gert Grammel (editor)
Juniper
Oskar-Schlemmer Str. 15
80807 Muenchen
Germany

Phone: +49 1725186386
Email: ggrammel@juniper.net

Gabriele Galimberti (editor)
Cisco
Via S. Maria Molgora, 48
20871 - Vimercate
Italy

Phone: +390392091462
Email: ggalimbe@cisco.com

Zafar Ali (editor)
Cisco
3000 Innovation Drive
KANATA
ONTARIO K2K 3E8

Email: zali@cisco.com

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

Phone: +49xxxxxxxxxxx
Email: RKunze@telekom.de

Dieter Beller (editor)
ALU
Lorenzstrasse, 10
70435 Stuttgart
Germany

Phone: +4971182143125
Email: Dieter.Beller@alcatel-lucent.com

Network Working Group
Internet-Draft
Intended status: Informational
Expires: January 22, 2015

O. Gonzalez de Dios, Ed.
Telefonica GCTO
J. Meuric, Ed.
Orange
D. Ceccarelli
Ericsson
July 21, 2014

Terminology and Models for Control of Traffic Engineered Networks with
Client-Server Relationship
draft-dios-ccamp-control-models-customer-provider-01

Abstract

Different kinds of relationships can be established among interconnected Traffic Engineered Networks. In particular, this document focuses on the case where there is a client-server relation between the network domains. The domain interconnection is a policy and administrative boundary. This informational document collects current terminology and provides a taxonomy for the possible control plane based operation models.

Each control model defines, on the one hand, the level of information that the domain acting as client receives by control plane means from the domain acting as server and, on the other hand, the control model will determine what can be requested from the client domain to the server domain.

Status of This Memo

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

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

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

This Internet-Draft will expire on January 22, 2015.

Copyright Notice

Copyright (c) 2014 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. Examples of Client-Server TE Network Domain Scenarios . .	3
2. Terminology	4
2.1. Routing domain	4
2.2. Overlay of routing domains	4
2.3. Multilayer	4
2.4. Policy	5
2.5. Client Domain - Server Domain Interface	5
2.5.1. UNI in IP over Optical Networks	5
2.5.2. ITU-T Definition of UNI	5
2.5.3. OIF Definition of UNI	6
2.5.4. Proposed Vocabulary	6
2.6. Reachability	7
2.6.1. Unqualified Reachability	7
2.6.2. Qualified Reachability	7
2.6.3. Qualified Reachability with associated potential TE path	8
3. Control Models	8
3.1. Signaling Only	8
3.1.1. Signaling with Requirements	9
3.1.2. Signaling with Collection	9
3.2. Signaling and Reachability Model	9
3.2.1. Signalling + Basic Reachability	10
3.2.2. Signalling + Qualified Reachability	10
3.2.3. Signalling + Qualified Reachability + Potential Services	10
3.3. Service Attributes vs service constraints	10
3.4. Other Models	11
3.4.1. Multi-Layer Networks / Multi-Region Networks	11
3.4.2. Management Model	11
4. Abstraction	11

5. Security Considerations	11
6. Contributing Authors	11
7. Acknowledgments	12
8. References	12
8.1. Normative References	12
8.2. Informative References	12
Authors' Addresses	12

1. Introduction

Traffic Engineered Networks can be interconnected, establishing different types of relationships among them. For example, both network can have a peering relation, where connections starting in one domain and end in the other domain. This document is focused on the case where the interconnected network domains have a client-server relationship among them. Such client-server relation comes from the two main points. On the one hand, end-to-end services in the client network can be set up using services of a network acting as server. On the other hand, the client-server relation comes from the fact that their interconnection is a policy and administrative boundary, limiting the amount of information allowed to be exchanged between networks. In the case of interconnected TE domains where there is no administrative nor strict policy boundary between client and server (typically, just a technology change), the MLN/MRN model can be applied.

The interface between the client and the server domain is typically called "User-to-Network Interface" (UNI), and regarded as signaling-only [RFC4208]. Due to the strict association of functionality to the UNI term, its exact scope has become highly controversial. This document compiles different definitions of the term used so far and propose some terminology to serve as a foundation to move the work forward.

What is more, the document compiles the possible operation models of client-server network from the control plane perspective. Each control model defines, on the one hand, the level of information of the domain acting as client provides through the control plane to the domain acting as server. On the other hand, the control model will determine what can be requested from the client domain to the server domain.

1.1. Examples of Client-Server TE Network Domain Scenarios

The most typical example of interconnected TE domains that follow a client-server relation is an IP/MPLS domain using the services of an optical OTN/WDM network. Note that the interconnected domain can be part of the same organization, but with different administration.

A particular network scenario that has attracted lot of attention from the industry is the IP/MPLS/OTN/WDM over WDM. The client network is based on multi-layer routers able to set up packet-based TE connections over wavelengths. The server network is a WDM optical network that provides the switching for the wavelengths as well as restoration capabilities of the optical channels.

Another example is MPLS over MPLS, where both client and server networks are able to set up packet based TE connections. This is the case, for example, of carrier-over-carrier scenarios.

Summing up, there number of applicable scenarios is wide.

2. Terminology

2.1. Routing domain

A routing domain is made of GMPLS enabled nodes (i.e., a network device including a GMPLS entity). These nodes can be either edge nodes (i.e., hosts, ingress LSRs or egress LSRs), or internal LSRs. An example of non-PSC host is an SONET/SDH Terminal Multiplexer (TM). Another example is an SONET/SDH interface card within an IP router or ATM switch.

A routing domain is characterized by being under the control of the same administration and by running a common set of protocols to exchange routing information

2.2. Overlay of routing domains

In an overlay environment we have a client routing domain and a server routing domain, each of which running its own routing protocol instance. Connectivity in the client routing domain can be made by connectivity services of the server domain.

2.3. Multilayer

As per RFC 5212 "UA data plane layer is a collection of network resources capable of terminating and/or switching data traffic of a particular format [RFC4397]. These resources can be used for establishing LSPs for traffic delivery. For example, VC-11 and VC4-64c represent two different layers."U

In a Multilayer network, each layer can be or not a routing domain. In fact, a multi-layer network can be controled with a single control plane instance in which all resources are adverstised in the same IGP instance

2.4. Policy

In an overlay network, policy is the set of rules that apply in the interface between two routing domains, and that restrict the level of information exchanged and the operations allowed. The policy decisions obey to confidentiality reasons (typically, the routing domains operate under the control of different administrations) and scalability (to avoid excessive flow of information that collapse the processing capacity of the nodes)

An example of policy example, visibility of the server domain could be restricted to the client domain.

2.5. Client Domain - Server Domain Interface

The interface between the client and the server domain is typically called "User-to-Network Interface" (UNI). However, the term "UNI" has been used in different contexts and SDOs. As a consequence, the exact definition of UNI and the functionalities included depend on the application. Bellow, as a reference, it is shown a set of the different definitions of UNI.

2.5.1. UNI in IP over Optical Networks

[RFC3717] says: "The client-optical internetwork interface (UNI) represents a service boundary between the client (e.g., IP router) and the optical network. The client and server (optical network) are essentially two different roles: the client role requests a service connection from a server; the server role establishes the connection to fulfill the service request -- provided all relevant admission control conditions are satisfied."

In other words, this definition refers to a signaling protocol between two administrative domains with a client-server relationship. It is agnostic to the existence of a data plane client-server relationship and to the side(s) of the boundary where it may happen, if any.

2.5.2. ITU-T Definition of UNI

ITU-T has defined the term UNI in the context of control plane. [G.807] [G.8081] (ITU-T): "User-Network Interface for the control plane (UNI): A bidirectional signaling interface between service requester and service server control plane entities."

The terms "requester/provider" are used to refer to the relationship.

2.5.3. OIF Definition of UNI

UNI: "The service control interface between a client device and the transport network."

UNI-C: "The logical entity that terminates UNI signalling on the client device side."

UNI-N: "The logical entity that terminates UNI signalling on the transport network side."

The terms "client/transport" and "client/network" are used to refer to the relationship.

2.5.4. Proposed Vocabulary

As listed above, the existing terminology is far from unique. To avoid overloaded concepts, this document proposes to use the "client/server" terms.

Unless stated, this document focuses on control protocol exchanges and their uses across administrative boundaries for client-server interconnection. Data plane transition and/or client-server relationship may not be aligned with the boundary.

2.5.4.1. Client network

A Client network is defined as a network domain able to request a connectivity service to a server network domain across an administrative boundary.

2.5.4.2. Server network

A Server network is defined as a network domain able to deliver connectivity services to a client network domain across an administrative boundary.

2.5.4.3. Client-Server Control Plane Interface

The control plane interface between the client network domain and the server network domain convey a set of control functionalities that help to operate such kind of networks. The exact functionalities of this Interface (and then the level of information exchanged) depend on the chosen control model. This document presents a taxonomy with the possible control models.

2.6. Reachability

In graph theory, reachability refers to the ability to get from one vertex to another within a graph. Thus, a vertex can reach another vertex if there exists a sequence of adjacent vertices which starts with the source vertex and ends with the destination vertex.

The document [draft-farrel-interconnected-te-info-exchange-04] provides the definition of what is reachability for client-server networks. [EDITOR's note: Text from draft-farrel-interconnected-te-info-exchange has been borrowed for this first version. Duplicated text will be deleted at later stages]

In an IP network, reachability is the ability to deliver a packet to a specific address or prefix. That is, the existence of an IP path to that address or prefix. TE reachability is the ability to reach a specific address along a TE path.

In the context of Traffic Engineered networks with client and server relationships, we can define several types of reachability:
[draft-farrel-interconnected-te-info-exchange-04]

2.6.1. Unqualified Reachability

Two client domain nodes are said to be reachable if, either there exists at least one path through the client domain that connects both nodes, or, in the case that there is no path exclusively through the client domain network, there exists at least one path connecting nodes of client and server domain by which both client nodes can be connected.

In the case of basic reachability, it is only known that it is possible to connect the nodes, but there is no notion of the details of such possible connections, such as, for example, bandwidth available or performance metrics. Also, the exact path to connect both nodes is not known to the client network. Note that, even if two nodes are reachable, there may not be enough resources for a desired TE connection with specific TE constraints.

2.6.2. Qualified Reachability

In this case, on top of the basic reachability, it is known some TE attributes of the possible connection (or connections). Examples of such attributes are: TE metrics, hop count, available bandwidth, delay, SRLG list. Note that this information is specific per connection. Thus, if there are several possible TE paths, there are a set of attributes.

2.6.3. Qualified Reachability with associated potential TE path

In this particular case, on top of the qualified reachability, there exists an associated potential TE path that satisfies the TE connection between two client nodes. Thus, in this case, the client Network has the information that there exists a TE path that can be set up at any time.

3. Control Models

The control of the networks formed by interconnected domains with a client-server relations between them can be done following different models. Each control model defines, on the one hand, the level of information that the domain acting as client receives by control plane means about the services given by the domain acting as server. This information, for example, can vary from a complete lack of information, so the client domain only knows that it could be possible to reach another point of its domain via the server network, to a detailed view on the possibilities offered by the server network. The level of detail of this information will determine which information is exchanged between both networks. On the other hand, the control model will determine what can be requested from the client domain to the server domain. As an example, the most basic use is specifying just the end-points to connect. Other cases may include the possibility to request a service specifying a set of constraints, like bandwidth, diversity, an optimization criteria, etc.

Which control model to choose depends on several factors. For the network operators, the main concern will be related to the level of trustness and relationship between client and server domains. Also, one key factor to take into account is the protocol interoperability. Note that, equipment in the interconnected domains may be from different technologies (but not necessarily) and are likely to use different implementations. The higher the level of functionality included in the control plane, the higher the protocol interoperability requirements, as it will force all implementations to support many functionality. Finally, scalability, that is, the ability of the control plane to provide the same functionality regarding the number of equipment, needs to be taken into account: the amount of information in each option will have different limits in terms on number of interconnected nodes.

3.1. Signaling Only

This first model considers that the sole functionality allowed in the control plane is signaling, that is the ability to request services from client to server domain.

In this model, the control plane does not provide a priori hints to the client domain about the state of the server domain (e.g., resource availability). This model does not preclude that, by other means like the management plane, the client domain knows what is possible or not. Such management actions are out of the scope of the control plane. Thus, it is perfectly feasible that the reachability information is provided either statically or by some management platform.

The most basic case relies on sending a loose ERO from the client, specifying the edges of the connection.

In a trusted interconnection mode, the signaling allows the client domain to provide a full ERO, given to the client network by external tools.

3.1.1. Signaling with Requirements

The control plane may allow to express complex requests to the server domain. That is, through the signaling protocol, it is allowed to not only request a connection between two points of the client domain, but also to include some constraints: e.g., minimum bandwidth, maximum delay, optimization criteria, or request diversity from another service. The policy at the edges of the server network will determine which constraints are accepted. Note the many of the requirements that can be expressed in the request are similar to what would be asked to a path computation function.

3.1.2. Signaling with Collection

Even though the only protocol enabled is signaling, it may be beneficial for the client domain to be able to know some updated information of the services that it has requested to the server. Thus, this case considers the possibility that, through the signaling protocol, the client domain can receive some information. What information it is allowed to collect will be determined by the policy of the server domain.

3.2. Signaling and Reachability Model

This second model considers that, in addition to signaling, the client domain receives some reachability information through a control plane mechanism.

3.2.1. Signalling + Basic Reachability

In this particular case, through control plane mechanisms, the client domain knows whether it is possible to reach a remote end point. The client domain should also remain aware of this information if there are failures in the server domain or if the associated capacity has been filled.

3.2.2. Signalling + Qualified Reachability

The control plane will provide information not only about the possibility to reach a remote end point, but also some TE information of possible connections. For example, the client domain will know that it is possible to reach another point with some bandwidth or delay. Note that, in this case, such information is sent by control plane mechanisms (not statically configured by management plane).

3.2.3. Signalling + Qualified Reachability + Potential Services

In addition to the TE information of the possible connections between two points, the control plane will also provide to the client domain information about potential server's services which could satisfy given requirements. By control plane procedures, the client domain can request, with respect to its needs, a service using such potential service and make high level path selection within the server domain.

3.3. Service Attributes vs service constraints

When asking for the setup of a service in the server domain, the client domain can put constraints on such request. Constraints can consist on the utilization of a path that minimizes a given metric (e.g. TE metric or end to end delay) or on a set of lower/upper bounds that must be followed (e.g. maximum number of hops or maximum end to end delay). Once the service has been provisioned (or just its paths computed), it is possible to identify (e.g. measure or collect) the attributes that characterize such service. For example the path has been computed so to meet the constraint of maximum end to end delay of 20ms, while one of its attributes is the effective end to end delay that is experimented along its path, which could be of e.g. 14 ms. Other examples of constraints and attributes can be found in path diversity. A typical constraint in LSP provisioning is diversity, which is a constraint, but then attributes of the two diverse LSPs like e.g. SRLGs can be collected. Both constraints and attributes need to be exchanged between a client and a server domain.

3.4. Other Models

3.4.1. Multi-Layer Networks / Multi-Region Networks

MLN/MRN extensions to control protocols have been defined. They are well scoped for client and server data plane domains without administrative boundary between them. This allows MLN nodes to participate in common control protocol instances. There is a full set of mechanisms to operate such networks [Editor's note: add refs to MLN/MRN)]. Typical use cases are switches combining both low- and high-order Sonet/SDH, or both ODUk and wavelengths.

However, MLN/MRN assumes no policy boundary between client and server domains. Thus, the level of information exchanged is not restricted, and full interoperability of both the signaling and routing protocols is required.

3.4.2. Management Model

In this particular case, the role of the control plane is limited to operate independently in each of the domains. [Editor's note: Common Control... WG => do we leave it?]

4. Abstraction

Abstraction:

- a physical topology is made of actual nodes interconnected by existing links, i.e. without abstraction;
- a virtual topology is made of nodes and/or links which may (or may not) exist or be instantiated to look the same as the advertised abstraction;
- a potential topology is made of nodes and/or links which are not existing at advertising time but could be instantiated on demand, i.e. a virtual topology which can be actually provided by a network.

5. Security Considerations

TBD

6. Contributing Authors

7. Acknowledgments

The authors would like to thank Lou Berger for pointing out the direction of the document and Dieter Beler for his review. The authors would like to specially thank all the authors of draft-farrel-interconnected-te-info-exchange-02

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.
- [RFC3107] Rekhter, Y. and E. Rosen, "Carrying Label Information in BGP-4", RFC 3107, May 2001.
- [RFC3717] Rajagopalan, B., Luciani, J., and D. Awduche, "IP over Optical Networks: A Framework", RFC 3717, March 2004.
- [RFC4208] Swallow, G., Drake, J., Ishimatsu, H., and Y. Rekhter, "Generalized Multiprotocol Label Switching (GMPLS) User-Network Interface (UNI): Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Support for the Overlay Model", RFC 4208, October 2005.

8.2. Informative References

- [draft-farrel-interconnected-te-info-exchange-04]
"Farrel, A., Drake, J., Bitar, N., Swallow, G.,
Ceccarelli, D. draft-farrel-interconnected-te-info-
exchange-04 Problem Statement and Architecture for
Information Exchange Between Interconnected Traffic
Engineered Networks", 2014.

Authors' Addresses

Oscar Gonzalez de Dios (editor)
Telefonica GCTO
Dis
Madrid 28045
Spain

Phone: +34913128832
Email: oscar.gonzalezdedios@telefonica.com

Julien Meuric (editor)
Orange
2 avenue Pierre Marzin
Lannion 22300
France

Email: julien.meuric@orange.com

Daniele Ceccarelli
Ericsson
Via Calda 5
Genova
Italy

Phone: +39 010 600 2512
Email: daniele.ceccarelli@ericsson.com

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: March 23, 2015

A. Farrel (Ed.)
J. Drake
Juniper Networks

N. Bitar
Verizon Networks

G. Swallow
Cisco Systems, Inc.

D. Ceccarelli
Ericsson

X. Zhang
Huawei
September 23, 2014

Problem Statement and Architecture for Information Exchange
Between Interconnected Traffic Engineered Networks

draft-farrel-interconnected-te-info-exchange-07.txt

Abstract

In Traffic Engineered (TE) systems, it is sometimes desirable to establish an end-to-end TE path with a set of constraints (such as bandwidth) across one or more network from a source to a destination. TE information is the data relating to nodes and TE links that is used in the process of selecting a TE path. The availability of TE information is usually limited to within a network (such as an IGP area) often referred to as a domain.

In order to determine the potential to establish a TE path through a series of connected networks, it is necessary to have available a certain amount of TE information about each network. This need not be the full set of TE information available within each network, but does need to express the potential of providing TE connectivity. This subset of TE information is called TE reachability information.

This document sets out the problem statement and architecture for the exchange of TE information between interconnected TE networks in support of end-to-end TE path establishment. For reasons that are explained in the document, this work is limited to simple TE constraints and information that determine TE reachability.

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."

Copyright Notice

Copyright (c) 2014 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	5
1.1. Terminology	6
1.1.1. TE Paths and TE Connections	6
1.1.2. TE Metrics and TE Attributes	6
1.1.3. TE Reachability	6
1.1.4. Domain	7
1.1.5. Aggregation	7
1.1.6. Abstraction	7
1.1.7. Abstract Link	7
1.1.8. Abstraction Layer Network	8
2. Overview of Use Cases	8
2.1. Peer Networks	8
2.1.1. Where is the Destination?	9
2.2. Client-Server Networks	10
2.3. Dual-Homing	12
2.4. Requesting Connectivity	13
2.4.1. Discovering Server Network Information	15
3. Problem Statement	15
3.1. Use of Existing Protocol Mechanisms	16
3.2. Policy and Filters	16
3.3. Confidentiality	17
3.4. Information Overload	17
3.5. Issues of Information Churn	18
3.6. Issues of Aggregation	19
3.7. Virtual Network Topology	20
4. Existing Work	21
4.1. Per-Domain Path Computation	21
4.2. Crankback	22
4.3. Path Computation Element	23
4.4. GMPLS UNI and Overlay Networks	24
4.5. Layer One VPN	25
4.6. VNT Manager and Link Advertisement	25
4.7. What Else is Needed and Why?	26
5. Architectural Concepts	26
5.1. Basic Components	26
5.1.1. Peer Interconnection	27
5.1.2. Client-Server Interconnection	27
5.2. TE Reachability	28
5.3. Abstraction not Aggregation	29
5.3.1. Abstract Links	30
5.3.2. The Abstraction Layer Network	30
5.3.3. Abstraction in Client-Server Networks.....	33
5.3.4. Abstraction in Peer Networks	34
5.4. Considerations for Dynamic Abstraction	40
5.5. Requirements for Advertising Links and Nodes	40
5.6. Addressing Considerations	40

6. Building on Existing Protocols	41
6.1. BGP-LS	41
6.2. IGPs	41
6.3. RSVP-TE	41
7. Applicability to Optical Domains and Networks	42
8. Modeling the User-to-Network Interface	43
9. Abstraction in L3VPN Multi-AS Environments	47
10. Scoping Future Work	49
10.1. Not Solving the Internet	49
10.2. Working With "Related" Domains	49
10.3. Not Finding Optimal Paths in All Situations	49
10.4. Not Breaking Existing Protocols	49
10.5. Sanity and Scaling	49
11. Manageability Considerations	50
12. IANA Considerations	50
13. Security Considerations	50
14. Acknowledgements	50
15. References	50
15.1. Informative References	50
Authors' Addresses	54
Contributors	55

1. Introduction

Traffic Engineered (TE) systems such as MPLS-TE [RFC2702] and GMPLS [RFC3945] offer a way to establish paths through a network in a controlled way that reserves network resources on specified links. TE paths are computed by examining the Traffic Engineering Database (TED) and selecting a sequence of links and nodes that are capable of meeting the requirements of the path to be established. The TED is constructed from information distributed by the IGP running in the network, for example OSPF-TE [RFC3630] or ISIS-TE [RFC5305].

It is sometimes desirable to establish an end-to-end TE path that crosses more than one network or administrative domain as described in [RFC4105] and [RFC4216]. In these cases, the availability of TE information is usually limited to within each network. Such networks are often referred to as Domains [RFC4726] and we adopt that definition in this document: viz.

For the purposes of this document, a domain is considered to be any collection of network elements within a common sphere of address management or path computational responsibility. Examples of such domains include IGP areas and Autonomous Systems.

In order to determine the potential to establish a TE path through a series of connected domains and to choose the appropriate domain connection points through which to route a path, it is necessary to have available a certain amount of TE information about each domain. This need not be the full set of TE information available within each domain, but does need to express the potential of providing TE connectivity. This subset of TE information is called TE reachability information. The TE reachability information can be exchanged between domains based on the information gathered from the local routing protocol, filtered by configured policy, or statically configured.

This document sets out the problem statement and architecture for the exchange of TE information between interconnected TE domains in support of end-to-end TE path establishment. The scope of this document is limited to the simple TE constraints and information (such as TE metrics, hop count, bandwidth, delay, shared risk) necessary to determine TE reachability: discussion of multiple additional constraints that might qualify the reachability can significantly complicate aggregation of information and the stability of the mechanism used to present potential connectivity as is explained in the body of this document.

1.1. Terminology

This section introduces some key terms that need to be understood to arrive at a common understanding of the problem space. Some of the terms are defined in more detail in the sections that follow (in which case forward pointers are provided) and some terms are taken from definitions that already exist in other RFCs (in which case references are given, but no apology is made for repeating or summarizing the definitions here).

1.1.1. TE Paths and TE Connections

A TE connection is a Label Switched Path (LSP) through an MPLS-TE or GMPLS network that directs traffic along a particular path (the TE path) in order to provide a specific service such as bandwidth guarantee, separation of traffic, or resilience between a well-known pair of end points.

1.1.2. TE Metrics and TE Attributes

TE metrics and TE attributes are terms applied to parameters of links (and possibly nodes) in a network that is traversed by TE connections. The TE metrics and TE attributes are used by path computation algorithms to select the TE paths that the TE connections traverse. Provisioning a TE connection through a network may result in dynamic changes to the TE metrics and TE attributes of the links and nodes in the network.

These terms are also sometimes used to describe the end-to-end characteristics of a TE connection and can be derived formulaically from the metrics and attributes of the links and nodes that the TE connection traverses. Thus, for example, the end-to-end delay for a TE connection is usually considered to be the sum of the delay on each link that the connection traverses.

1.1.3. TE Reachability

In an IP network, reachability is the ability to deliver a packet to a specific address or prefix. That is, the existence of an IP path to that address or prefix. TE reachability is the ability to reach a specific address along a TE path. More specifically, it is the ability to establish a TE connection in an MPLS-TE or GMPLS sense. Thus we talk about TE reachability as the potential of providing TE connectivity.

TE reachability may be unqualified (there is a TE path, but no information about available resources or other constraints is supplied) which is helpful especially in determining a path to a

destination that lies in an unknown domain, or may be qualified by TE attributes and TE metrics such as hop count, available bandwidth, delay, shared risk, etc.

1.1.4. Domain

As defined in [RFC4726], a domain is any collection of network elements within a common sphere of address management or path computational responsibility. Examples of such domains include Interior Gateway Protocol (IGP) areas and Autonomous Systems (ASes).

1.1.5. Aggregation

The concept of aggregation is discussed in Section 3.6. In aggregation, multiple network resources from a domain are represented outside the domain as a single entity. Thus multiple links and nodes forming a TE connection may be represented as a single link, or a collection of nodes and links (perhaps the whole domain) may be represented as a single node with its attachment links.

1.1.6. Abstraction

Section 5.3 introduces the concept of abstraction and distinguishes it from aggregation. Abstraction may be viewed as "policy-based aggregation" where the policies are applied to overcome the issues with aggregation as identified in section 3 of this document.

Abstraction is the process of applying policy to the available TE information within a domain, to produce selective information that represents the potential ability to connect across the domain. Thus, abstraction does not necessarily offer all possible connectivity options, but presents a general view of potential connectivity according to the policies that determine how the domain's administrator wants to allow the domain resources to be used.

1.1.7. Abstract Link

An abstract link is the representation of the characteristics of a path between two nodes in a domain produced by abstraction. The abstract link is advertised outside that domain as a TE link for use in signaling in other domains. Thus, an abstract link represents the potential to connect between a pair of nodes.

More details of abstract links are provided in Section 5.3.1.

1.1.8. Abstraction Layer Network

The abstraction layer network is introduced in Section 5.3.2. It may be seen as a brokerage layer network between one or more server networks and one or more client network. The abstraction layer network is the collection of abstract links that provide potential connectivity across the server network(s) and on which path computation can be performed to determine edge-to-edge paths that provide connectivity as links in the client network.

In the simplest case, the abstraction layer network is just a set of edge-to-edge connections (i.e., abstract links), but to make the use of server resources more flexible, the abstract links might not all extend from edge to edge, but might offer connectivity between server nodes to form a more complex network.

2. Overview of Use Cases

2.1. Peer Networks

The peer network use case can be most simply illustrated by the example in Figure 1. A TE path is required between the source (Src) and destination (Dst), that are located in different domains. There are two points of interconnection between the domains, and selecting the wrong point of interconnection can lead to a sub-optimal path, or even fail to make a path available.

For example, when Domain A attempts to select a path, it may determine that adequate bandwidth is available from Src through both interconnection points x1 and x2. It may pick the path through x1 for local policy reasons: perhaps the TE metric is smaller. However, if there is no connectivity in Domain Z from x1 to Dst, the path cannot be established. Techniques such as crankback (see Section 4.2) may be used to alleviate this situation, but do not lead to rapid setup or guaranteed optimality. Furthermore RSVP signalling creates state in the network that is immediately removed by the crankback procedure. Frequent events of such a kind impact scalability in a non-deterministic manner.

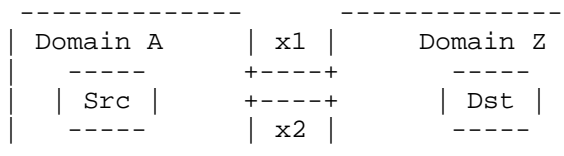


Figure 1 : Peer Networks

There are countless more complicated examples of the problem of peer networks. Figure 2 shows the case where there is a simple mesh of domains. Clearly, to find a TE path from Src to Dst, Domain A must not select a path leaving through interconnect x1 since Domain B has no connectivity to Domain Z. Furthermore, in deciding whether to select interconnection x2 (through Domain C) or interconnection x3 through Domain D, Domain A must be sensitive to the TE connectivity available through each of Domains C and D, as well the TE connectivity from each of interconnections x4 and x5 to Dst within Domain Z.

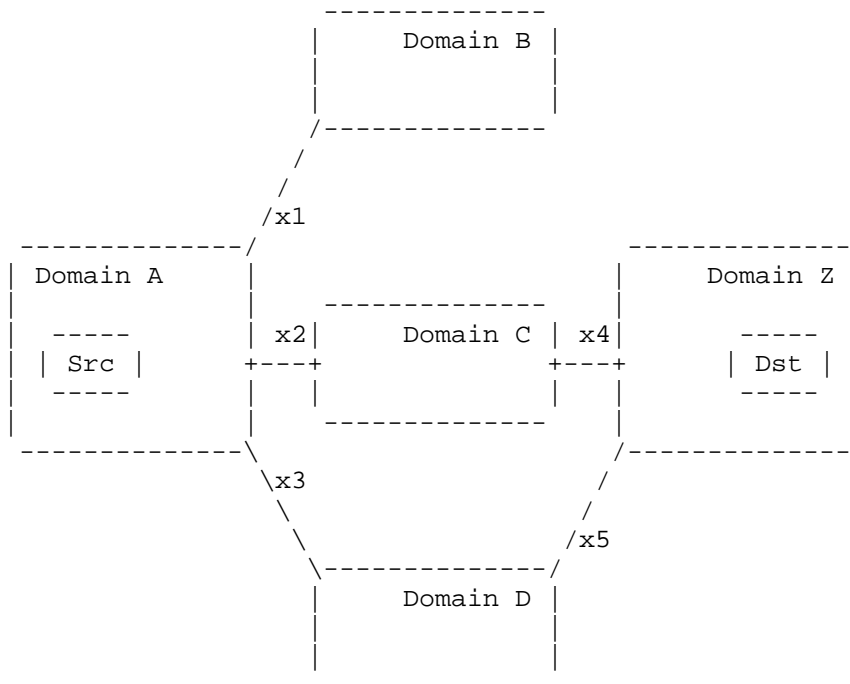


Figure 2 : Peer Networks in a Mesh

Of course, many network interconnection scenarios are going to be a combination of the situations expressed in these two examples. There may be a mesh of domains, and the domains may have multiple points of interconnection.

2.1.1.1. Where is the Destination?

A variation of the problems expressed in Section 2.1 arises when the source domain (Domain A in both figures) does not know where the

destination is located. That is, when the domain in which the destination node is located is not known to the source domain.

This is most easily seen in consideration of Figure 2 where the decision about which interconnection to select needs to be based on building a path toward the destination domain. Yet this can only be achieved if it is known in which domain the destination node lies, or at least if there is some indication in which direction the destination lies. This function is obviously provided in IP networks by inter-domain routing [RFC4271].

2.2. Client-Server Networks

Two major classes of use case relate to the client-server relationship between networks. These use cases have sometimes been referred to as overlay networks.

The first group of use case, shown in Figure 3, occurs when domains belonging to one network are connected by a domain belonging to another network. In this scenario, once connections (or tunnels) are formed across the lower layer network, the domains of the upper layer network can be merged into a single domain by running IGP adjacencies over the tunnels, and treating the tunnels as links in the higher layer network. The TE relationship between the domains (higher and lower layer) in this case is reduced to determining which tunnels to set up, how to trigger them, how to route them, and what capacity to assign them. As the demands in the higher layer network vary, these tunnels may need to be modified. Section 2.4 explains in a little more detail how connectivity may be requested

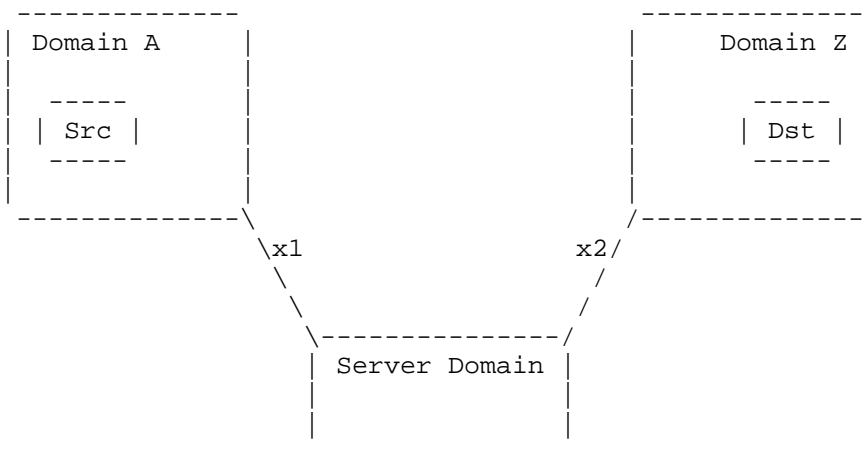


Figure 3 : Client-Server Networks

The second class of use case of client-server networking is for Virtual Private Networks (VPNs). In this case, as opposed to the former one, it is assumed that the client network has a different address space than that of the server layer where non-overlapping IP addresses between the client and the server networks cannot be guaranteed. A simple example is shown in Figure 4. The VPN sites comprise a set of domains that are interconnected over a core domain, the provider network.

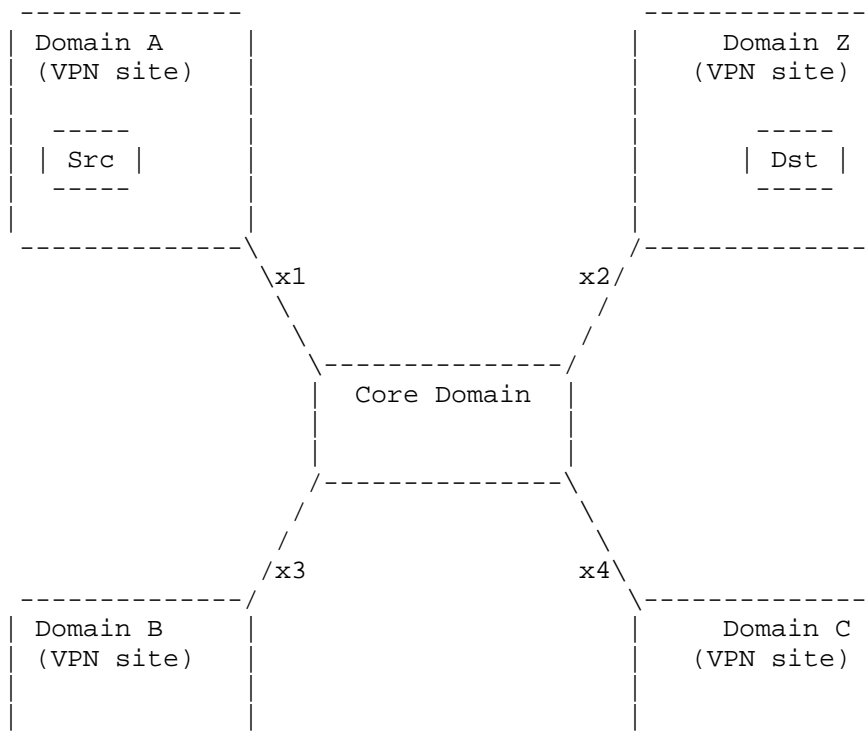


Figure 4 : A Virtual Private Network

Note that in the use cases shown in Figures 3 and 4 the client layer domains may (and, in fact, probably do) operate as a single connected network.

Both use cases in this section become "more interesting" when combined with the use case in Section 2.1. That is, when the connectivity between higher layer domains or VPN sites is provided by a sequence or mesh of lower layer domains. Figure 5 shows how this might look in the case of a VPN.

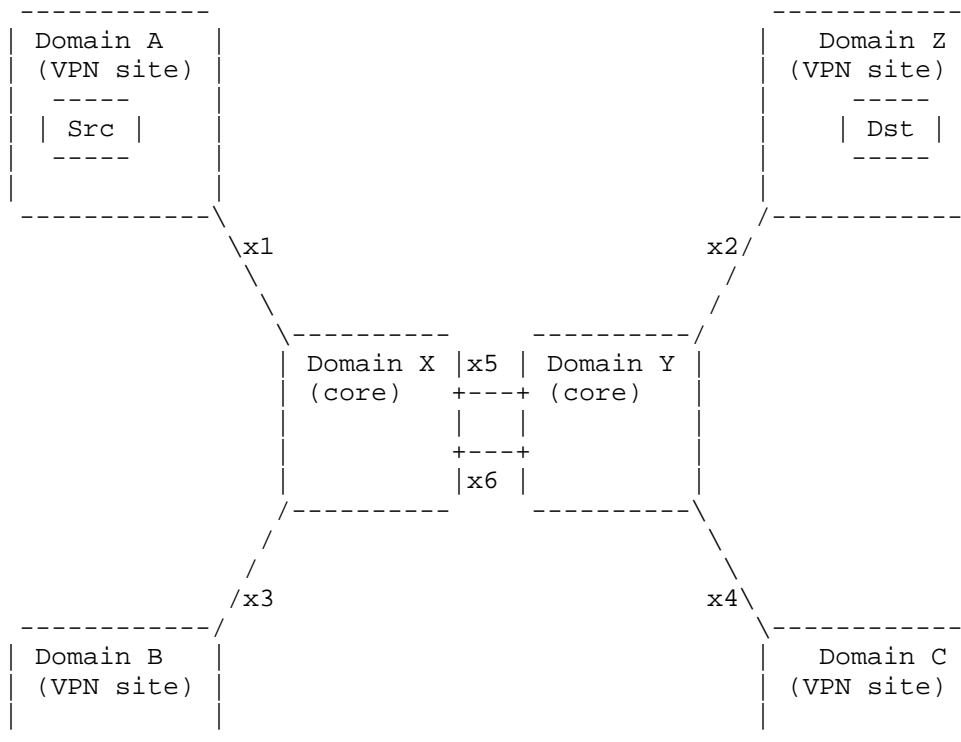


Figure 5 : A VPN Supported Over Multiple Server Domains

2.3. Dual-Homing

A further complication may be added to the client-server relationship described in Section 2.2 by considering what happens when a client domain is attached to more than one server domain, or has two points of attachment to a server domain. Figure 6 shows an example of this for a VPN.

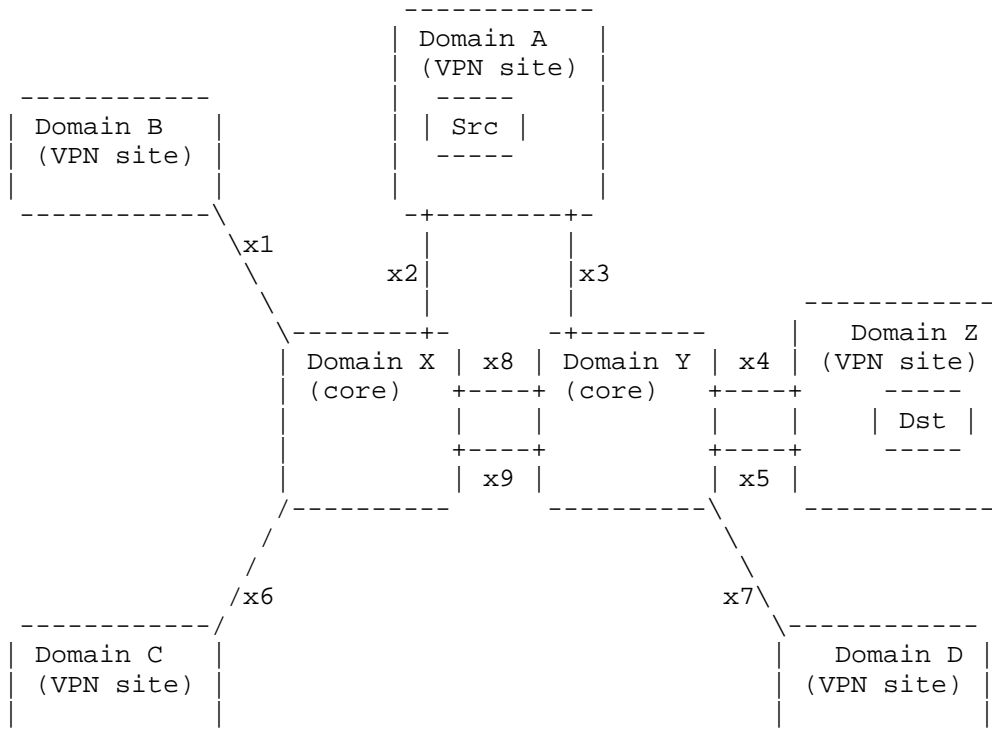


Figure 6 : Dual-Homing in a Virtual Private Network

2.4. Requesting Connectivity

This relationship between domains can be entirely under the control of management processes, dynamically triggered by the client network, or some hybrid of these cases. In the management case, the server network may be requested to establish a set of LSPs to provide client layer connectivity. In the dynamic case, the client may make a request to the server network exerting a range of controls over the paths selected in the server network. This range extends from no control (i.e., a simple request for connectivity), through a set of constraints (such as latency, path protection, etc.), up to and including full control of the path and resources used in the server network (i.e., the use of explicit paths with label subobjects).

There are various models by which a server network can be requested to set up the connections that support a service provided to the client network. These requests may come from management systems, directly from the client network control plane, or through some

intermediary broker such as the Virtual Network Topology Manager discussed in Section 4.6.

The trigger that causes the request to the server layer is also flexible. It could be that the client layer discovers a pressing need for server layer resources (such as the desire to provision an end-to-end connection in the client layer, or severe congestion on a specific path), or it might be that a planning application has considered how best to optimize traffic in the client network or how to handle a predicted traffic demand.

In all cases, the relationship between client and server networks is subject to policy so that server resources are under the administrative control of the operator or the server layer network and are only used to support a client layer network in ways that the server layer operator approves.

As just noted, connectivity requests issued to a server network may include varying degrees of constraint upon the choice of path that the server network can implement.

- o Basic Provisioning is a simple request for connectivity. The only constraints are the end points of the connection and the capacity (bandwidth) that the connection will support for the client layer. In the case of some server networks, even the bandwidth component of a basic provisioning request is superfluous because the server layer has no facility to vary bandwidth, but can offer connectivity only at a default capacity.
- o Basic Provisioning with Optimization is a service request that indicates one or more metrics that the server layer must optimize in its selection of a path. Metrics may be hop count, path length, summed TE metric, jitter, delay, or any number of technology-specific constraints.
- o Basic Provisioning with Optimization and Constraints enhances the optimization process to apply absolute constraints to functions of the path metrics. For example, a connection may be requested that optimizes for the shortest path, but in any case requests that the end-to-end delay be less than a certain value. Equally, optimization may be expressed in terms of the impact on the network. For example, a service may be requested in order to leave maximal flexibility to satisfy future service requests.
- o Fate Diversity requests ask for the server layer to provide a path that does not use any network resources (usually links and nodes) that share fate (i.e., can fail as the result of a single event) as the resources used by another connection. This allows the client

layer to construct protection services over the server layer network, for example by establishing virtual links that are known to be fate diverse. The connections that have diverse paths need not share end points.

- o Provisioning with Fate Sharing is the exact opposite of Fate Diversity. In this case two or more connections are requested to follow same path in the server network. This may be requested, for example, to create a bundled or aggregated link in the client layer where each component of the client layer composite link is required to have the same server layer properties (metrics, delay, etc.) and the same failure characteristics.
- o Concurrent Provisioning enables the inter-related connections requests described in the previous two bullets to be enacted through a single, compound service request.
- o Service Resilience requests the server layer to provide connectivity for which the server layer takes responsibility to recover from faults. The resilience may be achieved through the use of link-level protection, segment protection, end-to-end protection, or recovery mechanisms.

2.4.1.1. Discovering Server Network Information

Although the topology and resource availability information of a server network may be hidden from the client network, the service request interface may support features that report details about the services and potential services that the server network supports.

- o Reporting of path details, service parameters, and issues such as path diversity of LSPs that support deployed services allows the client network to understand to what extent its requests were satisfied. This is particularly important when the requests were made as "best effort".
- o A server network may support requests of the form "if I was to ask you for this service, would you be able to provide it?" That is, a service request that does everything except actually provision the service.

3. Problem Statement

The problem statement presented in this section is as much about the issues that may arise in any solution (and so have to be avoided) and the features that are desirable within a solution, as it is about the actual problem to be solved.

The problem can be stated very simply and with reference to the use cases presented in the previous section.

A mechanism is required that allows TE-path computation in one domain to make informed choices about the TE-capabilities and exit points from the domain when signaling an end-to-end TE path that will extend across multiple domains.

Thus, the problem is one of information collection and presentation, not about signaling. Indeed, the existing signaling mechanisms for TE LSP establishment are likely to prove adequate [RFC4726] with the possibility of minor extensions.

An interesting annex to the problem is how the path is made available for use. For example, in the case of a client-server network, the path established in the server network needs to be made available as a TE link to provide connectivity in the client network.

3.1. Use of Existing Protocol Mechanisms

TE information may currently be distributed in a domain by TE extensions to one of the two IGPs as described in OSPF-TE [RFC3630] and ISIS-TE [RFC5305]. TE information may be exported from a domain (for example, northbound) using link state extensions to BGP [I-D.ietf-idr-ls-distribution].

It is desirable that a solution to the problem described in this document does not require the implementation of a new, network-wide protocol. Instead, it would be advantageous to make use of an existing protocol that is commonly implemented on network nodes and is currently deployed, or to use existing computational elements such as Path Computation Elements (PCEs). This has many benefits in network stability, time to deployment, and operator training.

It is recognized, however, that existing protocols are unlikely to be immediately suitable to this problem space without some protocol extensions. Extending protocols must be done with care and with consideration for the stability of existing deployments. In extreme cases, a new protocol can be preferable to a messy hack of an existing protocol.

3.2. Policy and Filters

A solution must be amenable to the application of policy and filters. That is, the operator of a domain that is sharing information with another domain must be able to apply controls to what information is shared. Furthermore, the operator of a domain that has information shared with it must be able to apply policies and filters to the

received information.

Additionally, the path computation within a domain must be able to weight the information received from other domains according to local policy such that the resultant computed path meets the local operator's needs and policies rather than those of the operators of other domains.

3.3. Confidentiality

A feature of the policy described in Section 3.3 is that an operator of a domain may desire to keep confidential the details about its internal network topology and loading. This information could be construed as commercially sensitive.

Although it is possible that TE information exchange will take place only between parties that have significant trust, there are also use cases (such as the VPN supported over multiple server domains described in Section 2.4) where information will be shared between domains that have a commercial relationship, but a low level of trust.

Thus, it must be possible for a domain to limit the information share to just that which the computing domain needs to know with the understanding that less information that is made available the more likely it is that the result will be a less optimal path and/or more crankback events.

3.4. Information Overload

One reason that networks are partitioned into separate domains is to reduce the set of information that any one router has to handle. This also applies to the volume of information that routing protocols have to distribute.

Over the years routers have become more sophisticated with greater processing capabilities and more storage, the control channels on which routing messages are exchanged have become higher capacity, and the routing protocols (and their implementations) have become more robust. Thus, some of the arguments in favor of dividing a network into domains may have been reduced. Conversely, however, the size of networks continues to grow dramatically with a consequent increase in the total amount of routing-related information available. Additionally, in this case, the problem space spans two or more networks.

Any solution to the problems voiced in this document must be aware of the issues of information overload. If the solution was to simply

share all TE information between all domains in the network, the effect from the point of view of the information load would be to create one single flat network domain. Thus the solution must deliver enough information to make the computation practical (i.e., to solve the problem), but not so much as to overload the receiving domain. Furthermore, the solution cannot simply rely on the policies and filters described in Section 3.2 because such filters might not always be enabled.

3.5. Issues of Information Churn

As LSPs are set up and torn down, the available TE resources on links in the network change. In order to reliably compute a TE path through a network, the computation point must have an up-to-date view of the available TE resources. However, collecting this information may result in considerable load on the distribution protocol and churn in the stored information. In order to deal with this problem even in a single domain, updates are sent at periodic intervals or whenever there is a significant change in resources, whichever happens first.

Consider, for example, that a TE LSP may traverse ten links in a network. When the LSP is set up or torn down, the resources available on each link will change resulting in a new advertisement of the link's capabilities and capacity. If the arrival rate of new LSPs is relatively fast, and the hold times relatively short, the network may be in a constant state of flux. Note that the problem here is not limited to churn within a single domain, since the information shared between domains will also be changing. Furthermore, the information that one domain needs to share with another may change as the result of LSPs that are contained within or cross the first domain but which are of no direct relevance to the domain receiving the TE information.

In packet networks, where the capacity of an LSP is often a small fraction of the resources available on any link, this issue is partially addressed by the advertising routers. They can apply a threshold so that they do not bother to update the advertisement of available resources on a link if the change is less than a configured percentage of the total (or alternatively, the remaining) resources. The updated information in that case will be disseminated based on an update interval rather than a resource change event.

In non-packet networks, where link resources are physical switching resources (such as timeslots or wavelengths) the capacity of an LSP may more frequently be a significant percentage of the available link resources. Furthermore, in some switching environments, it is necessary to achieve end-to-end resource continuity (such as using

the same wavelength on the whole length of an LSP), so it is far more desirable to keep the TE information held at the computation points up-to-date. Fortunately, non-packet networks tend to be quite a bit smaller than packet networks, the arrival rates of non-packet LSPs are much lower, and the hold times considerably longer. Thus the information churn may be sustainable.

3.6. Issues of Aggregation

One possible solution to the issues raised in other sub-sections of this section is to aggregate the TE information shared between domains. Two aggregation mechanisms are often considered:

- Virtual node model. In this view, the domain is aggregated as if it was a single node (or router / switch). Its links to other domains are presented as real TE links, but the model assumes that any LSP entering the virtual node through a link can be routed to leave the virtual node through any other link (although recent work on "limited cross-connect switches" may help with this problem [I-D.ietf-ccamp-general-constraint-encode]).
- Virtual link model. In this model, the domain is reduced to a set of edge-to-edge TE links. Thus, when computing a path for an LSP that crosses the domain, a computation point can see which domain entry points can be connected to which other and with what TE attributes.

It is of the nature of aggregation that information is removed from the system. This can cause inaccuracies and failed path computation. For example, in the virtual node model there might not actually be a TE path available between a pair of domain entry points, but the model lacks the sophistication to represent this "limited cross-connect capability" within the virtual node. On the other hand, in the virtual link model it may prove very hard to aggregate multiple link characteristics: for example, there may be one path available with high bandwidth, and another with low delay, but this does not mean that the connectivity should be assumed or advertised as having both high bandwidth and low delay.

The trick to this multidimensional problem, therefore, is to aggregate in a way that retains as much useful information as possible while removing the data that is not needed. An important part of this trick is a clear understanding of what information is actually needed.

It should also be noted in the context of Section 3.5 that changes in the information within a domain may have a bearing on what aggregated data is shared with another domain. Thus, while the data shared in

reduced, the aggregation algorithm (operating on the routers responsible for sharing information) may be heavily exercised.

3.7. Virtual Network Topology

The terms "virtual topology" and "virtual network topology" have become overloaded in a relatively short time. We draw on [RFC5212] and [RFC5623] for inspiration to provide a definition for use in this document. Our definition is based on the fact that a topology at the client network layer is constructed of nodes and links. Typically, the nodes are routers in the client layer, and the links are data links. However, a layered network provides connectivity through the lower layer as LSPs, and these LSPs can provide links in the client layer. Furthermore, those LSPs may have been established in advance, or might be LSPs that could be set up if required. This leads to the definition:

A Virtual Network Topology (VNT) is made up of links in a network layer. Those links may be realized as direct data links or as multi-hop connections (LSPs) in a lower network layer. Those underlying LSPs may be established in advance or created on demand.

The creation and management of a VNT requires interaction with management and policy. Activity is needed in both the client and server layer:

- In the server layer, LSPs need to be set up either in advance in response to management instructions or in answer to dynamic requests subject to policy considerations.
- In the server layer, evaluation of available TE resources can lead to the announcement of potential connectivity (i.e., LSPs that could be set up on demand).
- In the client layer, connectivity (lower layer LSPs or potential LSPs) needs to be announced in the IGP as a normal TE link. Such links may or may not be made available to IP routing: but, they are never made available to IP routing until fully instantiated.
- In the client layer, requests to establish lower layer LSPs need to be made either when links supported by potential LSPs are about to be used (i.e., when a higher layer LSP is signalled to cross the link, the setup of the lower layer LSP is triggered), or when the client layer determines it needs more connectivity or capacity.

It is a fundamental of the use of a VNT that there is a policy point

at the lower-layer node responsible for the instantiation of a lower-layer LSP. At the moment that the setup of a lower-layer LSP is triggered, whether from a client-layer management tool or from signaling in the client layer, the server layer must be able to apply policy to determine whether to actually set up the LSP. Thus, fears that a micro-flow in the client layer might cause the activation of 100G optical resources in the server layer can be completely controlled by the policy of the server layer network's operator (and could even be subject to commercial terms).

These activities require an architecture and protocol elements as well as management components and policy elements.

4. Existing Work

This section briefly summarizes relevant existing work that is used to route TE paths across multiple domains.

4.1. Per-Domain Path Computation

The per-domain mechanism of path establishment is described in [RFC5152] and its applicability is discussed in [RFC4726]. In summary, this mechanism assumes that each domain entry point is responsible for computing the path across the domain, but that details of the path in the next domain are left to the next domain entry point. The computation may be performed directly by the entry point or may be delegated to a computation server.

This basic mode of operation can run into many of the issues described alongside the use cases in Section 2. However, in practice it can be used effectively with a little operational guidance.

For example, RSVP-TE [RFC3209] includes the concept of a "loose hop" in the explicit path that is signaled. This allows the original request for an LSP to list the domains or even domain entry points to include on the path. Thus, in the example in Figure 1, the source can be told to use the interconnection x2. Then the source computes the path from itself to x2, and initiates the signaling. When the signaling message reaches Domain Z, the entry point to the domain computes the remaining path to the destination and continues the signaling.

Another alternative suggested in [RFC5152] is to make TE routing attempt to follow inter-domain IP routing. Thus, in the example shown in Figure 2, the source would examine the BGP routing information to determine the correct interconnection point for forwarding IP packets, and would use that to compute and then signal a path for Domain A. Each domain in turn would apply the same

approach so that the path is progressively computed and signaled domain by domain.

Although the per-domain approach has many issues and drawbacks in terms of achieving optimal (or, indeed, any) paths, it has been the mainstay of inter-domain LSP set-up to date.

4.2. Crankback

Crankback addresses one of the main issues with per-domain path computation: what happens when an initial path is selected that cannot be completed toward the destination? For example, what happens if, in Figure 2, the source attempts to route the path through interconnection x2, but Domain C does not have the right TE resources or connectivity to route the path further?

Crankback for MPLS-TE and GMPLS networks is described in [RFC4920] and is based on a concept similar to the Acceptable Label Set mechanism described for GMPLS signaling in [RFC3473]. When a node (i.e., a domain entry point) is unable to compute a path further across the domain, it returns an error message in the signaling protocol that states where the blockage occurred (link identifier, node identifier, domain identifier, etc.) and gives some clues about what caused the blockage (bad choice of label, insufficient bandwidth available, etc.). This information allows a previous computation point to select an alternative path, or to aggregate crankback information and return it upstream to a previous computation point.

Crankback is a very powerful mechanism and can be used to find an end-to-end path in a multi-domain network if one exists.

On the other hand, crankback can be quite resource-intensive as signaling messages and path setup attempts may "wander around" in the network attempting to find the correct path for a long time. Since RSVP-TE signaling ties up networks resources for partially established LSPs, since network conditions may be in flux, and most particularly since LSP setup within well-known time limits is highly desirable, crankback is not a popular mechanism.

Furthermore, even if crankback can always find an end-to-end path, it does not guarantee to find the optimal path. (Note that there have been some academic proposals to use signaling-like techniques to explore the whole network in order to find optimal paths, but these tend to place even greater burdens on network processing.)

4.3. Path Computation Element

The Path Computation Element (PCE) is introduced in [RFC4655]. It is an abstract functional entity that computes paths. Thus, in the example of per-domain path computation (Section 4.1) the source node and each domain entry point is a PCE. On the other hand, the PCE can also be realized as a separate network element (a server) to which computation requests can be sent using the Path Computation Element Communication Protocol (PCEP) [RFC5440].

Each PCE has responsibility for computations within a domain, and has visibility of the attributes within that domain. This immediately enables per-domain path computation with the opportunity to off-load complex, CPU-intensive, or memory-intensive computation functions from routers in the network. But the use of PCE in this way does not solve any of the problems articulated in Sections 4.1 and 4.2.

Two significant mechanisms for cooperation between PCEs have been described. These mechanisms are intended to specifically address the problems of computing optimal end-to-end paths in multi-domain environments.

- The Backward-Recursive PCE-Based Computation (BRPC) mechanism [RFC5441] involves cooperation between the set of PCEs along the inter-domain path. Each one computes the possible paths from domain entry point (or source node) to domain exit point (or destination node) and shares the information with its upstream neighbor PCE which is able to build a tree of possible paths rooted at the destination. The PCE in the source domain can select the optimal path.

BRPC is sometimes described as "crankback at computation time". It is capable of determining the optimal path in a multi-domain network, but depends on knowing the domain that contains the destination node. Furthermore, the mechanism can become quite complicated and involve a lot of data in a mesh of interconnected domains. Thus, BRPC is most often proposed for a simple mesh of domains and specifically for a path that will cross a known sequence of domains, but where there may be a choice of domain interconnections. In this way, BRPC would only be applied to Figure 2 if a decision had been made (externally) to traverse Domain C rather than Domain D (notwithstanding that it could functionally be used to make that choice itself), but BRPC could be used very effectively to select between interconnections x1 and x2 in Figure 1.

- Hierarchical PCE (H-PCE) [RFC6805] offers a parent PCE that is responsible for navigating a path across the domain mesh and for

coordinating intra-domain computations by the child PCEs responsible for each domain. This approach makes computing an end-to-end path across a mesh of domains far more tractable. However, it still leaves unanswered the issue of determining the location of the destination (i.e., discovering the destination domain) as described in Section 2.1.1. Furthermore, it raises the question of who operates the parent PCE especially in networks where the domains are under different administrative and commercial control.

Further issues and considerations of the use of PCE can be found in [I-D.farrkingel-pce-questions].

4.4. GMPLS UNI and Overlay Networks

[RFC4208] defines the GMPLS User-to-Network Interface (UNI) to present a routing boundary between an overlay network and the core network, i.e. the client-server interface. In the client network, the nodes connected directly to the core network are known as edge nodes, while the nodes in the server network are called core nodes.

In the overlay model defined by [RFC4208] the core nodes act as a closed system and the edge nodes do not participate in the routing protocol instance that runs among the core nodes. Thus the UNI allows access to and limited control of the core nodes by edge nodes that are unaware of the topology of the core nodes. This respects the operational and layer boundaries while scaling the network.

[RFC4208] does not define any routing protocol extension for the interaction between core and edge nodes but allows for the exchange of reachability information between them. In terms of a VPN, the client network can be considered as the customer network comprised of a number of disjoint sites, and the edge nodes match the VPN CE nodes. Similarly, the provider network in the VPN model is equivalent to the server network.

[RFC4208] is, therefore, a signaling-only solution that allows edge nodes to request connectivity cross the core network, and leaves the core network to select the paths and set up the core LSPs. This solution is supplemented by a number of signaling extensions such as [RFC4874], [RFC5553], [I-D.ietf-ccamp-xro-lsp-subobject], [I-D.ietf-ccamp-rsvp-te-srlg-collect], and [I-D.ietf-ccamp-te-metric-recording] to give the edge node more control over the LSP that the core network will set up by exchanging information about core LSPs that have been established and by allowing the edge nodes to supply additional constraints on the core LSPs that are to be set up.

Nevertheless, in this UNI/overlay model, the edge node has limited

information of precisely what LSPs could be set up across the core, and what TE services (such as diverse routes for end-to-end protection, end-to-end bandwidth, etc.) can be supported.

4.5. Layer One VPN

A Layer One VPN (L1VPN) is a service offered by a core layer 1 network to provide layer 1 connectivity (TDM, LSC) between two or more customer networks in an overlay service model [RFC4847].

As in the UNI case, the customer edge has some control over the establishment and type of the connectivity. In the L1VPN context three different service models have been defined classified by the semantics of information exchanged over the customer interface: Management Based, Signaling Based (a.k.a. basic), and Signaling and Routing service model (a.k.a. enhanced).

In the management based model, all edge-to-edge connections are set up using configuration and management tools. This is not a dynamic control plane solution and need not concern us here.

In the signaling based service model [RFC5251] the CE-PE interface allows only for signaling message exchange, and the provider network does not export any routing information about the core network. VPN membership is known a priori (presumably through configuration) or is discovered using a routing protocol [RFC5195], [RFC5252], [RFC5523], as is the relationship between CE nodes and ports on the PE. This service model is much in line with GMPLS UNI as defined in [RFC4208].

In the enhanced model there is an additional limited exchange of routing information over the CE-PE interface between the provider network and the customer network. The enhanced model considers four different types of service models, namely: Overlay Extension, Virtual Node, Virtual Link and Per-VPN service models. All of these represent particular cases of the TE information aggregation and representation.

4.6. VNT Manager and Link Advertisement

As discussed in Section 3.7, operation of a VNT requires policy and management input. In order to handle this, [RFC5623] introduces the concept of the Virtual Network Topology Manager (VNTM). This is a functional component that applies policy to requests from client networks (or agents of the client network, such as a PCE) for the establishment of LSPs in the server network to provide connectivity in the client network.

The VNTM would, in fact, form part of the provisioning path for all

server network LSPs whether they are set up ahead of client network demand or triggered by end-to-end client network LSP signaling.

An important companion to this function is determining how the LSP set up across the server network is made available as a TE link in the client network. Obviously, if the LSP is established using management intervention, the subsequent client network TE link can also be configured manually. However, if the LSP is signaled dynamically there is need for the end points to exchange the link properties that they should advertise within the client network, and in the case of a server network that supports more than one client, it will be necessary to indicate which client or clients can use the link. This capability is provided in [RFC6107].

Note that a potential server network LSP that is advertised as a TE link in the client network might to be determined dynamically by the edge nodes. In this case there will need to be some effort to ensure that both ends of the link have the same view of the available TE resources, or else the advertised link will be asymmetrical.

4.7. What Else is Needed and Why?

As can be seen from Sections 4.1 through 4.6, a lot of effort has focused on client-server networks as described in Figure 3. Far less consideration has been given to network peering or the combination of the two use cases.

Various work has been suggested to extend the definition of the UNI such that routing information can be passed across the interface. However, this approach seems to break the architectural concept of network separation that the UNI facilitates.

Other approaches are working toward a flattening of the network with complete visibility into the server networks being made available in the client network. These approaches, while functional, ignore the main reasons for introducing network separation in the first place.

The remainder of this document introduces a new approach based on network abstraction that allows a server network to use its own knowledge of its resources and topology combined with its own policies to determine what edge-to-edge connectivity capabilities it will inform the client networks about.

5. Architectural Concepts

5.1. Basic Components

This section revisits the use cases from Section 2 to present the

Figure 7 shows the basic architectural concepts for connecting across peer networks. Nodes from four networks are shown: A1 and A2 come from one network; B1, B2, and B3 from another network; etc. The interfaces between the networks (sometimes known as External Network-to-Network Interfaces - ENNIs) are A2-B1, B3-C1, and C3-D1.

As shown in the figure, LSP tunnels that span the transit networks are used to achieve the required connectivity. These transit LSPs form the key building blocks of the end-to-end connectivity.

The transit tunnels can be used as hierarchical LSPs [RFC4206] to carry the end-to-end LSP, or can become stitching segments [RFC5150] of the end-to-end LSP. The transit tunnels B1-B3 and C-C3 can be as an abstract link as discussed in Section 5.3.

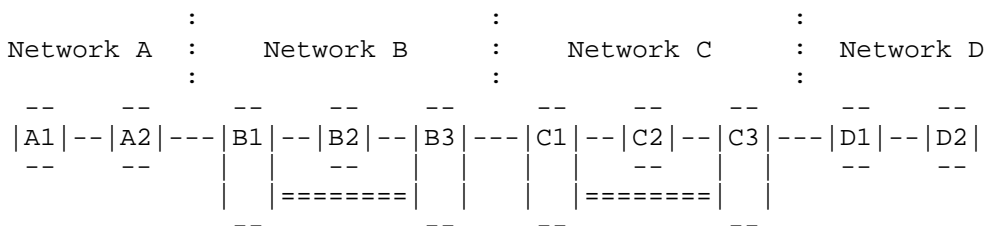


Figure 7 : Architecture for Peering

Figure 8 shows the basic architectural concepts for a client-server network. The client network nodes are C1, C2, CE1, CE2, C3, and C4. The core network nodes are CN1, CN2, CN3, and CN4. The interfaces CE1-CN1 and CE2-CN2 are the interfaces between the client and core networks.

The objective is to be able to support an end-to-end connection, C1-to-C4, in the client network. This connection may support TE or normal IP forwarding. To achieve this, CE1 is to be connected to CE2 by a link in the client layer that is supported by a core network LSP.

As shown in the figure, two LSPs are used to achieve the required connectivity. One LSP is set up across the core from CN1 to CN2. This core LSP then supports a three-hop LSP from CE1 to CE2 with its middle hop being the core LSP. It is this LSP that is presented as a link in the client network.

The practicalities of how the CE1-CE2 LSP is carried across the core LSP may depend on the switching and signaling options available in the core network. The LSP may be tunneled down the core LSP using the mechanisms of a hierarchical LSP [RFC4206], or the LSP segments CE1-CN1 and CN2-CE2 may be stitched to the core LSP as described in [RFC5150].

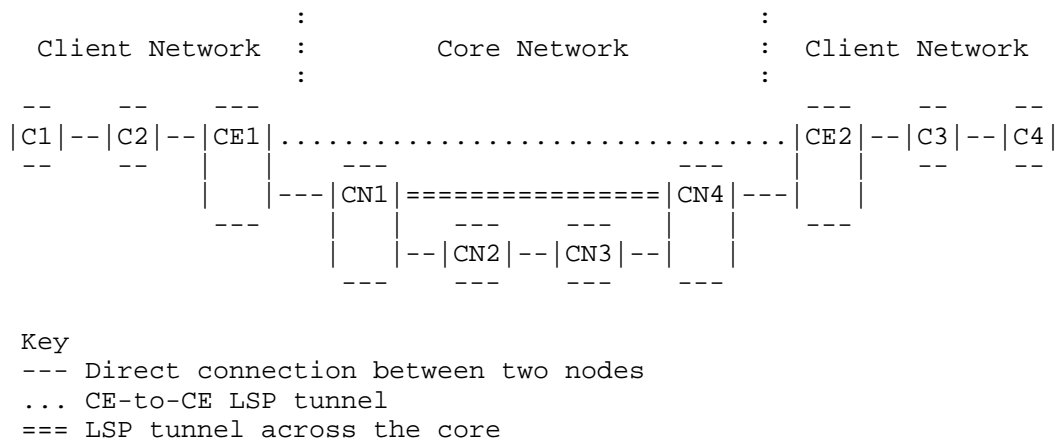


Figure 8 : Architecture for Client-Server Network

5.2. TE Reachability

As described in Section 1.1, TE reachability is the ability to reach a specific address along a TE path. The knowledge of TE reachability enables an end-to-end TE path to be computed.

In a single network, TE reachability is derived from the Traffic Engineering Database (TED) that is the collection of all TE information about all TE links in the network. The TED is usually built from the data exchanged by the IGP, although it can be supplemented by configuration and inventory details especially in

transport networks.

In multi-network scenarios, TE reachability information can be described as "You can get from node X to node Y with the following TE attributes." For transit cases, nodes X and Y will be edge nodes of the transit network, but it is also important to consider the information about the TE connectivity between an edge node and a specific destination node.

TE reachability may be unqualified (there is a TE path), or may be qualified by TE attributes such as TE metrics, hop count, available bandwidth, delay, shared risk, etc.

TE reachability information can be exchanged between networks so that nodes in one network can determine whether they can establish TE paths across or into another network. Such exchanges are subject to a range of policies imposed by the advertiser (for security and administrative control) and by the receiver (for scalability and stability).

5.3. Abstraction not Aggregation

Aggregation is the process of synthesizing from available information. Thus, the virtual node and virtual link models described in Section 3.6 rely on processing the information available within a network to produce the aggregate representations of links and nodes that are presented to the consumer. As described in Section 3, dynamic aggregation is subject to a number of pitfalls.

In order to distinguish the architecture described in this document from the previous work on aggregation, we use the term "abstraction" in this document. The process of abstraction is one of applying policy to the available TE information within a domain, to produce selective information that represents the potential ability to connect across the domain.

Abstraction does not offer all possible connectivity options (refer to Section 3.6), but does present a general view of potential connectivity. Abstraction may have a dynamic element, but is not intended to keep pace with the changes in TE attribute availability within the network.

Thus, when relying on an abstraction to compute an end-to-end path, the process might not deliver a usable path. That is, there is no actual guarantee that the abstractions are current or feasible.

While abstraction uses available TE information, it is subject to policy and management choices. Thus, not all potential connectivity

will be advertised to each client. The filters may depend on commercial relationships, the risk of disclosing confidential information, and concerns about what use is made of the connectivity that is offered.

5.3.1. Abstract Links

An abstract link is a measure of the potential to connect a pair of points with certain TE parameters. An abstract link may be realized by an existing LSP, or may represent the possibility of setting up an LSP.

When looking at a network such as that in Figure 8, the link from CN1 to CN4 may be an abstract link. If the LSP has already been set up, it is easy to advertise it as a link with known TE attributes: policy will have been applied in the server network to decide what LSP to set up. If the LSP has not yet been established, the potential for an LSP can be abstracted from the TE information in the core network subject to policy, and the resultant potential LSP can be advertised.

Since the client nodes do not have visibility into the core network, they must rely on abstraction information delivered to them by the core network. That is, the core network will report on the potential for connectivity.

5.3.2. The Abstraction Layer Network

Figure 9 introduces the Abstraction Layer Network. This construct separates the client layer resources (nodes C1, C2, C3, and C4, and the corresponding links), and the server layer resources (nodes CN1, CN2, CN3, and CN4 and the corresponding links). Additionally, the architecture introduces an intermediary layer called the Abstraction Layer. The Abstraction Layer contains the client layer edge nodes (C2 and C3), the server layer edge nodes (CN1 and CN4), the client-server links (C2-CN1 and CN4-C3) and the abstract link CN1-CN4.

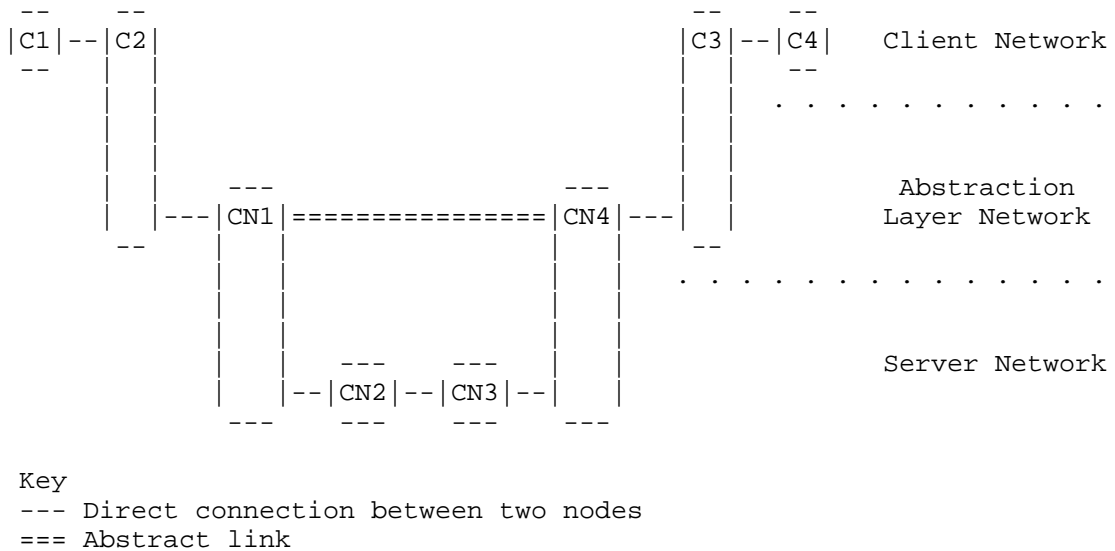


Figure 9 : Architecture for Abstraction Layer Network

The client layer network is able to operate as normal. Connectivity across the network can either be found or not found based on links that appear in the client layer TED. If connectivity cannot be found, end-to-end LSPs cannot be set up. This failure may be reported but no dynamic action is taken by the client layer.

The server network layer also operates as normal. LSPs across the server layer are set up in response to management commands or in response to signaling requests.

The Abstraction Layer consists of the physical links between the two networks, and also the abstract links. The abstract links are created by the server network according to local policy and represent the potential connectivity that could be created across the server network and which the server network is willing to make available for use by the client network. Thus, in this example, the diameter of the Abstraction Layer Network is only three hops, but an instance of an IGP could easily be run so that all nodes participating in the Abstraction Layer (and in particular the client network edge nodes) can see the TE connectivity in the layer.

When the client layer needs additional connectivity it can make a request to the Abstraction Layer Network. For example, the operator of the client network may want to create a link from C2 to C3. The Abstraction Layer can see the potential path C2-CN1-CN4-C3, and asks the server layer to realise the abstract link CN1-CN4. The server

layer provisions the LSP CN1-CN2-CN3-CN4 and makes the LSP available as a hierarchical LSP to turn the abstract link into a link that can be used in the client network. The Abstraction Layer can then set up an LSP C2-CN1-CN4-C3 using stitching or tunneling, and make the LSP available as a virtual link in the client network.

Sections 5.3.3 and 5.3.4 show how this model is used to satisfy the requirements for connectivity in client-server networks and in peer networks.

5.3.2.1. Nodes in the Abstraction Layer Network

Figure 9 shows a very simplified network diagram and the reader would be forgiven for thinking that only Client Network edge nodes and Server Network edge nodes may appear in the Abstraction Layer Network. But this is not the case: other nodes from the Server Network may be present. This allows the Abstraction Layer network to be more complex than a full mesh with access spokes.

Thus, as shown in Figure 10, a transit node in the Server Network (here the node is CN3) can be exposed as a node in the Abstraction Layer Network with Abstract Links connecting it to other nodes in the Abstraction Layer Network. Of course, in the network shown in Figure 10, there is little if any value in exposing CN3, but if it had other Abstract Links to other nodes in the Abstraction Layer Network and/or direct connections to Client Network nodes, then the resulting network would be richer.

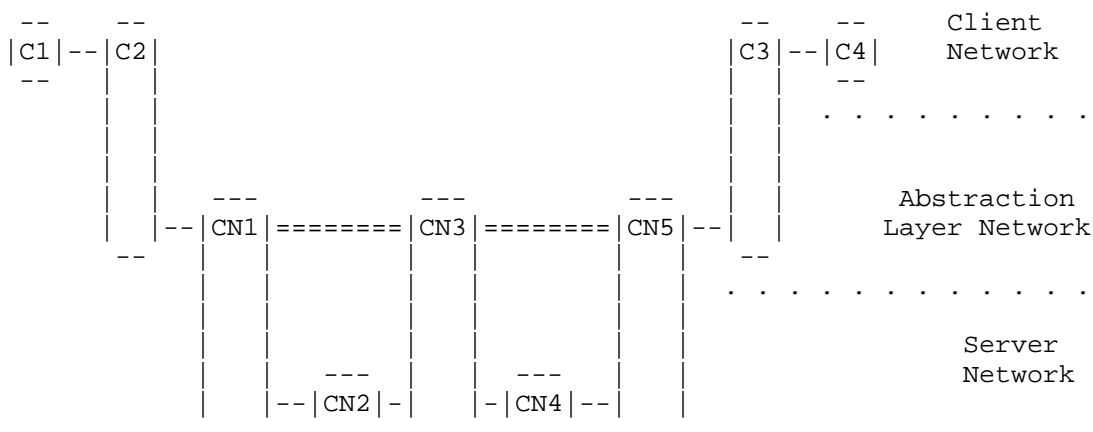


Figure 10 : Abstraction Layer Network with Additional Node

It should be noted that the nodes included in the Abstraction Layer

network in this way are not "Abstract Nodes" in the sense of a virtual node described in Section 3.6. While it is the case that the policy point responsible for advertising Server Network resources into the Abstraction Layer Network could choose to advertise Abstract Nodes in place of real physical nodes, it is believed that doing so would introduce significant complexity in terms of:

- Coordination between all of the external interfaces of the Abstract Node
- Management of changes in the Server Network that lead to limited capabilities to reach (cross-connect) across the Abstract Node. It may be noted that recent work on limited cross-connect capabilities such as exist in asymmetrical switches could be used to represent the limitations in an Abstract Node
[I-D.ietf-ccamp-general-constraint-encode],
[I-D.ietf-ccamp-gmpls-general-constraints-ospf-te].

5.3.3. Abstraction in Client-Server Networks

Section 5.3.2 has already introduced the concept of the Abstraction Layer Network through an example of a simple layered network. But it may be helpful to expand on the example using a slightly more complex network.

Figure 11 shows a multi-layer network comprising client nodes (labeled as Cn for n= 0 to 9) and server nodes (labeled as Sn for n = 1 to 9).

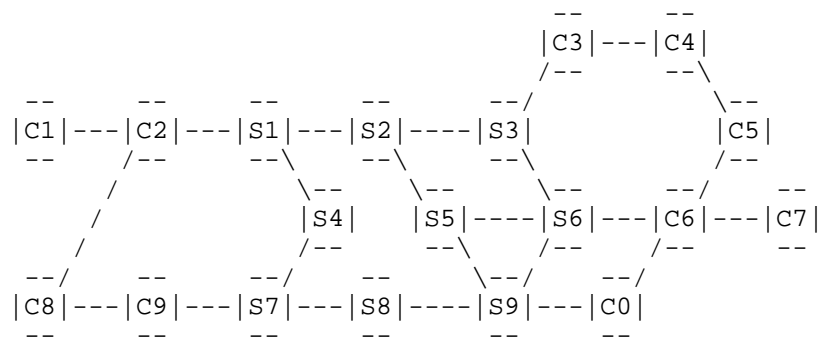


Figure 11 : An example Multi-Layer Network

If the network in Figure 11 is operated as separate client and server networks then the client layer topology will appear as shown in Figure 12. As can be clearly seen, the network is partitioned and there is no way to set up an LSP from a node on the left hand side

(say C1) to a node on the right hand side (say C7).

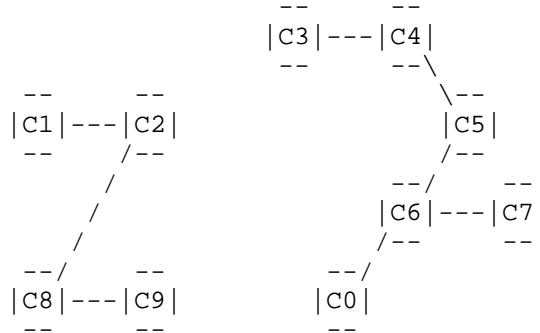


Figure 12 : Client Layer Topology Showing Partitioned Network

For reference, Figure 13 shows the corresponding server layer topology.

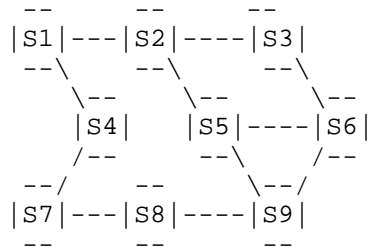


Figure 13 : Server Layer Topology

Operating on the TED for the server layer, a management entity or a software component may apply policy and consider what abstract links it might offer for use by the client layer. To do this it obviously needs to be aware of the connections between the layers (there is no point in offering an abstract link S2-S8 since this could not be of any use in this example).

In our example, after consideration of which LSPs could be set up in the server layer, four abstract links are offered: S1-S3, S3-S6, S1-S9, and S7-S9. These abstract links are shown as double lines on the resulting topology of the Abstraction Layer Network in Figure 14. As can be seen, two of the links must share part of a path (S1-S9 must share with either S1-S3 or with S7-S9). This could be achieved using distinct resources (for example, separate lambdas) where the paths are common, but it could also be done using resource sharing.

That would mean that when both S1-S3 and S7-S9 are realized as links carrying Abstraction Layer LSPs, the link S1-S9 can no longer be used.

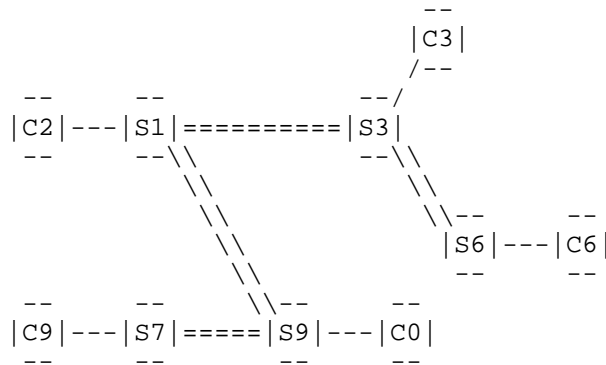


Figure 14 : Abstraction Layer Network with Abstract Links

The separate IGP instance running in the Abstraction Layer Network mean that this topology is visible at the edge nodes (C2, C3, C6, C9, and C0) as well as at a PCE if one is present.

Now the client layer is able to make requests to the Abstraction Layer Network to provide connectivity. In our example, it requests that C2 is connected to C3 and that C2 is connected to C0. This results in several actions:

1. The management component for the Abstraction Layer Network asks its PCE to compute the paths necessary to make the connections. This yields C2-S1-S3-C3 and C2-S1-S9-C0.
2. The management component for the Abstraction Layer Network instructs C2 to start the signaling process for the new LSPs in the Abstraction Layer.
3. C2 signals the LSPs for setup using the explicit routes C2-S1-S3-C3 and C2-S1-S9-C0.
4. When the signaling messages reach S1 (in our example, both LSPs traverse S1) the Abstraction Layer Network may find that the necessary underlying LSPs (S1-S2-S3 and S1-S2-S5-S9) have not been established since it is not a requirement that an abstract link be backed up by a real LSP. In this case, S1 computes the paths of the underlying LSPs and signals them.
5. Once the serve layer LSPs have been established, S1 can continue

to signal the Abstraction Layer LSPs either using the server layer LSPs as tunnels or as stitching segments.

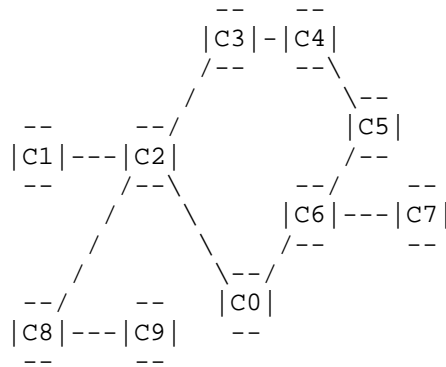


Figure 15 : Connected Client Layer Network with Additional Links

6. Finally, once the Abstraction Layer LSPs have been set up, the client layer can be informed and can start to advertise the new TE links C2-C3 and C2-C0. The resulting client layer topology is shown in Figure 15.

7. Now the client layer can compute an end-to-end path from C1 to C7.

5.3.3.1 Macro Shared Risk Link Groups

Network links often share fate with one or more other links. That is, a scenario that may cause a links to fail could cause one or more other links to fail. This may occur, for example, if the links are supported by the same fiber bundle, or if some links are routed down the same duct or in a common piece of infrastructure such as a bridge. A common way to identify the links that may share fate is to label them as belonging to a Shared Risk Link Group (SRLG) [RFC4202].

TE links created from LSPs in lower layers may also share fate, and it can be hard for a client network to know about this problem because it does not know the topology of the server network or the path of the server layer LSPs that are used to create the links in the client network.

For example, looking at the example used in Section 5.3.3 and considering the two abstract links S1-S3 and S1-S9 there is no way for the client layer to know whether the links C2-C0 and C2-C3 share fate. Clearly, if the client layer uses these links to provide a link-diverse end-to-end protection scheme, it needs to know that the links actually share a piece of network infrastructure (the server

layer link S1-S2).

Per [RFC4202], an SRLG represents a shared physical network resource upon which the normal functioning of a link depends. Multiple SRLGs can be identified and advertised for every TE link in a network. However, this can produce a scalability problem in a mutli-layer network that equates to advertising in the client layer the server layer route of each TE link.

Macro SRLGs (MSRLGs) address this scaling problem and are a form of abstraction performed at the same time that the abstract links are derived. In this way, only the links that actually links in the server layer need to be advertised rather than every link that potentially shares resources. This saving is possible because the abstract links are formulated on behalf of the server layer by a central management agency that is aware of all of the link abstractions being offered.

It may be noted that a less optimal alternative path for the abstract link S1-S9 exists in the server layer (S1-S4-S7-S8-S9). It is would be possible for the client layer request for connectivity C2-C0 to request that the path be maximally disjoint from the path C2-C3. While nothing can be done about the shared link C2-S1, the Abstraction Layer could request that the server layer instantiate the link S1-S9 to be diverse from the link S1-S3, and this request could be honored if the server layer policy allows.

5.3.3.2 A Server with Multiple Clients

A single server network may support multiple client networks. This is not an uncommon state of affairs for example when the server network provides connectivity for multiple customers.

In this case, the abstraction provided by the server layer may vary considerably according to the policies and commercial relationships with each customer. This variance would lead to a separate Abstraction Layer Network maintained to support each client network.

On the other hand, it may be that multiple clients are subject to the same policies and the abstraction can be identical. In this case, a single Abstraction Layer Network can support more than one client.

The choices here are made as an operational issue by the server layer network.

5.3.3.3 A Client with Multiple Servers

A single client network may be supported by multiple server networks. The server networks may provide connectivity between different parts of the client network or may provide parallel (redundant) connectivity for the client network.

In this case the Abstraction Layer Network should contain the abstract links from all server networks so that it can make suitable computations and create the correct TE links in the client network. That is, the relationship between client network and Abstraction Layer Network should be one-to-one.

Note that SRLGs and MSRLGs may be very hard to describe in the case of multiple server layer networks because the abstraction points will not know whether the resources in the various server layers share physical locations.

5.3.4. Abstraction in Peer Networks

Peer networks exist in many situations in the Internet. Packet networks may peer as IGP areas (levels) or as ASes. Transport networks (such as optical networks) may peer to provide concatenations of optical paths through single vendor environments (see Section 7). Figure 16 shows a simple example of three peer networks (A, B, and C) each comprising a few nodes.

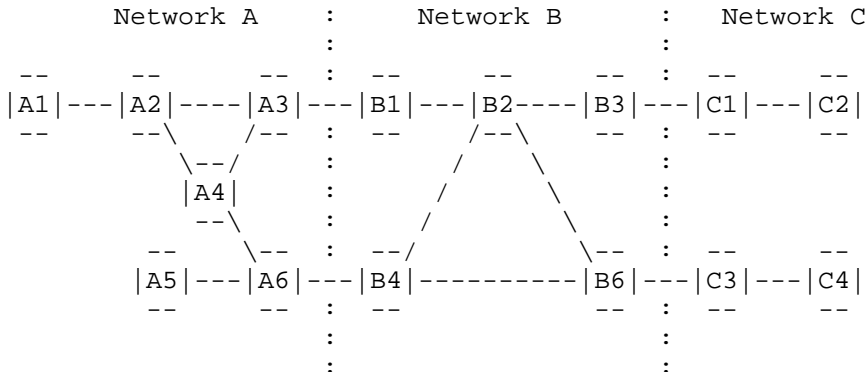


Figure 16 : A Network Comprising Three Peer Networks

As discussed in Section 2, peered networks do not share visibility of their topologies or TE capabilities for scaling and confidentiality reasons. That means, in our example, that computing a path from A1 to C4 can be impossible without the aid of cooperating PCEs or some form of crankback.

But it is possible to produce abstract links for the reachability across transit peer networks and instantiate an Abstraction Layer Network. That network can be enhanced with specific reachability information if a destination network is partitioned as is the case with Network C in Figure 16.

Suppose Network B decides to offer three abstract links B1-B3, B4-B3, and B4-B6. The Abstraction Layer Network could then be constructed to look like the network in Figure 17.

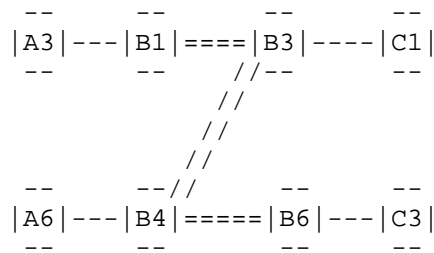


Figure 17 : Abstraction Layer Network for the Peer Network Example

Using a process similar to that described in Section 5.3.3, Network A can request connectivity to Network C and the abstract links can be instantiated as tunnels across the transit network, and edge-to-edge LSPs can be set up to join the two networks. Furthermore, if Network C is partitioned, reachability information can be exchanged to allow Network A to select the correct edge-to-edge LSP as shown in Figure 18.

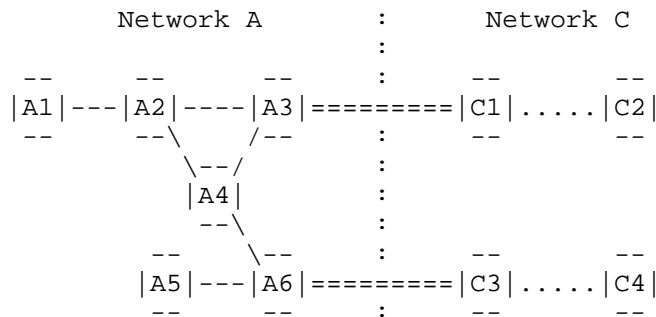


Figure 18 : Tunnel Connections to Network C with TE Reachability

Peer networking cases can be made far more complex by dual homing between network peering nodes (for example, A3 might connect to B1 and B4 in Figure 17) and by the networks themselves being arranged in a mesh (for example, A6 might connect to B4 and C1 in Figure 17).

These additional complexities can be handled gracefully by the Abstraction Layer Network model.

Further examples of abstraction in peer networks can be found in Sections 7 and 9.

5.4. Considerations for Dynamic Abstraction

<TBD>

5.5. Requirements for Advertising Links and Nodes

The Abstraction Layer Network is "just another network layer". The links and nodes in the network need to be advertised along with their associated TE information (metrics, bandwidth, etc.) so that the topology is disseminated and so that routing decisions can be made.

This requires a routing protocol running between the nodes in the Abstraction Layer Network. Note that this routing information exchange could be piggy-backed on an existing routing protocol instance, or use a new instance (or even a new protocol). Clearly, the information exchanged is only that which has been created as part of the abstraction function according to policy.

It should be noted that in some cases Abstract Link enablement is on-demand and all that is advertised in the topology for the Abstraction Layer Network is the potential for an Abstract Link to be set up. In this case we may ponder how the routing protocol will advertise topology information over a link that is not yet established. In other words, there must be a communication channel between the participating nodes so that the routing protocol messages can flow. The answer is that control plane connectivity exists in the Server Network and on the client-server edge links, and this can be used to carry the routing protocol messages for the Abstraction Layer Network. The same consideration applies to the advertisement, in the Client Network of the potential connectivity that the Abstraction Layer Network can provide.

5.6. Addressing Considerations

[Editor Note: Need to work up some text on addressing to cover the case of each domain having a different (potentially overlapping) address space and the need for inter-domain addressing. In fact, this should be quite simple but needs discussion.

Also needed is a discussion of the case where two client networks share an abstraction network (section 5.3.3.2). How does addressing work here? Are there security issues?]

6. Building on Existing Protocols

This section is not intended to prejudge a solutions framework or any applicability work. It does, however, very briefly serve to note the existence of protocols that could be examined for applicability to serve in realizing the model described in this document.

The general principle of protocol re-use is preferred over the invention of new protocols or additional protocol extensions as mentioned in Section 3.1.

6.1. BGP-LS

BGP-LS is a set of extensions to BGP described in [I-D.ietf-idr-ls-distribution]. It's purpose is to announce topology information from one network to a "north-bound" consumer. Application of BGP-LS to date has focused on a mechanism to build a TED for a PCE. However, BGP's mechanisms would also serve well to advertise Abstract Links from a Server Network into the Abstraction Layer Network, or to advertise potential connectivity from the Abstraction Layer Network to the Client Network.

6.2. IGPs

Both OSPF and IS-IS have been extended through a number of RFCs to advertise TE information. Additionally, both protocols are capable of running in a multi-instance mode either as ships that pass in the night (i.e., completely separate instances using different address) or as dual instances on the same address space. This means that either IGP could probably be used as the routing protocol in the Abstraction Layer Network.

6.3. RSVP-TE

RSVP-TE signaling can be used to set up traffic engineered LSPs to serve as hierarchical LSPs in the core network providing Abstract Links for the Abstraction Layer Network as described in [RFC4206]. Similarly, the CE-to-CE LSP tunnel across the Abstraction Layer Network can be established using RSVP-TE without any protocol extensions.

Furthermore, the procedures in [RFC6107] allow the dynamic signaling of the purpose of any LSP that is established. This means that when an LSP tunnel is set up, the two ends can coordinate into which routing protocol instance it should be advertised, and can also agree on the addressing to be said to identify the link that will be created.

7. Applicability to Optical Domains and Networks

Many optical networks are arranged a set of small domains. Each domain is a cluster of nodes, usually from the same equipment vendor and with the same properties. The domain may be constructed as a mesh or a ring, or maybe as an interconnected set of rings.

The network operator seeks to provide end-to-end connectivity across a network constructed from multiple domains, and so (of course) the domains are interconnected. In a network under management control such as through an Operations Support System (OSS), each domain is under the operational control of a Network Management System (NMS). In this way, an end-to-end path may be commissioned by the OSS instructing each NMS, and the NMSes setting up the path fragments across the domains.

However, in a system that uses a control plane, there is a need for integration between the domains.

Consider a simple domain, D1, as shown in Figure 19. In this case, the nodes A through F are arranged in a topological ring. Suppose that there is a control plane in use in this domain, and that OSPF is used as the TE routing protocol.

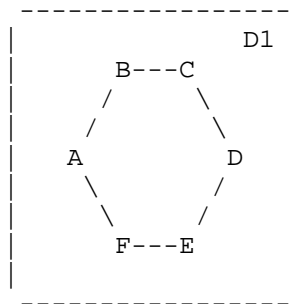


Figure 19 : A Simple Optical Domain

Now consider that the operator's network is built from a mesh of such domains, D1 through D7, as shown in Figure 20. It is possible that these domains share a single, common instance of OSPF in which case there is nothing further to say because that OSPF instance will distribute sufficient information to build a single TED spanning the whole network, and an end-to-end path can be computed. A more likely scenario is that each domain is running its own OSPF instance. In this case, each is able to handle the peculiarities (or rather, advanced functions) of each vendor's equipment capabilities.

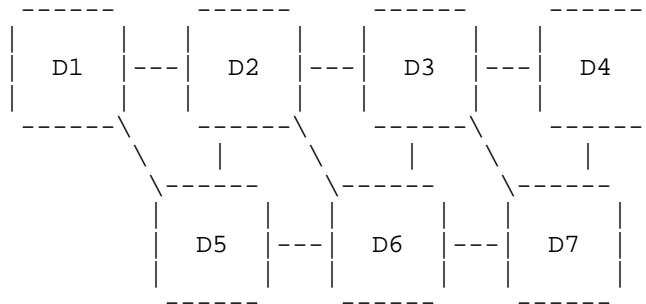


Figure 20 : A Simple Optical Domain

The question now is how to combine the multiple sets of information distributed by the different OSPF instances. Three possible models suggest themselves based on pre-existing routing practices.

- o In the first model (the Area-Based model) each domain is treated as a separate OSPF area. The end-to-end path will be specified to traverse multiple areas, and each area will be left to determine the path across the nodes in the area. The feasibility of an end-to-end path (and, thus, the selection of the sequence of areas and their interconnections) can be derived using hierarchical PCE.

This approach, however, fits poorly with established use of the OSPF area: in this form of optical network, the interconnection points between domains are likely to be links; and the mesh of domains is far more interconnected and unstructured than we are used to seeing in the normal area-based routing paradigm.

Furthermore, while hierarchical PCE may be able to solve this type of network, the effort involved may be considerable for more than a small collection of domains.

- o Another approach (the AS-Based model) treats each domain as a separate Autonomous System (AS). The end-to-end path will be specified to traverse multiple ASes, and each AS will be left to determine the path across the AS.

This model sits more comfortably with the established routing paradigm, but causes a massive escalation of ASes in the global Internet. It would, in practice, require that the operator used private AS numbers [RFC6996] of which there are plenty.

Then, as suggested in the Area-Based model, hierarchical PCE could be used to determine the feasibility of an end-to-end path and to derive the sequence of domains and the points of

interconnection to use. But, just as in that other model, the scalability of the hierarchical PCE approach must be questioned.

Furthermore, determining the mesh of domains (i.e., the inter-AS connections) conventionally requires the use of BGP as an inter-domain routing protocol. However, not only is BGP not normally available on optical equipment, but this approach indicates that the TE properties of the inter-domain links would need to be distributed and updated using BGP: something for which it is not well suited.

- o The third approach (the ASON model) follows the architectural model set out by the ITU-T [G.8080] and uses the routing protocol extensions described in [RFC6827]. In this model the concept of "levels" is introduced to OSPF. Referring back to Figure 20, each OSPF instance running in a domain would be construed as a "lower level" OSPF instance and would leak routes into a "higher level" instance of the protocol that runs across the whole network.

This approach handles the awkwardness of representing the domains as areas or ASes by simply considering them as domains running distinct instances of OSPF. Routing advertisements flow "upward" from the domains to the high level OSPF instance giving it a full view of the whole network and allowing end-to-end paths to be computed. Routing advertisements may also flow "downward" from the network-wide OSPF instance to any one domain so that it has visibility of the connectivity of the whole network.

While architecturally satisfying, this model suffers from having to handle the different characteristics of different equipment vendors. The advertisements coming from each low level domain would be meaningless when distributed into the other domains, and the high level domain would need to be kept up-to-date with the semantics of each new release of each vendor's equipment. Additionally, the scaling issues associated with a well-meshed network of domains each with many entry and exit points and each with network resources that are continually being updated reduces to the same problem as noted in the virtual link model. Furthermore, in the event that the domains are under control of different administrations, the domains would not want to distribute the details of their topologies and TE resources.

Practically, this third model turns out to be very close to the methodology described in this document. As noted in Section 7.1 of [RFC6827], there are policy rules that can be applied to define exactly what information is exported from or imported to a low level OSPF instance. The document even notes that some forms of aggregation may be appropriate. Thus, we can apply the following

simplifications to the mechanisms defined in RFC 6827:

- Zero information is imported to low level domains.
- Low level domains export only abstracted links as defined in this document and according to local abstraction policy and with appropriate removal of vendor-specific information.
- There is no need to formally define routing levels within OSPF.
- Export of abstracted links from the domains to the network-wide routing instance (the abstraction routing layer) can take place through any mechanism including BGP-LS or direct interaction between OSPF implementations.

With these simplifications, it can be seen that the framework defined in this document can be constructed from the architecture discussed in RFC 6827, but without needing any of the protocol extensions that that document defines. Thus, using the terminology and concepts already established, the problem may be solved as shown in Figure 21. The abstraction layer network is constructed from the inter-domain links, the domain border nodes, and the abstracted (cross-domain) links.

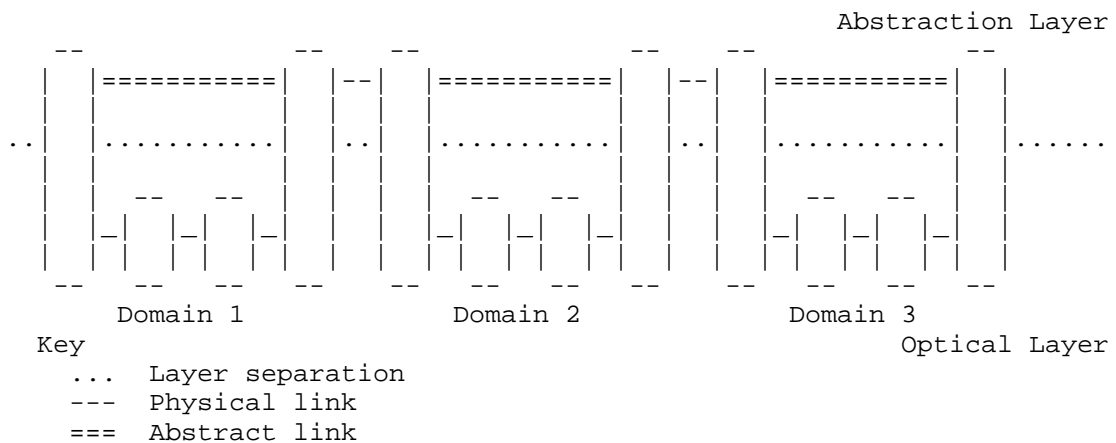


Figure 21 : The Optical Network Implemented Through the Abstraction Layer Network

8. Modeling the User-to-Network Interface

The User-to-Network Interface (UNI) is an important architectural concept in many implementations and deployments of client-server networks especially those where the client and server network have

different technologies. The UNI can be seen described in [G.8080], and the GMPLS approach to the UNI is documented in [RFC4208]. Other GMPLS-related documents describe the application of GMPLS to specific UNI scenarios: for example, [RFC6005] describes how GMPLS can support a UNI that provides access to Ethernet services.

Figure 1 of [RFC6005] is reproduced here as Figure 22. It shows the Ethernet UNI reference model, and that figure can serve as an example for all similar UNIs. In this case, the UNI is an interface between client network edge nodes and the server network. It should be noted that neither the client network nor the server network need be an Ethernet switching network.

There are three network layers in this model: the client network, the "Ethernet service network", and the server network. The so-called Ethernet service network consists of links comprising the UNI links and the tunnels across the server network, and nodes comprising the client network edge nodes and various server nodes. That is, the Ethernet service network is equivalent to the Abstraction Layer Network with the UNI links being the physical links between the client and server networks, and the client edge nodes taking the role of UNI Client-side (UNI-C) and the server edge nodes acting as the UNI Network-side (UNI-N) nodes.

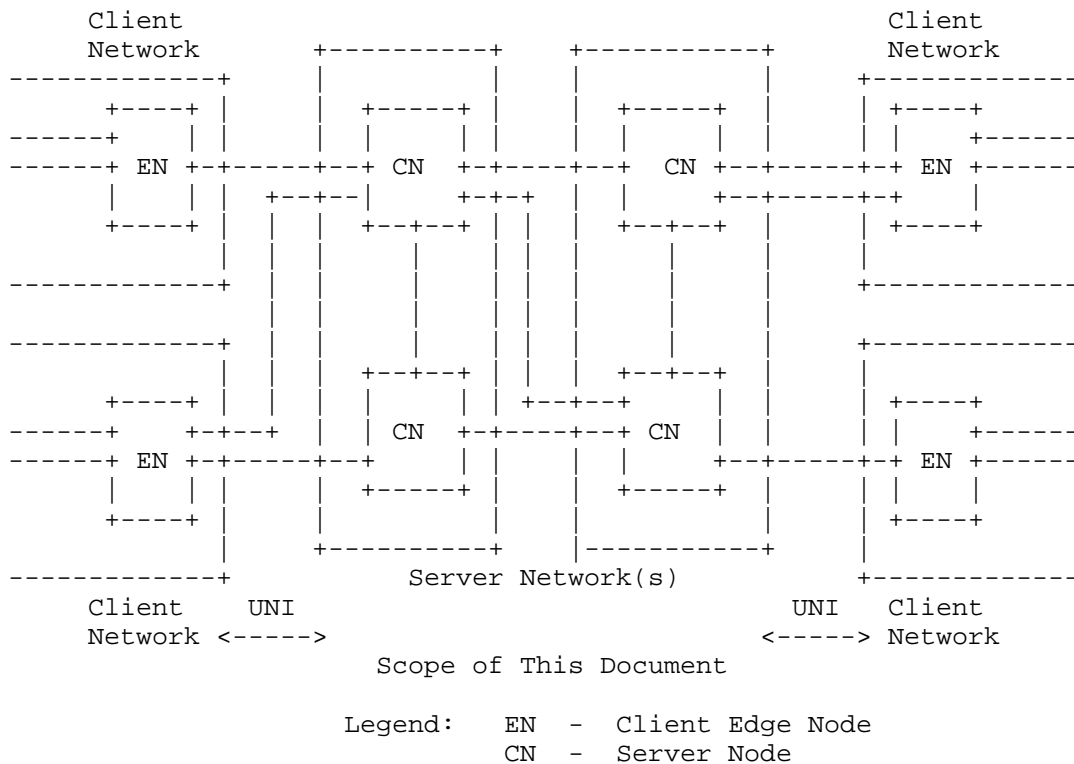


Figure 22 : Ethernet UNI Reference Model

An issue that is often raised concerns how a dual-homed client edge node (such as that shown at the bottom left-hand corner of Figure 22) can make determinations about how they connect across the UNI. This can be particularly important when reachability across the server network is limited or when two diverse paths are desired (for example, to provide protection). However, in the model described in this network, the edge node (the UNI-C) is part of the Abstraction Layer Network and can see sufficient topology information to make these decisions. There is, therefore, no need to enhance the signaling protocols at the GMPLS UNI nor to add routing exchanges at the UNI.

9. Abstraction in L3VPN Multi-AS Environments

Serving layer-3 VPNs (L3PVNs) across a multi-AS or multi-operator environment currently provides a significant planning challenge. Figure 6 shows the general case of the problem that needs to be solved. This section shows how the Abstraction Layer Network can address this problem.

10. Scoping Future Work

The section is provided to help guide the work on this problem and to ensure that oceans are not knowingly boiled.

10.1. Not Solving the Internet

The scope of the use cases and problem statement in this document is limited to "some small set of interconnected domains." In particular, it is not the objective of this work to turn the whole Internet into one large, interconnected TE network.

10.2. Working With "Related" Domains

Subsequent to Section 10.1, the intention of this work is to solve the TE interconnectivity for only "related" domains. Such domains may be under common administrative operation (such as IGP areas within a single AS, or ASes belonging to a single operator), or may have a direct commercial arrangement for the sharing of TE information to provide specific services. Thus, in both cases, there is a strong opportunity for the application of policy.

10.3. Not Finding Optimal Paths in All Situations

As has been well described in this document, abstraction necessarily involves compromises and removal of information. That means that it is not possible to guarantee that an end-to-end path over interconnected TE domains follows the absolute optimal (by any measure of optimality) path. This is taken as understood, and future work should not attempt to achieve such paths which can only be found by a full examination of all network information across all connected networks.

10.4. Not Breaking Existing Protocols

It is a clear objective of this work to not break existing protocols. The Internet relies on the stability of a few key routing protocols, and so it is critical that any new work must not make these protocols brittle or unstable.

10.5. Sanity and Scaling

All of the above points play into a final observation. This work is intended to bite off a small problem for some relatively simple use cases as described in Section 2. It is not intended that this work will be immediately (or even soon) extended to cover many large interconnected domains. Obviously the solution should as far as possible be designed to be extensible and scalable, however, it is

also reasonable to make trade-offs in favor of utility and simplicity.

11. Manageability Considerations

<TBD>

12. IANA Considerations

This document makes no requests for IANA action. The RFC Editor may safely remove this section.

13. Security Considerations

<TBD>

14. Acknowledgements

Thanks to Igor Bryskin for useful discussions in the early stages of this work.

Thanks to Gert Grammel for discussions on the extent of aggregation in abstract nodes and links.

Thanks to Deborah Brungard, Dieter Beller, and Vallinayakam Somasundaram for review and input.

Particular thanks to Vishnu Pavan Beeram for detailed discussions and white-board scribbling that made many of the ideas in this document come to life.

Text in Section 5.3.3 is freely adapted from the work of Igor Bryskin, Wes Doonan, Vishnu Pavan Beeram, John Drake, Gert Grammel, Manuel Paul, Ruediger Kunze, Friedrich Armbruster, Cyril Margaria, Oscar Gonzalez de Dios, and Daniele Ceccarelli in [I-D.beeram-ccamp-gmpls-enni] for which the authors of this document express their thanks.

15. References

15.1. Informative References

[G.8080] ITU-T, "Architecture for the automatically switched optical network (ASON)", Recommendation G.8080.

- [I-D.beeram-ccamp-gmpls-enni]
Bryskin, I., Beeram, V. P., Drake, J. et al., "Generalized Multiprotocol Label Switching (GMPLS) External Network Interface (E-NNI): Virtual Link Enhancements for the Overlay Model", draft-beeram-ccamp-gmpls-enni, work in progress.
- [I-D.farrkingel-pce-questions]
Farrel, A., and D. King, "Unanswered Questions in the Path Computation Element Architecture", draft-farrkingel-pce-questions, work in progress.
- [I-D.ietf-ccamp-general-constraint-encode]
Bernstein, G., Lee, Y., Li, D., and Imajuku, W., "General Network Element Constraint Encoding for GMPLS Controlled Networks", draft-ietf-ccamp-general-constraint-encode, work in progress.
- [I-D.ietf-ccamp-gmpls-general-constraints-ospf-te]
Zhang, F., Lee, Y., Han, J, Bernstein, G., and Xu, Y., "OSPF-TE Extensions for General Network Element Constraints", draft-ietf-ccamp-gmpls-general-constraints-ospf-te, work in progress.
- [I-D.ietf-ccamp-rsvp-te-srlg-collect]
Zhang, F. (Ed.) and O. Gonzalez de Dios (Ed.), "RSVP-TE Extensions for Collecting SRLG Information", draft-ietf-ccamp-rsvp-te-srlg-collect, work in progress.
- [I-D.ietf-ccamp-te-metric-recording]
Z. Ali, et al., "Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) extension for recording TE Metric of a Label Switched Path," draft-ali-ccamp-te-metric-recording, work in progress.
- [I-D.ietf-ccamp-xro-lsp-subobject]
Z. Ali, et al., "Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) LSP Route Diversity using Exclude Routes," draft-ali-ccamp-xro-lsp-subobject, work in progress.
- [I-D.ietf-idr-ls-distribution]
Gredler, H., Medved, J., Previdi, S., Farrel, A., and Ray, S., "North-Bound Distribution of Link-State and TE Information using BGP", draft-ietf-idr-ls-distribution, work in progress.

- [RFC2702] Awduche, D., Malcolm, J., Agogbua, J., O'Dell, M., and McManus, J., "Requirements for Traffic Engineering Over MPLS", RFC 2702, 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.
- [RFC3473] L. Berger, "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions", RC 3473, January 2003.
- [RFC3630] Katz, D., Kompella, and K., Yeung, D., "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.
- [RFC4105] Le Roux, J.-L., Vasseur, J.-P., and Boyle, J., "Requirements for Inter-Area MPLS Traffic Engineering", RFC 4105, June 2005.
- [RFC4202] Kompella, K. and Y. Rekhter, "Routing Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4202, October 2005.
- [RFC4206] Kompella, K. and Y. Rekhter, "Label Switched Paths (LSP) Hierarchy with Generalized Multi-Protocol Label Switching (GMPLS) Traffic Engineering (TE)", RFC 4206, October 2005.
- [RFC4208] Swallow, G., Drake, J., Ishimatsu, H., and Y. Rekhter, "User-Network Interface (UNI): Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Support for the Overlay Model", RFC 4208, October 2005.
- [RFC4216] Zhang, R., and Vasseur, J.-P., "MPLS Inter-Autonomous System (AS) Traffic Engineering (TE) Requirements", RFC 4216, November 2005.
- [RFC4271] Rekhter, Y., Li, T., and Hares, S., "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, January 2006.
- [RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, August 2006.

- [RFC4726] Farrel, A., Vasseur, J.-P., and Ayyangar, A., "A Framework for Inter-Domain Multiprotocol Label Switching Traffic Engineering", RFC 4726, November 2006.
- [RFC4847] T. Takeda (Ed.), "Framework and Requirements for Layer 1 Virtual Private Networks," RFC 4847, April 2007.
- [RFC4874] Lee, CY., Farrel, A., and S. De Chodder, "Exclude Routes - Extension to Resource ReserVation Protocol-Traffic Engineering (RSVP-TE)", RFC 4874, April 2007.
- [RFC4920] Farrel, A., Satyanarayana, A., Iwata, A., Fujita, N., and Ash, G., "Crankback Signaling Extensions for MPLS and GMPLS RSVP-TE", RFC 4920, July 2007.
- [RFC5150] Ayyangar, A., Kompella, K., Vasseur, JP., and A. Farrel, "Label Switched Path Stitching with Generalized Multiprotocol Label Switching Traffic Engineering (GMPLS TE)", RFC 5150, February 2008.
- [RFC5152] Vasseur, JP., Ayyangar, A., and Zhang, R., "A Per-Domain Path Computation Method for Establishing Inter-Domain Traffic Engineering (TE) Label Switched Paths (LSPs)", RFC 5152, February 2008.
- [RFC5195] Ould-Brahim, H., Fedyk, D., and Y. Rekhter, "BGP-Based Auto-Discovery for Layer-1 VPNs", RFC 5195, June 2008.
- [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.
- [RFC5251] Fedyk, D., Rekhter, Y., Papadimitriou, D., Rabbat, R., and L. Berger, "Layer 1 VPN Basic Mode", RFC 5251, July 2008.
- [RFC5252] Bryskin, I. and L. Berger, "OSPF-Based Layer 1 VPN Auto-Discovery", RFC 5252, July 2008.
- [RFC5305] Li, T., and Smit, H., "IS-IS Extensions for Traffic Engineering", RFC 5305, October 2008.
- [RFC5440] Vasseur, JP. and Le Roux, JL., "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, March 2009.
- [RFC5441] Vasseur, JP., Zhang, R., Bitar, N, and Le Roux, JL., "A Backward-Recursive PCE-Based Computation (BRPC) Procedure to Compute Shortest Constrained Inter-Domain Traffic Engineering Label Switched Paths", RFC 5441, April 2009.

- [RFC5523] L. Berger, "OSPFv3-Based Layer 1 VPN Auto-Discovery", RFC 5523, April 2009.
- [RFC5553] Farrel, A., Bradford, R., and JP. Vasseur, "Resource Reservation Protocol (RSVP) Extensions for Path Key Support", RFC 5553, May 2009.
- [RFC5623] Oki, E., Takeda, T., Le Roux, JL., and A. Farrel, "Framework for PCE-Based Inter-Layer MPLS and GMPLS Traffic Engineering", RFC 5623, September 2009.
- [RFC6005] Nerger, L., and D. Fedyk, "Generalized MPLS (GMPLS) Support for Metro Ethernet Forum and G.8011 User Network Interface (UNI)", RFC 6005, October 2010.
- [RFC6107] Shiimoto, K., and A. Farrel, "Procedures for Dynamically Signaled Hierarchical Label Switched Paths", RFC 6107, February 2011.
- [RFC6805] King, D., and A. Farrel, "The Application of the Path Computation Element Architecture to the Determination of a Sequence of Domains in MPLS and GMPLS", RFC 6805, November 2012.
- [RFC6827] Malis, A., Lindem, A., and D. Papadimitriou, "Automatically Switched Optical Network (ASON) Routing for OSPFv2 Protocols", RFC 6827, January 2013.
- [RFC6996] J. Mitchell, "Autonomous System (AS) Reservation for Private Use", BCP 6, RFC 6996, July 2013.

Authors' Addresses

Adrian Farrel
Juniper Networks
EMail: adrian@olddog.co.uk

John Drake
Juniper Networks
EMail: jdrake@juniper.net

Nabil Bitar
Verizon
40 Sylvan Road
Waltham, MA 02145
EMail: nabil.bitar@verizon.com

George Swallow
Cisco Systems, Inc.
1414 Massachusetts Ave
Boxborough, MA 01719
EMail: swallow@cisco.com

Daniele Ceccarelli
Ericsson
Via A. Negrone 1/A
Genova - Sestri Ponente
Italy
EMail: daniele.ceccarelli@ericsson.com

Xian Zhang
Huawei Technologies
Email: zhang.xian@huawei.com

Contributors

Gert Grammel
Juniper Networks
Email: ggrammel@juniper.net

Vishnu Pavan Beeram
Juniper Networks
Email: vbeeram@juniper.net

Oscar Gonzalez de Dios
Email: ogondio@tid.es

Fatai Zhang
Email: zhangfatai@huawei.com

Zafar Ali
Email: zali@cisco.com

Rajan Rao
Email: rrao@infinera.com

Sergio Belotti
Email: sergio.belotti@alcatel-lucent.com

Diego Caviglia
Email: diego.caviglia@ericsson.com

Jeff Tantsura
Email: jeff.tantsura@ericsson.com

Khuzema Pithewan
Email: kpithewan@infinera.com

Cyril Margaria
Email: cyril.margaria@gmail.com

Victor Lopez
Email: vlopez@tid.es

Internet Engineering Task Force
Internet-Draft
Intended status: Standards Track
Expires: January 7, 2016

G.Galimberti, Ed.
Cisco
R.Kunze, Ed.
Deutsche Telekom
Kam Lam, Ed.
Alcatel-Lucent
D. Hiremagalur, Ed.
Juniper
L.Fang, Ed.
G.Ratterree, Ed.
Microsoft
July 6, 2015

An SNMP MIB extension to RFC3591 to manage optical interface parameters
of "G.698.2 single channel" in DWDM applications
draft-galikunze-ccamp-g-698-2-snmp-mib-12

Abstract

This memo defines a module of the Management Information Base (MIB) used by Simple Network Management Protocol (SNMP) in TCP/IP- based internet. In particular, it defines objects for managing single channel optical interface parameters of DWDM applications, using the approach specified in G.698.2 [ITU.G698.2] . This interface, described in ITU-T G.872, G.709 and G.798, is one type of OTN multi-vendor Intra-Domain Interface (IaDI). This RFC is an extension of RFC3591 to support the optical parameters specified in ITU-T G.698.2 and application identifiers specified in ITU-T G.874.1 [ITU.G874.1]. Note that G.874.1 encompasses vendor-specific codes, which if used would make the interface a single vendor IaDI and could still be managed.

The MIB module defined in this memo can be used for Optical Parameters monitoring and/or configuration of the endpoints of the multi-vendor IaDI based on the Black Link approach.

Copyright Notice

Copyright (c) 2015 IETF Trust and the persons identified as the document authors. All rights reserved.

Status of This Memo

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

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

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

This Internet-Draft will expire on January 7, 2016.

Copyright Notice

Copyright (c) 2015 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. The Internet-Standard Management Framework	4
3. Conventions	5
4. Overview	5
4.1. Use Cases	6
4.2. Optical Parameters Description	13
4.2.1. Rs-Ss Configuration	13
4.2.2. Table of Application Identifiers	14
4.3. Use of ifTable	15
4.3.1. Use of ifTable for OPS Layer	16
4.3.2. Use of ifTable for OCh Layer	17
4.3.3. Use of ifStackTable	17
5. Structure of the MIB Module	18
6. Object Definitions	18
7. Relationship to Other MIB Modules	25
7.1. Relationship to the [TEMPLATE TODO] MIB	25
7.2. MIB modules required for IMPORTS	25
8. Definitions	25
9. Security Considerations	25

10. IANA Considerations	26
11. Contributors	26
12. References	27
12.1. Normative References	27
12.2. Informative References	29
Appendix A. Change Log	30
Appendix B. Open Issues	30
Authors' Addresses	30

1. Introduction

This memo defines a portion of the Management Information Base (MIB) used by Simple Network Management Protocol (SNMP) in TCP/IP-based internets. In particular, it defines objects for managing single channel optical interface parameters of DWDM applications, using the approach specified in G.698.2. This RFC is an extension of RFC3591 to support the optical parameters specified in ITU-T G.698.2 [ITU.G698.2] and application identifiers specified in ITU-T G.874.1 [ITU.G874.1]. Note that G.874.1 encompasses vendor-specific codes, which if used would make the interface a single vendor IaDI and could still be managed.

The Black Link approach allows supporting an optical transmitter/receiver pair of one vendor to inject an optical tributary signal and run it over an optical network composed of amplifiers, filters, add-drop multiplexers from a different vendor. In the OTN architecture, the 'black-link' represents a pre-certified network media channel conforming to G.698.2 specifications at the S and R reference points.

[Editor's note: In G.698.2 this corresponds to the optical path from point S to R; network media channel is also used and explained in draft-ietf-ccamp-flexi-grid-fwk-02]

Management will be performed at the edges of the network media channel (i.e., at the transmitters and receivers attached to the S and R reference points respectively) for the relevant parameters specified in G.698.2 [ITU.G698.2], G.798 [ITU.G798], G.874 [ITU.G874], and the performance parameters specified in G.7710/Y.1701 [ITU-T G.7710] and G.874.1 [ITU.G874.1].

G.698.2 [ITU.G698.2] is primarily intended for metro applications that include optical amplifiers. Applications are defined in G.698.2 [ITU.G698.2] using optical interface parameters at the single-channel connection points between optical transmitters and the optical multiplexer, as well as between optical receivers and the optical demultiplexer in the DWDM system. This Recommendation uses a methodology which does not explicitly specify the details of the optical network between reference point Ss and Rs, e.g., the passive

and active elements or details of the design. The Recommendation currently includes unidirectional DWDM applications at 2.5 and 10 Gbit/s (with 100 GHz and 50 GHz channel frequency spacing). Work is still under way for 40 and 100 Gbit/s interfaces. There is possibility for extensions to a lower channel frequency spacing. This document specifically refers to the "application code" defined in the G.698.2 [ITU.G698.2] and included in the Application Identifier defined in G.874.1 [ITU.G874.1] and G.872 [ITU.G872], plus a few optical parameters not included in the G.698.2 application code specification.

This draft refers and supports also the draft-kunze-g-698-2-management-control-framework

The building of an SNMP MIB describing the optical parameters defined in G.698.2 [ITU.G698.2], and reflected in G.874.1 [ITU.G874], allows the different vendors and operator to retrieve, provision and exchange information across the G.698.2 multi-vendor IaDI in a standardized way.

The MIB, reporting the Optical parameters and their values, characterizes the features and the performances of the optical components and allow a reliable black link design in case of multi vendor optical networks.

Although RFC 3591 [RFC3591] describes and defines the SNMP MIB of a number of key optical parameters, alarms and Performance Monitoring, as this RFC is over a decade old, it is primarily pre-OTN, and a more complete and up-to-date description of optical parameters and processes can be found in the relevant ITU-T Recommendations. The same considerations can be applied to the RFC 4054 [RFC4054]

2. The Internet-Standard Management Framework

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

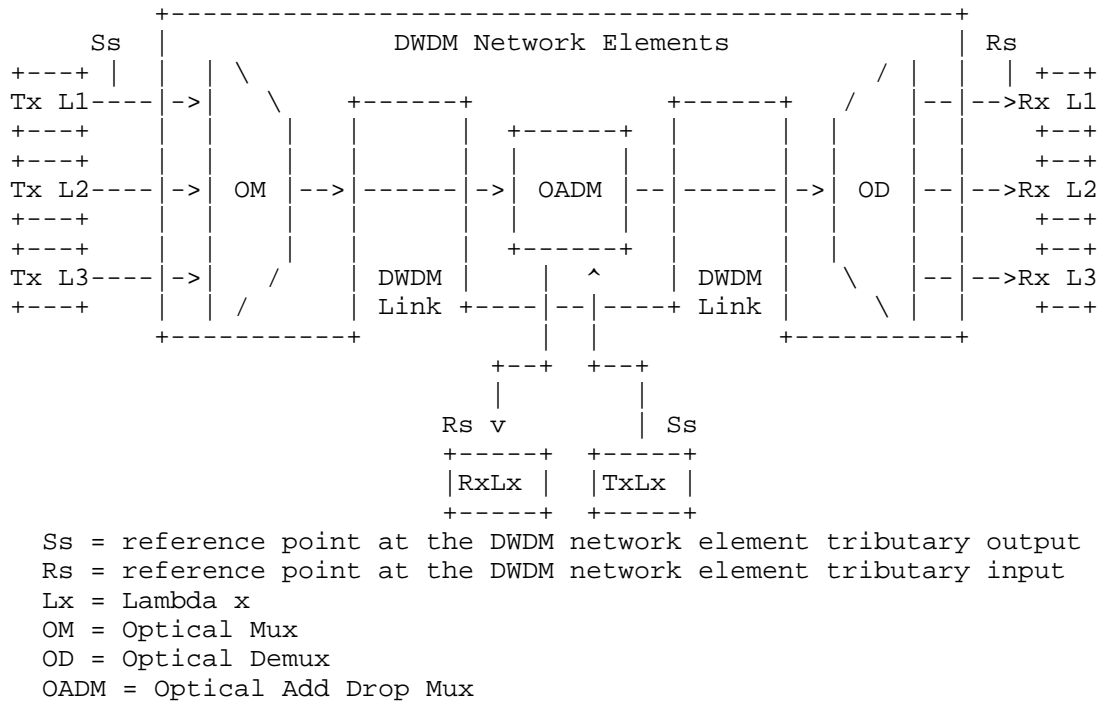
Managed objects are accessed via a virtual information store, termed the Management Information Base or MIB. MIB objects are generally accessed through the Simple Network Management Protocol (SNMP). Objects in the MIB are defined using the mechanisms defined in the Structure of Management Information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

3. Conventions

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119]. In the description of OIDs the convention: Set (S) Get (G) and Trap (T) conventions will describe the action allowed by the parameter.

4. Overview

Figure 1 shows a set of reference points, for the linear "black link" approach, for single-channel connection (Ss and Rs) between transmitters (Tx) and receivers (Rx). Here the DWDM network elements include an OM and an OD (which are used as a pair with the opposing element), one or more optical amplifiers and may also include one or more OADMs.



from Fig. 5.1/G.698.2

Figure 1: Linear Black Link approach

G.698.2 [ITU.G698.2] defines also Ring "Black Link" approach configurations [Fig. 5.2/G.698.2] and Linear "black link" approach for Bidirectional applications[Fig. 5.3/G.698.2]

4.1. Use Cases

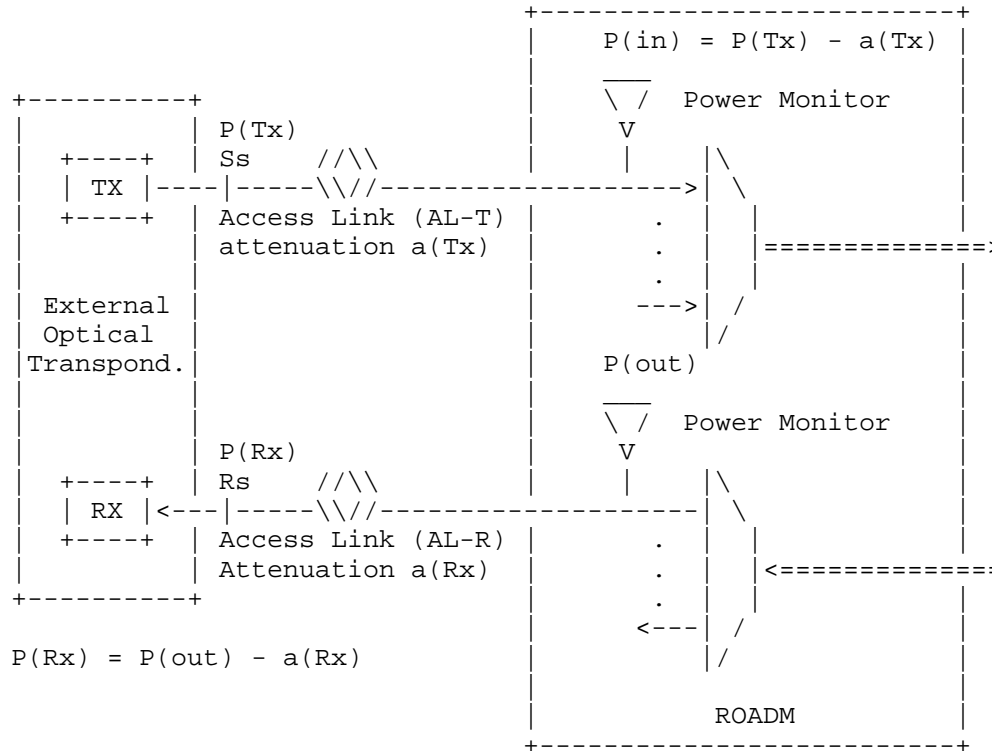
The use cases described below are assuming that power monitoring functions are available in the ingress and egress network element of the DWDM network, respectively. By performing link property correlation it would be beneficial to include the current transmit power value at reference point Ss and the current received power value at reference point Rs. For example if the Client transmitter power (OXC1) has a value of 0dBm and the ROADM interface measured power (at OLS1) is -6dBm the fiber patch cord connecting the two nodes may be pinched or the connectors are dirty. More, the interface characteristics can be used by the OLS network Control Plane in order to check the Optical Channels feasibility. Finally the OXC1 transceivers parameters (Application Code) can be shared with OXC2 using the LMP protocol to verify the Transceivers compatibility. The actual route selection of a specific wavelength within the allowed set is outside the scope of LMP. In GMPLS, the parameter selection (e.g. central frequency) is performed by RSVP-TE.

G.698.2 defines a single channel optical interface for DWDM systems that allows interconnecting network-external optical transponders across a DWDM network. The optical transponders are considered to be external to the DWDM network. This so-called 'black link' approach illustrated in Figure 5-1 of G.698.2 and a copy of this figure is provided below. The single channel fiber link between the Ss/Rs reference points and the ingress/egress port of the network element on the domain boundary of the DWDM network (DWDM border NE) is called access link in this contribution. Based on the definition in G.698.2 it is considered to be part of the DWDM network. The access link typically is realized as a passive fiber link that has a specific optical attenuation (insertion loss). As the access link is an integral part of the DWDM network, it is desirable to monitor its attenuation. Therefore, it is useful to detect an increase of the access link attenuation, for example, when the access link fiber has been disconnected and reconnected (maintenance) and a bad patch panel connection (connector) resulted in a significantly higher access link attenuation (loss of signal in the extreme case of an open connector or a fiber cut). In the following section, two use cases are presented and discussed:

- 1) pure access link monitoring
- 2) access link monitoring with a power control loop

These use cases require a power monitor as described in G.697 (see section 6.1.2), that is capable to measure the optical power of the incoming or outgoing single channel signal. The use case where a power control loop is in place could even be used to compensate an increased attenuation as long as the optical transmitter can still be operated within its output power range defined by its application code.

Figure 2 Access Link Power Monitoring



- For AL-T monitoring: $P(Tx)$ and $a(Tx)$ must be known
- For AL-R monitoring: $P(Rx)$ and $a(Rx)$ must be known

An alarm shall be raised if $P(in)$ or $P(Rx)$ drops below a configured threshold (t [dB]):

- $P(in) < P(Tx) - a(Tx) - t$ (Tx direction)
- $P(Rx) < P(out) - a(Rx) - t$ (Rx direction)
- $a(Tx) = | a(Rx)$

Figure 2: Extended LMP Model

Pure Access Link (AL) Monitoring Use Case

Figure 4 illustrates the access link monitoring use case and the different physical properties involved that are defined below:

- S_s, R_s : G.698.2 reference points
- $P(Tx)$: current optical output power of transmitter Tx
- $a(Tx)$: access link attenuation in Tx direction (external transponder point of view)
- $P(in)$: measured current optical input power at the input port of border DWDM NE
- t : user defined threshold (tolerance)
- $P(out)$: measured current optical output power at the output port of border DWDM NE
- $a(Rx)$: access link attenuation in Rx direction (external transponder point of view)
- $P(Rx)$: current optical input power of receiver Rx

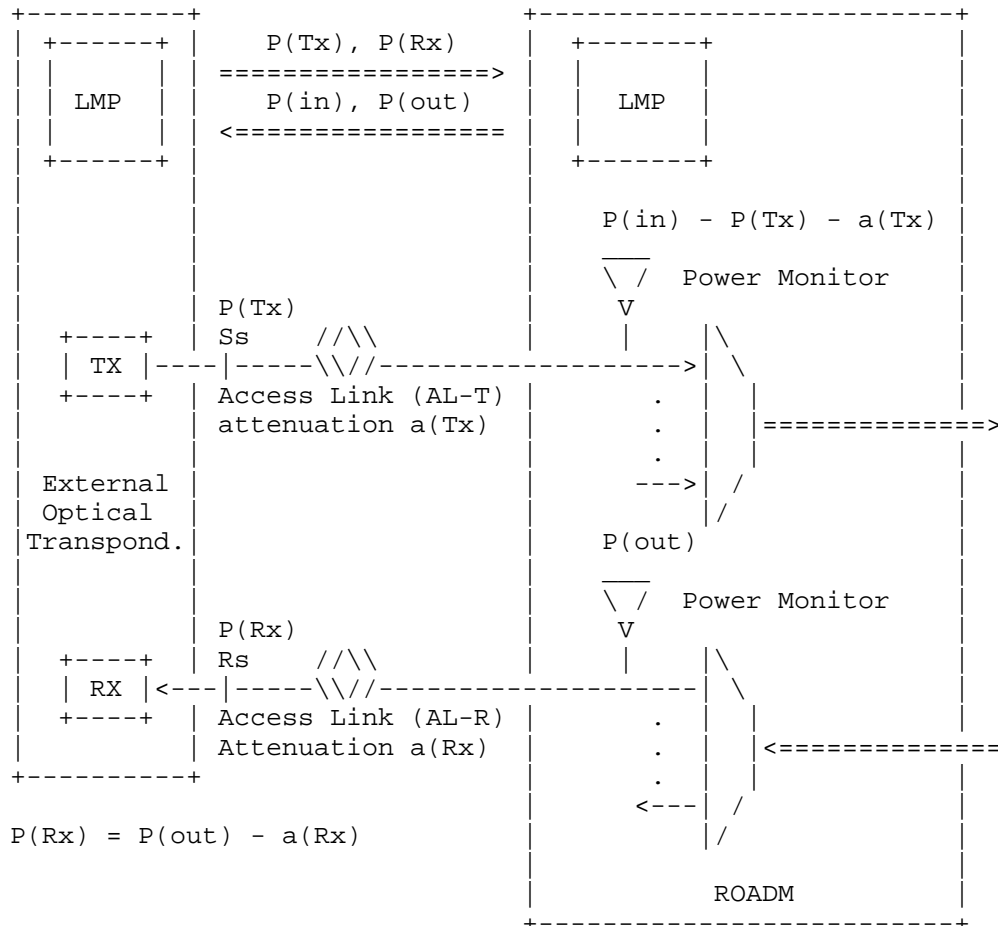
Assumptions:

- The access link attenuation in both directions ($a(Tx)$, $a(Rx)$) is known or can be determined as part of the commissioning process. Typically, both values are the same.
- A threshold value t has been configured by the operator. This should also be done during commissioning.
- A control plane protocol is in place that allows to periodically send the optical power values $P(Tx)$ and $P(Rx)$ to the control plane protocol instance on the DWDM border NE. This is illustrated in Figure 3.
- The DWDM border NE is capable to periodically measure the optical power P_{in} and P_{out} as defined in G.697 by power monitoring points depicted as yellow triangles in the figures below.

AL monitoring process:

- Tx direction: the measured optical input power P_{in} is compared with the expected optical input power $P(Tx) - a(Tx)$. If the measured optical input power P_{in} drops below the value $(P(Tx) - a(Tx) - t)$ a low power alarm shall be raised indicating that the access link attenuation has exceeded $a(Tx) + t$.
- Rx direction: the measured optical input power $P(Rx)$ is compared with the expected optical input power $P(out) - a(Rx)$. If the measured optical input power $P(Rx)$ drops below the value $(P(out) - a(Rx) - t)$ a low power alarm shall be raised indicating that the access link attenuation has exceeded $a(Rx) + t$.

Figure 3 Use case 1: Access Link power monitoring



- For AL-T monitoring: $P(Tx)$ and $a(Tx)$ must be known
 - For AL-R monitoring: $P(Rx)$ and $a(Rx)$ must be known
- An alarm shall be raised if $P(in)$ or $P(Rx)$ drops below a configured threshold (t [dB]):
- $P(in) < P(Tx) - a(Tx) - t$ (Tx direction)
 - $P(Rx) < P(out) - a(Rx) - t$ (Rx direction)
 - $a(Tx) = a(Rx)$

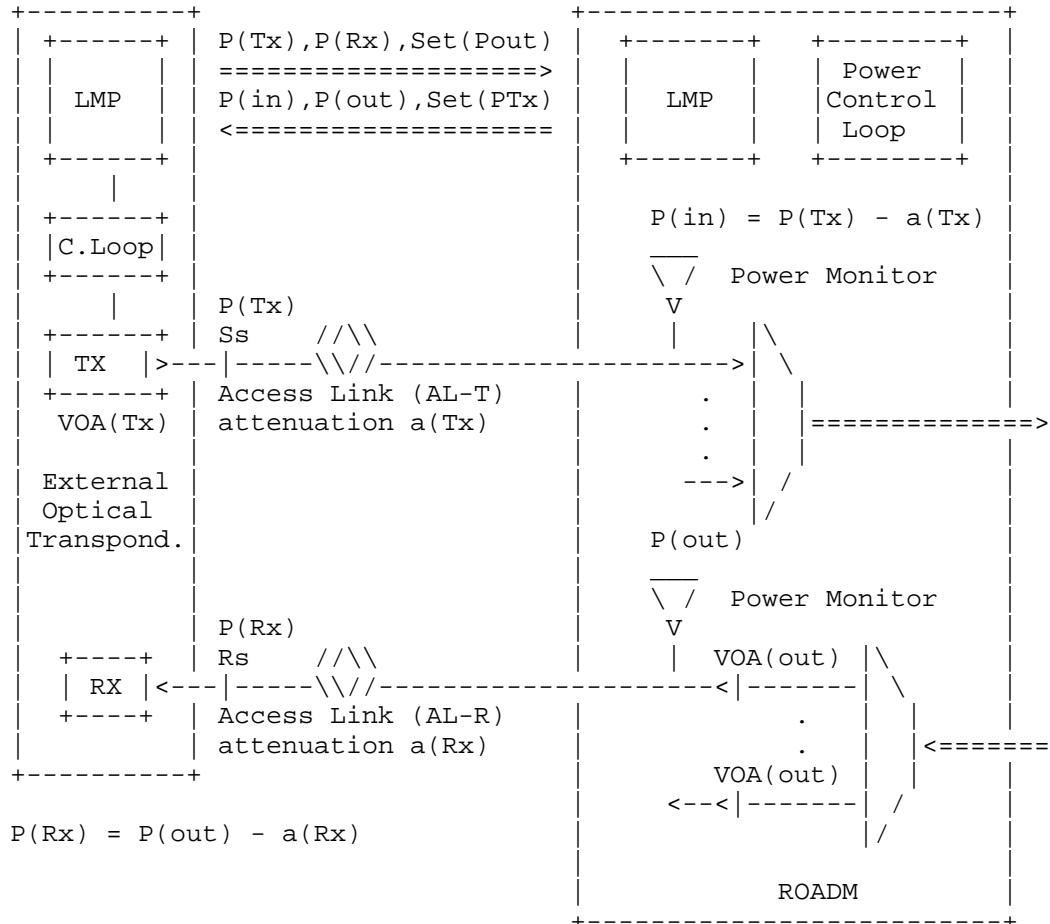
Figure 3: Extended LMP Model

Power Control Loop Use Case

This use case is based on the access link monitoring use case as described above. In addition, the border NE is running a power control application that is capable to control the optical output power of the single channel tributary signal at the output port of the border DWDM NE (towards the external receiver Rx) and the optical output power of the single channel tributary signal at the external transmitter Tx within their known operating range. The time scale of this control loop is typically relatively slow (e.g. some 10s or minutes) because the access link attenuation is not expected to vary much over time (the attenuation only changes when re-cabling occurs).

From a data plane perspective, this use case does not require additional data plane extensions. It does only require a protocol extension in the control plane (e.g. this LMP draft) that allows the power control application residing in the DWDM border NE to modify the optical output power of the DWDM domain-external transmitter Tx within the range of the currently used application code. Figure 5 below illustrates this use case utilizing the LMP protocol with extensions defined in this draft.

Figure 4 Use case 2: Power Control Loop



The Power Control Loops in Transponder and ROADM regulate the Variable Optical Attenuators (VOA) to adjust the proper power in base of the ROADM and Receiver characteristics and the Access Link attenuation

Figure 4: Extended LMP Model

4.2. Optical Parameters Description

The G.698.2 pre-certified network media channels are managed at the edges, i.e. at the transmitters (Tx) and receivers (Rx) attached to the S and R reference points respectively. The set of parameters that could be managed are specified in G.698.2 [ITU.G698.2] section 5.3 referring the "application code" notation

The definitions of the optical parameters are provided below to increase the readability of the document, where the definition is ended by (G) the parameter can be retrieve with a GET, when (S) it can be provisioned by a SET, (G,S) can be either GET and SET.

To support the management of these parameters, the SNMP MIB in RFC 3591 [RFC3591] is extended with a new MIB module defined in section 6 of this document. This new MIB module includes the definition of new configuration table of the OCh Layer for the parameters at Tx (S) and Rx (R).

4.2.1. Rs-Ss Configuration

The Rs-Ss configuration table allows configuration of Central Frequency, Power and Application identifiers as described in [ITU.G698.2] and G.694.1 [ITU.G694.1]

This parameter report the current Transceiver Output power, it can be either a setting and measured value (G, S).

Central frequency (see G.694.1 Table 1):

This parameter indicates the central frequency value that Ss and Rs will be set, to work (in THz), in particular Section 6/G.694.1 (G, S).

Single-channel application identifiers (see G.698.2):

This parameter indicates the transceiver application identifier at Ss and Rs as defined in [ITU.G698.2] Chapter 5.4 - this parameter can be called Optical Interface Identifier OII as per [draft-martinelli-wson-interface-class] (G).

Number of Single-channel application identifiers Supported

This parameter indicates the number of Single-channel application codes supported by this interface (G).

Current Laser Output power:

This parameter report the current Transceiver Output power, see RFC3591.

Current Laser Input power:

This parameter report the current Transceiver Input power see RFC3591.

PARAMETERS	Get/Set	Reference
Central Frequency	G,S	G.694.1 S.6
Single-channel Application Identifier number in use	G	G.874.1
Single-channel Application Identifier Type in use	G	G.874.1
Single-channel Application Identifier in use	G	G.874.1
Number of Single-channel Application Identifiers Supported	G	N.A.
Current Output Power	G,S	RFC3591
Current Input Power	G	RFC3591

Table 1: Rs-Ss Configuration

4.2.2. Table of Application Identifiers

This table has a list of Application Identifiers supported by this interface at point R are defined in G.698.2.

Application Identifier Number:

The number that uniquely identifies the Application Identifier.

Application Identifier Type:

Type of application Identifier: STANDARD / PROPRIETARY in G.874.1

Note: if the A.I. type = PROPRIETARY, the first 6 Octets of the Application Identifier (PrintableString) must contain the Hexadecimal representation of an OUI (organizationally unique identifier) assigned to the vendor whose implementation generated the Application Identifier; the remaining octets of the PrintableString are unspecified.

Application Identifier:

This is the application Identifier that is defined in G.874.1.

4.3. Use of ifTable

This section specifies how the MIB II interfaces group, as defined in RFC 2863 [RFC2863], is used for the link ends of a black link. Only the ifGeneralInformationGroup will be supported for the ifTable and the ifStackTable to maintain the relationship between the OCh and OPS layers. The OCh and OPS layers are managed in the ifTable using IfEntries that correlate to the layers depicted in Figure 1.

For example, a device with TX and/or RX will have an Optical Physical Section (OPS) layer, and an OCh layer. There is a one to n relationship between the OPS and OCh layers.

EDITOR NOTE: Reason for changing from OChr to OCh: Edition 3 of G.872 removed OChr from the architecture and G.709 was subsequently updated to account for this architectural change.

Figure 5 In the following figures, opticalPhysicalSection are abbreviated as OPS.

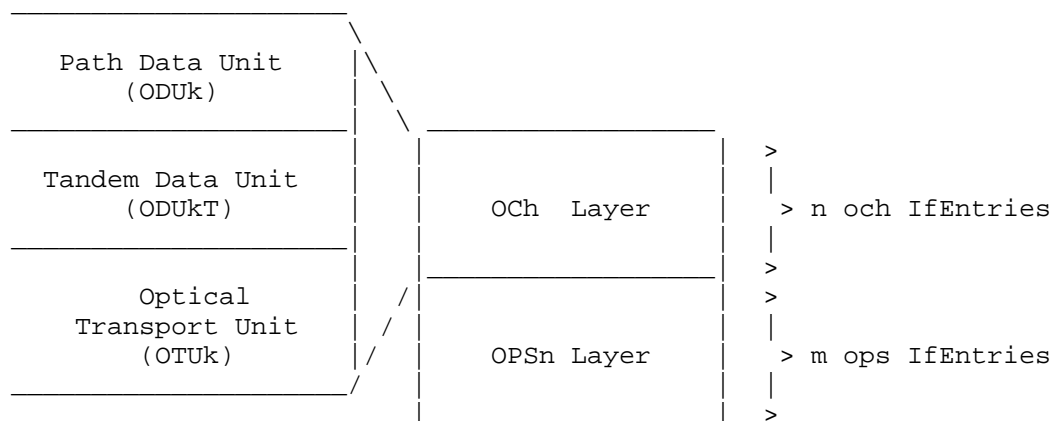


Figure 5: OTN Layers for OPS and OCh

Each opticalChannel IfEntry is mapped to one of the m opticalPhysicalSection IfEntries, where m is greater than or equal to 1. Conversely, each opticalTransPhysicalSection port entry is mapped to one of the n opticalChannel IfEntries, where n is greater than or equal to 1.

The design of the Optical Interface MIB provides the option to model an interface either as a single bidirectional object containing both sink and source functions or as a pair of unidirectional objects, one containing sink functions and the other containing source functions.

If the sink and source for a given protocol layer are to be modelled as separate objects, then there need to be two ifTable entries, one that corresponds to the sink and one that corresponds to the source, where the directionality information is provided in the configuration tables for that layer via the associated Directionality objects. The agent is expected to maintain consistent directionality values between ifStackTable layers (e.g., a sink must not be stacked in a 1:1 manner on top of a source, or vice-versa), and all protocol layers that are represented by a given ifTable entry are expected to have the same directionality.

When separate ifTable entries are used for the source and sink functions of a given physical interface, association between the two uni-directional ifTable entries (one for the source function and the other for the sink functions) should be provided. It is recommended that identical ifName values are used for the two ifTable entries to indicate such association. An implementation shall explicitly state what mechanism is used to indicate the association, if ifName is not used.

4.3.1. Use of ifTable for OPS Layer

Only the ifGeneralInformationGroup needs to be supported.

ifTable Object	Use for OTN OPS Layer
=====	
ifIndex	The interface index.
ifDescr	Optical Transport Network (OTN) Optical Physical Section (OPS)
ifType	opticalPhysicalSection (xxx)
<<<Editor Note: Need new IANA registration value for xxx. >>>	
ifSpeed	Actual bandwidth of the interface in bits per second. If the bandwidth of the interface is greater than the maximum value of 4,294,967,295 then the maximum value is reported and ifHighSpeed must be used to report the interface's speed.

ifPhysAddress	An octet string with zero length. (There is no specific address associated with the interface.)
ifAdminStatus	The desired administrative state of the interface. Supports read-only access.
ifOperStatus	The operational state of the interface. The value lowerLayerDown(7) is not used, since there is no lower layer interface. This object is set to notPresent(6) if a component is missing, otherwise it is set to down(2) if either of the objects optIfOPSnCurrentStatus indicates that any defect is present.
ifLastChange	The value of sysUpTime at the last change in ifOperStatus.
ifName	Enterprise-specific convention (e.g., TL-1 AID) to identify the physical or data entity associated with this interface or an OCTET STRING of zero length. The enterprise-specific convention is intended to provide the means to reference one or more enterprise-specific tables.
ifLinkUpDownTrapEnable	Default value is enabled(1). Supports read-only access.
ifHighSpeed	Actual bandwidth of the interface in Mega-bits per second. A value of n represents a range of 'n-0.5' to 'n+0.499999'.
ifConnectorPresent	Set to true(1).
ifAlias	The (non-volatile) alias name for this interface as assigned by the network manager.

4.3.2. Use of ifTable for OCh Layer

Use of ifTable for OCh Layer See RFC 3591 [RFC3591] section 2.4

4.3.3. Use of ifStackTable

Use of the ifStackTable and ifInvStackTable to associate the opticalPhysicalSection and opticalChannel interface entries is best illustrated by the example shown in Figure 3. The example assumes an

ops interface with ifIndex i that carries two multiplexed OCh interfaces with ifIndex values of j and k, respectively. The example shows that j and k are stacked above (i.e., multiplexed into) i. Furthermore, it shows that there is no layer lower than i and no layer higher than j and/or k.

Figure 6

HigherLayer	LowerLayer

0	j
0	k
j	i
k	i
i	0

Figure 6: Use of ifStackTable for an OTN port

For the inverse stack table, it provides the same information as the interface stack table, with the order of the Higher and Lower layer interfaces reversed.

5. Structure of the MIB Module

EDITOR NOTE: text will be provided based on the MIB module in Section 6

6. Object Definitions

EDITOR NOTE: Once the scope in Section 1 and the parameters in Section 4 are finalized, a MIB module will be defined. It could be an extension to the OPT-IF-MIB module of RFC 3591. >>>

```
OPT-IF-698-MIB DEFINITIONS ::= BEGIN
```

```
IMPORTS
```

```
    MODULE-IDENTITY,
    OBJECT-TYPE,
    Gauge32,
    Integer32,
    Unsigned32,
    Counter64,
    transmission,
    NOTIFICATION-TYPE
        FROM SNMPv2-SMI
    TEXTUAL-CONVENTION,
    RowPointer,
    RowStatus,
    TruthValue,
    DisplayString,
    DateAndTime
        FROM SNMPv2-TC
    SnmpAdminString
        FROM SNMP-FRAMEWORK-MIB
    MODULE-COMPLIANCE, OBJECT-GROUP
        FROM SNMPv2-CONF
    ifIndex
        FROM IF-MIB
    optIfMibModule
        FROM OPT-IF-MIB;
```

```
-- This is the MIB module for the optical parameters -
-- Application codes associated with the black link end points.
```

```
optIfXcvrMibModule MODULE-IDENTITY
    LAST-UPDATED "201401270000Z"
    ORGANIZATION "IETF Ops/Camp MIB Working Group"
    CONTACT-INFO
        "WG charter:
         http://www.ietf.org/html.charters/

        Mailing Lists:
        Editor: Gabriele Galimberti
        Email: ggalimbe@cisco.com"
    DESCRIPTION
        "The MIB module to describe Black Link tranceiver
        characteristics to rfc3591.
```

Copyright (C) The Internet Society (2014). This version of this MIB module is an extension to rfc3591; see the RFC itself for full legal notices."

REVISION "201305050000Z"

DESCRIPTION

"Draft version 1.0"

REVISION "201305050000Z"

DESCRIPTION

"Draft version 2.0"

REVISION "201302270000Z"

DESCRIPTION

"Draft version 3.0"

REVISION "201307020000Z"

DESCRIPTION

"Draft version 4.0"

Changed the draft to include only the G.698 parameters."

REVISION "201311020000Z"

DESCRIPTION

"Draft version 5.0"

Mib has a table of application code/vendor transceivercode G.698"

REVISION "201401270000Z"

DESCRIPTION

"Draft version 6.0"

REVISION "201407220000Z"

DESCRIPTION

"Draft version 8.0"

Removed Vendor transceiver code"

REVISION "201502220000Z"

DESCRIPTION

"Draft version 11.0"

Added reference to OUI in the first 6 Octets of a proprietary Application code

Added a Length field for the Application code

Changed some names"

REVISION "201507060000Z"

DESCRIPTION

"Draft version 12.0"

Added Power Measurement Use Cases

and ITU description" "

::= { optIfMibModule 4 }

::= { optIfMibModule 4 }

-- Addition to the RFC 3591 objects

optIfOChSsRsGroup OBJECT IDENTIFIER ::= { optIfXcvrMibModule 1 }

```
-- OCh Ss/Rs config table
-- The application code/vendor transceiver class for the Black Link
-- Ss-Rs will be added to the OchConfigTable
```

```
optIfOchSsRsConfigTable OBJECT-TYPE
    SYNTAX SEQUENCE OF OptIfOchSsRsConfigEntry
    MAX-ACCESS not-accessible
    STATUS current
    DESCRIPTION
        "A table of Och General config extension parameters"
    ::= { optIfOchSsRsGroup 1 }
```

```
optIfOchSsRsConfigEntry OBJECT-TYPE
    SYNTAX OptIfOchSsRsConfigEntry
    MAX-ACCESS not-accessible
    STATUS current
    DESCRIPTION
        "A conceptual row that contains G.698 parameters for an
        interface."
    INDEX { ifIndex }
    ::= { optIfOchSsRsConfigTable 1 }
```

```
OptIfOchSsRsConfigEntry ::=
    SEQUENCE {
        optIfOchCentralFrequency                Unsigned32,
        optIfOchCfgApplicationIdentifierNumber   Unsigned32,
        optIfOchCfgApplicationIdentifierType     Unsigned32,
        optIfOchCfgApplicationIdentifierLength   Unsigned32,
        optIfOchCfgApplicationIdentifier         DisplayString,
        optIfOchNumberApplicationCodesSupported Unsigned32
    }
```

```
optIfOchCentralFrequency OBJECT-TYPE
    SYNTAX Unsigned32
    MAX-ACCESS read-write
    UNITS "THz"
    STATUS current
    DESCRIPTION
        " This parameter indicates the frequency of this interface.
        "
    ::= { optIfOchSsRsConfigEntry 1 }
```

```
optIfOchCfgApplicationIdentifierNumber OBJECT-TYPE
    SYNTAX Unsigned32
    MAX-ACCESS read-write
    STATUS current
    DESCRIPTION
        "This parameter uniquely indicates the transceiver
```

```
        application code at Ss and Rs as defined in [ITU.G874.1],
        that is used by this interface.
        The optIfOChSrcApplicationIdentifierTable has all the
        application codes supported by this interface. "
 ::= { optIfOChSsRsConfigEntry 2 }

optIfOChCfgApplicationIdentifierType OBJECT-TYPE
    SYNTAX Unsigned32
    MAX-ACCESS read-write
    STATUS current
    DESCRIPTION
        "This parameter indicates the transceiver type of
        application code at Ss and Rs as defined in [ITU.G874.1],
        that is used by this interface.
        The optIfOChSrcApplicationIdentifierTable has all the
        application codes supported by this interface
        Standard = 0, PROPRIETARY = 1. "
 ::= { optIfOChSsRsConfigEntry 3 }

optIfOChCfgApplicationIdentifierLenght OBJECT-TYPE
    SYNTAX Unsigned32
    MAX-ACCESS read-write
    STATUS current
    DESCRIPTION
        "This parameter indicates the number of octets in the
        Application Identifier.
        "
 ::= { optIfOChSsRsConfigEntry 4 }

optIfOChCfgApplicationIdentifier OBJECT-TYPE
    SYNTAX DisplayString
    MAX-ACCESS read-write
    STATUS current
    DESCRIPTION
        "This parameter indicates the transceiver application code
        at Ss and Rs as defined in [ITU.G698.2] Chapter 5.3, that
        is used by this interface. The
        optIfOChSrcApplicationCodeTable has all the application
        codes supported by this interface.
        If the optIfOChCfgApplicationIdentifierType is 1
        (Proprietary), then the first 6 octets of the printable
        string will be the OUI (organizationally unique identifier)
        assigned to the vendor whose implementation generated the
        Application Identifier."
 ::= { optIfOChSsRsConfigEntry 5 }

optIfOChNumberApplicationIdentifiersSupported OBJECT-TYPE
```



```

SYNTAX Unsigned32
MAX-ACCESS read-only
STATUS current
DESCRIPTION
    " Number of Application codes supported by this interface."
 ::= { optIfOChSsRsConfigEntry 6 }

-- Table of Application codes supported by the interface
-- OptIfOChSrcApplicationCodeEntry

optIfOChSrcApplicationIdentifierTable OBJECT-TYPE
    SYNTAX SEQUENCE OF OptIfOChSrcApplicationIdentifierEntry
    MAX-ACCESS not-accessible
    STATUS current
    DESCRIPTION
        "A Table of Application codes supported by this interface."
    ::= { optIfOChSsRsGroup 2 }

optIfOChSrcApplicationIdentifierEntry OBJECT-TYPE
    SYNTAX OptIfOChSrcApplicationIdentifierEntry
    MAX-ACCESS not-accessible
    STATUS current
    DESCRIPTION
        "A conceptual row that contains the Application code for
         this interface."
    INDEX { ifIndex, optIfOChApplicationIdentifierNumber }
    ::= { optIfOChSrcApplicationIdentifierTable 1 }

OptIfOChSrcApplicationIdentifierEntry ::=
    SEQUENCE {
        optIfOChApplicationIdentifierNumber      Integer32,
        optIfOChApplicationIdentifierType         Integer32,
        optIfOChApplicationIdentifierLength       Integer32,
        optIfOChApplicationIdentifier             DisplayString
    }

optIfOChApplicationIdentifierNumber OBJECT-TYPE
    SYNTAX Integer32 (1..255)
    MAX-ACCESS not-accessible
    STATUS current
    DESCRIPTION
        " The number/identifier of the application code supported at
         this interface. The interface can support more than one
         application codes.
        "
    ::= { optIfOChSrcApplicationIdentifierEntry 1}

```

```
optIfOChApplicationIdentifierType OBJECT-TYPE
    SYNTAX Integer32 (1..255)
    MAX-ACCESS read-only
    STATUS current
    DESCRIPTION
        " The type of identifier of the application code supported at
          this interface. The interface can support more than one
          application codes.
          Standard = 0, PROPRIETARY = 1
        "
    ::= { optIfOChSrcApplicationIdentifierEntry 2}

optIfOChApplicationIdentifierLength OBJECT-TYPE
    SYNTAX Integer32 (1..255)
    MAX-ACCESS read-only
    STATUS current
    DESCRIPTION
        " This parameter indicates the number of octets in the
          Application Identifier.
        "
    ::= { optIfOChSrcApplicationIdentifierEntry 3}

optIfOChApplicationIdentifier OBJECT-TYPE
    SYNTAX DisplayString
    MAX-ACCESS read-only
    STATUS current
    DESCRIPTION
        " The application code supported by this interface DWDM
          link.
          If the optIfOChApplicationIdentifierType is 1 (Proprietary),
          then the first 6 octets of the printable string will be
          the OUI (organizationally unique identifier) assigned to
          the vendor whose implementation generated the Application
          Identifier."
    ::= { optIfOChSrcApplicationIdentifierEntry 4}

-- Notifications

-- Central Frequency Change Notification
optIfOChCentralFrequencyChange NOTIFICATION-TYPE
    OBJECTS { optIfOChCentralFrequency }
    STATUS current
    DESCRIPTION
        "Notification of a change in the central frequency."
```

```
::= { optIfXcvrMibModule 1 }
```

END

7. Relationship to Other MIB Modules

7.1. Relationship to the [TEMPLATE TODO] MIB

7.2. MIB modules required for IMPORTS

8. Definitions

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

9. Security Considerations

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

o

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

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

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

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

10. IANA Considerations

Option #1:

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

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

Option #2:

Editor's Note (to be removed prior to publication): the IANA is requested to assign a value for "XXX" under the 'mib-2' subtree and to record the assignment in the SMI Numbers registry. When the assignment has been made, the RFC Editor is asked to replace "XXX" (here and in the MIB module) with the assigned value and to remove this note.

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

Option #3:

This memo includes no request to IANA.

11. Contributors

Arnold Mattheus
Deutsche Telekom
Darmstadt
Germany
email a.mattheus@telekom.de

Manuel Paul
Deutsche Telekom
Berlin
Germany
email Manuel.Paul@telekom.de

Frank Luennemann
Deutsche Telekom
Munster
Germany
email Frank.Luennemann@telekom.de

Scott Mansfield
Ericsson Inc.
email scott.mansfield@ericsson.com

Najam Saquib
Cisco
Ludwig-Erhard-Strasse 3
ESCHBORN, HESSEN 65760
GERMANY
email nasaquib@cisco.com

Walid Wakim
Cisco
9501 Technology Blvd
ROSEMONT, ILLINOIS 60018
UNITED STATES
email wwakim@cisco.com

Ori Gerstel
Sedona System
ISRAEL
email orig@sedonasys.com

12. References

12.1. Normative References

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

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC2578] McCloghrie, K., Ed., Perkins, D., Ed., and J. Schoenwaelder, Ed., "Structure of Management Information Version 2 (SMIv2)", STD 58, RFC 2578, April 1999.
- [RFC2579] McCloghrie, K., Ed., Perkins, D., Ed., and J. Schoenwaelder, Ed., "Textual Conventions for SMIv2", STD 58, RFC 2579, April 1999.
- [RFC2580] McCloghrie, K., Perkins, D., and J. Schoenwaelder, "Conformance Statements for SMIv2", STD 58, RFC 2580, April 1999.
- [RFC3591] Lam, H-K., Stewart, M., and A. Huynh, "Definitions of Managed Objects for the Optical Interface Type", RFC 3591, September 2003.
- [RFC6205] Otani, T. and D. Li, "Generalized Labels for Lambda-Switch-Capable (LSC) Label Switching Routers", RFC 6205, March 2011.
- [ITU.G698.2] International Telecommunications Union, "Amplified multichannel dense wavelength division multiplexing applications with single channel optical interfaces", ITU-T Recommendation G.698.2, November 2009.
- [ITU.G709] International Telecommunications Union, "Interface for the Optical Transport Network (OTN)", ITU-T Recommendation G.709, February 2012.
- [ITU.G872] International Telecommunications Union, "Architecture of optical transport networks", ITU-T Recommendation G.872 and Amd.1, October 2012.
- [ITU.G798] International Telecommunications Union, "Characteristics of optical transport network hierarchy equipment functional blocks", ITU-T Recommendation G.798 and Amd.1, December 2012.

- [ITU.G874]
International Telecommunications Union, "Management aspects of optical transport network elements", ITU-T Recommendation G.874, August 2013.
- [ITU.G874.1]
International Telecommunications Union, "Optical transport network (OTN): Protocol-neutral management information model for the network element view", ITU-T Recommendation G.874.1, October 2012.
- [ITU.G959.1]
International Telecommunications Union, "Optical transport network physical layer interfaces", ITU-T Recommendation G.959.1, November 2009.
- [ITU.G826]
International Telecommunications Union, "End-to-end error performance parameters and objectives for international, constant bit-rate digital paths and connections", ITU-T Recommendation G.826, November 2009.
- [ITU.G8201]
International Telecommunications Union, "Error performance parameters and objectives for multi-operator international paths within the Optical Transport Network (OTN)", ITU-T Recommendation G.8201, April 2011.
- [ITU.G694.1]
International Telecommunications Union, "Spectral grids for WDM applications: DWDM frequency grid", ITU-T Recommendation G.694.1, February 2012.
- [ITU.G7710]
International Telecommunications Union, "Common equipment management function requirements", ITU-T Recommendation G.7710, February 2012.

12.2. Informative References

- [RFC3410] Case, J., Mundy, R., Partain, D., and B. Stewart, "Introduction and Applicability Statements for Internet-Standard Management Framework", RFC 3410, December 2002.
- [RFC2629] Rose, M., "Writing I-Ds and RFCs using XML", RFC 2629, June 1999.

[RFC4181] Heard, C., "Guidelines for Authors and Reviewers of MIB Documents", BCP 111, RFC 4181, September 2005.

[I-D.kunze-g-698-2-management-control-framework]
Kunze, R., "A framework for Management and Control of optical interfaces supporting G.698.2", draft-kunze-g-698-2-management-control-framework-00 (work in progress), July 2011.

[RFC4054] Strand, J. and A. Chiu, "Impairments and Other Constraints on Optical Layer Routing", RFC 4054, May 2005.

Appendix A. Change Log

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

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

Appendix B. Open Issues

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

Authors' Addresses

Gabriele Galimberti (editor)
Cisco
Via Santa Maria Molgora, 48 c
20871 - Vimercate
Italy

Phone: +390392091462
Email: ggalimbe@cisco.com

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

Phone: +49xxxxxxxxxxx
Email: RKunze@telekom.de

Hing-Kam Lam (editor)
Alcatel-Lucent
600-700 Mountain Avenue, Murray Hill
New Jersey, 07974
USA

Phone: +17323313476
Email: kam.lam@alcatel-lucent.com

Dharini Hiremagalur (editor)
Juniper
1194 N Mathilda Avenue
Sunnyvale - 94089 California
USA

Phone: +1408
Email: dharinih@juniper.net

Luyuan Fang (editor)
Microsoft
5600 148th Ave NE
Redmond, WA 98502
USA

Email: lufang@microsoft.com

Gary Ratterree (editor)
Microsoft
5600 148th Ave NE
Redmond, WA 98502
USA

Email: gratt@microsoft.com

CCAMP Working Group
Internet-Draft
Intended status: Informational
Expires: October 25, 2014

Rakesh Gandhi, Ed.
Zafar Ali
Gabriele Maria Galimberti
Cisco Systems, Inc.
Xian Zhang
Huawei
April 23, 2014

RSVP-TE Signaling For GMPLS Restoration LSP
draft-gandhi-ccamp-gmpls-restoration-lsp-04

Abstract

In transport networks, there are requirements where Generalized Multi-Protocol Label Switching (GMPLS) end-to-end recovery scheme needs to employ restoration Label Switched Path (LSP) while keeping resources for the working and/or protecting LSPs reserved in the network after the failure.

This document reviews how the LSP association is to be provided using Resource Reservation Protocol - Traffic Engineering (RSVP-TE) signaling in the context of GMPLS end-to-end recovery when using restoration LSP where failed LSP is not torn down. No new procedures or mechanisms are defined by this document, and it is strictly informative in nature.

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."

Copyright Notice

Copyright (c) 2014 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. Signaling Restoration LSP Association	5
3. IANA Considerations	5
4. Security Considerations	5
5. Acknowledgement	5
6. References	6
6.1. Normative References	6
6.2. Informative References	6
Authors' Addresses	7

1. Introduction

Generalized Multi-Protocol Label Switching (GMPLS) [RFC3473] extends Multi-Protocol Label Switching (MPLS) to include support for different switching technologies. These switching technologies provide several protection schemes [RFC4426][RFC4427] (e.g., 1+1, 1:N and M:N). Resource Reservation Protocol - Traffic Engineering (RSVP-TE) signaling has been extended to support various GMPLS recovery schemes [RFC4872][RFC4873], to establish Label Switched Paths (LSPs), typically for working LSP and protecting LSP. [RFC4427] Section 7 specifies various schemes for GMPLS recovery.

In GMPLS recovery schemes generally considered, restoration LSP is signaled after the failure has been detected and notified on the working LSP. In non-revertive recovery mode, working LSP is assumed to be removed from the network before restoration LSP is signaled. For revertive recovery mode, a restoration LSP is signaled while working LSP and/or protecting LSP are not torn down in control plane due to a failure. In transport networks, as working LSPs are typically signaled over a nominal path, service providers would like to keep resources associated with the working LSPs reserved. This is to make sure that the service (working LSP) can use the nominal path when the failure is repaired to provide deterministic behaviour and guaranteed Service Level Agreement (SLA). Consequently, revertive recovery mode is usually preferred by recovery schemes used in transport networks.

As defined in [RFC4872] and being considered in this document, "fully dynamic rerouting switches normal traffic to an alternate LSP that is not even partially established only after the working LSP failure occurs. The new alternate route is selected at the LSP head-end node, it may reuse resources of the failed LSP at intermediate nodes and may include additional intermediate nodes and/or links."

One example of the recovery scheme considered in this document is 1+R recovery. The 1+R recovery is exemplified in Figure 1. In this example, working LSP on path A-B-C-Z is pre-established. Typically after a failure detection and notification on the working LSP, a second LSP on path A-H-I-J-Z is established as a restoration LSP. Unlike protection LSP, restoration LSP is signaled per need basis.

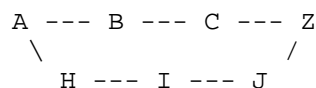


Figure 1: An example of 1+R recovery scheme

During failure switchover with 1+R recovery scheme, in general, working LSP resources are not released and working and restoration LSPs coexist in the network. Nonetheless, working and restoration LSPs can share network resources. Typically when failure is recovered on the working LSP, restoration LSP is no longer required and torn down (e.g., revertive mode).

Another example of the recovery scheme considered in this document is 1+1+R. In 1+1+R, a restoration LSP is signaled for the working LSP and/or the protecting LSP after the failure has been detected and notified on the working LSP or the protecting LSP. The 1+1+R recovery is exemplified in Figure 2. In this example, working LSP on path A-B-C-Z and protecting LSP on path A-D-E-F-Z are pre-established. After a failure detection and notification on a working LSP or protecting LSP, a third LSP on path A-H-I-J-Z is established as a restoration LSP. The restoration LSP in this case provides protection against a second order failure. Restoration LSP is torn down when the failure on the working or protecting LSP is repaired.

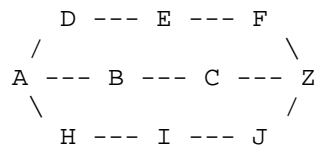


Figure 2: An example of 1+1+R recovery scheme

[RFC4872] Section 14 defines PROTECTION object for GMPLS recovery signaling. As defined, the PROTECTION object is used to identify primary and secondary LSPs using S bit and protecting and working LSPs using P bit. Furthermore, [RFC4872] defines the usage of ASSOCIATION object for associating GMPLS working and protecting LSPs.

[RFC6689] Section 2.2 reviews the procedure for providing LSP associations for GMPLS end-to-end recovery and covers the schemes where the failed working LSP and/or protecting LSP are torn down.

This document reviews how the LSP association is to be provided for GMPLS end-to-end recovery when using restoration LSP where working and protecting LSP resources are kept reserved in the network after the failure.

2. Signaling Restoration LSP Association

Where GMPLS end-to-end recovery scheme needs to employ restoration LSP while keeping resources for the working and/or protecting LSPs reserved in the network after the failure, restoration LSP is signaled with ASSOCIATION object with the association ID set to the LSP ID of the LSP it is restoring. For example, when a restoration LSP is signaled for a working LSP, the ASSOCIATION object in the restoration LSP contains the association ID set to the LSP ID of the working LSP. Similarly, when a restoration LSP is signaled for a protecting LSP, the ASSOCIATION object in the restoration LSP contains the association ID set to the LSP ID of the protecting LSP.

The procedure for signaling the PROTECTION object is specified in [RFC4872]. Specifically, restoration LSP being used as a working LSP is signaled with P bit cleared and being used as a protecting LSP is signaled with P bit set.

As discussed in Section 1 of this document, [RFC6689] Section 2.2 reviews the procedure for providing LSP associations for the GMPLS end-to-end recovery scheme using restoration LSP where the failed working LSP and/or protecting LSP are torn down.

3. IANA Considerations

This document makes no request for IANA action.

4. Security Considerations

This document reviews procedures defined in [RFC4872] and [RFC6689] and does not define any new procedure. As such, no new security considerations are introduced in this document.

5. Acknowledgement

The authors would like to thank George Swallow for the discussions on the GMPLS restoration.

6. References

6.1. Normative References

- [RFC3473] Berger, L., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions", RFC 3473, January 2003.
- [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.
- [RFC6689] Berger, L., "Usage of the RSVP ASSOCIATION Object", RFC 6689, July 2012.

6.2. Informative References

- [RFC4426] Lang, J., Rajagopalan, B., and Papadimitriou, D., "Generalized Multiprotocol Label Switching (GMPLS) Recovery Functional Specification", RFC 4426, March 2006.
- [RFC4427] Mannie, E., and Papadimitriou, D., "Recovery (Protection and Restoration) Terminology for Generalized Multi-Protocol Label Switching", RFC 4427, March 2006.
- [RFC4873] Berger, L., Bryskin, I., Papadimitriou, D., and Farrel, A., "GMPLS Segment Recovery", RFC 4873, May 2007.

Authors' Addresses

Rakesh Gandhi (editor)
Cisco Systems, Inc.

Email: rgandhi@cisco.com

Zafar Ali
Cisco Systems, Inc.

Email: zali@cisco.com

Gabriele Maria Galimberti
Cisco Systems, Inc.

Email: ggalimbe@cisco.com

Xian Zhang
Huawei Technologies
Research Area F3-1B,
Huawei Industrial Base,
Shenzhen, 518129, China

Email: zhang.xian@huawei.com

CCAMP Working Group
Internet Draft
Intended status: Standard Track
Expires: April 26, 2015

Zafar Ali, Ed.
George Swallow, Ed.
Cisco Systems
F. Zhang, Ed.
Huawei
D. Beller, Ed.
Alcatel-Lucent
October 27, 2014

Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Path
Diversity using Exclude Route

draft-ietf-ccamp-lsp-diversity-05.txt

Status of this Memo

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

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). 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 26, 2015.

Copyright Notice

Copyright (c) 2014 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

RFC 4874 specifies methods by which path exclusions can be communicated during RSVP-TE signaling in networks where precise explicit paths are not computed by the LSP source node. This document specifies procedures for additional route exclusion subobject based on Paths currently existing or expected to exist within the network.

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
1.1. Client-Initiated Identifier	5
1.2. PCE-allocated Identifier	6
1.3. Network-Assigned Identifier	7
2. RSVP-TE signaling extensions	9
2.1. Diversity XRO Subobject	9
2.1.1. IPv4 Diversity XRO Subobject	9
2.1.2. IPv6 Diversity XRO Subobject	14
2.2. Processing rules for the Diversity XRO subobject	17
2.3. Diversity EXRS Subobject	20
3. Security Considerations	22
4. IANA Considerations	22
4.1. New XRO subobject types	22
4.2. New EXRS subobject types	23
4.3. New RSVP error sub-codes	23
5. Acknowledgements	23
6. References	24
6.1. Normative References	24
6.2. Informative References	24

1. Introduction

Path diversity for multiple connections is a well-known Service Provider requirement. Diversity constraints ensure that Label-Switched Paths (LSPs) can be established without sharing resources, thus greatly reducing the probability of simultaneous connection failures.

When a source node has full topological knowledge and is permitted to signal an Explicit Route Object, diverse paths for LSPs can be computed by this source node. However, there are scenarios when

path computations are performed by different nodes, and there is therefore a need for relevant diversity constraints to be communicated to those nodes. These include (but are not limited to):

- . LSPs with loose hops in the Explicit Route Object (ERO), e.g. inter-domain LSPs;
- . Generalized Multi-Protocol Label Switching (GMPLS) User-Network Interface (UNI), where path computation may be performed by the core node [RFC4208].

[RFC4874] introduced a means of specifying nodes and resources to be excluded from a route, using the eXclude Route Object (XRO) and Explicit Exclusion Route Subobject (EXRS). It facilitates the calculation of diverse paths for LSPs based on known properties of those paths including addresses of links and nodes traversed, and Shared Risk Link Groups (SRLGs) of traversed links. Employing these mechanisms requires that the source node that initiates signaling knows the relevant properties of the path(s) from which diversity is desired. However, there are circumstances under which this may not be possible or desirable, including (but not limited to):

- . Exclusion of a path which does not originate, terminate or traverse the source node of the diverse LSP, in which case the addresses of links and SRLGs of the path from which diversity is required are unknown to the source node.
- . Exclusion of a path which is known to the source node of the diverse LSP for which the node has incomplete or no path information, e.g. due to operator policy. In this case, the existence of the reference path is known to the source node but the information required to construct an XRO object to guarantee diversity from the reference path is not fully known. Inter-domain and GMPLS overlay networks can present such restrictions.

This is exemplified in the Figure 1, where overlay reference model from [RFC4208] is shown.

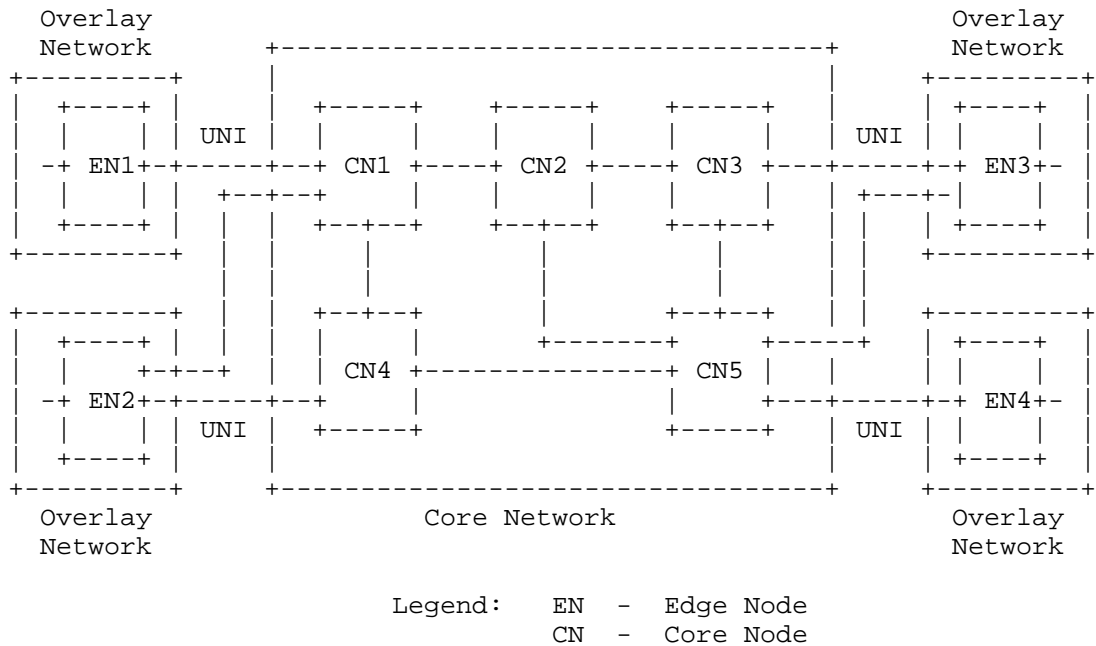


Figure 1: Overlay Reference Model [RFC4208]

Figure 1 depicts two types of UNI connectivity: single-homed and dual-homed ENs (which also applies to higher order multi-homed connectivity.). Single-homed EN devices are connected to a single CN device via a single UNI link. This single UNI link may constitute a single point of failure. UNI connection between EN1 and CN1 is an example of single-homed UNI connectivity.

A single point of failure caused by a single-homed UNI can be avoided when the EN device is connected to two different CN devices, as depicted for EN2 in Figure 1. For the dual-homing case, it is possible to establish two different UNI connections from the same source EN device to the same destination EN device. For example, two connections from EN2 to EN3 may use the two UNI links EN2-CN1 and EN2-CN4. To avoid single points of failure within the provider network, it is necessary to also ensure path (LSP) diversity within the core network.

In a UNI network such as that shown in Figure 1, the CNs typically perform path computation. Information sharing across

the UNI boundary is restricted based on the policy rules imposed by the core network. Typically, the core network topology information is not exposed to the ENs. In the network shown in Figure 1, consider a use case where an LSP from EN2 to EN4 needs to be SRLG diverse from an LSP from EN1 to EN3. In this case, EN2 may not know SRLG attributes of the EN1- EN3 LSP and hence cannot construct an XRO to exclude these SRLGs. In this example EN2 cannot use the procedures described in [RFC4874]. Similarly, an LSP from EN2 to EN3 traversing CN1 needs to be diverse from an LSP from EN2 to EN3 going via CN4. Again in this case, exclusions based on [RFC4874] cannot be used.

This document addresses these diversity requirements by introducing the notion of excluding the path taken by particular LSP(s). The reference LSP(s) or route(s) from which diversity is required is/are identified by an "identifier". The type of identifier to use is highly dependent on the networking deployment scenario; it could be client-initiated, allocated by the (core) network or managed by a PCE. This document defines three different types of identifiers corresponding to these three cases: a client initiated identifier, a PCE allocated Identifier and CN ingress node (UNI-N) allocated Identifier.

1.1. Client-Initiated Identifier

There are scenarios in which the ENs have the following requirements for the diversity identifier:

- The identifier is controlled by the client side and is specified as part of the service request.
- Both client and server understand the identifier.
- It is necessary to be able to reference the identifier even if the LSP referenced by it is not yet signaled.
- The identifier is to be stable for a long period of time.
- The identifier is to be stable even when the referenced tunnel is rerouted.
- The identifier is to be human-readable.

These requirements are met by using the Resource ReserVation Protocol (RSVP) tunnel/ LSP Forwarding Equivalence Class (FEC) as the identifier.

The usage of the client-initiated identifier is illustrated by using Figure 1. Suppose a tunnel from EN2 to EN4 needs to be diverse with respect to a tunnel from EN1 to EN3. The tunnel FEC of the EN1-EN3 tunnel is FEC1, where FEC1 is defined by the tuple (tunnel-id = T1, source address = EN1.ROUTE Identifier (RID), destination address = EN3.RID, extended tunnel-id = EN1.RID). Similarly, tunnel FEC of the EN2-EN3 tunnel is FEC2, where FEC2 is defined by the tuple (tunnel-id = T2, source address = EN2.RID, destination address = EN4.RID, extended tunnel-id = EN2.RID). The EN1-EN3 tunnel is signaled with an exclusion requirement from FEC2, and the EN2-EN3 tunnel is signaled with an exclusion requirement from FEC1. In order to maintain diversity between these two connections within the core network, it is assumed that the core network implements Crankback Signaling [RFC4920]. Note that crankback signaling is known to lead to slower setup times and sub-optimal paths under some circumstances as described by [RFC4920].

1.2. PCE-allocated Identifier

In scenarios where a PCE is deployed and used to perform path computation, the core edge node (e.g., node CN1 in Figure 1) could consult a PCE to allocate identifiers, which are used to signal path diversity constraints. In other scenarios a PCE is deployed in each border node or a PCE is part of a Network Management System (NMS). In all these cases, the Path Key as defined in [RFC5520] can be used in RSVP signaling as the identifier to ensure diversity.

An example of specifying LSP diversity using a Path Key is shown in Figure 2, where a simple network with two domains is shown. It is desired to set up a pair of path-disjoint LSPs from the source in Domain 1 to the destination in Domain 2, but the domains keep strict confidentiality about all path and topology information.

The first LSP is signaled by the source with ERO {A, B, loose Dst} and is set up with the path {Src, A, B, U, V, W, Dst}. However, when sending the RRO out of Domain 2, node U would normally strip the path and replace it with a loose hop to the destination. With this limited information, the source is unable to include enough detail in the ERO of the second LSP to avoid it taking, for example, the path {Src, C, D, X, V, W, Dst} for path-disjointness.

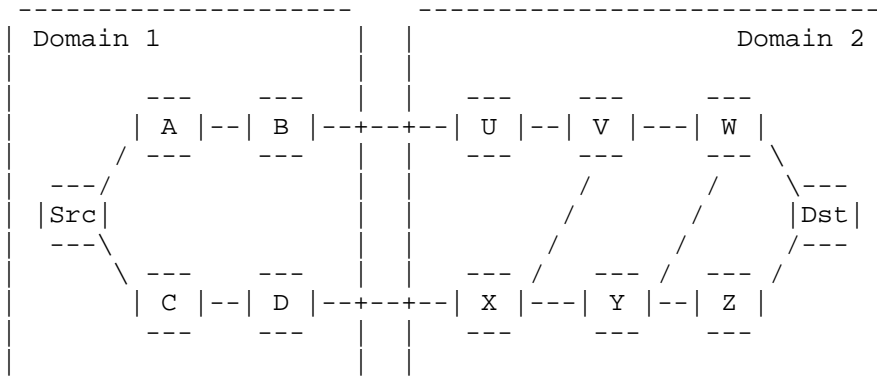


Figure 1: A Simple Multi-Domain Network

In order to improve the situation, node U performs the PCE function and replaces the path segment {U, V, W} in the RRO with a Path Key Subobject. The Path Key Subobject assigns an "identifier" to the key. The PCE ID in the message indicates that it was node U that made the replacement.

With this additional information, the source is able to signal the subsequent LSPs with the ERO set to {C, D, exclude Path Key(EXRS), loose Dst}. When the signaling message reaches node X, it can consult node U to expand the Path Key and know how to avoid the path of the first LSP. Alternatively, the source could use an ERO of {C, D, loose Dst} and include an XRO containing the Path Key.

This mechanism can work with all the Path-Key resolution mechanisms, as detailed in [RFC5553] section 3.1. A PCE, co-located or not, may be used to resolve the Path-Key, but the node (i.e., a Label Switching Router (LSR)) can also use the Path Key information to index a Path Segment previously supplied to it by the entity that originated the Path-Key, for example the LSR that inserted the Path-Key in the RRO or a management system.

1.3. Network-Assigned Identifier

There are scenarios in which the network provides diversity-related information for a service that allows the client device to include this information in the signaling message. If the Shared Resource Link Group (SRLG) identifier information is both available and shareable (by policy) with the ENs, the procedure

defined in [DRAFT-SRLG-RECORDING] can be used to collect SRLG identifiers associated with an LSP (LSP1). When a second LSP (LSP2) needs to be diverse with respect to LSP1, the EN constructing the RSVP signaling message for setting up LSP2 can insert the SRLG identifiers associated with LSP1 as diversity constraints into the XRO using the procedure described in [RFC4874]. However, if the core network SRLG identifiers are either not available or not shareable with the ENs based on policies enforced by core network, existing mechanisms cannot be used.

In this draft, a signaling mechanism is defined where information signaled to the CN via the UNI does not require shared knowledge of core network SRLG information. For this purpose, the concept of a Path Affinity Set (PAS) is used for abstracting SRLG information. The motive behind the introduction of the PAS is to minimize the exchange of diversity information between the core network (CNs) and the client devices (ENs). The PAS contains an abstract SRLG identifier associated with a given path rather than a detailed SRLG list. The PAS is a single identifier that can be used to request diversity and associate diversity. The means by which the processing node determines the path corresponding to the PAS is beyond the scope of this document.

A CN on the core network boundary interprets the specific PAS identifier (e.g. "123") as meaning to exclude the core network SRLG information (or equivalent) that has been allocated by LSPs associated with this PAS identifier value. For example, if a Path exists for the LSP with the identifier "123", the CN would use local knowledge of the core network SRLGs associated with the "123" LSPs and use those SRLGs as constraints for path computation. If a PAS identifier is included for exclusion in the connection request, the CN (UNI-N) in the core network is assumed to be able to determine the existing core network SRLG information and calculate a path that meets the determined diversity constraints.

When a CN satisfies a connection setup for a (SRLG) diverse signaled path, the CN may optionally record the core network SRLG information for that connection in terms of CN based parameters and associates that with the EN addresses in the Path message. Specifically for Layer-1 Virtual Private Networks (L1VPNs), Port Information Tables (PIT) [RFC5251] can be leveraged to translate between client (EN) addresses and core network addresses.

The PAS and the associated SRLG information can be distributed within the core network by an Interior Gateway Protocol (IGP) or

by other means such as configuration. They can then be utilized by other CNs when other ENs are requesting paths to be setup that would require path/connection diversity. In the VPN case, this information is distributed on a VPN basis and contains a PAS identifier, CN addresses and SRLG information. In this way, on a VPN basis, the core network can have additional opaque records for the PAS values for various Paths along with the SRLG list associated with the Path. This information is internal to the core network and is known only to the core network.

2. RSVP-TE signaling extensions

This section describes the signaling extensions required to address the aforementioned requirements and use cases.

2.1. Diversity XRO Subobject

New Diversity XRO subobjects are defined by this document as follows.

2.1.1. IPv4 Diversity XRO Subobject

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 XRO Type										Length										DI Type A-Flags E-Flags										Resvd									
										IPv4 Diversity Identifier source address																													
										Diversity Identifier Value																													
//										...																				//									

L:

The L-flag is used as for the XRO subobjects defined in [RFC4874], i.e.,

0 indicates that the attribute specified MUST be excluded.

1 indicates that the attribute specified SHOULD be avoided.

XRO Type

Type for IPv4 diversity XRO subobject (to be assigned by IANA; suggested value: 37).

Length

The Length contains the total length of the subobject in bytes, including the Type and Length fields. The Length is variable, depending on the diversity identifier value.

Diversity Identifier Type (DI Type)

Diversity Identifier Type (DI Type) indicates the way the reference LSP(s) or route(s) with which diversity is required is identified. Three values are defined in this document:

IPv4 Client Initiated Identifier	1 (to be assigned by IANA)
IPv4 PCE Allocated Identifier	2 (to be assigned by IANA)
IPv4 Network Assigned Identifier	3 (to be assigned by IANA)

Attribute Flags (A-Flags):

The Attribute Flags (A-Flags) are used to communicate desirable attributes of the LSP being signaled. The following flags are defined. Each flag acts independently. Any combination of flags is permitted.

0x01 = Destination node exception

Indicates that the exclusion does not apply to the destination node of the LSP being signaled.

0x02 = Processing node exception

Indicates that the exclusion does not apply to the border node(s) performing ERO expansion for the LSP being signaled. An ingress UNI-N node is an example of such a node.

0x04 = Penultimate node exception

Indicates that the penultimate node of the LSP being signaled MAY be shared with the excluded path even when this violates the exclusion flags.

0x08 = LSP ID to be ignored

This flag is only applicable when the diversity is specified using the client-initiated identifier, the flag indicates tunnel level exclusion, as detailed in section 2.2.

Exclusion Flags (E-Flags):

The Exclusion-Flags are used to communicate the desired type(s) of exclusion. The following flags are defined. Any combination of these flags is permitted.

0x01 = SRLG exclusion

Indicates that the path of the LSP being signaled is requested to be SRLG-diverse from the excluded path specified by the Diversity XRO subobject.

0x02 = Node exclusion

Indicates that the path of the LSP being signaled is requested to be node-diverse from the excluded path specified by the Diversity XRO subobject.

(Note: the meaning of this flag may be modified by the value of the Attribute-flags.)

0x04 = Link exclusion

Indicates that the path of the LSP being signaled is requested to be link-diverse from the path specified by the Diversity XRO subobject.

Resvd

This field is reserved. It SHOULD be set to zero on transmission, and MUST be ignored on receipt.

IPv4 Diversity Identifier source address:

This field is set to the IPv4 address of the node that assigns the diversity identifier. Depending on the diversity identifier type, the diversity identifier source may be a client node, PCE entity or network node. Specifically:

- o When the diversity identifier type is set to "IPv4 Client Initiated Identifier", the value is set to IPv4 tunnel sender address of the reference LSP against which diversity is desired. IPv4 tunnel sender address is as defined in [RFC3209].
- o When the diversity identifier type is set to "IPv4 PCE Allocated Identifier", the value indicates the IPv4 address of the node that assigned the Path Key identifier and that can return an expansion of the Path Key or use the Path Key as exclusion in a path computation. The Path Key is defined in [RFC5553].
- o When the diversity identifier type is set to "IPv4 Network Assigned Identifier", the value indicates the IPv4 address of the node publishing the Path Affinity Set (PAS).

Diversity Identifier Value:

Encoding for this field depends on the diversity identifier type, as defined in the following.

When the diversity identifier type is set to "IPv4 Client Initiated Identifier", the diversity identifier value is encoded as follows:

2.1.2. IPv6 Diversity XRO Subobject

```

      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|  XRO Type  |      Length      |DI Type|A-Flags|E-Flags| Resvd |
+-----+-----+-----+-----+-----+-----+-----+-----+
|               IPv6 Diversity Identifier source address          |
+-----+-----+-----+-----+-----+-----+-----+-----+
|               IPv6 Diversity Identifier source address (cont.)   |
+-----+-----+-----+-----+-----+-----+-----+-----+
|               IPv6 Diversity Identifier source address (cont.)   |
+-----+-----+-----+-----+-----+-----+-----+-----+
|               IPv6 Diversity Identifier source address (cont.)   |
+-----+-----+-----+-----+-----+-----+-----+-----+
|               Diversity Identifier Value                          |
//                               ...                               //
|                               |
+-----+-----+-----+-----+-----+-----+-----+-----+

```

L:

The L-flag is used as for the XRO subobjects defined in [RFC4874], i.e.,

0 indicates that the attribute specified MUST be excluded.

1 indicates that the attribute specified SHOULD be avoided.

XRO Type

Type for IPv6 diversity XRO subobject (to be assigned by IANA; suggested value: 38).

Length

The Length contains the total length of the subobject in bytes, including the Type and Length fields. The Length is variable, depending on the diversity identifier value.

Attribute Flags (A-Flags):

As defined in Section 2.1.1 for the IPv4 counterpart.

Exclusion Flags (E-Flags):

As defined in Section 2.1.1 for the IPv4 counterpart.

Resvd

This field is reserved. It SHOULD be set to zero on transmission, and MUST be ignored on receipt.

Diversity Identifier Type (DI Type)

This field is defined in the same fashion as its IPv4 counterpart described in Section 2.1.1. The DI Types associated with IPv6 addresses are defined, as follows:

IPv6 Client Initiated Identifier	4 (to be assigned by IANA)
IPv6 PCE Allocated Identifier	5 (to be assigned by IANA)
IPv6 Network Assigned Identifier	6 (to be assigned by IANA)

These identifier are assigned and used as defined in Section 2.1.1.

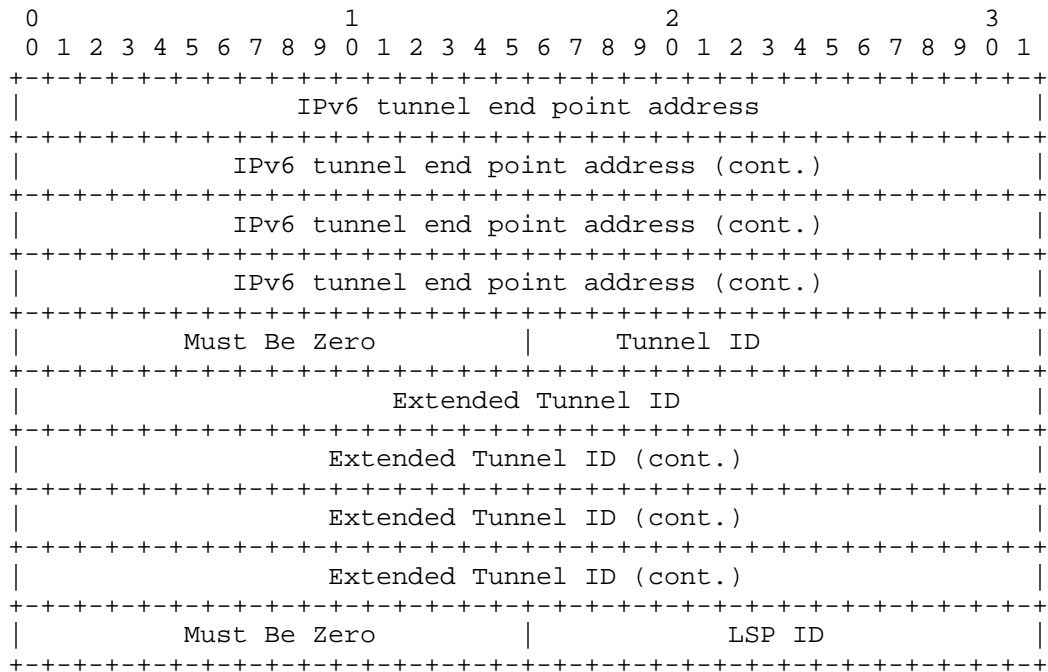
IPv4 Diversity Identifier source address:

This field is set to IPv6 address of the node that assigns the diversity identifier. How identity of node for various diversity types is determined is as described in Section 2.1.1 for the IPv4 counterpart.

Diversity Identifier Value:

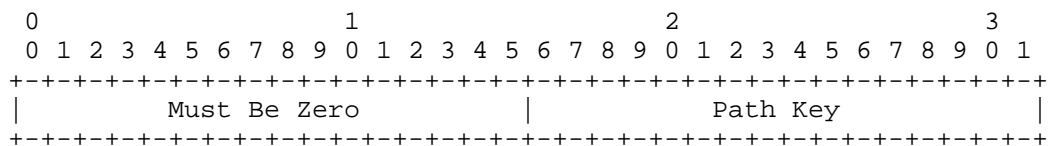
Encoding for this field depends on the diversity identifier type, as defined in the following.

When the diversity identifier type is set to "IPv6 Client Initiated Identifier", the diversity identifier value is encoded as follows:



The IPv6 tunnel end point address, Tunnel ID, IPv6 Extended Tunnel ID and LSP ID are as defined in [RFC3209].

When the diversity identifier type is set to "IPv6 PCE Allocated Identifier", the diversity identifier value is encoded as follows:



The Path Key is defined in [RFC5553].

When the diversity identifier type is set to "IPv6 Network Assigned Identifier", the diversity identifier value is 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
+-----+-----+-----+-----+-----+-----+-----+-----+
|                                     Path Affinity Set (PAS) identifier                                     |
+-----+-----+-----+-----+-----+-----+-----+-----+

```

The Path affinity Set (PAS) identifier is as defined in Section 2.1.1.

2.2. Processing rules for the Diversity XRO subobject

The procedure defined in [RFC4874] for processing XRO and EXRS is not changed by this document. If the processing node cannot recognize the IPv4/ IPv6 Diversity XRO subobject, the node is expected to follow the procedure defined in [RFC4874].

An XRO object MAY contain multiple Diversity subobjects. E.g., In order to exclude multiple Path Keys, an EN may include multiple Diversity XRO subobjects each with a different Path Key. Similarly, in order to exclude multiple PAS identifiers, an EN may include multiple Diversity XRO subobjects each with a different PAS identifier. However, all Diversity subobjects in an XRO SHOULD contain the same Diversity Identifier Type. If a Path message contains an XRO with Diversity subobjects with multiple Diversity Identifier Types, the processing node SHOULD return a PathErr with the error code "Routing Problem" (24) and error sub-code "XRO Too Complex" (68).

The attribute-flags affect the processing of the Diversity XRO subobject as follows:

- o When the "destination node exception" flag is set, the exclusion SHOULD be ignored for the destination node.
- o When the "processing node exception" flag is set, the exclusion SHOULD be ignored for the processing node. The processing node is the node performing path calculation.

- o When the "penultimate node exception" flag is set, the exclusion SHOULD be ignored for the penultimate node on the path of the LSP being established.
- o The "LSP ID to be ignored" flag is only defined for the "IPv4/ IPv6 Client Initiated Identifier" diversity types. When the Diversity Identifier Type is set to any other value, this flag SHOULD NOT be set on transmission and MUST be ignored in processing. When this flag is not set, the lsp-id is not ignored and the exclusion applies only to the specified LSP (i.e., LSP level exclusion).

If the L-flag of the diversity XRO subobject is not set, the processing node proceeds as follows.

- "IPv4/ IPv6 Client Initiated Identifiers" Diversity Type: the processing node MUST ensure that any path calculated for the signaled LSP is diverse from the RSVP TE FEC identified by the client in the XRO subobject.
- "IPv4/ IPv6 PCE Allocated Identifiers" Diversity Type: the processing node MUST ensure that any path calculated for the signaled LSP is diverse from the route identified by the Path-Key. The processing node MAY use the PCE identified by the IPv4 Diversity Identifier source address in the subobject for route computation. The processing node MAY use the Path-Key resolution mechanisms described in [RFC5553].
- "IPv4/ IPv6 Network Assigned Identifiers" Diversity Type: the processing node MUST ensure that the path calculated for the signaled LSP respects the requested PAS exclusion. .
- Regardless of whether the path computation is performed locally or at a remote node (e.g., PCE), the processing node MUST ensure that any path calculated for the signaled LSP respects the requested exclusion flags with respect to the excluded path referenced by the subobject, including local resources.
- If the excluded path referenced in the XRO subobject is unknown to the processing node, the processing node SHOULD ignore the diversity XRO subobject and SHOULD proceed with the signaling request. After sending the ResvErr for the signaled LSP, the processing node SHOULD return a PathErr with the error code "Notify Error" (25) and error sub-code "Route reference in diversity XRO identifier unknown" (value to be assigned by IANA, suggested value: 13) for the signaled LSP.

- If the processing node fails to find a path that meets the requested constraint, the processing node MUST return a PathErr with the error code "Routing Problem" (24) and error sub-code "Route blocked by Exclude Route" (67).

If the L-flag of the diversity XRO subobject is set, the processing node proceeds as follows:

- "IPv4/ IPv6 Client Initiated Identifiers" Diversity Type: the processing node SHOULD ensure that the path calculated for the signaled LSP is diverse from the RSVP TE FEC identified by the client in the XRO subobject.
- "IPv4/ IPv6 PCE Allocated Identifiers" Diversity Type: the processing node SHOULD ensure that the path calculated for the signaled LSP is diverse from the route identified by the Path-Key.
- "IPv4/ IPv6 Network Assigned Identifiers" Diversity Type: the processing node SHOULD ensure that the path calculated for the signaled LSP respects the requested PAS exclusion. The means by which the processing node determines the path corresponding to the PAS is beyond the scope of this document.
- The processing node SHOULD respect the requested exclusion flags with respect to the excluded path to the extent possible.
- If the processing node fails to find a path that meets the requested constraint, it SHOULD proceed with signaling using a suitable path that meets the constraint as far as possible. After sending the Resv for the signaled LSP, it SHOULD return a PathErr message with error code "Notify Error" (25) and error sub-code "Failed to respect Exclude Route" (value: to be assigned by IANA, suggest value: 14) to the source node.

If, subsequent to the initial signaling of a diverse LSP:

- An excluded path referenced in the XRO subobject becomes known to the processing node, or a change in the excluded path becomes known to the processing node, the processing node SHOULD re-evaluate the exclusion and diversity constraints requested by the diverse LSP to determine whether they are still satisfied.
- If the requested exclusion constraints for the diverse LSP are no longer satisfied and an alternative path for the diverse LSP that can satisfy those constraints exists, then:

- o If the L-flag was not set in the original exclusion, the processing node MUST send a PathErr message for the diverse LSP with the error code "Routing Problem" (24) and error sub-code "Route blocked by Exclude Route" (67). The PSR flag SHOULD NOT be set. A source node receiving a PathErr message with this error code and sub-code combination SHOULD take appropriate actions to migrate the compliant path.
- o If the L-flag was set in the original exclusion, the processing node SHOULD send a PathErr message for the diverse LSP with the error code "Notify Error" (25) and a new error sub-code "compliant path exists" (value: to be assigned by IANA, suggest value: 15). The PSR flag SHOULD NOT be set. A source node receiving a PathErr message with this error code and sub-code combination MAY signal a new LSP to migrate the compliant path.
- If the requested exclusion constraints for the diverse LSP are no longer satisfied and no alternative path for the diverse LSP that can satisfy those constraints exists, then:
 - o If the L-flag was not set in the original exclusion, the processing node MUST send a PathErr message for the diverse LSP with the error code "Routing Problem" (24) and error sub-code "Route blocked by Exclude Route" (67). The PSR flag SHOULD be set.
 - o If the L-flag was set in the original exclusion, the processing node SHOULD send a PathErr message for the diverse LSP with the error code error code "Notify Error" (25) and error sub-code "Failed to respect Exclude Route" (value: to be assigned by IANA, suggest value: 14). The PSR flag SHOULD NOT be set.

The following rules apply whether or not the L-flag is set:

- A source node receiving a PathErr message with the error code "Notify Error" (25) and error sub-codes "Route of XRO tunnel identifier unknown" or "Failed to respect Exclude Route" MAY take no action.

2.3. Diversity EXRS Subobject

[RFC4874] defines the EXRS ERO subobject. An EXRS is used to identify abstract nodes or resources that must not or should not be used on the path between two inclusive abstract nodes or

resources in the explicit route. An EXRS contains one or more subobjects of its own, called EXRS subobjects [RFC4874].

An EXRS MAY include Diversity subobject as specified in this document. In this case, the IPv4 EXRS format is 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
+-----+-----+-----+-----+-----+-----+-----+-----+
|L|      Type      |      Length      |      Reserved      |
+-----+-----+-----+-----+-----+-----+-----+-----+
|L|  XRO Type  |      Length  |DI Type|A-Flags|E-Flags| Resvd |
+-----+-----+-----+-----+-----+-----+-----+-----+
|      IPv4 Diversity Identifier source address      |
+-----+-----+-----+-----+-----+-----+-----+-----+
|      Diversity Identifier Value      |
//                                     ...                                     //
|
+-----+-----+-----+-----+-----+-----+-----+-----+

```

Similarly, the IPv6 EXRS format is 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
+-----+-----+-----+-----+-----+-----+-----+-----+
|L|      Type      |      Length      |      Reserved      |
+-----+-----+-----+-----+-----+-----+-----+-----+
|L|  XRO Type  |      Length  |DI Type|A-Flags|E-Flags| Resvd |
+-----+-----+-----+-----+-----+-----+-----+-----+
|      IPv6 Diversity Identifier source address      |
+-----+-----+-----+-----+-----+-----+-----+-----+
|      IPv6 Diversity Identifier source address (cont.)      |
+-----+-----+-----+-----+-----+-----+-----+-----+
|      IPv6 Diversity Identifier source address (cont.)      |
+-----+-----+-----+-----+-----+-----+-----+-----+
|      IPv6 Diversity Identifier source address (cont.)      |
+-----+-----+-----+-----+-----+-----+-----+-----+
|      Diversity Identifier Value      |
//                                     ...                                     //
|
+-----+-----+-----+-----+-----+-----+-----+-----+

```

The meanings of respective fields in EXRS header are as defined in [RFC4874]. The meanings of respective fields in the Diversity subobject are as defined earlier in this document for the XRO subobject.

The processing rules for the EXRS object are unchanged from [RFC4874]. When the EXRS contains one or more Diversity subobject(s), the processing rules specified in Section 2.2 apply to the node processing the ERO with the EXRS subobject.

If a loose-hop expansion results in the creation of another loose-hop in the outgoing ERO, the processing node MAY include the EXRS in the newly created loose hop for further processing by downstream nodes.

The processing node exception for the EXRS subobject applies to the node processing the ERO.

The destination node exception for the EXRS subobject applies to the explicit node identified by the ERO subobject that identifies the next abstract node. This flag is only processed if the L bit is set in the ERO subobject that identifies the next abstract node.

The penultimate node exception for the EXRS subobject applies to the node before the explicit node identified by the ERO subobject that identifies the next abstract node. This flag is only processed if the L bit is set in the ERO subobject that identifies the next abstract node.

3. Security Considerations

This document does not introduce any additional security issues above those identified in [RFC5920], [RFC2205], [RFC3209], [RFC3473] and [RFC4874].

4. IANA Considerations

4.1. New XRO subobject types

IANA registry: RSVP PARAMETERS

Subsection: Class Names, Class Numbers, and Class Types

This document introduces two new subobjects for the EXCLUDE_ROUTE object [RFC4874], C-Type 1.

Subobject Description -----	Subobject Type -----
IPv4 Diversity subobject	To be assigned by IANA (suggested value: 37)
IPv6 Diversity subobject	To be assigned by IANA (suggested value: 38)

4.2. New EXRS subobject types

The diversity XRO subobjects are also defined as new EXRS subobjects.

4.3. New RSVP error sub-codes

IANA registry: RSVP PARAMETERS
Subsection: Error Codes and Globally Defined Error Value Sub-Codes

For Error Code "Notify Error" (25) (see [RFC3209]) the following sub-codes are defined.

Sub-code -----	Value -----
Route of XRO tunnel identifier unknown	To be assigned by IANA. Suggested Value: 13.
Failed to respect Exclude Route	To be assigned by IANA. Suggested Value: 14.
Compliant path exists	To be assigned by IANA. Suggested Value: 15.

5. Acknowledgements

The authors would like to thank Luyuan Fang and Walid Wakim for their review comments.

6. References

6.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, December 2001.
- [RFC3473] Berger, L., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions", RFC 3473, January 2003.
- [RFC4874] Lee, CY., Farrel, A., and S. De Cnodder, "Exclude Routes - Extension to Resource ReserVation Protocol-Traffic Engineering (RSVP-TE)", RFC 4874, April 2007.
- [RFC5553] Farrel, A., Ed., Bradford, R., and JP. Vasseur, "Resource Reservation Protocol (RSVP) Extensions for Path Key Support", RFC 5553, May 2009.

6.2. Informative References

- [RFC4208] Swallow, G., Drake, J., Ishimatsu, H., and Y. Rekhter, "Generalized Multiprotocol Label Switching (GMPLS) User-Network Interface (UNI): Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Support for the Overlay Model", RFC 4208, October 2005.
- [RFC4920] Farrel, A., Ed., Satyanarayana, A., Iwata, A., Fujita, N., and G. Ash, "Crankback Signaling Extensions for MPLS and GMPLS RSVP-TE", RFC 4920, July 2007.
- [RFC5520] Bradford, R., Ed., Vasseur, JP., and A. Farrel, "Preserving Topology Confidentiality in Inter-Domain Path Computation Using a Path-Key-Based Mechanism", RFC 5520, April 2009.
- [DRAFT-SRLG-RECORDING] F. Zhang, D. Li, O. Gonzalez de Dios, C. Margaria, "RSVP-TE Extensions for Collecting SRLG Information", draft-ietf-ccamp-rsvp-te-srlg-collect.txt, work in progress.

- [RFC2205] Braden, R. (Ed.), Zhang, L., Berson, S., Herzog, S. and S. Jamin, "Resource ReserVation Protocol -- Version 1 Functional Specification", RFC 2205, September 1997.
- [RFC4026] Andersson, L. and T. Madsen, "Provider Provisioned Virtual Private Network (VPN) Terminology", RFC 4026, March 2005.
- [RFC5253] Takeda, T., Ed., "Applicability Statement for Layer 1 Virtual Private Network (L1VPN) Basic Mode", RFC 5253, July 2008.
- [RFC5920] Fang, L., Ed., "Security Framework for MPLS and GMPLS Networks", RFC 5920, July 2010.

Contributors' Addresses

Igor Bryskin
ADVA Optical Networking
Email: ibryskin@advaoptical.com

Daniele Ceccarelli
Ericsson
Email: Daniele.Ceccarelli@ericsson.com

Dhruv Dhody
Huawei Technologies
EMail: dhruv.ietf@gmail.com

Oscar Gonzalez de Dios
Telefonica I+D
Email: ogondio@tid.es

Don Fedyk
Hewlett-Packard
Email: don.fedyk@hp.com

Clarence Filsfils
Cisco Systems, Inc.
Email: cfilsfil@cisco.com

Xihua Fu
ZTE

Email: fu.xihua@zte.com.cn

Gabriele Maria Galimberti
Cisco Systems
Email: ggalimbe@cisco.com

Ori Gerstel
SDN Solutions Ltd.
Email: origerstel@gmail.com

Matt Hartley
Cisco Systems
Email: mhartley@cisco.com

Kenji Kumaki
KDDI Corporation
Email: ke-kumaki@kddi.com

Rudiger Kunze
Deutsche Telekom AG
Email: Ruediger.Kunze@telekom.de

Lieven Levrau
Alcatel-Lucent
Email: Lieven.Levrau@alcatel-lucent.com

Cyril Margaria
cyril.margaria@gmail.com

Julien Meuric
France Telecom Orange
Email: julien.meuric@orange.com

Yuji Tochio
Fujitsu
Email: tochio@jp.fujitsu.com

Xian Zhang
Huawei Technologies
Email: zhang.xian@huawei.com

Authors' Addresses

Zafar Ali
Cisco Systems.
Email: zali@cisco.com

Dieter Beller
Alcatel-Lucent
Email: Dieter.Beller@alcatel-lucent.com

George Swallow
Cisco Systems
Email: swallow@cisco.com

Fatai Zhang
Huawei Technologies
Email: zhangfatai@huawei.com

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: April 30, 2015

F. Zhang, Ed.
Huawei
O. Gonzalez de Dios, Ed.
Telefonica Global CTO
D. Li
Huawei
C. Margaria

M. Hartley
Z. Ali
Cisco
October 27, 2014

RSVP-TE Extensions for Collecting SRLG Information
draft-ietf-ccamp-rsvp-te-srlg-collect-09

Abstract

This document provides extensions for the Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) to support automatic collection of Shared Risk Link Group (SRLG) information for the TE link formed by a Label Switched Path (LSP).

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 30, 2015.

Copyright Notice

Copyright (c) 2014 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	2
1.1. Applicability Example: Dual Homing	3
2. Requirements Language	4
3. RSVP-TE Requirements	5
3.1. SRLG Collection Indication	5
3.2. SRLG Collection	5
3.3. SRLG Update	5
4. Encodings	5
4.1. SRLG Collection Flag	5
4.2. SRLG sub-object	6
5. Signaling Procedures	7
5.1. SRLG Collection	7
5.2. SRLG Update	9
5.3. Compatibility	9
6. Manageability Considerations	9
6.1. Policy Configuration	9
6.2. Coherent SRLG IDs	9
7. Security Considerations	10
8. IANA Considerations	10
8.1. RSVP Attribute Bit Flags	10
8.2. ROUTE_RECORD Object	11
8.3. Policy Control Failure Error subcodes	11
9. Acknowledgements	11
10. References	11
10.1. Normative References	11
10.2. Informative References	12
Authors' Addresses	12

1. Introduction

It is important to understand which TE links in the network might be at risk from the same failures. In this sense, a set of links can constitute a 'shared risk link group' (SRLG) if they share a resource whose failure can affect all links in the set [RFC4202].

On the other hand, as described in [RFC4206] and [RFC6107], H-LSP (Hierarchical LSP) or S-LSP (stitched LSP) can be used for carrying one or more other LSPs. Both of the H-LSP and S-LSP can be formed as

a TE link. In such cases, it is important to know the SRLG information of the LSPs that will be used to carry further LSPs.

This document provides a mechanism to collect the SRLGs used by a LSP, which can then be advertised as properties of the TE-link formed by that LSP. Note that specification of the the use of the collected SRLGs is outside the scope of this document.

1.1. Applicability Example: Dual Homing

An interesting use case for the SRLG collection procedures defined in this document is achieving LSP diversity in a dual homing scenario. The use case is illustrated in Figure 1, when the overlay model is applied as defined in RFC 4208 [RFC4208] . In this example, the exchange of routing information over the User-Network Interface (UNI) is prohibited by operator policy.

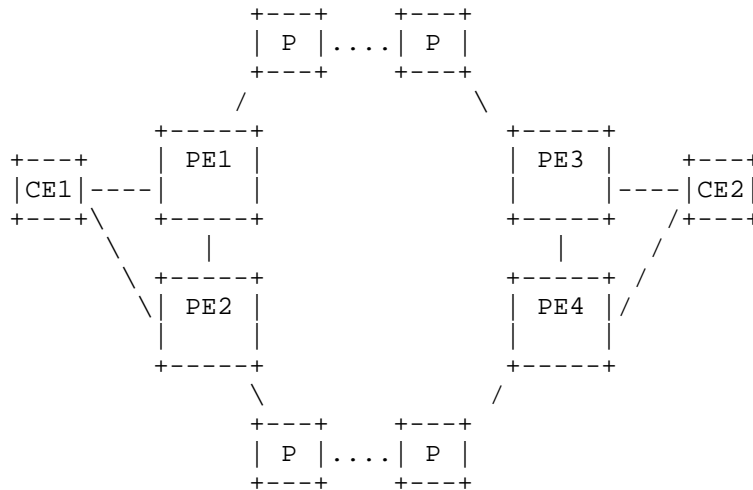


Figure 1: Dual Homing Configuration

Single-homed customer edge (CE) devices are connected to a single provider edge (PE) device via a single UNI link (which could be a bundle of parallel links, typically using the same fiber cable). This single UNI link can constitute a single point of failure. Such a single point of failure can be avoided if the CE device is connected to two PE devices via two UNI interfaces as depicted in Figure 1 above for CE1 and CE2, respectively.

For the dual-homing case, it is possible to establish two connections (LSPs) from the source CE device to the same destination CE device where one connection is using one UNI link to PE1, for example, and

the other connection is using the UNI link to PE2. In order to avoid single points of failure within the provider network, it is necessary to also ensure path (LSP) diversity within the provider network in order to achieve end-to-end diversity for the two LSPs between the two CE devices CE1 and CE2. This use case describes how it is possible to achieve path diversity within the provider network based on collected SRLG information. As the two connections (LSPs) enter the provider network at different PE devices, the PE device that receives the connection request for the second connection needs to know the additional path computation constraints such that the path of the second LSP is disjoint with respect to the already established first connection.

As SRLG information is normally not shared between the provider network and the client network, i.e., between PE and CE devices, the challenge is how to solve the diversity problem when a CE is dual-homed. For example, CE1 in Figure 1 may have requested an LSP1 to CE2 via PE1 that is routed via PE3 to CE2. CE1 can then subsequently request an LSP2 to CE2 via PE2 with the constraint that it needs to be maximally SRLG disjoint with respect to LSP1. PE2, however, does not have any SRLG information associated with LSP1, which is needed as input for its constraint-based path computation function. If CE1 is capable of retrieving the SRLG information associated with LSP1 from PE1, it can pass this information to PE2 as part of the LSP2 setup request (RSVP PATH message), and PE2 can now calculate a path for LSP2 that is SRLG disjoint with respect to LSP1. The SRLG information associated with LSP1 can already be retrieved when LSP1 is setup or at any time before LSP2 is setup.

The RSVP extensions for collecting SRLG information defined in this document make it possible to retrieve SRLG information for an LSP and hence solve the dual-homing LSP diversity problem. When CE1 sends the setup request for LSP2 to PE2, it can also request the collection of SRLG information for LSP2 and send that information to PE1. This will ensure that the two paths for the two LSPs remain mutually diverse, which is important, when the provider network is capable to restore connections that failed due to a network failure (fiber cut) in the provider network.

Note that the knowledge of SRLG information even for multiple LSPs does not allow a CE devices to derive the provider network topology based on the collected SRLG information.

2. 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].

3. RSVP-TE Requirements

3.1. SRLG Collection Indication

The ingress node of the LSP SHOULD be capable of indicating whether the SRLG information of the LSP is to be collected during the signaling procedure of setting up an LSP. SRLG information SHOULD NOT be collected without an explicit request for it being made by the ingress node.

3.2. SRLG Collection

If requested, the SRLG information SHOULD be collected during the setup of an LSP. The endpoints of the LSP can use the collected SRLG information, for example, for routing, sharing and TE link configuration purposes.

3.3. SRLG Update

When the SRLG information of an existing LSP for which SRLG information was collected during signaling changes, the relevant nodes of the LSP SHOULD be capable of updating the SRLG information of the LSP. This means that the signaling procedure SHOULD be capable of updating the new SRLG information.

4. Encodings

4.1. SRLG Collection Flag

In order to indicate nodes that SRLG collection is desired, this document defines a new flag in the Attribute Flags TLV (see RFC 5420 [RFC5420]), which MAY be carried in an LSP_REQUIRED_ATTRIBUTES or LSP_ATTRIBUTES Object:

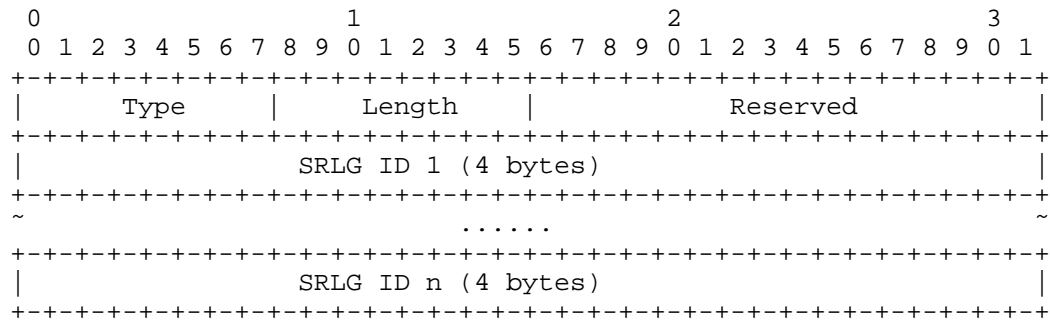
- o Bit Number (temporarily 12, an early allocation has been made by IANA, see Section 8.1 for more details): SRLG Collection flag

The SRLG Collection flag is meaningful on a Path message. If the SRLG Collection flag is set to 1, it means that the SRLG information SHOULD be reported to the ingress and egress node along the setup of the LSP.

The rules of the processing of the Attribute Flags TLV are not changed.

4.2. SRLG sub-object

This document defines a new RRO sub-object (ROUTE_RECORD sub-object) to record the SRLG information of the LSP. Its format is modeled on the RRO sub-objects defined in RFC 3209 [RFC3209].



Type

The type of the sub-object. The value is temporarily 34. An early allocation has been made by IANA (see Section 8.2 for more details).

Length

The Length field contains the total length of the sub-object in bytes, including the Type and Length fields. The Length depends on the number of SRLG IDs.

Reserved

This 2 byte field is reserved. It SHOULD be set to zero on transmission and MUST be ignored on receipt.

SRLG ID

This 4 byte field contains one SRLG ID. There is one SRLG ID field per SRLG collected. There MAY be multiple SRLG ID fields in an SRLG sub-object

As described in RFC 3209 [RFC3209], the RECORD_ROUTE object is managed as a stack. The SRLG sub-object SHOULD be pushed by the node before the node IP address or link identifier. The SRLG-sub-object SHOULD be pushed after the Attribute subobject, if present, and after the LABEL subobject, if requested.

RFC 5553 [RFC5553] describes mechanisms to carry a PKS (Path Key Sub-object) in the RRO so as to facilitate confidentiality in the

signaling of inter-domain TE LSPs, and allows the path segment that needs to be hidden (that is, a Confidential Path Segment (CPS)) to be replaced in the RRO with a PKS. If the CPS contains SRLG Sub-objects, these MAY be retained in the RRO by adding them again after the PKS Sub-object in the RRO. The CPS is defined in RFC 5520 [RFC5520]

A node MUST NOT push a SRLG sub-object in the RECORD_ROUTE without also pushing either a IPv4 sub-object, a IPv6 sub-object, a Unnumbered Interface ID sub-object or a Path Key sub-object.

The rules of the processing of the LSP_REQUIRED_ATTRIBUTES, LSP_ATTRIBUTE and ROUTE_RECORD Objects are not changed.

5. Signaling Procedures

5.1. SRLG Collection

Per RFC 3209 [RFC3209], an ingress node initiates the recording of the route information of an LSP by adding a RRO to a Path message. If an ingress node also desires SRLG recording, it MUST set the SRLG Collection Flag in the Attribute Flags TLV which MAY be carried either in an LSP_REQUIRED_ATTRIBUTES Object when the collection is mandatory, or in an LSP_ATTRIBUTES Object when the collection is desired, but not mandatory

When a node receives a Path message which carries an LSP_REQUIRED_ATTRIBUTES Object and the SRLG Collection Flag set, if local policy determines that the SRLG information is not to be provided to the endpoints, it MUST return a PathErr message with Error Code 2 (policy) and Error subcode "SRLG Recording Rejected" (value 31, an early allocation of the value has been done by IANA, see Section 8.3 for more details) to reject the Path message.

When a node receives a Path message which carries an LSP_ATTRIBUTES Object and the SRLG Collection Flag set, if local policy determines that the SRLG information is not to be provided to the endpoints, the Path message SHOULD NOT be rejected due to SRLG recording restriction and the Path message SHOULD be forwarded without any SRLG sub-object(s) in the RRO of the corresponding outgoing Path message.

If local policy permits the recording of the SRLG information, the processing node SHOULD add local SRLG information, as defined below, to the RRO of the corresponding outgoing Path message. The processing node MAY add multiple SRLG sub-objects to the RRO if necessary. It then forwards the Path message to the next node in the downstream direction.

If the addition of SRLG information to the RRO would result in the RRO exceeding its maximum possible size or becoming too large for the Path message to contain it, the requested SRLGs MUST NOT be added. If the SRLG collection request was contained in an LSP_REQUIRED_ATTRIBUTES Object, the processing node MUST behave as specified by RFC 3209 [RFC3209] and drop the RRO from the Path message entirely. If the SRLG collection request was contained in an LSP_ATTRIBUTES Object, the processing node MAY omit some or all of the requested SRLGs from the RRO; otherwise it MUST behave as specified by RFC 3209 [RFC3209] and drop the RRO from the Path message entirely.

Following the steps described above, the intermediate nodes of the LSP can collect the SRLG information in the RRO during the processing of the Path message hop by hop. When the Path message arrives at the egress node, the egress node receives SRLG information in the RRO.

Per RFC 3209 [RFC3209], when issuing a Resv message for a Path message which contains an RRO, an egress node initiates the RRO process by adding an RRO to the outgoing Resv message. The processing for RROs contained in Resv messages then mirrors that of the Path messages.

When a node receives a Resv message for an LSP for which SRLG Collection is specified, then when local policy allows recording SRLG information, the node SHOULD add SRLG information, to the RRO of the corresponding outgoing Resv message, as specified below. When the Resv message arrives at the ingress node, the ingress node can extract the SRLG information from the RRO in the same way as the egress node.

Note that a link's SRLG information for the upstream direction cannot be assumed to be the same as that in the downstream.

- o For Path and Resv messages for a unidirectional LSP, a node SHOULD include SRLG sub-objects in the RRO for the downstream data link only.
- o For Path and Resv messages for a bidirectional LSP, a node SHOULD include SRLG sub-objects in the RRO for both the upstream data link and the downstream data link from the local node. In this case, the node MUST include the information in the same order for both Path messages and Resv messages. That is, the SRLG sub-object for the upstream link is added to the RRO before the SRLG sub-object for the downstream link.

Based on the above procedure, the endpoints can get the SRLG information automatically. Then the endpoints can for instance

advertise it as a TE link to the routing instance based on the procedure described in [RFC6107] and configure the SRLG information of the FA automatically.

5.2. SRLG Update

When the SRLG information of a link is changed, the LSPs using that link need to be aware of the changes. The procedures defined in Section 4.4.3 of RFC 3209 [RFC3209] MUST be used to refresh the SRLG information if the SRLG change is to be communicated to other nodes according to the local node's policy. If local policy is that the SRLG change SHOULD be suppressed or would result in no change to the previously signaled SRLG-list, the node SHOULD NOT send an update.

5.3. Compatibility

A node that does not recognize the SRLG Collection Flag in the Attribute Flags TLV is expected to proceed as specified in RFC 5420 [RFC5420]. It is expected to pass the TLV on unaltered if it appears in a LSP_ATTRIBUTES object, or reject the Path message with the appropriate Error Code and Value if it appears in a LSP_REQUIRED_ATTRIBUTES object.

A node that does not recognize the SRLG RRO sub-object is expected to behave as specified in RFC 3209 [RFC3209]: unrecognized subobjects are to be ignored and passed on unchanged.

6. Manageability Considerations

6.1. Policy Configuration

In a border node of inter-domain or inter-layer network, the following SRLG processing policy SHOULD be capable of being configured:

- o Whether the SRLG IDs of the domain or specific layer network can be exposed to the nodes outside the domain or layer network, or whether they SHOULD be summarized, mapped to values that are comprehensible to nodes outside the domain or layer network, or removed entirely.

A node using RFC 5553 [RFC5553] and PKS MAY apply the same policy.

6.2. Coherent SRLG IDs

In a multi-layer multi-domain scenario, SRLG ids can be configured by different management entities in each layer/domain. In such scenarios, maintaining a coherent set of SRLG IDs is a key

requirement in order to be able to use the SRLG information properly. Thus, SRLG IDs SHOULD be unique. Note that current procedure is targeted towards a scenario where the different layers and domains belong to the same operator, or to several coordinated administrative groups. Ensuring the aforementioned coherence of SRLG IDs is beyond the scope of this document.

Further scenarios, where coherence in the SRLG IDs cannot be guaranteed are out of the scope of the present document and are left for further study.

7. Security Considerations

This document builds on the mechanisms defined in [RFC3473], which also discusses related security measures. In addition, [RFC5920] provides an overview of security vulnerabilities and protection mechanisms for the GMPLS control plane. The procedures defined in this document permit the transfer of SRLG data between layers or domains during the signaling of LSPs, subject to policy at the layer or domain boundary. It is recommended that domain/layer boundary policies take the implications of releasing SRLG information into consideration and behave accordingly during LSP signaling.

8. IANA Considerations

8.1. RSVP Attribute Bit Flags

IANA has created a registry and manages the space of the Attribute bit flags of the Attribute Flags TLV, as described in section 11.3 of RFC 5420 [RFC5420], in the "Attribute Flags" section of the "Resource Reservation Protocol-Traffic Engineering (RSVP-TE) Parameters" registry located in <http://www.iana.org/assignments/rsvp-te-parameters>. IANA has made an early allocation in the "Attribute Flags" section of the mentioned registry that expires on 2015-09-11.

This document introduces a new Attribute Bit Flag:

Bit No	Name	Attribute Flags Path	Attribute Flags Resv	RRO	Reference
-----	-----	-----	-----	---	-----
12 (temporary expires 2015-09-11)	SRLG collection Flag	Yes	Yes	Yes	This I-D

8.2. ROUTE_RECORD Object

IANA manages the "RSVP PARAMETERS" registry located at <http://www.iana.org/assignments/rsvp-parameters>. IANA has made an early allocation in the Sub-object type 21 ROUTE_RECORD - Type 1 Route Record registry. The early allocation expires on 2015-09-11.

This document introduces a new RRO sub-object:

Value	Description	Reference
-----	-----	-----
34 (temporary, expires 2015-09-11)	SRLG sub-object	This I-D

8.3. Policy Control Failure Error subcodes

IANA manages the assignments in the "Error Codes and Globally-Defined Error Value Sub-Codes" section of the "RSVP PARAMETERS" registry located at <http://www.iana.org/assignments/rsvp-parameters>. IANA has made an early allocation in the "Sub-Codes -2 Policy Control Failure" subsection of the the "Error Codes and Globally-Defined Error Value Sub-Codes" section of the "RSVP PARAMETERS" registry. The early allocation expires on 2015-09-11.

This document introduces a new Policy Control Failure Error sub-code:

Value	Description	Reference
-----	-----	-----
21 (temporary, expires 2015-09-11)	SRLG Recording Rejected	This I-D

9. Acknowledgements

The authors would like to thank Igor Bryskin, Ramon Casellas, Lou Berger, Alan Davey, Dhruv Dhody and Dieter Beller for their useful comments and improvements to the document.

10. References

10.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [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.
- [RFC5420] Farrel, A., Papadimitriou, D., Vasseur, JP., and A. Ayyangarps, "Encoding of Attributes for MPLS LSP Establishment Using Resource Reservation Protocol Traffic Engineering (RSVP-TE)", RFC 5420, February 2009.
- [RFC5520] Bradford, R., Vasseur, JP., and A. Farrel, "Preserving Topology Confidentiality in Inter-Domain Path Computation Using a Path-Key-Based Mechanism", RFC 5520, April 2009.
- [RFC5553] Farrel, A., Bradford, R., and JP. Vasseur, "Resource Reservation Protocol (RSVP) Extensions for Path Key Support", RFC 5553, May 2009.

10.2. Informative References

- [RFC4202] Kompella, K. and Y. Rekhter, "Routing Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4202, October 2005.
- [RFC4206] Kompella, K. and Y. Rekhter, "Label Switched Paths (LSP) Hierarchy with Generalized Multi-Protocol Label Switching (GMPLS) Traffic Engineering (TE)", RFC 4206, October 2005.
- [RFC4208] Swallow, G., Drake, J., Ishimatsu, H., and Y. Rekhter, "Generalized Multiprotocol Label Switching (GMPLS) User-Network Interface (UNI): Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Support for the Overlay Model", RFC 4208, October 2005.
- [RFC5920] Fang, L., "Security Framework for MPLS and GMPLS Networks", RFC 5920, July 2010.
- [RFC6107] Shiimoto, K. and A. Farrel, "Procedures for Dynamically Signaled Hierarchical Label Switched Paths", RFC 6107, February 2011.

Authors' Addresses

Fatai Zhang (editor)
Huawei
F3-5-B RD Center
Bantian, Longgang District, Shenzhen 518129
P.R.China

Email: zhangfatai@huawei.com

Oscar Gonzalez de Dios (editor)
Telefonica Global CTO
Distrito Telefonica, edificio sur, Ronda de la Comunicacion 28045
Madrid 28050
Spain

Phone: +34 913129647
Email: oscar.gonzalezdedios@telefonica.com

Dan Li
Huawei
F3-5-B RD Center
Bantian, Longgang District, Shenzhen 518129
P.R.China

Email: danli@huawei.com

Cyril Margaria
Suite 4001, 200 Somerset Corporate Blvd.
Bridgewater, NJ 08807
US

Email: cyril.margaria@gmail.com

Matt Hartley
Cisco

Email: mhartley@cisco.com

Zafar Ali
Cisco

Email: zali@cisco.com

Network Working Group
Internet Draft

Intended status: Standards Track

Expires: November 2015

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

May 18, 2015

Signaling Extensions for Wavelength Switched Optical Networks
draft-ietf-ccamp-wson-signaling-12.txt

Abstract

This document provides extensions to Generalized Multi-Protocol Label Switching (GMPLS) signaling for control of Wavelength Switched Optical Networks (WSON). Such extensions are applicable 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. This document provides mechanisms to support distributed wavelength assignment with choice in distributed wavelength assignment algorithms.

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 9, 2015.

Copyright Notice

Copyright (c) 2015 IETF Trust and the persons identified as the document authors. All rights reserved.

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

Conventions used in this document

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

Table of Contents

1. Introduction.....	3
2. Terminology.....	3
3. Requirements for WSON Signaling.....	4
3.1. WSON Signal Characterization.....	4
3.2. Per Node Processing Configuration.....	5
3.3. Bidirectional WSON LSPs.....	6
3.4. Distributed Wavelength Assignment Selection Method.....	6
3.5. Optical Impairments.....	6
4. WSON Signal Traffic Parameters, Attributes and Processing.....	6
4.1. Traffic Parameters for Optical Tributary Signals.....	7
4.2. WSON Processing Hop Attribute TLV.....	7
4.2.1. Resource Block Information Sub-TLV.....	8
4.2.2. Wavelength Selection Sub-TLV.....	9
5. Security Considerations.....	11
6. IANA Considerations.....	12
7. Acknowledgments.....	13
8. References.....	14
8.1. Normative References.....	14

8.2. Informative References.....	15
Author's Addresses.....	15

1. Introduction

This document 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 bidirectional wavelength assignment while more advanced extensions are given to support the networks described in [RFC6163] which feature connections requiring configuration of input, output, and general signal processing capabilities at a node along a Label Switched Path (LSP).

These extensions build on previous work for the control of lambda and G.709 based networks.

Related references with this document are [RFC7446] that provides a high-level information model and [WSON-Encode] that provides common encodings that can be applicable to other protocol extensions such as routing.

2. Terminology

CWDM: Coarse Wavelength Division Multiplexing.

DWDM: Dense Wavelength Division Multiplexing.

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 information bearing optical signal centered at a given frequency (wavelength) to one with "equivalent" content centered at a different wavelength. Wavelength conversion can be implemented via an optical-electronic-optical (OEO) process or via a strictly optical process.

WDM: Wavelength Division Multiplexing.

Wavelength Switched Optical Networks (WSON): WDM based optical networks in which switching is performed selectively based on the frequency 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 Receiver: 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 receiver.

Optical End Node: The end of a wavelength (optical lambdas) lightpath in the data plane. It may be equipped with some optical/electronic devices such as wavelength multiplexers/demultiplexer (e.g. AWG), optical transponder, etc., which are employed to transmit/terminate the optical signals for data transmission.

FEC: Forward Error Correction. Forward error correction (FEC) is a digital signal processing technique used to enhance data reliability. It does this by introducing redundant data, called error correcting code, prior to data transmission or storage. FEC provides the receiver with the ability to correct errors without a reverse channel to request the retransmission of data.

3R Regeneration: The process of amplifying (correcting loss), reshaping (correcting noise and dispersion), retiming (synchronizing with the network clock), and retransmitting an optical signal.

3. Requirements for WSON Signaling

The following requirements for GMPLS based WSON signaling are in addition to the functionality already provided by existing GMPLS signaling mechanisms.

3.1. WSON Signal Characterization

WSON signaling needs to 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 [RFC6163] we have the following list of WSON signal characteristic information

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. These parameters are summarized in the Optical Interface Class as defined in the [RFC7446] and the assumption is that a class always includes signal compatibility information.

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].

3.2. Per Node Processing Configuration

In addition to configuring a node along an LSP to input or output a signal with specific attributes, we may need to signal the node to perform specific processing, such as 3R regeneration, on the signal at a particular node. [RFC6163] discussed three types of processing:

- (A) Regeneration (possibly different types)
- (B) Fault and Performance Monitoring
- (C) Attribute Conversion

The extensions here provide for the configuration of these types of processing at nodes along an LSP.

3.3. Bidirectional WSON LSPs

WSON signaling can support LSP setup consistent with the wavelength continuity constraint for bidirectional connections. The following

cases need to be separately supported:

- (a) Where the same wavelength is used for both upstream and downstream directions
- (b) Where different wavelengths can be used for both upstream and downstream directions.

This document will review existing GMPLS bidirectional solutions according to WSON case.

3.4. Distributed Wavelength Assignment Selection Method

WSON signaling can support the selection of a specific distributed wavelength assignment method.

This method is beneficial in cases of equipment failure, etc., where fast provisioning used in quick recovery is critical to protect carriers/users against system loss. This requires efficient signaling which supports distributed wavelength assignment, in particular when the wavelength assignment capability is not available.

As discussed in the [RFC6163] different computational approaches for wavelength assignment are available. One method is the use of distributed wavelength assignment. This feature would allow the specification of a particular approach when more than one is implemented in the systems along the path.

3.5. Optical Impairments

This draft does not address signaling information related to optical impairments.

4. WSON Signal Traffic Parameters, Attributes and Processing

As discussed in [RFC6163] 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 [RFC6163].

LSPs signaled through extensions provided in this document MUST apply the following signaling parameters:

- . Switching Capability = WSON-LSC ([WSON-OSPF]).
- . Encoding Type = Lambda ([RFC3471])
- . Label Format = as defined in [RFC6205]

[RFC6205] defines the label format as applicable to LSC capable device.

4.1. Traffic Parameters for Optical Tributary Signals

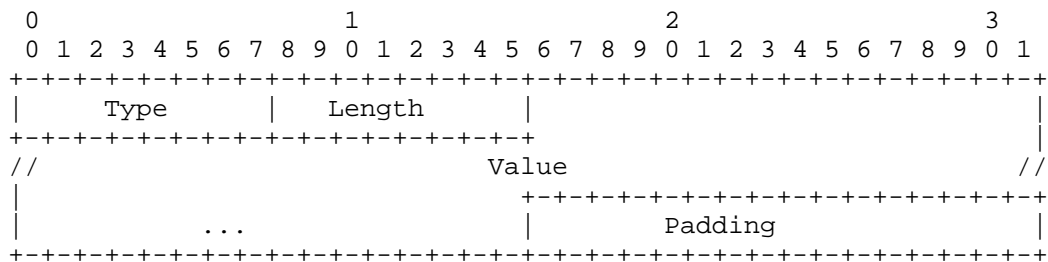
In [RFC3471] we see that the G-PID (client signal type) and bit rate (byte rate) of the signals are defined as parameters and in [RFC3473] they are conveyed in the Generalized Label Request object and the RSVP SENDER_TSPEC/FLOWSPEC objects respectively.

4.2. WSON Processing Hop Attribute TLV

Section 3.1. provided the requirements for signaling to indicate to a particular node along an LSP what type of processing to perform on an optical signal or how to configure that node to accept or transmit an optical signal with particular attributes.

To target a specific node, this section defines a WSON Processing Hop Attribute TLV. This TLV is encoded as an attributes TLV, see [RFC5420]. The TLV is carried in the ERO and RRO LSP Attribute Subobjects, and processed according to the procedures, defined in [RSVP-RO]. The type value of the WSON Processing Hop Attribute TLV is TBD by IANA.

The WSON Processing Hop Attribute TLV carries one or more sub-TLVs with the following format:



Type

The identifier of the sub-TLV.

Length

Indicates the total length of the sub-TLV in octets. That is, the combined length of the Type, Length, and Value fields, i.e., two plus the length of the Value field in octets.

The entire sub-TLV MUST be padded with zeros to ensure four-octet alignment of the sub-TLV.

Value

Zero or more octets of data carried in the sub-TLV.

Padding: Variable

Padding is used to ensure that the length of the WSON Processing Hop Attribute TLV meets the multiple of 4 byte size requirement.

Sub-TLV ordering is significant and MUST be preserved. Error processing follows [RSVP-RO].

The following sub-TLV types are defined in this document:

Sub-TLV Name	Type	Length
ResourceBlockInfo	1	variable
WavelengthSelection	2	8 octets (2 octet padding)

The TLV can be represented in Reduced Backus-Naur Form (RBNF) [RFC5511] syntax as:

```
<WSON Processing Hop Attribute> ::= <ResourceBlockInfo>  
[<ResourceBlockInfo>] [<WavelengthSelection>]
```

4.2.1. ResourceBlockInfo Sub-TLV

The format of the ResourceBlockInfo sub-TLV value field is defined in Section 4 of [WSON-Encode]. It is a list of available Optical Interface Classes and processing capabilities.

At least one ResourceBlockInfo sub-TLV MUST be present in the WSON_ Processing Hop Attribute TLV. No more than two ResourceBlockInfo sub-TLVs SHOULD be present. Any present ResourceBlockInfo sub-TLVs MUST be processed in the order received, and extra (unprocessed) SHOULD be ignored.

The ResourceBlockInfo field contains several information elements as defined by [WSON-Encode]. The following rules apply to the sub-TLV:

- o RB Set Field can carry one or more RB Identifier. Only the first of RB Identifier listed in the RB Set Field SHALL be processed, any others SHOULD be ignored.
- o In the case of unidirectional LSPs, only one ResourceBlockInfo sub-TLV SHALL be processed and the I and O bits can be safely ignored.
- o In the case of a bidirectional LSP, there MUST be either:
 - (a) only one ResourceBlockInfo sub-TLV present in a WSON_Processing Hop Attribute TLV, and the bits I and O both set to 1, or
 - (b) two ResourceBlockInfo sub-TLVs present, one of which has only the I bit set and the other of which has only the O bit set.
- o The rest of information carried within the ResourceBlockInfo sub-TLV includes Optical Interface Class List, Input Bit Rate List and Processing Capability List. These lists MAY contain one or more elements. These elements apply equally to both bidirectional and unidirectional LSPs.

Any violation of these rules detected by a transit or egress node SHALL be treated as an error and be processed per [RSVP-RO].

A ResourceBlockInfo sub-TLV can be constructed by a node and added to a ERO_Hop_ATTRIBUTE subobject in order to be processed by downstream nodes (transit and egress). As defined in [RSVP-RO], the R bit reflects the LSP_REQUIRED_ATTRIBUTE and LSP_ATTRIBUTE semantic defined in [RFC5420] and SHOULD be set accordingly.

Once a node properly parses a ResourceBlockInfo Sub-TLV received in an ERO_Hop_ATTRIBUTE subobject (according to the rules stated above and in [RSVP-RO]), the node allocates the indicated resources, e.g., the selected regeneration pool, for the LSP. In addition, the node SHOULD report compliance by adding a RRO_Hop_ATTRIBUTE subobject with the WSON Processing Hop Attribute TLV (and its sub-TLVs) indicating the utilized resources. ResourceBlockInfo Sub-TLVs carried in a RRO_Hop_ATTRIBUTE subobject are subject to [RSVP-RO] and standard RRO processing, see [RFC3209].

4.2.2. WavelengthSelection Sub-TLV

Routing + Distributed Wavelength Assignment (R+DWA) is one of the options defined by the [RFC6163]. The output from the routing

function will be a path but the wavelength will be selected on a hop-by-hop basis.

As discussed in [RFC6163], the wavelength assignment can be either for a unidirectional lightpath or for a bidirectional lightpath constrained to use the same lambda in both directions.

In order to indicate wavelength assignment directionality and wavelength assignment method, the Wavelength Selection, or WavelengthSelection, sub-TLV is defined to be carried in the WSON Processing Hop Attribute TLV defined above.

The WavelengthSelection sub-TLV value field is defined as:

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
W										WA Method										Reserved																			

Where:

W (1 bit): 0 denotes requiring the same wavelength in both directions, 1 denotes that different wavelengths on both directions are allowed.

Wavelength Assignment (WA) Method (7 bits):

0 - unspecified (any); This does not constrain the WA method used by a specific node. This value is implied when the WavelengthSelection Sub-TLV is absent.

1 - First-Fit. All the wavelengths are numbered and this WA method chooses the available wavelength with the lowest index.

2 - Random. This WA method chooses an available wavelength randomly.

3 - Least-Loaded (multi-fiber). This WA method selects the wavelength that has the largest residual capacity on the most loaded link along the route. This method is used in multi-fiber networks. If used in single-fiber networks, it is equivalent to the FF WA method.

4- 127: Unassigned.

The processing rules of this TLV are as follows:

If a receiving node does not support the attribute(s), its behaviors are specified below:

- W bit not supported: a PathErr MUST be generated with the Error Code "Routing Problem" (24) with error sub-code "Unsupported WavelengthSelection Symmetry value" (value to be assigned by IANA, suggested value: 107).
- WA method not supported: a PathErr MUST be generated with the Error Code "Routing Problem" (24) with error sub-code "Unsupported Wavelength Assignment value" (value to be assigned by IANA, suggested value: 108).

A WavelengthSelection sub-TLV can be constructed by a node and added to a ERO_Hop_ATTRIBUTE subobject in order to be processed by downstream nodes (transit and egress). As defined in [RSVP-RO], the R bit reflects the LSP_REQUIRED_ATTRIBUTE and LSP_ATTRIBUTE semantic defined in [RFC5420] and SHOULD be set accordingly.

Once a node properly parses the WavelengthSelection Sub-TLV received in an ERO_Hop_ATTRIBUTE subobject, the node use the indicated wavelength assignment method (at that hop) for the LSP. In addition, the node SHOULD report compliance by adding a RRO_Hop_ATTRIBUTE subobject with the WSON Processing Hop Attribute TLV (and its sub-TLVs) indicated the utilized method. WavelengthSelection Sub-TLVs carried in a RRO_Hop_ATTRIBUTE subobject are subject to [RSVP-RO] and standard RRO processing, see [RFC3209].

5. Security Considerations

This document is built on the mechanisms defined in [RFC3473], and only differs in specific information communicated. As such, this document introduces no new security considerations to the existing GMPLS signaling protocols. See [RFC3473], for details of the supported security measures. Additionally, [RFC5920] provides an overview of security vulnerabilities and protection mechanisms for the GMPLS control plane.

6. IANA Considerations

Upon approval of this document, IANA is requested to make the assignment of a new value for the existing "Attributes TLV Space" registry located at <http://www.iana.org/assignments/rsvp-te-parameters/>, as updated by [RSVP-RO]:

Type	Name	Allowed on LSP ATTRIBUTES	Allowed on LSP REQUIRED ATTRIBUTES	Allowed on RO LSP Attribute Subobject	Reference
TBA	WSON Processing Hop Attribute TLV	No	No	Yes	[This.I-D]

Upon approval of this document, IANA is requested to create a new registry named "Sub-TLV Types for WSON Processing Hop Attribute TLV" located at <http://www.iana.org/assignments/rsvp-te-parameters/>.

The following entries are to be added:

Value	Sub-TLV Type	Reference
1	ResourceBlockInfo	[This.I-D]
2	WavelengthSelection	[This.I-D]

All assignments are to be performed via Standards Action or Specification Required policies as defined in [RFC5226 <<http://tools.ietf.org/html/rfc5226>>].

Upon approval of this document, IANA is requested to create a new registry named "Values for Wavelength Assignment Method field in WavelengthSelection Sub-TLV" located at <http://www.iana.org/assignments/rsvp-te-parameters/>.

The following entries are to be added:

Value	Meaning	Reference
0	unspecified	[This.I-D]
1	First-Fit	[This.I-D]

2	Random	[This.I-D]
3	Least-Loaded (multi-fiber)	[This.I-D]
4-127	unassigned	

All assignments are to be performed via Standards Action or Specification Required policies as defined in [RFC5226]. The assignment policy chosen for any specific code point must be clearly stated in the document that describes the code point so that IANA can apply the correct policy.

Upon approval of this document, IANA is requested to make the assignment of a new value for the existing "Sub-Codes . 24 Routing Problem" registry located at <http://www.iana.org/assignments/rsvp-parameters/>:

Value	Description	Reference
107	Unsupported WavelengthSelection symmetry value	[This.I-D]
108	Unsupported Wavelength Assignment value	[This.I-D]

7. Acknowledgments

Authors would like to thanks Lou Berger, Cyril Margaria and Xian Zhang for comments and suggestions.

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.
- [RFC5226] Narten, T., H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", RFC 5226, May 2008.

- [RFC6205] T. Otani, H. Guo, K. Miyazaki, D. Caviglia, "Generalized Labels for Lambda-Switch-Capable Label Switching Routers", RFC 6205, March 2011.
- [WSON-Encode] Bernstein G., Lee Y., Li D., and W. Imajuku, "Routing and Wavelength Assignment Information Encoding for Wavelength Switched Optical Networks", draft-ietf-ccamp-rwa-wson-encode, work in progress.
- [WSON-OSPF] Lee, Y, Bernstein G., "GMPLS OSPF Enhancement for Signal and Network Element Compatibility for Wavelength Switched Optical Networks", draft-ietf-ccamp-wson-signal-compatibility-ospf, work in progress.
- [RFC5511] Farrel, A., "Routing Backus-Naur Form (RBNF): A Syntax Used to Form Encoding Rules in Various Routing Protocol Specifications", RFC 5511, April 2009.
- [RFC3209] Awduche, D., et al., "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.
- [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 2009.
- [RSVP-RO] Margaria, C., et al, "LSP Attribute in ERO", draft-ietf-ccamp-lsp-attribute-ro, work in progress.

8.2. Informative References

- [RFC5920] Fang, L., Ed., "Security Framework for MPLS and GMPLS

Networks", RFC 5920, July 2010.

[RFC6163] Y. Lee, G. Bernstein, W. Imajuku, "Framework for GMPLS and PCE Control of Wavelength Switched Optical Networks", RFC 6163, February 2008.

[RFC7446] G. Bernstein, Y. Lee, D. Li, W. Imajuku, "Routing and Wavelength Assignment Information Model for Wavelength Switched Optical Networks", RFC 7446, February 2015.

9. Contributors

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

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

Authors' Addresses

Greg M. Bernstein (editor)
Grotto Networking
Fremont California, USA

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

Young Lee (editor)
Huawei Technologies
5340 Legacy Dr. Building 3
Plano, TX 75024
USA

Phone: (469) 277-5838
Email: leeyoung@huawei.com

Sugang Xu
National Institute of Information and Communications Technology
4-2-1 Nukui-Kitamachi, Koganei,
Tokyo, 184-8795 Japan

Phone: +81 42-327-6927
Email: xsg@nict.go.jp

Giovanni Martinelli
Cisco
Via Philips 12
20052 Monza, IT

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

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

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: December 19, 2014

Z. Li
M. Chen
Huawei
G. Mirsky
Ericsson
June 17, 2014

Routing Extensions for Discovery of Role-based MPLS Label Switching
Router (MPLS LSR) Traffic Engineering (TE) Mesh Membership
draft-li-ccamp-role-based-automesh-02

Abstract

A Traffic Engineering (TE) mesh-group is defined as a group of Label Switch Routers (LSRs) that are connected by a full mesh of TE LSPs. Routing (OSPF and IS-IS) extensions for discovery Multiprotocol Label Switching (MPLS) LSR TE mesh membership has been defined to automate the creation of mesh of TE LSPs.

This document introduces a role-based TE mesh-group that applies to the scenarios where full mesh TE LSPs is not necessary and TE LSPs setup depends on the roles of LSRs in a TE mesh-group. Interior Gateway Protocol (IGP) routing extensions for automatic discovery of role-based TE mesh membership are defined accordingly.

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].

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 December 19, 2014.

Copyright Notice

Copyright (c) 2014 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	2
1.1. Terminology	3
2. Role-based TE Mesh Group	4
3. IGP Role-based TE Mesh-group Extensions	4
3.1. OSPF TE-ROLE-MESH-GROUP TLV Format	4
3.2. IS-IS TE-ROLE-MESH-GROUP Sub-TLV Format	7
4. Elements of Procedure	10
4.1. OSPF	10
4.2. IS-IS	11
5. Backward Compatibility	12
6. IANA Considerations	12
6.1. OSPF	12
6.2. IS-IS	13
7. Security Considerations	13
8. Acknowledgements	13
9. References	13
9.1. Normative References	13
9.2. Informative References	14
Authors' Addresses	14

1. Introduction

A TE mesh-group [RFC4972] is defined as a group of LSRs that are connected by a full mesh of TE LSPs. [RFC4972] specifies Intermediate System-to-Intermediate System (IS-IS) and Open Shortest Path First (OSPF) extensions to provide an automatic discovery of the set of LSR members of a TE mesh-group in order to automate the creation of such mesh of TE LSPs. This is called "auto-mesh TE" or "auto-mesh". The auto-mesh TE significantly simplifies the configuration and deployment of TE LSPs.

In some scenarios, it may not be necessary to establish full mesh TE LSPs among all the LSRs of a TE mesh-group. An example of the use case of non-full mesh of TE LSPs in the mobile backhaul (MBH) networks is presented in ([I-D.li-mpls-seamless-mpls-mbb]). In MBH network TE LSPs are usually setup between the Cell Site Gateways(CSGs) and the Radio Network Controller (RNC) Site Gateways(RSGs). TE LSPs interconnecting CSGs and TE LSPs interconnecting RSGs are not necessary. In most deployments the number of CSGs is very large and there are much more CSGs than RSGs in an MBH domain. With the auto-mesh mechanism defined[RFC4972] full mesh of TE LSPs will be established interconnecting CSGs and RSGs. As result large number of unnecessary TE LSPs will be established interconnecting CSGs and interconnecting RSGs. This likely will not scale well with addition of more CSG devices, would stress control plane with unwarranted RSVP state.

Thus there are requirements to optimize the auto-mesh TE and to reduce the number of unnecessary TE LSPs. This document introduces a "role-based auto-mesh TE" or "role-based auto-mesh" where the setup of the TE LSPs is based on the role of the LSRs within a particular TE mesh-group. Therefore, besides the discovery of the membership of a TE mesh-group, it needs to discover the role of each node in the TE mesh-group.

Another scenario to which the role-based auto-mesh TE can apply is the Resource Reservation Protocol-Traffic Engineering (RSVP-TE) Point-to-Multipoint (P2MP) TE LSP[RFC4875] scenario. For a RSVP-TE P2MP TE LSP, the root LSR has to know all the leaf LSRs before signalling the P2MP TE LSP. The automatic discovery mechanisms defined in this document can be used to discover the leaf LSRs for P2MP TE LSPs.

This document defines IGP routing extensions to automatically discover of the members and their roles of a TE mesh-group.

1.1. Terminology

RSVP-TE - Resource Reservation Protocol-Traffic Engineering

P2MP - Point-to-Multipoint

IS-IS - Intermediate System-to-Intermediate System

OSPF - Open Shortest Path First

CSG - Cell Site Gateway

RNC - Radio Network Controller

MBH - Mobile Backhaul

MPLS - Multiprotocol Label Switching

LSP - Label Switched Path

TE LSP - Traffic Engineered LSP

2. Role-based TE Mesh Group

A role-based TE mesh-group is a special TE mesh-group where TE LSPs will not be established among all member LSRs. In a role-based TE mesh-group LSRs will have different roles. TE LSPs setup depends on the roles of the LSRs in a TE mesh-group. This document introduces two types of role-based TE mesh group: Hub-Spoke and Root-Leaf.

For a Hub-Spoke TE mesh-group, an LSR can be a Hub, Spoke or both Hub and Spoke LSR in a group. The rules for Hub-Spoke TE mesh-group are as follows:

TE LSPs SHOULD only be setup between Spoke and Hub LSRs.

TE LSPs MUST NOT be setup between/among Spoke LSRs.

TE LSPs MUST NOT be setup between/among Hub LSRs.

For a Root-Leaf TE mesh-group, an LSR can be a Root, a Leaf or both a Root and Leaf LSR. Once the membership and roles are determined, the root LSRs can signal the P2MP TE LSPs toward all the Leaf LSRs. There may be multiple P2MP TE LSPs within a TE mesh-group.

3. IGP Role-based TE Mesh-group Extensions

3.1. OSPF TE-ROLE-MESH-GROUP TLV Format

The OSPF TE-ROLE-MESH-GROUP TLV is used to advertise that an LSR joins/leaves a role-based TE mesh-group and the role of the LSR in the TE mesh-group. The OSPF TE-ROLE-MESH-GROUP TLV format for IPv4 (Figure 2) and IPv6 (Figure 3) is as follows:

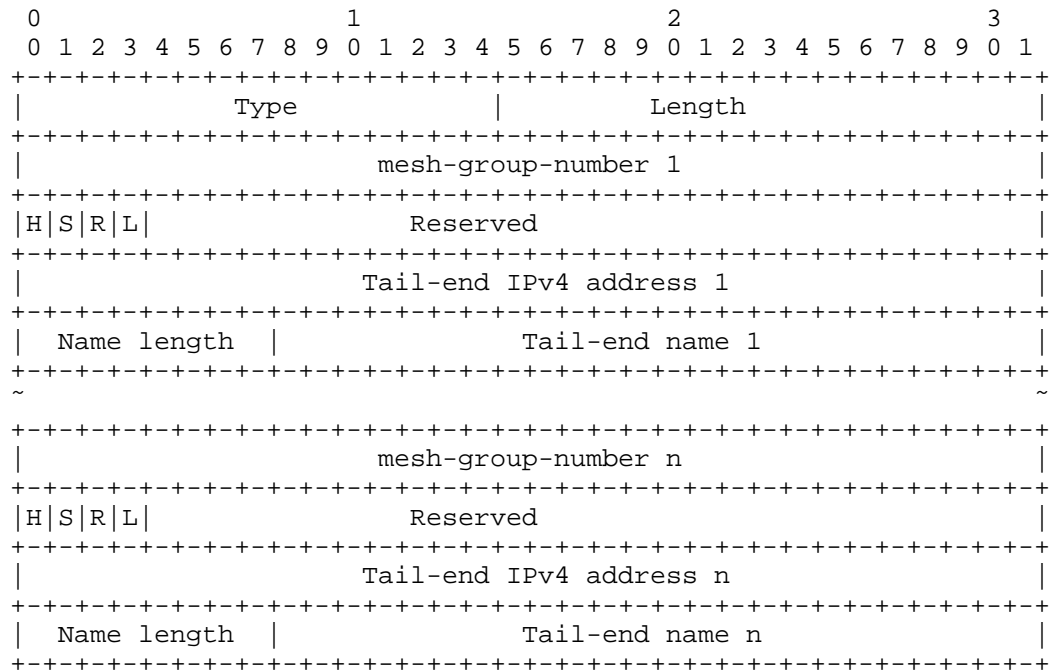


Figure 2 - OSPF TE-ROLE-MESH-GROUP TLV format (IPv4 Address)

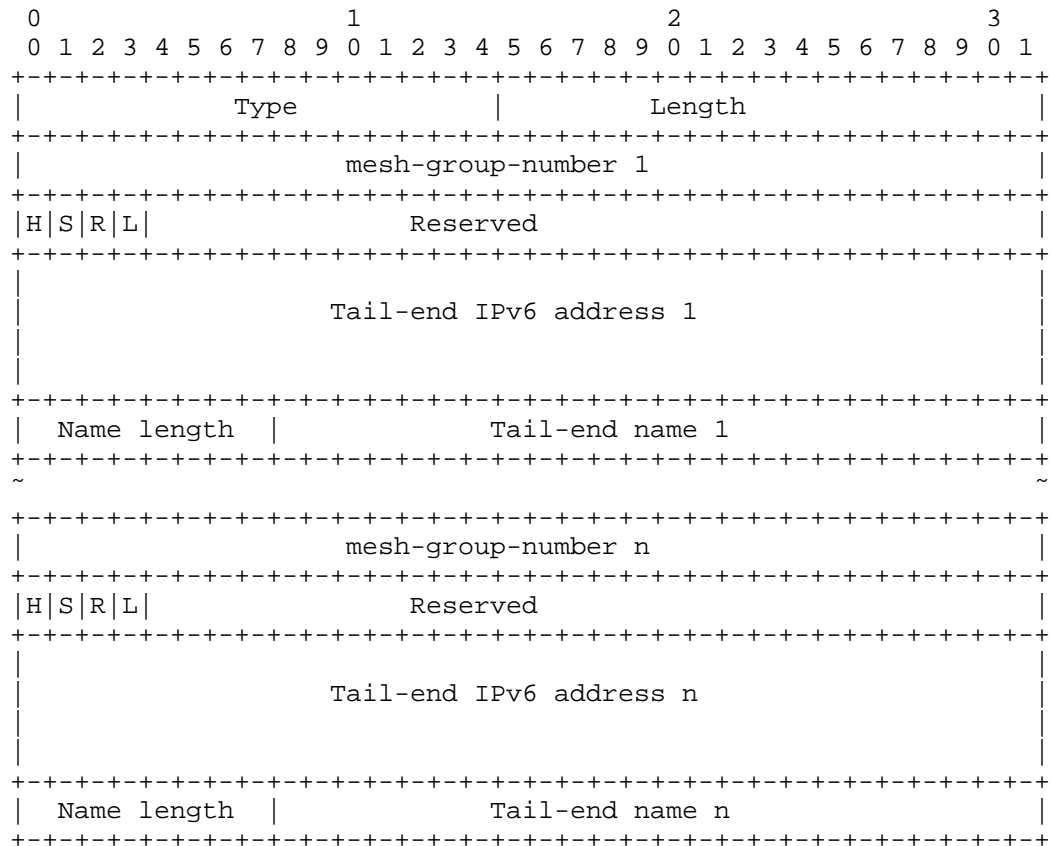


Figure 3 - OSPF TE-ROLE-MESH-GROUP TLV format (IPv6 Address)

The Type of OSPF TE-ROLE-MESH-GROUP TLV for IPv4 is TBD1, the value of the Length is variable.

The Type of OSPF TE-ROLE-MESH-GROUP TLV for IPv6 is TBD2, the value of the Length is variable.

The OSPF TE-ROLE-MESH-GROUP TLV may contain one or more role-based mesh-group entries. Each entry corresponds to a role-based TE mesh-group. The definition of the mesh-group-number, the Tail-end address, the Name length and the Tail-end name in each role-based mesh group entry is the same as that of OSPF TE-MESH-GROUP TLV defined in [RFC4972].

In addition, for each mesh group entry, an four-octet flag field is introduced and four flags are defined in this document. Other bits

are reserved for future use and MUST be set to zero when sent, and MUST be ignored when received.

The H (Hub) bit, when set, it indicates the LSR is a Hub LSR.

The S (Spoke) bit, when set, it indicates the LSR is a Spoke LSR.

The R (Root) bit, when set, it indicates an LSR is a Root LSR.

The L (Leaf) bit, when set, it indicates an LSR is a Leaf.

The H and S bit are dedicated for Hub-Spoke TE mesh-group and can be both set. When both bits set, it indicates that an LSR has both the Hub and Spoke role in the group.

The R and Leaf bit can be both set, when both bits set, it indicates an LSR is a Root and Leaf LSR. The R bit and Leaf bit are only used for Root-Leaf TE mesh-group, for other TE mesh-groups, it MUST be set to zero and MUST be ignored when received.

3.2. IS-IS TE-ROLE-MESH-GROUP Sub-TLV Format

The IS-IS TE-ROLE-MESH-GROUP sub-TLV is used to advertise that an LSR joins/leaves a TE mesh-group and the role of the LSR in the TE mesh-group. The IS-IS TE-ROLE-MESH-GROUP sub-TLV format for IPv4 (Figure 4) and IPv6 (Figure 5) is as follows:

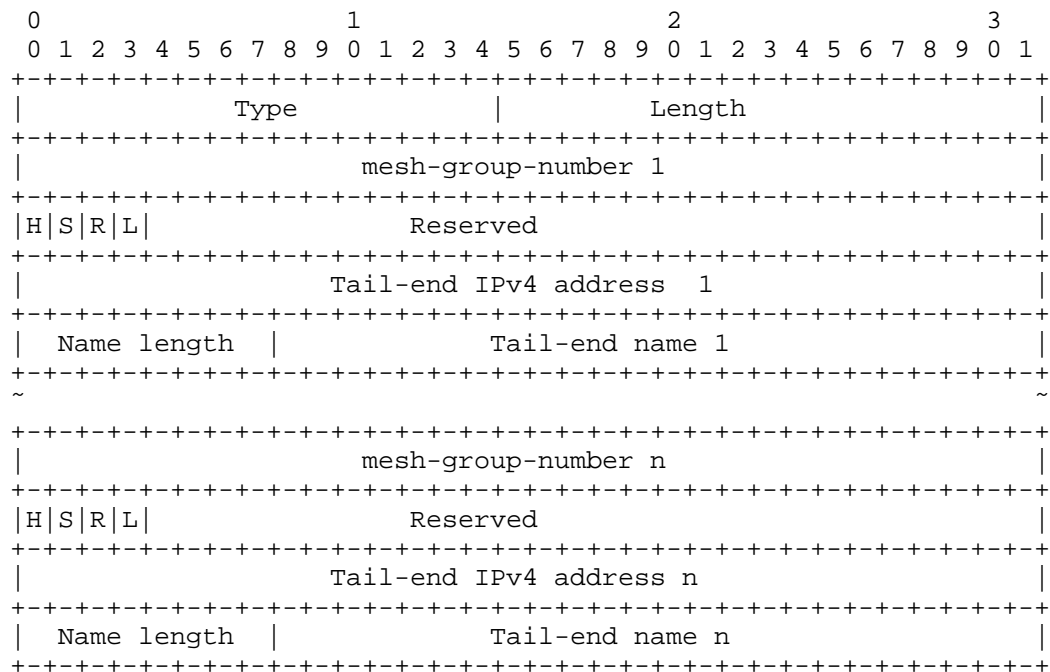


Figure 4 - IS-IS TE-ROLE-MESH-GROUP sub-TLV format (IPv4 Address)

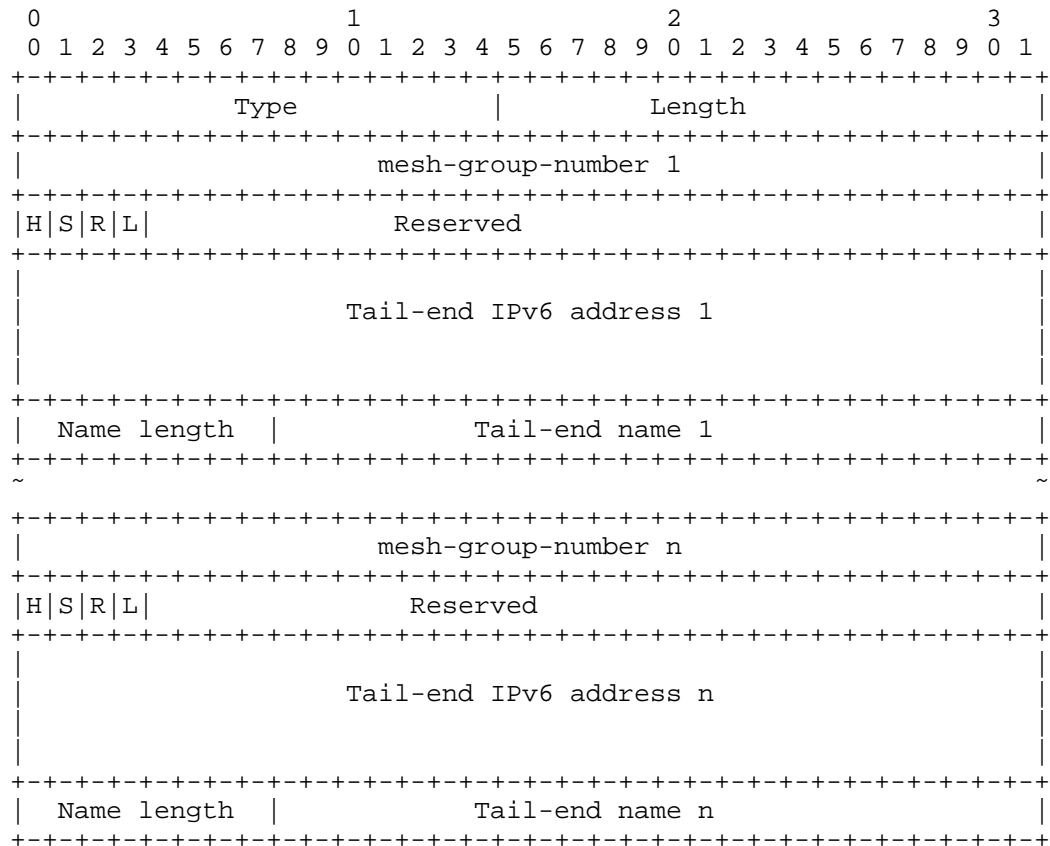


Figure 5 - IS-IS TE-ROLE-MESH-GROUP sub-TLV format (IPv6 Address)

The Type of IS-IS TE-ROLE-MESH-GROUP sub-TLV for IPv4 is TBD3, the value of the Length is variable.

The Type of IS-IS TE-ROLE-MESH-GROUP sub-TLV for IPv6 is TBD4, the value of the Length is variable.

The IS-IS Role-based TE-ROLE-MESH-GROUP sub-TLV may contain one or more role-based mesh-group entries. Each entry corresponds to a role-based TE mesh-group. The definition of the fields, mesh-group-number, Tail-end address, Name length and Tail-end name in each role-based mesh group entry is the same as that of IS-IS TE-MESH-GROUP sub-TLV defined in [RFC4972].

The H, S, R and L bits are defined as in Section 3.1 of this document.

4. Elements of Procedure

The OSPF TE-ROLE-MESH-GROUP TLV is carried within the OSPF Routing Information LSA, and the IS-IS TE-ROLE-MESH-GROUP sub-TLV is carried within the IS-IS Router capability TLV. As such, elements of procedure are inherited from those defined in [RFC4970] and [RFC4971] for OSPF and IS-IS respectively. Specifically, a router **MUST** originate a new LSA/LSP whenever the content of this information changes, or whenever required by regular routing procedure (e.g., updates).

The TE-ROLE-MESH-GROUP TLV is **OPTIONAL** and **MUST NOT** include more than one of each of the IPv4 instances or the IPv6 instance. If either the IPv4 or the IPv6 OSPF TE-ROLE-MESH-GROUP TLV occurs more than once within the OSPF Router Information LSA, only the first instance is processed, subsequent TLV(s) **MUST** be ignored. Similarly, if either the IPv4 or the IPv6 IS-IS TE-ROLE-MESH-GROUP sub-TLV occurs more than once within the IS-IS Router capability TLV, only the first instance is processed, subsequent TLV(s) **MUST** be ignored.

4.1. OSPF

The TE-ROLE-MESH-GROUP TLV is advertised within an OSPF Router Information opaque LSA (opaque type of 4, opaque ID of 0) for OSPFv2 [RFC2328] and within a new LSA (Router Information LSA) for OSPFv3 [RFC5340]. The Router Information LSAs for OSPFv2 and OSPFv3 are defined in [RFC4970].

A router **MUST** originate a new OSPF router information LSA whenever the content of any of the advertised TLV changes or whenever required by the regular OSPF procedure (LSA update (every LSRefreshTime)). If an LSR desires to join or leave a particular role-based TE mesh group or an LSR desires to change its role in a mesh group, it **MUST** originate a new OSPF Router Information LSA comprising the updated TE-ROLE-MESH-GROUP TLV. In the case of a join, a new entry will be added to the TE-ROLE-MESH-GROUP TLV; if the LSR leaves a mesh-group, the corresponding entry will be removed from the TE-ROLE-MESH-GROUP TLV; if the LSR changes its role in the role-based mesh group, the corresponding entry will be updated in the TE-ROLE-MESH-GROUP TLV. Note that these operations can be performed in the context of a single LSA update. An implementation **SHOULD** be able to detect any change to a previously received TE-ROLE-MESH-GROUP TLV from a specific LSR.

As defined in [RFC5250] for OSPFv2 and in [RFC5340] for OSPFv3, the flooding scope of the Router Information LSA is determined by the LSA Opaque type for OSPFv2 and the values of the S1/S2 bits for OSPFv3.

For OSPFv2 Router Information opaque LSA:

- Link-local scope: type 9;
- Area-local scope: type 10;
- Routing-domain scope: type 11. In this case, the flooding scope is equivalent to the Type 5 LSA flooding scope.

For OSPFv3 Router Information LSA:

- Link-local scope: OSPFv3 Router Information LSA with the S1 and S2 bits cleared;
- Area-local scope: OSPFv3 Router Information LSA with the S1 bit set and the S2 bit cleared;
- Routing-domain scope: OSPFv3 Router Information LSA with S1 bit cleared and the S2 bit set.

A router may generate multiple OSPF Router Information LSAs with different flooding scopes.

The Role-based TE-MESH-GROUP TLV may be advertised within an Area-local or Routing-domain scope Router Information LSA, depending on the MPLS TE mesh group profile:

- If the MPLS TE mesh-group is contained within a single area (all the LSRs of the mesh-group are contained within a single area), the TE-ROLE-MESH-GROUP TLV MUST be generated within an Area-local Router Information LSA.
- If the MPLS TE mesh-group spans multiple OSPF areas, the TE Role-based mesh-group TLV MUST be generated within a Routing-domain scope router information LSA.

When the router receives TE-ROLE-MESH-GROUP TLV, it SHOULD setup MPLS TE LSPs according rules which defined in the Section 3.

4.2. IS-IS

The TE-ROLE-MESH-GROUP sub-TLV is advertised within the IS-IS Router CAPABILITY TLV defined in [RFC4971].

An IS-IS router MUST originate a new IS-IS LSP whenever the content of any of the advertised sub-TLV changes or whenever required by regular IS-IS procedure (LSP updates). If an LSR desires to join or leave a particular role-based TE mesh group or an LSR desires to

change its role in a mesh group, it MUST originate a new LSP comprising the refreshed IS-IS Router capability TLV comprising the updated TE-ROLE-MESH-GROUP sub-TLV. In the case of a join, a new entry will be added to the TE-ROLE-MESH-GROUP sub-TLV; if the LSR leaves a mesh-group, the corresponding entry will be deleted from the TE-ROLE-MESH-GROUP sub-TLV; if the LSR changes its role in the role-based mesh group, the corresponding entry will be updated in the TE-ROLE-MESH-GROUP sub-TLV. Note that these operations can be performed in the context of a single update. An implementation SHOULD be able to detect any change to a previously received TE-ROLE-MESH-GROUP sub-TLV from a specific LSR.

If the flooding scope of a TE-ROLE-MESH-GROUP sub-TLV is limited to an IS-IS level/area, the sub-TLV MUST NOT be leaked across level/area and the S flag of the Router CAPABILITY TLV MUST be cleared. Conversely, if the flooding scope of a TE-ROLE-MESH-GROUP sub-TLV is the entire routing domain, the TLV MUST be leaked across IS-IS levels/areas, and the S flag of the Router CAPABILITY TLV MUST be set. In both cases, the flooding rules specified in [RFC4971] apply.

As specified in [RFC4971], a router may generate multiple IS-IS Router CAPABILITY TLVs within an IS-IS LSP with different flooding scopes.

When the router receives TE-ROLE-MESH-GROUP sub-TLV, it SHOULD setup MPLS TE LSPs according rules which defined in the Section 3.

5. Backward Compatibility

For a role-based TE mesh-group, if there are some LSRs only supporting mechanisms defined [RFC4972], all the LSRs of the mesh-group MUST process as defined in [RFC4972]. The operators should avoid to add an LSR that does not support role-based auto-mesh TE to a role-based TE mesh-group.

6. IANA Considerations

6.1. OSPF

The registry for the Router Information LSA is defined in [RFC4970]. IANA assigned a new OSPF TLV code-point for the TE-ROLE-MESH-GROUP TLVs carried within the Router Information LSA.

Value	TLV	References
-----	-----	-----
TBD1	TE-ROLE-MESH-GROUP TLV (IPv4)	this document
TBD2	TE-ROLE-MESH-GROUP TLV (IPv6)	this document

6.2. IS-IS

The registry for the Router Capability TLV is defined in [RFC4971]. IANA assigned a new IS-IS sub-TLV code-point for the TE-ROLE-MESH-GROUP sub-TLVs carried within the IS-IS Router Capability TLV.

Value -----	Sub-TLV -----	References -----
TBD3	TE-ROLE-MESH-GROUP sub-TLV (IPv4)	this document
TBD4	TE-ROLE-MESH-GROUP sub-TLV (IPv6)	this document

7. Security Considerations

The function described in this document does not create any new security issues for the OSPF and IS-IS protocols, the security considerations described in [RFC4972] apply here.

8. Acknowledgements

The authors would like to thank Loa Andersson for his valuable comments.

The authors would also like to thank the authors of [RFC4972] from where we have taken most of the elements procedures.

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.
- [RFC2328] Moy, J., "OSPF Version 2", STD 54, RFC 2328, April 1998.
- [RFC4970] Lindem, A., Shen, N., Vasseur, JP., Aggarwal, R., and S. Shaffer, "Extensions to OSPF for Advertising Optional Router Capabilities", RFC 4970, July 2007.
- [RFC4971] Vasseur, JP., Shen, N., and R. Aggarwal, "Intermediate System to Intermediate System (IS-IS) Extensions for Advertising Router Information", RFC 4971, July 2007.
- [RFC4972] Vasseur, JP., Leroux, JL., Yasukawa, S., Previdi, S., Psenak, P., and P. Mabbey, "Routing Extensions for Discovery of Multiprotocol (MPLS) Label Switch Router (LSR) Traffic Engineering (TE) Mesh Membership", RFC 4972, July 2007.

- [RFC5250] Berger, L., Bryskin, I., Zinin, A., and R. Coltun, "The OSPF Opaque LSA Option", RFC 5250, July 2008.
- [RFC5340] Coltun, R., Ferguson, D., Moy, J., and A. Lindem, "OSPF for IPv6", RFC 5340, July 2008.

9.2. Informative References

- [I-D.li-mpls-seamless-mpls-mbb]
Li, Z., Li, L., Morillo, M., and T. Yang, "Seamless MPLS for Mobile Backhaul", draft-li-mpls-seamless-mpls-mbb-01 (work in progress), February 2014.
- [RFC4875] Aggarwal, R., Papadimitriou, D., and S. Yasukawa, "Extensions to Resource Reservation Protocol - Traffic Engineering (RSVP-TE) for Point-to-Multipoint TE Label Switched Paths (LSPs)", RFC 4875, May 2007.

Authors' Addresses

Zhenbin Li
Huawei

Email: lizhenbin@huawei.com

Mach(Guoyi) Chen
Huawei

Email: mach.chen@huawei.com

Greg Mirsky
Ericsson

Email: gregory.mirsky@ericsson.com

Network Working Group
Internet Draft
Intended status: Standards Track

H. Long, M.Ye
Huawei Technologies Co., Ltd
G. Mirsky
Ericsson
A Alessandro
Telecom Italia S.p.A
H. Shah
Ciena
July 4, 2014

Expires: January 2015

OSPF Routing Extension for Links with Variable Discrete Bandwidth
draft-long-ccamp-ospf-availability-extension-04.txt

Abstract

Packet switching network MAY contain links with variable discrete bandwidth, e.g., copper, radio, etc. The bandwidth of such link MAY change discretely in reaction to changing external environment. Availability is typically used for describing such links during network planning. This document describes an extension for OSPF routing for route computation in a Packet Switched Network (PSN) which contains links with variable discrete bandwidth by introducing an OPTIONAL Availability sub-TLV.

Status of this Memo

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

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

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

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

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

This Internet-Draft will expire on January 4, 2014.

Copyright Notice

Copyright (c) 2014 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. Overview	3
3. Extension to OSPF Routing Protocol.....	4
3.1. Interface Switching Capacity Descriptor	4
3.2. ISCD Availability sub-TLV.....	5
3.3. Signaling Process.....	6
4. Security Considerations.....	6
5. IANA Considerations	6
6. References	6
6.1. Normative References.....	6
6.2. Informative References.....	7
7. Acknowledgments	7

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 following acronyms are used in this draft:

OSPF	Open Shortest Path First
PSN	Packet Switched Network
SNR	Signal-to-noise Ratio
LSP	Label Switched Path
ISCD	Interface Switching Capacity Descriptor

PE Provider Edge

LSA Link State Advertisement

1. Introduction

Some data communication technologies allow seamless change of maximum physical bandwidth through a set of known discrete values. For example, in mobile backhaul network, microwave links are very popular for providing connection of last hops. In case of heavy rain, to maintain the link connectivity, the microwave link MAY lower the modulation level since demodulating lower modulation level need lower signal-to-noise ratio (SNR). This is called adaptive modulation technology [EN 302 217]. However, lower modulation level also means lower link bandwidth. When link bandwidth reduced because of modulation down-shifting, high priority traffic can be maintained, while lower priority traffic is dropped. Similarly the cooper links MAY change their effective link bandwidth due to external interference.

The parameter availability [G.827, F.1703, P.530] is often used to describe the link capacity during network planning. Assigning different availability classes to different types of service over such kind of links provides more efficient planning of link capacity. To set up an LSP across these links, availability information is required for the nodes to verify bandwidth satisfaction and make bandwidth reservation. The availability information SHOULD be inherited from the availability requirements of the services expected to be carried on the LSP. For example, voice service usually needs ''five nines'' availability, while non-real time services MAY adequately perform at four or three nines availability.

For the route computation, the availability information SHOULD be provided along with bandwidth resource information. In this document, an extension on Interface Switching Capacity Descriptor (ISCD) [RFC4202] for availability information is defined to support in routing signaling. The extension reuses the reserved field in the ISCD and also introduces an OPTIONAL Availability sub-TLV.

If there is a hop that cannot support the Availability sub-TLV, the Availability sub-TLV SHOULD be ignored.

2. Overview

A node which has link(s) with variable bandwidth attached SHOULD contain a <bandwidth, availability> information list in its OSPF TE LSA messages. The list provides the information that how much

bandwidth a link can support for a specified availability. This information is used for path calculation by the PE node(s).

To setup an label switching path (LSP), a PE node MAY collect link information which is spread in OSPF TE LSA messages by network nodes to get know about the network topology, calculate out an LSP route based on the network topology and send the calculated LSP route to signaling to initiate a PATH/RESV message for setting up the LSP.

Availability information is required to carry in the signaling message to better utilize the link bandwidth. The signaling extension for availability can be found in [ASTE].

3. Extension to OSPF Routing Protocol

3.1. Interface Switching Capacity Descriptor

The Interface Switching Capacity Descriptor (ISCD) sub-TLV [RFC 4203] has the following 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								
Type										Length																													
Switching Cap										Encoding										AI										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																																							
Switching Capacity-specific Information																																							
(variable)																																							

A new AI field is defined in this document.

AI: ISCD Availability sub-TLV index, 8 bits

This new field is the index of Availability sub-TLV for this ISCD sub-TLV.

3.2. ISCD Availability sub-TLV

When the Switching Capability field is PSC-1, PSC-2, PSC-3, PSC-4, the Switching Capability specific information field MAY include one or more ISCD Availability sub-TLV(s). The ISCD Availability sub-TLV has the following 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	2	3	4	5	6	7	8	9																								
Type																Length																																															
Index																Reserved																																															
Availability level																																																															
LSP Bandwidth at Availability level n																																																															

Type: 0x01, 16 bits;

Length: 16 bits;

Index: 8 bits

This field is the index of this Availability sub-TLV, referred by the AI field of the ISCD sub-TLV.

Availability level: 32 bits

This field is a 32-bit IEEE floating point number which describes the decimal value of availability guarantee of the switching capacity in the ISCD object which has the AI value equal to Index of this sub-TLV. The value MUST be less than 1.

LSP Bandwidth at Availability level n: 32 bits

This field is a 32-bit IEEE floating point number which describes the LSP Bandwidth at a certain Availability level which was described in the Availability field.

3.3. Signaling Process

A node which has link(s) with variable bandwidth attached SHOULD contain one or more ISCD Availability sub-TLVs in its OSPF TE LSA messages. Each ISCD Availability sub-TLV provides the information about how much bandwidth a link can support for a specified availability. This information SHOULD be used for path calculation by the PE node(s).

A node that doesn't support ISCD Availability sub-TLV SHOULD ignore ISCD Availability sub-TLV.

4. Security Considerations

This document does not introduce new security considerations to the existing OSPF protocol.

5. IANA Considerations

This document introduces an Availability sub-TLV of the ISCD sub-TLV of the TE Link TLV in the TE Opaque LSA for OSPF v2. This document proposes a suggested value for the Availability sub-TLV; it is recommended that the suggested value be granted by IANA. Initial values are as follows:

Type	Length	Format	Description
---	----	-----	-----
0	-	Reserved	Reserved value
0x01	8	see Section 3.2	Availability sub-TLV

6. References

6.1. Normative References

- [RFC2210] Wroclawski, J., ''The Use of RSVP with IETF Integrated Services'', RFC 2210, September 1997.
- [RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, December 2001.

- [RFC3473] Berger, L., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions", RFC 3473, January 2003.
- [RFC4202] Kompella, K. and Rekhter, Y. (Editors), "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.
- [G.827] ITU-T Recommendation, "Availability performance parameters and objectives for end-to-end international constant bit-rate digital paths", September, 2003.
- [F.1703] ITU-R Recommendation, "Availability objectives for real digital fixed wireless links used in 27 500 km hypothetical reference paths and connections", January, 2005.
- [P.530] ITU-R Recommendation, "Propagation data and prediction methods required for the design of terrestrial line-of-sight systems", February, 2012
- [EN 302 217] ETSI standard, "Fixed Radio Systems; Characteristics and requirements for point-to-point equipment and antennas", April, 2009
- [ASTE] H., Long, M., Ye, Mirsky, G., Alessandro, A., Shah, H., "RSVP-TE Signaling Extension for Links with Variable Discrete Bandwidth", Work in Progress, February, 2014

6.2. Informative References

- [MCOS] Minei, I., Gan, D., Kompella, K., and X. Li, "Extensions for Differentiated Services-aware Traffic Engineered LSPs", Work in Progress, June 2006.

7. Acknowledgments

Authors' Addresses

Hao Long
Huawei Technologies Co., Ltd.
No.1899, Xiyuan Avenue, Hi-tech Western District
Chengdu 611731, P.R.China

Phone: +86-18615778750
Email: longhao@huawei.com

Min Ye
Huawei Technologies Co., Ltd.
No.1899, Xiyuan Avenue, Hi-tech Western District
Chengdu 611731, P.R.China

Email: amy.yemin@huawei.com

Greg Mirsky
Ericsson

Email: gregory.mirsky@ericsson.com

Alessandro D'Alessandro
Telecom Italia S.p.A

Email: alessandro.dalessandro@telecomitalia.it

Himanshu Shah
Ciena Corp.
3939 North First Street
San Jose, CA 95134
US

Email: hshah@ciena.com

CCAMP
Internet-Draft
Intended status: Informational
Expires: February 13, 2015

G. Martinelli, Ed.
Cisco
X. Zhang, Ed.
Huawei Technologies
G. Galimberti
Cisco
A. Zanardi
D. Siracusa
F. Pederzoli
CREATE-NET
Y. Lee
F. Zhang
Huawei Technologies
August 12, 2014

Information Model for Wavelength Switched Optical Networks (WSONs) with
Impairments Validation
draft-martinelli-ccamp-wson-iv-info-05

Abstract

This document defines an information model to support Impairment-Aware (IA) Routing and Wavelength Assignment (RWA) functionality. This information model extends the information model for impairment-free RWA process in WSON to facilitate computation of paths where optical impairment constraints need to be considered.

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 February 13, 2015.

Copyright Notice

Copyright (c) 2014 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. Definitions, Applicability and Properties	3
2.1. Definitions	4
2.2. Applicability	4
2.3. Properties	5
3. ITU-T List of Optical Parameters	6
4. Background from WSON-RWA Information Model	8
5. Optical Impairment Information Model	9
5.1. The Optical Impairment Vector	10
5.2. Node Information	10
5.2.1. Impairment Matrix	10
5.2.2. Impairment Resource Block Information	13
5.3. Link Information	13
5.4. Path Information	13
6. Encoding Considerations	14
7. Control Plane Architectures	14
7.1. IV-Centralized	15
7.2. IV-Distributed	15
8. Acknowledgements	15
9. IANA Considerations	15
10. Security Considerations	16
11. References	16
11.1. Normative References	16
11.2. Informative References	16
Appendix A. FAQ	17
A.1. Why the Application Code does not suffice for Optical Impairment Validation?	17
A.2. Are DWDM network multivendor?	17
Authors' Addresses	17

1. Introduction

In the context of Wavelength Switched Optical Network (WSON), [RFC6163] describes the basic framework for a GMPLS and PCE-based Routing and Wavelength Assignment (RWA) control plane. The associated information model [I-D.ietf-ccamp-rwa-info] defines information/parameters required by an RWA process without optical impairment considerations.

There are cases of WSON where optical impairments play a significant role and are considered as important constraints. The framework document [RFC6566] defines the problem scope and related control plane architectural options for the Impairment Aware RWA (IA-RWA) operation. Options include different combinations of Impairment Validation (IV) and RWA functions in term of different combination of control plane functions (i.e., PCE, Routing, Signaling).

A Control Plane with RWA-IA will not be able to solve the optical impairment problem in a detailed and exhaustive way, however, it may take advantage of some data plane knowledge to make better decisions during its path computing phase. The final outcome will be a path, instantiated through a wavelength in the data plane, that has a "better chance" to work than that path were calculated without IA information. "Better chance" means that path setup may still fail and the GMPLS control plane will follow its usual procedures upon errors and failures. A control plane will not replace a the network design phase that remains a fundamental step for DWDM Optical Networks. As the non-linear impairments which need to be considered in the calculation of an optical path will be vendor-dependent, the parameters considered in this document is not an exhaustive list.

This document provides an information model for the impairment aware case to allow the impairment validation function implemented in the control plane or enabled by control plane available information. This model goes in addition to [I-D.ietf-ccamp-rwa-info] and shall support any control plane architectural option described by the framework document (see sections 4.2 and 4.3 of [RFC6566]) where a set of combinations of control plane functions vs. IV function is provided.

2. Definitions, Applicability and Properties

This section provides some concepts to help understand the model and to make a clear separation from data plane definitions (ITU-T recommendations). The first sub-section provides definitions while the Applicability sections uses the defined definitions to scope this document.

2.1. Definitions

- o Computational Model / Optical Computational Model.
Defined by ITU standard documents. In this context we look for models able to compute optical impairments for a given lightpath.
- o Information Model.
Defined by IETF (this document) and provides the set of information required by control plane to apply the Computational Model.
- o Level of Approximation.
This concept refers to the Computational Model as it may compute optical impairment with a certain level of uncertainty. This level is generally not measured but [RFC6566] Section 4.1.1 provides a rough classification about it.
- o Feasible Path.
It is the output of the C-SPF with RWA-IV capability. It's an optical path that satisfies optical impairment constraints. The path, instantiated through wavelength(s), may actually work or not work depending of the level of approximation.
- o Existing Service Disruption.
An effect known to optical network designers is the cross-interaction among spectrally adjacent wavelengths: an existing wavelength may experience increased BER due to the setup of an adjacent wavelength. Solving this problem is a typical optical network design activity. Just as an example, a simple solution is adding optical margins (e.g., additional OSNR), although complex and detailed methods exist.
- o DWDM Line Segments.
[ITU.G680] provides definition and picture for the "Situation 1" DWDM Line segments: " Situation 1 - The optical path between two consecutive 3R regenerators is composed of DWDM line segments from a single vendor and OADMs and PXC's from another vendor". Document [RFC6566] Figure 1 shows an LSP composed by two DWDM line segments according to [ITU.G680] definition.

2.2. Applicability

This document targets at Scenario C defined in [RFC6566] section 4.1.1. as approximate impairment estimation. The Approximate concept refer to the fact that this Information Model covers information mainly provided by [ITU.G680] Computational Model.

Computational models having no or little approximation, referred as IV-Detailed in the [RFC6566], currently does not exist in term of ITU-T recommendation. They generally deal with non-linear optical impairment and are usually vendor specific.

The Information Model defined in this document does not speculate about the mathematical formulas used to fill up information model parameters, hence it does not preclude changing the computational model. At the same time, the authors do not believe this Information Model is exhaustive and if necessary further documents will cover additional models after they become available.

The result of RWA-IV process implementing this Information Model is a path (and a wavelength in the data plane) that has better chance to be feasible than if it was computed without any IV function. The Existing Service Disruption, as per the definition above, would still be a problem left to a network design phase.

2.3. Properties

An information model may have several attributes or properties that need to be defined for each optical parameter made available to the control plane. The properties will help to determine how the control plane can deal with a specific impairment parameter, depending on architectural options chosen within the overall impairment framework [RFC6566]. In some case, properties value will help to identify the level of approximation supported by the IV process.

- o Time Dependency

This identifies how an impairment parameter may vary with time. There could be cases where there is no time dependency, while in other cases there may be need of re-evaluation after a certain time. In this category, variations in impairments due to environmental factors such as those discussed in [G.sup47] are considered. In some cases, an impairment parameter that has time dependency may be considered as a constant for approximation. In this information model, we do neglect this property.

- o Wavelength Dependency

This property identifies if an impairment parameter can be considered as constant over all the wavelength spectrum of interest or not. 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. In this information model, we consider both case: dependency / no dependency on a specific wavelength. This property appears directly in the information model definitions and related encoding.

- o Linearity

As impairments are representation of physical effects, there are some that have a linear behaviour while other are non-linear. Linear approximation is in scope of Scenario C of [RFC6566]. During the impairment validation process, this property implies that the optical effect (or quantity) satisfies the superposition principle, thus a final result can be calculated by the sum of each component. The linearity implies the additivity of optical quantities considered during an impairment validation process. The non-linear effects in general do not satisfy this property. The information model presented in this document however, easily allow introduction of non-linear optical effects with a linear approximated contribution to the linear ones.

- o Multi-Channel

There are cases where a channel's impairments take different values depending on the aside wavelengths already in place, this is mostly due to non-linear impairments. The result would be a dependency among different LSPs sharing the same path. This information model do not consider this kind of property.

The following table summarise the above considerations where in the first column reports the list of properties to be considered for each optical parameter, while the second column states if this property is taken into account or not by this information model.

Property	Info Model Awareness
Time Dependency	no
Wavelength Dependency	yes
Linearity	yes
Multi-channel	no

Table 1: Optical Impairment Properties

3. ITU-T List of Optical Parameters

As stated by Section 2.2 this Information Model does not intend to be exhaustive and targets an approximate computational model although not precluding future evolutions towards more detailed or different impairments estimation methods.

On the same line, ITU SG15/Q6 provides (through [LS78]) a list of optical parameters with following observations:

- (a) the problem of calculating the non-linear impairments in a multi-vendor environment is not solved. The transfer functions works only for the so called [ITU.G680] "Situation 1".
- (b) The generated list of parameters is not definitive or exhaustive.

In particular, [ITU.G680] contains many parameters that would be required to estimate linear impairments. Some of the Computational Models defined within [ITU.G680] requires parameters defined in other documents like [ITU.G671]. The purpose of the list here below makes this match between the two documents.

[ITU.G697] defines parameters can be monitored in an optical network. This Information Model and associated encoding document will reuse [ITU.G697] parameters identifiers and encoding for the purpose of path computation.

The list of optical parameters starts from [ITU.G680] Section 9 which provides the optical computational models for the following p:

G-1 OSNR. Section 9.1

G-2 Chromatic Dispersion (CD). Section 9.2

G-3 Polarization Mode Dispersion (PMD). Section 9.3

G-4 Polarization Dependent Loss (PDL). Section 9.3

In addition to the above, the following list of parameters has been mentioned by [LS78]:

- L-1 "Channel frequency range", [ITU.G671]. This parameter is part of the application code and encoded through Optical Interface Class as defined in [I-D.ietf-ccamp-rwa-info].
- L-2 "Modulation format and rate". This parameter is part of the application code and encoded through Optical Interface Class as defined in [I-D.ietf-ccamp-rwa-info].
- L-3 "Channel power". Required by G-1.
- L-4 "Ripple". According to [ITU.G680], this parameter can be taken into account as additional OSNR penalty.
- L-5 "Channel signal-spontaneous noise figure", [ITU.G680]. Required by OSNR calculation (see G-1) above.

- L-6 "Channel chromatic dispersion (for fibre segment or network element)". Already in G-2 above.
- L-7 "Channel local chromatic dispersion (for a fibre segment)". Already in G-2 above (since consider both local and fiber dispersions).
- L-8 "Differential group delay (for a network element)", [ITU.G671]. Required by G-3.
- L-9 "Polarisation mode dispersion (for a fibre segment)", [ITU.G650.2, ITU.G680]. Defined above as G-3.
- L-10 "Polarization dependent loss (for a network element)", [ITU.G671, ITU.G680]. Defined above as G-4.
- L-11 "Reflectance", [ITU.G671].
- L-12 "Isolation", [ITU.G671] and [ITU.GSUP39].
- L-13 "Channel extinction", [ITU.G671] and [ITU.GSUP39].
- L-14 "Attenuation coefficient (for a fibre segment)", [ITU.G650.1].
- L-15 "Non-linear coefficient (for a fibre segment)", [ITU.G650.2]. Required for Non-Linear Optical Impairment Computational Models. Neglected by this document.

The final list of parameters is G-1, G-2, G-3, G-4, L-3, L-4, L-5, L-8, L-11, L-12, L-13, L-14.

4. Background from WSOON-RWA Information Model

In this section we report terms already defined for the WSOON-RWA (impairment free) as in [I-D.ietf-ccamp-rwa-info] and [I-D.ietf-ccamp-general-constraint-encode]. The purpose is to provide essential information that will be reused or extended for the impairment case.

In particular [I-D.ietf-ccamp-rwa-info] defines the connectivity matrix as the following:

ConnectivityMatrix ::= <MatrixID> <ConnType> <Matrix>

According to [I-D.ietf-ccamp-general-constraint-encode], this definition is further detailed as:


```
ConnectivityMatrix ::=  
    <MatrixID> <ConnType> ((<LinkSet> <LinkSet>) ...)
```

This second formula highlights how the connectivity matrix is built by pairs of LinkSet objects identifying the internal connectivity capability due to internal optical node constraint(s). It's essentially binary information and tell if a wavelength or a set of wavelengths can go from an input port to an output port.

As an additional note, connectivity matrix belongs to node information and is purely static. Dynamic information related to the actual usage of the connections is available through specific extension to link information.

Furthermore [I-D.ietf-ccamp-rwa-info] define the resource block as follow:

```
ResourceBlockInfo ::= <ResourceBlockSet> [<InputConstraints>]  
    [<ProcessingCapabilities>] [<OutputConstraints>]
```

Which is an efficient way to model constraints of a WSON node.

5. Optical Impairment Information Model

The idea behind this information model is to categorize the impairment parameters into three types and extend the information model already defined for impairment-free WSONs. The three categories are:

- o Node Information. The concept of connectivity matrix is reused and extended to introduce an impairment matrix, which represents the impairments suffered on the internal path between two ports. In addition, the concept of Resource Block is also reused and extended to provide an efficient modelization of per-port impairment.
- o Link Information representing impairment information related to a specific link or hop.
- o Path Information representing the impairment information related to the whole path.

All the above three categories will make use of a generic container, the Impairment Vector, to transport optical impairment information.

This information model however will allow however to add additional parameters beyond the one defined by [ITU.G680] in order to support additional computational models. This mechanism could eventually be applicable to both linear and non-linear parameters.

This information model makes the assumption that each optical node in the network is able to provide the control plane protocols with its own parameter values. However, no assumption is made on how the optical nodes get those value information (e.g., internally computed, provisioned by a network management system, etc.). To this extent, the information model intentionally ignores all internal detailed parameters that are used by the formulas of the Optical Computational Model (i.e., "transfer function") and simply provides the object containers to carry results of the formulas.

5.1. The Optical Impairment Vector

Optical Impairment Vector (OIV) is defined as a list of optical parameters to be associated to a WSON node or a WSON link. It is defined as:

```
<OIV> ::= ([<LabelSet>] <OPTICAL_PARAM>) ...
```

The optional LabelSet object enables wavelength dependency property as per Table 1. LabelSet has its definition in [I-D.ietf-ccamp-general-constraint-encode].

OPTICAL_PARAM. This object represents an optical parameter. The Impairment vector can contain a set of parameters as identified by [ITU.G697] since those parameters match the terms of the linear impairments computational models provided by [ITU.G680]. This information model does not speculate about the set of parameters (since defined elsewhere, e.g. ITU-T), however it does not preclude extensions by adding new parameters.

5.2. Node Information

5.2.1. Impairment Matrix

Impairment matrix describes a list of the optical parameters that applies to a network element as a whole or ingress/egress port pairs of a network element. Wavelength dependency property of optical parameters is also considered.

```
ImpairmentMatrix ::= <MatrixID> <ConnType>  
  ((<LinkSet> <LinkSet> <OIV>) ...)
```

Where:

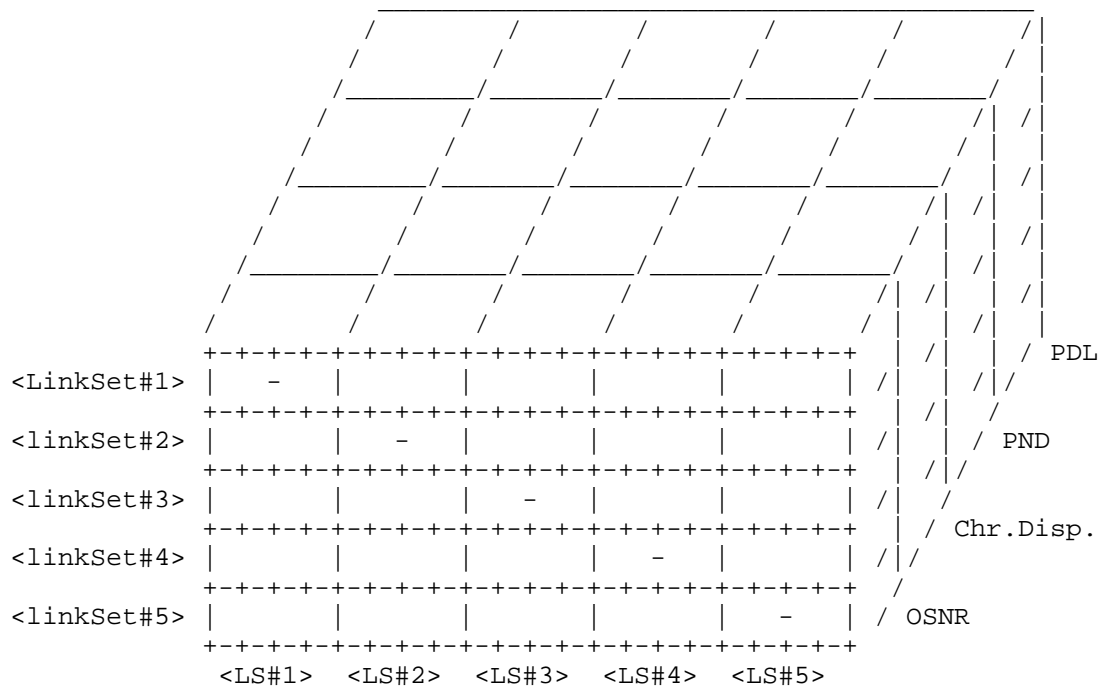
MatrixID. This ID is a unique identifier for the matrix. It shall be unique in scope among connectivity matrices defined in [I-D.ietf-ccamp-rwa-info] and impairment matrices defined here.

ConnType. This number identifies the type of matrix and it shall be unique in scope with other values defined by impairment-free WSON documents.

LinkSet. Same object definition and usage as [I-D.ietf-ccamp-general-constraint-encode]. The pairs of LinkSet identify one or more internal node constrain.

OIV. The Optical Impairment Vector defined above.

The model can be represented as a multidimensional matrix shown in the following picture



The connectivity matrix from [I-D.ietf-ccamp-general-constraint-encode] is only a two dimensional matrix, containing only binary information, through the LinkSet pairs. In this model, a third dimension is added by generalizing the binary information through the Optical Impairment Vector associated with each LinkSet pair. Optical parameters in the picture are reported just as examples while details go into specific encoding draft [I-D.martinelli-ccamp-wson-iv-encode].

This representation shows the most general case however, the total amount of information transported by control plane protocols can be greatly reduced by proper encoding when the same set of values apply to all LinkSet pairs.

[EDITOR NODE: first run of the information model does looks for generality not for optimizing the quantity of information. We'll deal with optimization in a further step.]

5.2.2. Impairment Resource Block Information

This information model reuses the definition of Resource Block Information adding the associated impairment vector.

```
ResourceBlockInfo ::= <ResourceBlockSet> [<InputConstraints>]  
                        [<ProcessingCapabilities>] [<OutputConstraints>] [<OIV>]
```

The object ResourceBlockInfo is then used as specified within [I-D.ietf-ccamp-rwa-info].

5.3. Link Information

For the list of optical parameters associated to the link, the same approach used for the node-specific impairment information can be applied. The link-specific impairment information is extended from [I-D.ietf-ccamp-rwa-info] as the following:

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

DynamicLinkInfo is already defined in [I-D.ietf-ccamp-rwa-info] while OIV is the Optical Impairment Vector is defined in the previous section.

5.4. Path Information

There are cases where the optical impairments can only be described as a constraint on the overall end to end path. In such case, the optical impairment and/or parameter, cannot be derived (using a simple function) from the set of node / link contributions.

An equivalent case is the option reported by [RFC6566] on IV-Candidate paths where, the control plane knows a list of optically feasible paths so a new path setup can be selected among that list. Independent from the protocols and functions combination (i.e. RWA vs. Routing vs. PCE), the IV-Candidates imply a path property stating that a path is optically feasible.

```
<PathInfo> ::= <OIV>
```

[EDITOR NOTE: section to be completed, especially to evaluate protocol implications. Likely resemble to RSVP ADSPEC].

6. Encoding Considerations

Details about encoding will be defined in a separate document [I-D.martinelli-ccamp-wson-iv-encode] however worth remembering that, within [ITU.G697] Appending V, ITU already provides a guideline for encoding some optical parameters.

In particular [ITU.G697] indicates that each parameter shall be represented by a 32 bit floating point number.

Values for optical parameters are provided by optical node and it could provide by direct measurement or from some internal computation starting from indirect measurement. In such cases, it could be useful to understand the variance associated with the value of the optical parameter hence, the encoding shall provide the possibility to include a variance as well.

This kind of information will enable IA-RWA process to make some additional considerations on wavelength feasibility. [RFC6566] Section 4.1.3 reports some considerations regarding this degree of confidence during the impairment validation process.

7. Control Plane Architectures

This section briefly describes how the definitions contained in this information model will match the architectural options described by [RFC6566].

The first assumption is that the WSON GMPLS extensions are available and operational. To such extent, the WSON-RWA will provide the following information through its path computation (and RWA process):

- o The wavelengths connectivity, considering also the connectivity constraints limited by reconfigurable optics, and wavelengths availability.
- o The interface compatibility at the physical level.
- o The Optical-Electro-Optical (OEO) availability within the network (and related physical interface compatibility). As already stated by the framework this information it's very important for impairment validation:
 - A. If the IV functions fail (path optically infeasible), the path computation function may use an available OEO point to find a

feasible path. In normally operated networks OEO are mainly used to support optically unfeasible path than mere wavelength conversion.

- B. The OEO points reset the optical impairment information since a new light is generated.

7.1. IV-Centralized

Centralized IV process is performed by a single entity (e.g., a PCE). Given sufficient impairment information, it can either be used to provide a list of paths between two nodes, which are valid in terms of optical impairments. Alternatively, it can help validate whether a particular selected path and wavelength is feasible or not. This requires distribution of impairment information to the entity performing the IV process.

This Information Model doesn't make any hypothesis on distribution method for optical parameters but only defines the essential building blocks. A centralized entity may get knowledge of required information through routing protocols or other mechanism such as BGP-LS.

7.2. IV-Distributed

Assuming the information model is implemented through a routing protocol, every node in the WSOON network shall be able to perform an RWA-IV function.

The signalling phase may provide additional checking as other traffic engineering parameters.

8. Acknowledgements

Authors would like to acknowledge Greg Bernstein and Moustafa Kattan as authors of a previous similar draft whose content partially converged here.

Authors would like to thank ITU SG15/Q6 and in particular Peter Stassar and Pete Anslow for providing useful information and text to CCAMP through joint meetings and liaisons.

9. IANA Considerations

This document does not contain any IANA requirement.

10. 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.

11. References

11.1. Normative References

[ITU.G671]

International Telecommunications Union, "Transmission characteristics of optical components and subsystems", ITU-T Recommendation G.671, February 2012.

[ITU.G680]

International Telecommunications Union, "Physical transfer functions of optical network elements", ITU-T Recommendation G.680, July 2007.

[ITU.G697]

International Telecommunications Union, "Optical monitoring for dense wavelength division multiplexing systems", ITU-T Recommendation G.697, February 2012.

11.2. Informative References

[I-D.ietf-ccamp-general-constraint-encode]

Bernstein, G., Lee, Y., Li, D., and W. Imajuku, "General Network Element Constraint Encoding for GMPLS Controlled Networks", draft-ietf-ccamp-general-constraint-encode-15 (work in progress), August 2014.

[I-D.ietf-ccamp-rwa-info]

Lee, Y., Bernstein, G., Li, D., and W. Imajuku, "Routing and Wavelength Assignment Information Model for Wavelength Switched Optical Networks", draft-ietf-ccamp-rwa-info-21 (work in progress), February 2014.

[I-D.martinelli-ccamp-wson-iv-encode]

Martinelli, G., Zhang, X., Galimberti, G., Siracusa, D., Zanardi, A., Pederzolli, F., Lee, Y., and F. Zhang, "Information Encoding for WSON with Impairments Validation", draft-martinelli-ccamp-wson-iv-encode-04 (work in progress), July 2014.

- [LS78] International Telecommunications Union SG15/Q6, "LS/s on CCAMP Liaison to ITU-T SG15 Q6 and Q12 on WSON", LS <https://datatracker.ietf.org/liaison/1288/>, October 2013.
- [RFC6163] Lee, Y., Bernstein, G., and W. Imajuku, "Framework for GMPLS and Path Computation Element (PCE) Control of Wavelength Switched Optical Networks (WSNs)", RFC 6163, April 2011.
- [RFC6566] Lee, Y., Bernstein, G., Li, D., and G. Martinelli, "A Framework for the Control of Wavelength Switched Optical Networks (WSNs) with Impairments", RFC 6566, March 2012.

Appendix A. FAQ

A.1. Why the Application Code does not suffice for Optical Impairment Validation?

Application Codes are encoded within GMPLS WSON protocol through the Optical Interface Class as defined in [I-D.ietf-ccamp-rwa-info].

The purpose of the Application Code in RWA is simply to assess the interface compatibility: same Application Code means that two interfaces can have an LSP connecting the two.

Application Codes contain other information useful for IV process (e.g., see the list of parameters) so they are required however Computational Models requires more parameters to assess the path feasibility.

A.2. Are DWDM network multivendor?

According to [ITU.G680] "Situation 1" the DWDM line segments are single are single vendor but an LSP can make use of different data planes entities from different vendors. For example: DWDM interfaces (represented in the control plane through the Optical Interface Class) from a vendor and network elements described by Stutation 1 from another vendor.

Authors' Addresses

Giovanni Martinelli (editor)
Cisco
via Santa Maria Molgora, 48/C
Vimercate 20871
Italy

Phone: +39 039 2092044
Email: giomarti@cisco.com

Xian Zhang (editor)
Huawei Technologies
F3-5-B R&D Center, Huawei Base
Bantian, Longgang District
Shenzhen 518129
P.R. China

Phone: +86 755 28972465
Email: zhang.xian@huawei.com

Gabriele M. Galimberti
Cisco
Via Santa Maria Molgora, 48/C
Vimercate 20871
Italy

Phone: +39 039 2091462
Email: ggalimbe@cisco.com

Andrea Zanardi
CREATE-NET
via alla Cascata 56/D, Povo
Trento 38123
Italy

Email: andrea.zanardi@create-net.org

Domenico Siracusa
CREATE-NET
via alla Cascata 56/D, Povo
Trento 38123
Italy

Email: domenico.siracusa@create-net.org

Federico Pederzolli
CREATE-NET
via alla Cascata 56/D, Povo
Trento 38123
Italy

Email: federico.perdezolli@create-net.org

Young Lee
Huawei Technologies
1700 Alma Drive, Suite 100
Plano, TX 75075
U.S.A

Email: ylee@huawei.com

Fatai Zhang
Huawei Technologies
F3-5-B R&D Center, Huawei Base
Bantian, Longgang District
Shenzhen 518129
P.R. China

Email: zhang.fatai@huawei.com

CCAMP Working Group
Internet-Draft
Intended Status: Informational
Expires: April 11, 2015

Xian Zhang
Haomian Zheng, Ed.
Huawei
Rakesh Gandhi, Ed.
Zafar Ali
Gabriele Maria Galimberti
Cisco Systems, Inc.
Pawel Brzozowski
ADVA Optical
October 8, 2014

RSVP-TE Signaling Procedure for GMPLS Restoration and Resource Sharing-
based LSP Setup and Teardown

draft-zhang-ccamp-gmpls-resource-sharing-proc-03

Abstract

In transport networks, there are requirements where Generalized Multi-Protocol Label Switching (GMPLS) end-to-end recovery scheme needs to employ restoration Label Switched Path (LSP) while keeping resources for the working and/or restoration LSPs reserved in the network after the failure occurs. This document reviews how the LSP association is to be provided using Resource Reservation Protocol - Traffic Engineering (RSVP-TE) signaling in the context of GMPLS end-to-end recovery when using restoration LSP where failed LSP is not torn down.

This document compliments existing standards by explaining the missing pieces of information during the RSVP-TE signaling procedure in support of resource sharing-based LSP setup/teardown in GMPLS-controlled circuit networks. No new procedures or mechanisms are defined by this document, and it is strictly informative in nature.

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>.

Copyright Notice

Copyright (c) 2014 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. Problem Statement	4
2.1. GMPLS Restoration	4
2.1.1. 1+R Restoration	4
2.1.2. 1+1+R Restoration	5
2.2. Resource Sharing-based LSP Setup/Teardown	6
3. RSVP-TE Signaling For Restoration LSP Association	7
4. RSVP-TE Signaling For Resource Sharing During LSP Setup/Teardown	8
4.1. LSPs with Identical Tunnel ID	8
4.1.1. Restoration LSP Setup	8
4.1.2. LSP Reversion	10
4.1.2.1. Make-while-break Reversion	11
4.1.2.2. Make-before-break Reversion	13
4.1.3. Re-optimization LSP Setup and Reversion	15
4.2. LSPs with Different Tunnel IDs	15
5. Security Considerations	16
6. IANA Considerations	16
7. Acknowledgement	16
8. References	17
8.1. Normative References	17
8.2. Informative References	17
9. Authors' Addresses	19

1. Introduction

Generalized Multi-Protocol Label Switching (GMPLS) [RFC3945] defines a set of protocols, including Open Shortest Path First - Traffic Engineering (OSPF-TE) [RFC4203] and Resource Reservation Protocol - Traffic Engineering (RSVP-TE) [RFC3473]. These protocols can be used to create Label Switched Paths (LSPs) in a number of deployment scenarios with various transport technologies. The GMPLS protocol set extends MPLS, which supports only Packet Switch Capable (PSC) and Layer 2 Switch Capable interfaces (L2SC), to also cater for interfaces capable of Time Division Multiplexing (TDM), Lambda Switching (LSC) and Fiber Switching (FSC). These switching technologies provide several protection schemes [RFC4426][RFC4427] (e.g., 1+1, 1:N and M:N). Resource Reservation Protocol - Traffic Engineering (RSVP-TE) signaling has been extended to support various GMPLS recovery schemes [RFC4872][RFC4873], to establish Label Switched Paths (LSPs), typically for working LSP and protecting LSP. [RFC4427] Section 7 specifies various schemes for GMPLS recovery.

In GMPLS recovery schemes generally considered, restoration LSP is signaled after the failure has been detected and notified on the working LSP. In non-revertive recovery mode, working LSP is assumed to be removed from the network before restoration LSP is signaled. For revertive recovery mode, a restoration LSP is signaled while working LSP and/or protecting LSP are not torn down in control plane due to a failure. In transport networks, as working LSPs are typically signaled over a nominal path, service providers would like to keep resources associated with the working LSPs reserved. This is to make sure that the service (working LSP) can use the nominal path when the failure is repaired to provide deterministic behavior and guaranteed Service Level Agreement (SLA). Consequently, revertive recovery mode is usually preferred by recovery schemes used in transport networks.

The Make-Before-Break (MBB) mechanisms exploiting the Shared-Explicit (SE) reservation style can be employed in MPLS networks to avoid double booking of resource during the process of LSP re-optimization as specified in [RFC3209]. This method is also used in GMPLS-controlled networks [RFC4872] [RFC4873] for end-to-end and segment recovery of LSPs. This was further generalized to support resource sharing oriented applications in MPLS networks as well as non-LSP contexts, as specified in [RFC6780].

Due to the fact that the features of GMPLS-controlled networks (specifically for TDM, LSC and FSC), are not identical to that of the MPLS networks, additional considerations for resource sharing based LSP association are needed. As defined in [RFC4872] and being considered in this document, "fully dynamic rerouting switches normal

traffic to an alternate LSP that is not even partially established only after the working LSP failure occurs. The new alternate route is selected at the LSP head-end node, it may reuse resources of the failed LSP at intermediate nodes and may include additional intermediate nodes and/or links". During the signaling procedure for resource sharing based LSP setup/teardown, the behaviors of the nodes along the path may be different from that in the MPLS networks as well as the effect it may have on the traffic delivery.

As described in [RFC6689], ASSOCIATION Object is used to identify the LSPs for restoration using association type "Recovery" [RFC4872] and for resource sharing using association type "Resource Sharing" [RFC4873].

Following section describes the problem statements for the GMPLS restoration and resource sharing based LSP setup and teardown.

2. Problem Statement

Problem statements for the GMPLS restoration schemes and resource sharing-based LSP setup and teardown are described in this section.

2.1. GMPLS Restoration

2.1.1. 1+R Restoration

One example of the recovery scheme considered in this document is 1+R recovery. The 1+R recovery is exemplified in Figure 1. In this example, working LSP on path A-B-C-Z is pre-established. Typically after a failure detection and notification on the working LSP, a second LSP on path A-H-I-J-Z is established as a restoration LSP. Unlike protection LSP, restoration LSP is signaled per need basis.

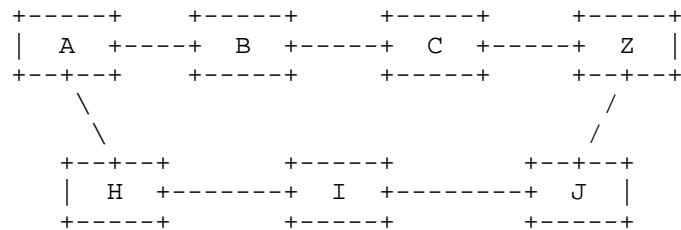


Figure 1: An Example of 1+R Recovery Scheme

During failure switchover with 1+R recovery scheme, in general, working LSP resources are not released and working and restoration LSPs coexist in the network. Nonetheless, working and restoration

LSPs can share network resources. Typically when failure is recovered on the working LSP, restoration LSP is no longer required and torn down (e.g., revertive mode).

2.1.2. 1+1+R Restoration

Another example of the recovery scheme considered in this document is 1+1+R. In 1+1+R, a restoration LSP is signaled for the working LSP and/or the protecting LSP after the failure has been detected and notified on the working LSP or the protecting LSP. The 1+1+R recovery is exemplified in Figure 2.

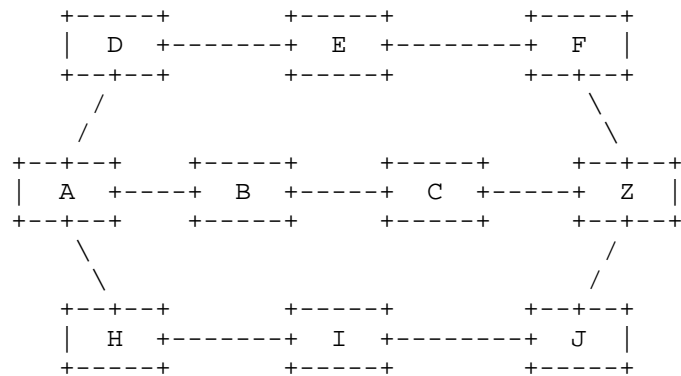


Figure 2: An Example of 1+1+R Recovery Scheme

In this example, working LSP on path A-B-C-Z and protecting LSP on path A-D-E-F-Z are pre-established. After a failure detection and notification on a working LSP or protecting LSP, a third LSP on path A-H-I-J-Z is established as a restoration LSP. The restoration LSP in this case provides protection against a second order failure. Restoration LSP is torn down when the failure on the working or protecting LSP is repaired.

[RFC4872] Section 14 defines PROTECTION Object for GMPLS recovery signaling. As defined, the PROTECTION Object is used to identify primary and secondary LSPs using S bit and protecting and working LSPs using P bit. Furthermore, [RFC4872] defines the usage of ASSOCIATION Object for associating GMPLS working and protecting LSPs.

[RFC6689] Section 2.2 reviews the procedure for providing LSP associations for GMPLS end-to-end recovery and covers the schemes where the failed working LSP and/or protecting LSP are torn down.

This document reviews how the LSP association is to be provided for GMPLS end-to-end recovery when using restoration LSP where working

and protecting LSP resources are kept reserved in the network after the failure.

2.2. Resource Sharing-based LSP Setup/Teardown

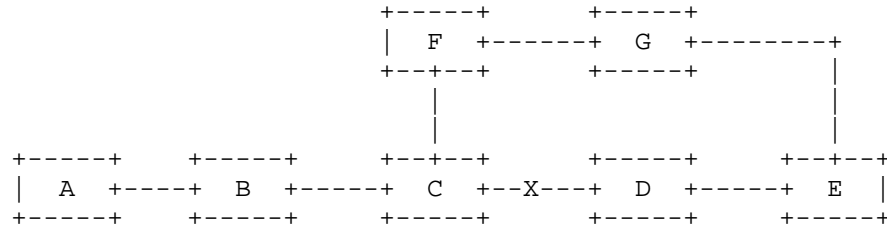


Figure 3: A Simple OTN Network

Using the Optical Transport Network (OTN) topology shown in Figure 3 as an example, GMPLS-controlled circuit LSP1 (A-B-C-D-E) is the working LSP and it allows for resource sharing when the LSP is dynamically rerouted due to link failure. Upon detecting the failure of a link along the LSP1, e.g. Link C-D, node A needs to decide on which alternate path it will establish an LSP to reroute the traffic.

In this case, A-B-C-F-G-E is chosen as the alternative path for the LSP and the resources on the path segment A-B-C are re-used by this LSP. Since this is an OTN network, which is different from the packet-switching network, the label has a mapping into the data plane resource used (e.g. wavelength) and also the nodes along the path need to send triggering commands to data plane nodes for setting up cross-connection accordingly during the RSVP-TE signaling process. In this case, the following issues are left un-described in the existing standards for resource sharing based LSP setup/teardown in GMPLS-controlled circuit networks:

- Reservation style Shared-Explicit (SE) as defined in [RFC3209] may not be applicable due to the nature of the GMPLS-controlled circuits. It is not clear how reservation style is to be used by the GMPLS LSPs for resource sharing.

- As described in [RFC3209], the purpose of Make-Before-Break (MBB) is to "not disrupt traffic or adversely impact network operations while TE tunnel rerouting is in progress". Due to the nature of the GMPLS-controlled circuit networks, this may not be fulfilled under certain scenarios. Thus, the name "Make-Before-Break" may no longer hold true.

- The existing MBB method may not be sufficient to support LSP setup and teardown with resource sharing.

- In [RFC3209], the MBB method assumes the old and new LSPs share the same tunnel ID (i.e., sharing the same source and destination nodes).

[RFC4873] does not impose this constraint but limit the resource sharing usage in LSP recoveries only. [RFC6780] generalizes the resource sharing application, based on the ASSOCIATION Object, to be useful in MPLS networks as well as in non-LSP association such as Voice Call-Waiting. Recently, there are also requirements to generalize resource sharing of LSPs with different tunnel IDs, such as the one mentioned in [PCEP-RSO] and LSPs with LSP-stitching across multi-domains. In this case, how the signaling process can make intermediate nodes aware of the resource sharing constraint and behave accordingly is an issue that needs to be described.

- The node behavior during traffic reversion in the GMPLS-controlled circuit network is missing and should be clarified.

This document reviews the signaling procedure for resource sharing-based LSP setup and teardown for GMPLS-based circuits in OTN networks. This includes the node behavior description, besides clarifying some un-discussed points for this process. Two typical examples mentioned in this document are LSP restoration and LSP re-optimization, where it is desirable to share resources. This document does not define any RSVP-TE signaling extensions. If necessary, discussion is provided to identify potential extensions to the existing RSVP-TE protocol. It is expected that the extensions, if there are any, will be addressed in separate documents.

3. RSVP-TE Signaling For Restoration LSP Association

Where GMPLS end-to-end recovery scheme needs to employ restoration LSP while keeping resources for the working and/or protecting LSPs reserved in the network after the failure, restoration LSP is signaled with ASSOCIATION Object that has association type set to "Recovery" [RFC4872] with the association ID set to the LSP ID of the LSP it is restoring. For example, when a restoration LSP is signaled for a working LSP, the ASSOCIATION Object in the restoration LSP contains the association ID set to the LSP ID of the working LSP. Similarly, when a restoration LSP is signaled for a protecting LSP, the ASSOCIATION Object in the restoration LSP contains the association ID set to the LSP ID of the protecting LSP.

The procedure for signaling the PROTECTION Object is specified in [RFC4872]. Specifically, restoration LSP being used as a working LSP is signaled with P bit cleared and being used as a protecting LSP is signaled with P bit set.

As discussed in Section 2 of this document, [RFC6689] Section 2.2 reviews the procedure for providing LSP associations for the GMPLS end-to-end recovery scheme using restoration LSP where the failed working LSP and/or protecting LSP are torn down.

4. RSVP-TE Signaling For Resource Sharing During LSP Setup/Teardown

For LSP restoration upon failure, as explained in Section 11 of [RFC4872], the purpose of using MBB is to re-use existing resources. Thus, the behavior of the intermediate nodes during rerouting process will not further impact traffic since it has been interrupted due to the already broken working LSP. However, for the following two cases, the behavior of intermediate nodes may impact the traffic delivery: (1) LSP reversion; (2) LSP re-optimization.

Another dimension that needs separate attention is how to correlate the two LSPs sharing resource. For the LSPs with the same Tunnel ID, [RFC4872] and reviewed in this section. For the LSPs with different Tunnel IDs, signaling procedure is clarified in Section 4.2 of this document.

4.1. LSPs with Identical Tunnel ID

For resource sharing among LSPs with identical Tunnel IDs, SE flag and ASSOCIATION Object are used together. The SE flag is to enable resource sharing and the ASSOCIATION Object with association type "Resource Sharing" [RFC4873] is to identify the associated LSPs.

As a first step, in order to allow resource sharing, the original LSP setup should explicitly carry the SE flag in the SESSION_ATTRIBUTE Object during the initial LSP setup, irrespective of the purpose of resource sharing.

The basic signaling procedure for alternative LSP setup has been described by the existing standards. In [RFC3209], it describes the basic MBB signaling flow for MPLS-TE networks. [RFC4872] adds additional information when using MBB for LSP rerouting.

As mentioned before, for LSP setup/teardown in GMPLS-controlled circuit networks, the network elements along the path need to send cross-connection setup/teardown commands to data plane node(s) either during the PATH message forwarding phase or the RESV message forwarding phase.

4.1.1. Restoration LSP Setup

For LSP restoration, the complete signaling flow processes for both

LSP restorations upon failure and LSP reversion upon link failure recovery are described in this section.

Table 1: Node Behavior during Restoration LSP Setup

Category	Node Behavior during Restoration LSP setup
C1	<ul style="list-style-type: none"> + Reusing existing resource on both input and output interfaces. + This type of nodes only needs to book the existing resource when receiving the PATH message and no cross-connection setup command is needed when receiving the RESV message.
C2	<ul style="list-style-type: none"> + Reusing existing resource only on one of the interfaces, either input or output interfaces and need to use new resource on the other interface. + This type of nodes needs to book the resources on the interface where new resource are needed and re-use the existing resource on the other interface when it receives the PATH message. Upon receiving the RESV message, it needs to send the re-configuration the cross-connection command to its corresponding data plane node.
C3	<ul style="list-style-type: none"> + Using new resource on both interfaces. + This type of nodes needs to book the new resource when receiving PATH and send the cross-connection setup command upon receiving RESV.

For LSP rerouting upon working LSP failure, using the network shown in Figure 3 as an example.

Working LSP: A-B-C-D-E

Restoration LSP: A-B-C-F-G-E

The restoration LSP may be calculated by the head-end node or a Path Computation Element (PCE) [RFC4655]. Assuming that the cross-connection configuration command is sent by the control plane nodes during the RESV forwarding phase, the node behavior for setting up the alternative LSP can be classified into the following three categories as shown in Table 1.

- o Make-while-break reversion, where resources associated with working or protecting LSP are reconfigured while removing reservations for restoration LSP.
- o Make-before-break reversion, where resources associated with working or protecting LSP are reconfigured before removing restoration LSP.

It is worth mentioning that in GMPLS-controlled circuit OTN networks both reversion types will result in a short traffic disruption.

4.1.2.1. Make-while-break Reversion

In this technique, restoration LSP is simply requested to be deleted. Removing reservations for restoration LSP triggers reconfiguration of resources associated with working or protecting LSP on every node where resources are shared. Hence, whenever reservation for restoration LSP is removed from a node, data plane configuration changes to reflect reservations of working or protection LSP as signaling progresses. Eventually, after the whole restoration LSP is deleted, data plane configuration will fully match working or protecting LSP reservations on the whole path. Thus reversion is complete.

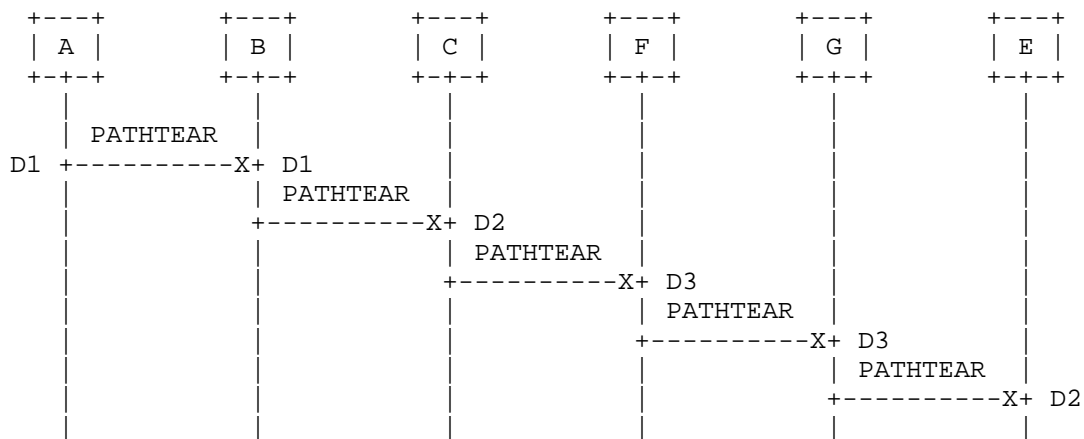


Figure 5: Signaling Procedure for LSP Make-while-break Reversion

Figure 5 shows signaling process of make-while-break reversion of LSP PathTear message. For alarm-free LSP deletion, the mechanisms described in Section 6 of [RFC4208] should be followed. Resource sharing between working and restoration LSP takes place on nodes A, B, C and E. These are the nodes where reconfiguration of resources associated with working LSP can take place.

Node behavior upon removing reservation for restoration LSP depends on how resources are shared with working or protecting LSP:

Table 2: Node behavior during LSP make-while-break reversion

Category	Node behavior during LSP make-while-break reversion
D1	<ul style="list-style-type: none"> + Working and restoration LSP share resources on both incoming and outgoing interface. + CP change: Reservation for restoration LSP is removed. + DP change: None, as data plane configuration already reflects working LSP reservation.
D2	<ul style="list-style-type: none"> + Working and restoration LSP share resources on one of the interfaces. + CP change: Reservation for restoration LSP is removed. + DP change: Resource on the interface that is not shared between working and restoration LSP is freed. + Cross-connection is updated to reflect working LSP reservation.
D3	<ul style="list-style-type: none"> + Working and restoration LSP do not share resources. + CP change: Reservation for restoration LSP is removed. + DP change: Resources associated with restoration LSP are freed.

Make-while-break, while being relatively simple in its logic, has a few limitations which may be not acceptable in some implementations:

- o No rollback

Deletion of a LSP is not a revertive process. If for some reason reconfiguration of data plane on one of the nodes to match working or protection LSP reservations fails, falling back to restoration LSP is no longer an option, as its state might have already been removed from other nodes.

- o No completion guarantee

Deletion of a LSP provides no guarantees of completion. In particular, if RSVP packets are lost due to nodal or DCN failures it is probable for a LSP to be only partially deleted. To mitigate this, RSVP could maintain soft state reservations

and hence eventually remove remaining reservations due to refresh timeouts. This approach is not feasible in circuit networks however, since control and data channels are often separated and hence soft state reservations are not used.

Finally, one could argue that graceful LSP deletion [RFC3473] would provide guarantee of completion. While this is true for most cases, many implementations will timeout graceful deletion if LSP is not removed within certain amount of time, e.g. due to a transit node fault. After that, deletion procedures that provide no completion guarantees will be attempted. Hence in corner cases completion guarantee cannot be provided.

- o No explicit notification of completion to ingress node

In some cases it may be useful for ingress node to know when the data plane has been reconfigured to match working or protection LSP reservations. This knowledge could be used for initiating operations like enabling alarm monitoring, power equalization and others. Unfortunately, for the reasons mentioned above, make-while-break reversion lacks such explicit notification.

4.1.2.2. Make-before-break Reversion

MBB reversion can be used to overcome limitations of make-while-break reversion. It is similar in spirit to MBB concept used for restoration. Instead of relying on deletion of restoration LSP, it chooses to establish a new LSP to reconfigure resources on the working or protection LSP path. Only if setup of this LSP is successful will other LSPs be deleted. MBB reversion consists of two parts:

A) Make part:

Creating a new reversion LSP following working or protection LSP's path - see Figure 6. Reversion LSP is sharing resources both with working and restoration LSPs. As reversion LSP is created, resources are reconfigured to match its reservations - nodes follow procedures described in Table 1. Hence after reversion LSP is created, data plane configuration essentially reflects working or protecting LSP reservations.

B) Break part:

After 'make' part is finished, working and restoration LSPs are torn down. Removing reservations for working and restoration LSPs does not cause any resource reconfiguration on reversion LSP's path - nodes follow same procedures as for 'break' part of any MBB operation. Hence after working and restoration LSPs are removed, data plane configuration is exactly the same as before

LSP setup is resilient against RSVP message loss, as PATH and RESV messages are refreshed periodically. Hence, given that network recovers its DCN eventually, setup is guaranteed to finish with either success or failure.

- o Explicit notification of completion to ingress node

Ingress knows that data plane has been reconfigured to match working or protection LSP reservations when it receives RESV for the reversion LSP.

4.1.3. Re-optimization LSP Setup and Reversion

For LSP re-optimization where the new LSP and old LSPs share resource, the signaling flow for new LSP setup and old LSP teardown is similar to those shown in Figures 4 and 5.

The issue that should be noted is the traffic will be disrupted if the new path setup process changes the cross-connection configuration of the nodes along the old LSP. If no traffic interruption is desirable, it should either ensure that the old and new LSP do not share the resource other than the source and destination nodes or use other mechanisms. This is out the scope of this document.

Similarly, if LSP re-optimization fails and there is a need for LSP reversion, the traffic may be disrupted when resources are shared and cross-connections need to be reconfigured and reverted.

4.2. LSPs with Different Tunnel IDs

For two LSPs with different Tunnel IDs, the ASSOCIATION Object is used to specify that they are sharing resource (by setting ASSOCIATION type as "Resource Sharing" (value 2) as well as to identify these correlated LSPs. There are two types:

- (1) Sharing the common nodes, such as segment recovery, the source and destination nodes of the segment recovery LSP is the intermediate nodes along the working LSPs;

- (2) Resource sharing is used in a generalized context (such as multi-layer or multi-domain networks); it may result in either sharing source nodes in common, or destination nodes in common, or non end-points in common, if viewed from one domain's perspective.

The path computation can either be performed by the source node or edge nodes for the path/path segment or carried out by the PCE, such as the one explained in [PCEP-RSO]. This document does not impose any constraint with regard to path computation.

[RFC4873] considers resource sharing for LSP segment recovery. The ASSOCIATION Object usage is limited. [RFC6780] extends the usage of ASSOCIATION Object to cover generalized resource sharing applications. The extended ASSOCIATION Object is primarily defined for MPLS-TP, but it can be applied in a wider scope [RFC6780]. It can be used in the second types mentioned above. The configuration and processing rules of extended ASSOCIATION Object defined in [RFC6780] should be followed. The only issue that need pay attention to is that uniqueness of LSP association for the second type should be guaranteed when crossing the layer or domain boundary. The mechanisms for how to ensure this are outside the scope of this document.

Other than this, the signaling flow for this type of resource sharing is similar to the description provided in Section 4.1.1. Similar to what is discussed in previous sections, the traffic delivery may be interrupted. Depending on whether the short traffic interruption is acceptable or not, additional mechanisms may be needed and are outside the scope of this document.

5. Security Considerations

This document reviews procedures defined in [RFC4872] and [RFC6689] and does not define any new procedure. This document does not incur any new security issues other than those already covered in [RFC3209] [RFC4872] [RFC4873] and [RFC6780].

6. IANA Considerations

This informational document does not make any requests for IANA action.

7. Acknowledgement

The authors would like to thank George Swallow for the discussions on the GMPLS restoration.

8. References

8.1. Normative References

- [RFC3209] D. Awduche et al, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, December 2001.
- [RFC3473] L. Berger, Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions", RFC 3473, January 2003.
- [RFC3945] Mannie, E., "Generalized Multi-Protocol Label Switching (GMPLS) Architecture", RFC 3945, October 2004.
- [RFC4203] Kompella, K., and Rekhter, Y., "OSPF Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4203, October 2005.
- [RFC4872] J.P. Lang et al, "RSVP-TE Extensions in Support of End-to-End Generalized Multi-Protocol Label Switching (GMPLS) Recovery", RFC 4872, May 2007.
- [RFC4873] L. Berger et al, "GMPLS Segment Recovery", RFC 4873, May 2007.
- [RFC6689] L. Berger, "Usage of the RSVP ASSOCIATION Object", RFC 6689, July 2012.
- [RFC6780] L. Berger et al, "RSVP ASSOCIATION Object Extensions", RFC 6780, October 2012.

8.2. Informative References

- [PCEP-RSO] X. Zhang, et al, "Extensions to Path Computation Element Protocol (PCEP) to Support Resource Sharing-based Path Computation", work in progress, February 2014.
- [RFC4426] Lang, J., Rajagopalan, B., and Papadimitriou, D., "Generalized Multiprotocol Label Switching (GMPLS) Recovery Functional Specification", RFC 4426, March 2006.
- [RFC4427] Mannie, E., and Papadimitriou, D., "Recovery (Protection and Restoration) Terminology for Generalized Multi-Protocol Label Switching", RFC 4427, March 2006.

- [RFC4655] A. Farrel et al, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, August 2006.

- [RFC4208] Swallow, G., Drake, J., Ishimatsu, H., Rekhter, Y.,
 "Generalized Multiprotocol Label Switching (GMPLS)
 User-Network Interface (UNI): Resource ReserVation
 Protocol-Traffic Engineering (RSVP-TE) Support for the
 Overlay Model", RFC 4208, October 2005.

9. Authors' Addresses

Xian Zhang
Huawei Technologies
F3-1-B R&D Center, Huawei Base
Bantian, Longgang District
Shenzhen 518129 P.R.China

Email: zhang.xian@huawei.com

Haomian Zheng (editor)
Huawei Technologies
F3-1-B R&D Center, Huawei Base
Bantian, Longgang District
Shenzhen 518129 P.R.China

Email: zhenghaomian@huawei.com

Rakesh Gandhi (editor)
Cisco Systems, Inc.

Email: rgandhi@cisco.com

Zafar Ali
Cisco Systems, Inc.

Email: zali@cisco.com

Gabriele Maria Galimberti
Cisco Systems, Inc.

Email: ggalimbe@cisco.com

Pawel Brzozowski
ADVA Optical

Email: PBrzozowski@advaoptical.com

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: January 4, 2015

Z. Li
L. Zhang
Huawei Technologies
G. Yang
China Telecom
July 3, 2014

RSVP-TE Extensions for Bit Error Rate (BER) Measurement
draft-zhang-ccamp-rsvpte-ber-measure-02

Abstract

In the mobile backhaul network, the mobile service is sensitive to Bit Error Rate (BER). When the BER value of the service exceeds the threshold, the cell site equipments will stop working and the mobile terminal users cannot obtain voice and data services anymore. Now the mobile backhaul tends to be IP/MPLS network and MPLS TE LSP is used to bear the mobile service which may be encapsulated in PW or L3VPN end to end. Then the ingress Label Switched Router (LSR) of the MPLS TE LSP needs to get information on BER along the path of the LSP. This document proposes new extensions of RSVP-TE to advertise the BER measurement requirement of the specific LSP to all of the transit LSRs and the egress LSR, and to report the BER measurement result from any transit or egress LSR towards the ingress LSR.

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].

Status of This Memo

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

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

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

This Internet-Draft will expire on January 4, 2015.

Copyright Notice

Copyright (c) 2014 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	2
2. Terminology	3
3. BER_REQUEST TLV	3
3.1. Format of BER_REQUEST	3
3.2. Procedures for BER_REQUEST TLV	5
4. BER Measurement Result Report	5
4.1. Error Code for BER measurement report	5
4.2. Procedures for BER Error Code	6
5. IANA Considerations	6
6. Security Considerations	6
7. References	6
7.1. Normative References	6
7.2. Informative References	7
Authors' Addresses	7

1. Introduction

Bit Error Rate (BER) is a significant parameter for the mobile service, which can cause the cell site equipment to stop working when its value exceeds the threshold that the mobile service can tolerate. In IP/MPLS based mobile backhaul network, PW and L3VPN are adopted to bear the mobile service, and MPLS TE LSP is adopted as the transport tunnel for which Hot-standby (MPLS TE HSB) or fast reroute (MPLS TE FRR) technologies is used to meet the SLA(Service Level Agreement). There are different kinds of failure detection methods, such as BFD[RFC5884] or MPLS OAM[RFC4378], to trigger fast traffic switch when failure happens. But as to BER, even if the BER value exceeds the threshold, the detection mechanisms may be not able to detect the failure to trigger traffic switch to the backup path. In order to solve the issue, RSVP-TE extensions can be introduced to notify the ingress LSR of the BER measurement result when a specific LSR along

the LSP detects that the BER value exceeds the threshold or restores to the normal value below the threshold. When the ingress LSP receives the BER measurement result, it can switch the traffic between the primary path and the backup path which is policy specific and out of scope of the document.

This document defines new extensions of RSVP-TE for BER measurement: One extension is to advertise the BER measurement requirement to all of the transit LSRs and the egress LSR, then these LSRs along the path will start BER measurement for the LSP. The other extension is to report the BER measurement result from any transit LSR or the egress LSR towards the ingress LSR.

2. Terminology

BER: Bit Error Rate

RAN: Radio Access Network

LSR: Label Switch Router

LSP: Label Switch Path

3. BER_REQUEST TLV

3.1. Format of BER_REQUEST

Path Message of RSVP-TE is used to signal the BER measurement requirement of the LSP, and the LSP_ATTRIBUTES object will be included in the Path Message. The LSP_ATTRIBUTES object which is defined in [RFC5420] is used to signal attributes required in support of an LSP, or to indicate the nature or use of an LSP. The LSP_ATTRIBUTES object format is as below (refer to[RFC5420]):

```
LSP ATTRIBUTES class = 197, C-Type = 1
```

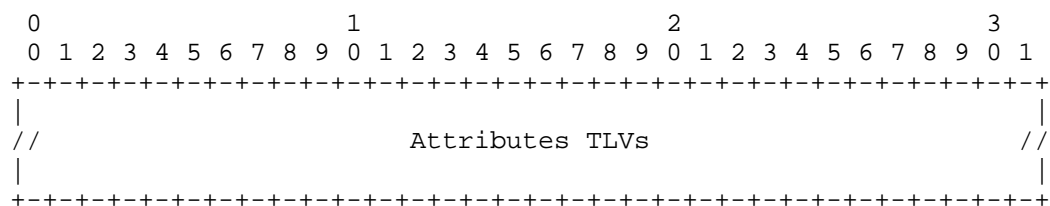


Figure 1: LSP_ATTRIBUTES object

The LSP_ATTRIBUTES object class is 197 of the form 11bbbbbbb. This C-Num value (see [RFC2205], Section 3.10) ensures that LSRs that do

not recognize the object pass it on transparently. One C-Type is defined, C-Type = 1 for LSP Attributes.

The Attributes TLVs are encoded as below:

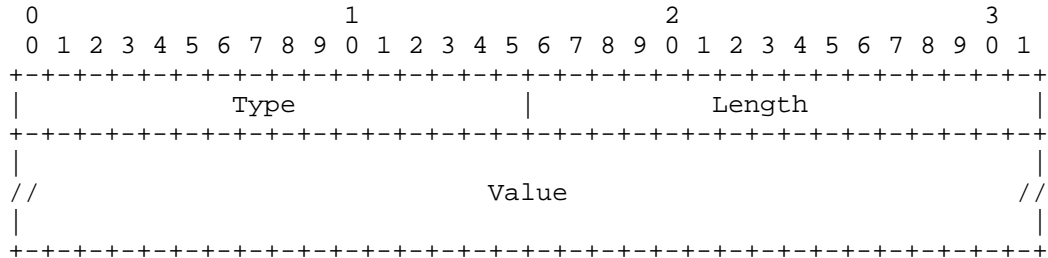


Figure 2: Attributes TLVs format

Here we define the BER_REQUEST TLV which is a new type of Attribute TLV to indicate the BER measurement requirement of the LSP. The format of BER_REQUEST TLV is as below:

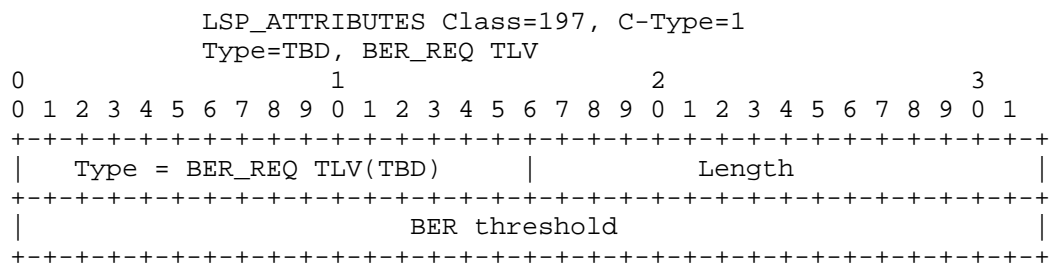


Figure 3:BER_REQUEST TLV

Type

The identifier of the BER_REQUEST TLV which should be allocated by IANA.

Length

Indicates the total length of the TLV in two octets.

Value

The BER threshold for the service. It is a 32-bit IEEE floating point number. The format of IEEE floating-point numbers is further summarized in [RFC1832].

3.2. Procedures for BER_REQUEST TLV

BER_REQUEST TLV is one type of attribute TLVs of the LSP_ATTRIBUTE object. It is optional and MAY be placed in Path messages to advertise the BER measurement requirement of the LSP. The process of the LSP_ATTRIBUTE object can refer to section 4.2 in [RFC5420].

When a RSVP-TE LSP requires the BER measurement of the path, the ingress LSR MUST send a Path Message with BER_REQUEST TLV in which the BER threshold value is set according to the service requirement.

When an LSR receives a Path Message with the BER_REQUEST TLV, the LSR SHOULD start the BER measurement for the LSP. The LSR MUST pass the Path Message with BER_REQUEST TLV unchanged to the next LSR. If the measured BER value exceeds the BER threshold value set in the BER_REQUEST TLV, the LSR MUST report the bit error result towards the ingress LSR of the LSP.

If an LSR cannot support the BER_REQUEST TLV, the LSR SHOULD ignore this TLV and pass the Path Message with BER_REQUEST TLV unchanged to the next LSR.

For an LSR which has started the BER measurement on receiving the Path Message with the BER_REQUEST TLV, if the LSR receives the updated Path Message without BER_REQUEST TLV, it MUST stop the BER measurement for this LSP and pass the Path Message unchanged to the next LSR.

4. BER Measurement Result Report

For an LSR that starts the BER measurement, When the BER measurement value exceeds the threshold for the service, the LSR MUST report the indication of bit error towards the ingress LSR of the LSP. If the LSR has already reported the indication of bit error, it MUST report the elimination of bit error towards the ingress LSR when it measures that the BER is below the specified threshold. A new type of Error Code and its Error Value of the ERROR_SPEC object are defined to report the BER measurement result within PathErr Message.

4.1. Error Code for BER measurement report

The ERROR_SPEC object is defined in [RFC2205] and [RFC3209]. The BER Error Code and its Error Values are defined as below:

Error Code	Error Value	Description

TBD	0	Bit Error Elimination
	1	Bit Error Indication

4.2. Procedures for BER Error Code

The BER measurement result is reported through a new Error Code and the corresponding Error Value of the ERROR_SPEC object which is placed in PathErr Message.

For a LSR which has started the BER measurement for the specific LSP, if the BER measurement value exceeds the threshold of the service, a PathErr Message MUST be sent towards the ingress LSR of this LSP. The PathErr Message MUST include an ERROR_SPEC object with the BER Error code and Error Value 1 for Bit Error Indication. When the BER measurement value becomes less than the BER threshold value after report the Bit Error Indication, the LSR MUST send a PathErr Message including an ERROR_SPEC object with the BER Error Code and Error Value 0 for Bit Error Elimination.

5. IANA Considerations

IANA should allocate the type value of the BER_REQUEST TLV and the BER Error Code which are defined in this document.

6. Security Considerations

The extensions of RSVP TE for BER in this document do not introduce any new security issues, and the reader is referred to the security considerations expressed in [RFC2205], [RFC3209], and [RFC5420].

7. References

7.1. Normative References

- [RFC1832] Srinivasan, R., "XDR: External Data Representation Standard", RFC 1832, August 1995.
- [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.
- [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.

- [RFC5420] Farrel, A., Papadimitriou, D., Vasseur, JP., and A. Ayyangarps, "Encoding of Attributes for MPLS LSP Establishment Using Resource Reservation Protocol Traffic Engineering (RSVP-TE)", RFC 5420, February 2009.

7.2. Informative References

- [RFC4378] Allan, D. and T. Nadeau, "A Framework for Multi-Protocol Label Switching (MPLS) Operations and Management (OAM)", RFC 4378, February 2006.
- [RFC5884] Aggarwal, R., Kompella, K., Nadeau, T., and G. Swallow, "Bidirectional Forwarding Detection (BFD) for MPLS Label Switched Paths (LSPs)", RFC 5884, June 2010.

Authors' Addresses

Zhenbin Li
Huawei Technologies
Huawei Bld., No.156 Beiqing Rd.
Beijing 100095
China

Email: lizhenbin@huawei.com

Li Zhang
Huawei Technologies
Huawei Bld., No.156 Beiqing Rd.
Beijing 100095
China

Email: monica.zhangli@huawei.com

Guangming Yang
China Telecom
No. 109, Zhongshan Road West, Tianhe District
Guangzhou 510630
China

Email: yanggm@gsta.com