```
Internet Engineering Task Force                       G.Galimberti, Ed.
Internet-Draft                                                    Cisco
Intended status: Standards Track                          R.Kunze, Ed.
Expires: January 7, 2016                               Deutsche Telekom
                                                           K. Lam, Ed.
                                                         Alcatel-Lucent
                                                    D. Hiremagalur, Ed.
                                                      G. G.Grammel, Ed.
                                                                Juniper
                                                           L.Fang, Ed.
                                                       G.Ratterree, Ed.
                                                              Microsoft
                                                           July 6, 2015
```

A YANG model to manage the optical interface parameters of "G.698.2 single channel" in DWDM applications
draft-dharini-netmod-g-698-2-yang-04

Abstract

This memo defines a Yang model that translates the SNMP mib module defined in draft-galikunze-ccamp-g-698-2-snmp-mib for managing single channel optical interface parameters of DWDM applications, using the approach specified in G.698.2. This model is 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 Yang model 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

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/.

Copyright Notice

Table of Contents

1. Introduction

This memo defines a Yang model that translates the SNMP mib module defined in draft-galikunze-ccamp-g-698-2-snmp-mib for managing single channel optical interface parameters of DWDM applications, using the approach specified in G.698.2. This model is to support the optical parameters specified in ITU-T G.698.2 [ITU.G698.2], application identifiers specified in ITU-T G.874.1 [ITU.G874.1] and the Optical Power at Transmitter and Receiver side. 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 the draft-kunze-g-698-2-management-control-framework

The building of a yang model describing the optical parameters defined in G.698.2 [ITU.G698.2], and reflected in G.874.1 [ITU.G874.1], allows the different vendors and operator to retrieve, provision and exchange information across the G.698.2 multi-vendor IaDI in a standardized way. In addition to the parameters specified in ITU recommendations the Yang models support also the "vendor specifica application identifier", the Tx and Rx power at the Ss and Rs points and the channel frequency.

The Yang Model, 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], which draft-galikunze-ccamp-g-698-2-snmp-mib is extending, 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].

This memo specifies a Yang model for optical interfaces.

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.

```
          +-------------------------------------------------+
       Ss |              DWDM Network Elements              | Rs
   +--+ | |  | \                                       / |  |  | +--+
   Tx L1--|->|   \    +------+            +------+   /   |--|-->Rx L1
   +---+  |  |    |   |      |  +------+  |      |  |    |  |    +--+
   +---+  |  |    |   |      |  |      |  |      |  |    |  |    +--+
   Tx L2--|->| OM |-->|------|->| OADM |--|------|->| OD |--|-->Rx L2
   +---+  |  |    |   |      |  |      |  |      |  |    |  |    +--+
   +---+  |  |    |   |      |  +------+  |      |  |    |  |    +--+
   Tx L3--|->|   /    | DWDM |    |  ^    | DWDM |   \   |--|-->Rx L3
   +---+  |  | /      | Link +----|--|----+ Link |     \ |  |    +--+
          +-----------+           |  |           +----------+
                               +--+  +--+
                               |        |
                            Rs v        | Ss
                            +-----+  +-----+
                            |RxLx |  |TxLx |
                            +-----+  +-----+
```

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

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. 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 (R) the parameter can be retrieve with a read, when (W) it can be provisioned by a write, (R,W) can be either read or written.

4.1.1. Rs-Ss Configuration

The Rs-Ss configuration table allows configuration of Central Frequency, Power and Application codes 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) (see G.694.1 Table 1):
This parameter indicates the Central frequency value that Ss and Rs will be set to work (in THz). See the details in Section 6/ G.694.1 (G, S).

Single-channel application codes(see G.698.2):
This parameter indicates the transceiver application code 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 codes 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, it can be either a setting and measured value (G, S).

Current Laser Input power:
This parameter report the current Transceiver Input power (G).

PARAMETERS	Get/Set	Reference
Central frequency Value	G,S	G.694.1 S.6
Single-channel application codes	G	G.698.2 S.5.3
Number of Single-channel application codes Supported	G	N.A.
Current Output Power	G,S	N.A.
Current Input Power	G	N.A.

Table 1: Rs-Ss Configuration

4.1.2. Table of Application Codes

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

Application code Identifier:
The Identifier for the Application code.

Application code Type:
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 Standard = 0, PROPRIETARY = 1
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 Code.

Application code Length:
The number of octets in the Application Code.

Application code:
This is the application code that is defined in G.698.2 or the vendor generated code which has the OUI.

4.2. 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 2 Access Link Power Monitoring


                                       +--------------------------+
                                       |    P(in) = P(Tx) - a(Tx) |
                                       |    ___                   |
       +----------+                    |    \ /  Power Monitor    |
       |          | P(Tx)              |     V                    |
       |  +----+  | Ss    //\\         |     |    |\              |
       |  | TX |----|-----\\//------------------->| \             |
       |  +----+  | Access Link (AL-T) |       .  |  |            |
       |          | attenuation a(Tx)  |       .  |  |==============>
       |          |                    |       .  |  |            |
       | External |                    |      --->| /             |
       | Optical  |                    |          |/              |
       |Transpond.|                    |    P(out)                |
       |          |                    |    ___                   |
       |          |                    |    \ /  Power Monitor    |
       |          | P(Rx)              |     V                    |
       |  +----+  | Rs    //\\         |     |    |\              |
       |  | RX |<---|-----\\//--------------------| \             |
       |  +----+  | Access Link (AL-R) |       .  |  |            |
       |          | Attenuation a(Rx)  |       .  |  |<==============
       +----------+                    |       .  |  |            |
                                       |      <---| /             |
       P(Rx) = P(out) - a(Rx)          |          |/              |
                                       |                          |
                                       |          ROADM           |
                                       +--------------------------+

        -  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:

```
 - Ss, Rs: 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 llustrated in Figure 3.
- The DWDM border NE is capable to periodically measure the optical power Pin and Pout 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 Pin 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

```
     +----------+                    +--------------------------+
     | +------+ |    P(Tx), P(Rx)    |  +-------+               |
     | |      | | =================> |  |       |               |
     | | LMP  | |    P(in), P(out)   |  |  LMP  |               |
     | |      | | <================= |  |       |               |
     | +------+ |                    |  +-------+               |
     |          |                    |                          |
     |          |                    |    P(in) - P(Tx) - a(Tx) |
     |          |                    |    ___                   |
     |          |                    |    \ /  Power Monitor    |
     |          | P(Tx)              |     V                    |
     |  +----+  | Ss    //\\         |     |    |\              |
     |  | TX |----|-----\\//------------------->| \             |
     |  +----+  | Access Link (AL-T) |       .  |  |            |
     |          | attenuation a(Tx)  |       .  |  |==============>
     |          |                    |       .  |  |            |
     | External |                    |      --->| /             |
     | Optical  |                    |          |/              |
     |Transpond.|                    |    P(out)                |
     |          |                    |    ___                   |
     |          |                    |    \ /  Power Monitor    |
     |          | P(Rx)              |     V                    |
     |  +----+  | Rs    //\\         |     |    |\              |
     |  | RX |<---|-----\\//--------------------| \             |
     |  +----+  | Access Link (AL-R) |       .  |  |            |
     |          | Attenuation a(Rx)  |       .  |  |<==============
     +----------+                    |       .  |  |            |
                                     |      <---| /             |
     P(Rx) = P(out) - a(Rx)          |          |/              |
                                     |                          |
                                     |           ROADM          |
                                     +--------------------------+

     - 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

```
    +----------+                       +--------------------------+
    | +------+ | P(Tx),P(Rx),Set(Pout) |  +-------+   +--------+  |
    | |      | | ====================> |  |       |   | Power  |  |
    | | LMP  | | P(in),P(out),Set(PTx) |  |  LMP  |   |Control |  |
    | |      | | <==================== |  |       |   | Loop   |  |
    | +------+ |                       |  +-------+   +--------+  |
    |     |    |                       |                          |
    | +------+ |                       |    P(in) = P(Tx) - a(Tx) |
    | |C.Loop| |                       |    ___                   |
    | +------+ |                       |    \ /  Power Monitor    |
    |     |    | P(Tx)                 |     V                    |
    | +------+ | Ss    //\\            |     |    |\              |
    | | TX  |>---|-----\\//----------------------->| \             |
    | +------+ | Access Link (AL-T)    |       .  |  |            |
    |  VOA(Tx) | attenuation a(Tx)     |       .  |  |==============>
    |          |                       |       .  |  |            |
    | External |                       |      --->| /             |
    | Optical  |                       |          |/              |
    |Transpond.|                       |    P(out)                |
    |          |                       |    ___                   |
    |          |                       |    \ /  Power Monitor    |
    |          | P(Rx)                 |     V                    |
    |  +----+  | Rs    //\\            |     |  VOA(out) |\       |
    |  | RX |<---|-----\\//---------------------<|-------| \      |
    |  +----+  | Access Link (AL-R)    |             .   |  |     |
    |          | attenuation a(Rx)     |             .   |  |<=======
    +----------+                       |        VOA(out) |  |     |
                                       |     <--<|-------| /      |
    P(Rx) = P(out) - a(Rx)             |                 |/       |
                                       |                          |
                                       |            ROADM         |
                                       +--------------------------+
```

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 caracteristics and the Access Link attenuation

Figure 4: Extended LMP Model

4.3. Optical Interface for G.698.2

The ietf-opt-if-g698-2 is an augment to the ietf-interface. It allows the user to set the application code/vendor transceiver class/ Central frequency and the output power. The module can also be used to get the list of supported application codes/transceiver class and also the Central frequency/output power/input power of the interface.

```
     module: ietf-opt-if-g698-2
     augment /if:interfaces/if:interface:
         +--rw optIfOChRsSs
               +--rw ifCurrentApplicationCode
               |  +--rw applicationCodeId    uint8
               |  +--rw applicationCodeType  uint8
               |  +--rw applicationCodeLength uint8
               |  +--rw applicationCode?     string
               +--ro ifSupportedApplicationCodes
               |  +--ro numberApplicationCodesSupported?   uint32
               |  +--ro applicationCodesList* [applicationCodeId]
               |     +--ro applicationCodeId   uint8
               |     +--rw applicationCodeType  uint8
               |     +--rw applicationCodeLength uint8
               |     +--ro applicationCode?    string
               +--rw outputPower?                      int32
               +--ro inputPower?                       int32
               +--rw centralFrequency?                 uint32


    notifications:
   +---n optIfOChCentralFrequencyChange
   |  +--ro if-name?      leafref
   |  +--ro newCentralFrequency
   |     +--ro centralFrequency?   uint32
   +---n optIfOChApplicationCodeChange
   |  +--ro if-name?              leafref
   |  +--ro newApplicationCode
   |     +--ro applicationCodeId?   uint8
   |     +--rw applicationCodeType  uint8
   |     +--rw applicationCodeLength uint8
   |     +--ro applicationCode?     string
```

5. Structure of the Yang Module

ietf-opt-if-g698-2 is a top level model for the support of this feature.

6. Yang Module

The ietf-opt-if-g698-2 is defined as an extension to ietf interfaces.

```
<CODE BEGINS> file "ietf-opt-if-g698-2.yang"

module ietf-opt-if-g698-2 {
     namespace "urn:ietf:params:xml:ns:yang:ietf-opt-if-g698-2";
     prefix ietf-opt-if-g698-2;

     import ietf-interfaces {
       prefix if;
     }

     organization
       "IETF NETMOD (NETCONF Data Modelling Language)
       Working Group";

     contact
       "WG Web:   <http://tools.ietf.org/wg/netmod/>
        WG List:  <mailto:netmod@ietf.org>

        WG Chair: Thomas Nadeau
                  <mailto:tnadeau@lucidvision.com>

        WG Chair: Juergen Schoenwaelder
                  <mailto:j.schoenwaelder@jacobs-university.de>

        Editor:   Dharini Hiremagalur
                  <mailto:dharinih@juniper.net>";

     description
       "This module contains a collection of YANG definitions for
        configuring Optical interfaces.

        Copyright (c) 2013 IETF Trust and the persons identified
        as authors of the code.  All rights reserved.

        Redistribution and use in source and binary forms, with or
        without modification, is permitted pursuant to, and
        subject to the license terms contained in, the Simplified
        BSD License set forth in Section 4.c of the IETF Trust’s
        Legal Provisions Relating to IETF Documents
        (http://trustee.ietf.org/license-info).";

    revision "2015-06-24" {
           description
```

```
                 "Revision 4.0";

              reference
                 " draft-dharini-netmod-dwdm-if-yang 3.0";
       }
       revision "2015-02-24" {
              description
                 "Revision 3.0";

              reference
                 " draft-dharini-netmod-dwdm-if-yang 3.0";
       }
       revision "2014-11-10" {
              description
                 "Revision 2.0";
              reference
                 " ";
       }
       revision "2014-10-14" {
              description
                 "Revision 1.0";
               reference
                 " ";
       }
       revision "2014-05-10" {
              description
                  "Initial revision.";
              reference
                  "RFC XXXX: A YANG Data Model for Optical
                   Management of an Interface for g.698.2
                   support";
       }




         grouping optIfOChApplicationCode  {
                description "Application code entity.";
                leaf applicationCodeId {
                   type uint8 {
                         range "1..255";
                   }
                   description
                         "Id for the Application code";
                }
                leaf applicationCodeType {
                   type uint8 {
                         range "0..1";
                }
                  description
                         "Type for the Application code
                           0 - Standard, 1 - Proprietory
                           When the Type is Proprietory, then the
                           first 6 octets of the applicationCode
                           will be the OUI (organizationally unique
                           identifier)";

                }
                leaf applicationCodeLength {
                   type uint8 {
                         range "1..255";
                   }
                   description
                         "Number of octets in the Application code";

                }
                leaf applicationCode {
                   type string {
                         length "1..255";
                   }
                   description "This parameter indicates the
                        transceiver application code at Ss and Rs as
                        defined in [ITU.G698.2] Chapter 5.3, that
                        is/should be used by this interface.
                        The optIfOChApplicationsCodeList has all the
                        application codes supported by this
                        interface.";

                }
         }


         grouping optIfOChApplicationCodeList {
           description "List of Application codes group.";
           leaf numberApplicationCodesSupported {
               type uint32;
               description "Number of Application codes
                            supported by this interface";
            }
           list applicationCodeList {
               key "applicationCodeId";
               description "List of the application codes";
               uses optIfOChApplicationCode;
           }
         }
```

```
        grouping optIfOChPower {
           description "Interface optical Power";
           leaf outputPower {
               type int32;
               units ".01dbm";
               description "The output power for this interface in
                             .01 dBm.";
           }

           leaf inputPower {
                type int32;
                units ".01dbm";
                config false;
                description "The current input power of this
                      interface";
           }
        }

         grouping optIfOChCentralFrequency {
           description "Interface Central Frequency";
             leaf  centralFrequency {
               type uint32;
               description "This parameter indicate This parameter
                       indicates the frequency of this interface ";

              }
        }

         notification optIfOChCentralFrequencyChange {
            description "A change of Central Frequency has been
                         detected.";
            leaf "if-name" {
                type leafref {
                    path "/if:interfaces/if:interface/if:name";
                }
                description "Interface name";
            }
            container newCentralFrequency {
                   description "The new Central Frequency of the
                                interface";
                   uses optIfOChCentralFrequency;
            }
         }

         notification optIfOChApplicationCodeChange {
            description "A change of Application code has been
                         detected.";
            leaf "if-name" {
                type leafref {
                    path "/if:interfaces/if:interface/if:name";
                }
                description "Interface name";
            }
            container newApplicationCode {
               description "The new application code for the
                   interface";
               uses optIfOChApplicationCode;
            }
         }



         augment "/if:interfaces/if:interface" {
           description "Parameters for an optical interface";
           container optIfOChRsSs {
              description "RsSs path configuration for an interface";
              container ifCurrentApplicationCode {
                   description "Current Application code of the
                                interface";
                   uses optIfOChApplicationCode;
               }

               container ifSupportedApplicationCodes {
                   config false;
                   description "Supported Application codes of
                                the interface";
                   uses optIfOChApplicationCodeList;
               }

               uses optIfOChPower;

               uses optIfOChCentralFrequency;

            }
         }
      }

   <CODE ENDS>
```

7. Security Considerations

The YANG module defined in this memo is designed to be accessed via the NETCONF protocol [RFC6241]. he lowest NETCONF layer is the secure transport layer and the mandatory-to-implement secure transport is SSH [RFC6242]. The NETCONF access control model [RFC6536] provides the means to restrict access for particular NETCONF users to a pre-configured subset of all available NETCONF protocol operation and content.

8. IANA Considerations

This document registers a URI in the IETF XML registry [RFC3688]. Following the format in [RFC3688], the following registration is requested to be made:

URI: urn:ietf:params:xml:ns:yang:ietf-interfaces:ietf-opt-if-g698-2

Registrant Contact: The IESG.

XML: N/A, the requested URI is an XML namespace.

This document registers a YANG module in the YANG Module Names registry [RFC6020].

This document registers a YANG module in the YANG Module Names registry [RFC6020].

prefix: ietf-opt-if-g698-2 reference: RFC XXXX

9. Acknowledgements

Gert Grammel is partly funded by European Union Seventh Framework Programme under grant agreement 318514 CONTENT.

10. Contributors

Dean Bogdanovic
Juniper Networks
Westford
U.S.A.
email deanb@juniper.net

Bernd Zeuner
Deutsche Telekom
Darmstadt
Germany
email B.Zeuner@telekom.de

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

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

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

11. References

11.1. 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, March 2003.

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

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

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

[ITU.G874.1] International Telecommunications Union, "Optical transport network (OTN): Protocol-neutral management information model for the network element view", ITU-T Recommendation G.874.1, January 2002.

[ITU.G959.1]
International Telecommunications Union, "Optical transport network physical layer interfaces", ITU-T Recommendation G.959.1, November 2009.

[ITU.G826]
International Telecommunications Union, "End-to-end error performance parameters and objectives for international, constant bit-rate digital paths and connections", ITU-T Recommendation G.826, November 2009.

[ITU.G8201]
International Telecommunications Union, "Error performance parameters and objectives for multi-operator international paths within the Optical Transport Network (OTN)", ITU-T Recommendation G.8201, April 2011.

[ITU.G694.1]
International Telecommunications Union, "Spectral grids for WDM applications: DWDM frequency grid", ITU-T Recommendation G.694.1, June 2002.

[ITU.G7710]
International Telecommunications Union, "Common equipment management function requirements", ITU-T Recommendation G.7710, May 2008.

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.

[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: +49xxxxxxxxxx
Email: RKunze@telekom.de

Kam Lam (editor)
Alcatel-Lucent
USA

Phone: +1 732 331 3476
Email: kam.lam@alcatel-lucent.com

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

Email: dharinih@juniper.net

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

Phone: +49 1725186386
Email: ggrammel@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

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

Status of This Memo

Copyright Notice

Table of Contents

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 |              DWDM Network Elements              | Rs
+--+ | |  | \                                       / |  |  | +--+
Tx L1--|->|   \    +------+            +------+   /   |--|-->Rx L1
+---+  |  |    |   |      |  +------+  |      |  |    |  |    +--+
+---+  |  |    |   |      |  |      |  |      |  |    |  |    +--+
Tx L2--|->| OM |-->|------|->| OADM |--|------|->| OD |--|-->Rx L2
+---+  |  |    |   |      |  |      |  |      |  |    |  |    +--+
+---+  |  |    |   |      |  +------+  |      |  |    |  |    +--+
Tx L3--|->|   /    | DWDM |    |  ^    | DWDM |   \   |--|-->Rx L3
+---+  |  | /      | Link +----|--|----+ Link |     \ |  |    +--+
       +-----------+           |  |           +----------+
                            +--+  +--+
                            |        |
                         Rs v        | Ss
                         +-----+  +-----+
                         |RxLx |  |TxLx |
                         +-----+  +-----+
```

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])

```
         +------+ Ss    +------+       +------+    Rs +------+
         |      | ----- |      |       |      | ----- |      |
         | OXC1 | ----- | OLS1 | ===== | OLS2 | ----- | OXC2 |
         |      | ----- |      |       |      | ----- |      |
         +------+       +------+       +------+       +------+
           ^  ^             ^              ^             ^  ^
           |  |             |              |             |  |
           |  +-----LMP-----+              +-----LMP-----+  |
           |                                                |
           +---------------------LMP----------------------+
```

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

```
                                       +--------------------------+
                                       |    P(in) = P(Tx) - a(Tx) |
                                       |    ___                   |
       +----------+                    |    \ /   Power Monitor   |
       |          | P(Tx)              |     V    P(in)           |
       |  +----+  | Ss    //\\         |     |    |\              |
       |  | TX |----|-----\\//------------------->| \             |
       |  +----+  | Access Link (AL-T) |       .  |  |            |
       |          | attenuation a(Tx)  |       .  |  |==============>
       |          |                    |       .  |  |            |
       | External |                    |      --->| /             |
       | Optical  |                    |          |/              |
       |Transpond.|                    |    P(out)                |
       |          |                    |    ___                   |
       |          |                    |    \ /   Power Monitor   |
       |          | P(Rx)              |     V    P(out)          |
       |  +----+  | Rs    //\\         |     |    |\              |
       |  | RX |<---|-----\\//--------------------| \             |
       |  +----+  | Access Link (AL-R) |       .  |  |            |
       |          | Attenuation a(Rx)  |       .  |  |<==============
       +----------+                    |       .  |  |            |
                                       |      <---| /             |
       P(Rx) = P(out) - a(Rx)          |          |/              |
                                       |                          |
                                       |           ROADM          |
                                       +--------------------------+
```

- 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:

```
 - Ss, Rs: 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 llustrated in Figure 3.
- The DWDM border NE is capable to periodically measure the optical power Pin and Pout 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 Pin 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

```
     +----------+                    +--------------------------+
     | +------+ |    P(Tx), P(Rx)    |  +-------+               |
     | |      | | =================> |  |       |               |
     | | LMP  | |    P(in), P(out)   |  |  LMP  |               |
     | |      | | <================= |  |       |               |
     | +------+ |                    |  +-------+               |
     |          |                    |                          |
     |          |                    |    P(in) - P(Tx) - a(Tx) |
     |          |                    |    ___                   |
     |          |                    |    \ /  Power Monitor    |
     |          | P(Tx)              |     V                    |
     |  +----+  | Ss    //\\         |     |    |\              |
     |  | TX |----|-----\\//------------------->| \             |
     |  +----+  | Access Link (AL-T) |       .  |  |            |
     |          | attenuation a(Tx)  |       .  |  |==============>
     |          |                    |       .  |  |            |
     | External |                    |      --->| /             |
     | Optical  |                    |          |/              |
     |Transpond.|                    |    P(out)                |
     |          |                    |    ___                   |
     |          |                    |    \ /  Power Monitor    |
     |          | P(Rx)              |     V                    |
     |  +----+  | Rs    //\\         |     |    |\              |
     |  | RX |<---|-----\\//--------------------| \             |
     |  +----+  | Access Link (AL-R) |       .  |  |            |
     |          | Attenuation a(Rx)  |       .  |  |<==============
     +----------+                    |       .  |  |            |
                                     |      <---| /             |
     P(Rx) = P(out) - a(Rx)          |          |/              |
                                     |                          |
                                     |           ROADM          |
                                     +--------------------------+
```

- 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 use case utilizing the LMP protocol with extensions defined in this draft.

Figure 5 Use case 2: Power Control Loop

```
     +----------+                       +--------------------------+
     | +------+ | P(Tx),P(Rx),Set(Pout) |  +-------+   +--------+  |
     | |      | | ====================> |  |       |   | Power  |  |
     | | LMP  | | P(in),P(out),Set(PTx) |  |  LMP  |   |Control |  |
     | |      | | <==================== |  |       |   | Loop   |  |
     | +------+ |                       |  +-------+   +--------+  |
     |     |    |                       |                          |
     | +------+ |                       |    P(in) = P(Tx) - a(Tx) |
     | |C.Loop| |                       |    ___                   |
     | +------+ |                       |    \ /  Power Monitor    |
     |     |    | P(Tx)                 |     V                    |
     | +------+ | Ss    //\\            |     |    |\              |
     | | TX |>----|-----\\//----------------------->| \             |
     | +------+ | Access Link (AL-T)    |       .  |  |            |
     |  VOA(Tx) | attenuation a(Tx)     |       .  |  |==============>
     |          |                       |       .  |  |            |
     | External |                       |      --->| /             |
     | Optical  |                       |          |/              |
     |Transpond.|                       |    P(out)                |
     |          |                       |    ___                   |
     |          |                       |    \ /  Power Monitor    |
     |          | P(Rx)                 |     V                    |
     |  +----+  | Rs    //\\            |     |  VOA(out) |\        |
     |  | RX |<---|-----\\//---------------------<|-------| \       |
     |  +----+  | Access Link (AL-R)    |              .  |  |      |
     |          | attenuation a(Rx)     |              .  |  |<=======
     +----------+                       |        VOA(out) |  |      |
                                        |     <--<|-------| /      |
     P(Rx) = P(out) - a(Rx)             |                 |/       |
                                        |                          |
                                        |            ROADM         |
                                        +--------------------------+
```

- 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 caracteristics 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_ApplicationIdentier	(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:

```
 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     |         (Reserved)            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                      Central Frequency                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Number of Application              |                        |
|   Identifiers Supported              |     (Reserved)         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Single-channel|  A.I. Type    |         A.I. length           |
| Application   |   in use      |                               |
| Identifier    |               |                               |
| Number in use |               |                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           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 identifiers this entry - 8 bits

3. Application Indentifier 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:

```
     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     |         (Reserved)            |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |   Number of Application                |                      |
    |   Identifiers Supported                |     (Reserved)       |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    | Single-channel|  A.I. Type    |         A.I. length           |
    | Application   |               |                               |
    | Identifier    |               |                               |
    | Number        |               |                               |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |               Single-channel Application Identifier           |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |               Single-channel Application Identifier           |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |               Single-channel Application Identifier           |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    //              ....                                           //
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    | Single-channel|               |         A.I. length           |
    | Application   |   A.I. Type   |                               |
    | 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.

```
     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 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:

```
     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     |         (Reserved)            |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                       Output Power                            |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

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:

```
 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     |                   (Reserved)  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                   Current Input Power                         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

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: +49xxxxxxxxxx
Email: RKunze@telekom.de

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

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

```
Internet Engineering Task Force                        G.Galimberti, Ed.
Internet-Draft                                                     Cisco
Intended status: Standards Track                           R.Kunze, Ed.
Expires: January 7, 2016                                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

Status of This Memo

Copyright Notice

Table of Contents

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.

```
             +-------------------------------------------------+
         Ss  |              DWDM Network Elements              | Rs
    +---+ |  |  | \                                       / |  |  | +--+
    Tx L1----|->|   \    +-------+            +-------+   /   |--|-->Rx L1
    +---+    |  |    |   |       |  +-------+  |       |  |    |  |    +--+
    +---+    |  |    |   |       |  |       |  |       |  |    |  |    +--+
    Tx L2----|->| OM |-->|------|->| OADM |--|------|->| OD |--|-->Rx L2
    +---+    |  |    |   |       |  |       |  |       |  |    |  |    +--+
    +---+    |  |    |   |       |  +-------+  |       |  |    |  |    +--+
    Tx L3----|->|   /    | DWDM |    |  ^    | DWDM |   \   |--|-->Rx L3
    +---+    |  | /      | Link +----|--|----+ Link |     \ |  |    +--+
             +-----------+           |  |           +----------+
                                  +--+  +--+
                                  |        |
                               Rs v        | Ss
                               +-----+  +-----+
                               |RxLx |  |TxLx |
                               +-----+  +-----+
      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

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

```
                                      +--------------------------+
                                      |    P(in) = P(Tx) - a(Tx) |
                                      |    ___                   |
      +----------+                    |    \ /  Power Monitor    |
      |          | P(Tx)              |     V                    |
      |  +----+  | Ss    //\\         |     |    |\              |
      |  | TX |----|-----\\//------------------->| \             |
      |  +----+  | Access Link (AL-T) |       .  |  |            |
      |          | attenuation a(Tx)  |       .  |  |==============>
      |          |                    |       .  |  |            |
      | External |                    |      --->| /             |
      | Optical  |                    |          |/              |
      |Transpond.|                    |    P(out)                |
      |          |                    |    ___                   |
      |          |                    |    \ /  Power Monitor    |
      |          | P(Rx)              |     V                    |
      |  +----+  | Rs    //\\         |     |    |\              |
      |  | RX |<---|-----\\//--------------------| \             |
      |  +----+  | Access Link (AL-R) |       .  |  |            |
      |          | Attenuation a(Rx)  |       .  |  |<==============
      +----------+                    |       .  |  |            |
                                      |      <---| /             |
      P(Rx) = P(out) - a(Rx)          |          |/              |
                                      |                          |
                                      |          ROADM           |
                                      +--------------------------+
```

- 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:

```
 - Ss, Rs: 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 llustrated in Figure 3.
- The DWDM border NE is capable to periodically measure the optical power Pin and Pout 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 Pin 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

```
     +----------+                    +--------------------------+
     | +------+ |    P(Tx), P(Rx)    |  +-------+               |
     | |      | | =================> |  |       |               |
     | | LMP  | |    P(in), P(out)   |  |  LMP  |               |
     | |      | | <================= |  |       |               |
     | +------+ |                    |  +-------+               |
     |          |                    |                          |
     |          |                    |    P(in) - P(Tx) - a(Tx) |
     |          |                    |    ___                   |
     |          |                    |    \ /  Power Monitor    |
     |          | P(Tx)              |     V                    |
     |  +----+  | Ss    //\\         |     |     |\             |
     |  | TX |----|-----\\//------------------->| \            |
     |  +----+  | Access Link (AL-T) |       .  |  |            |
     |          | attenuation a(Tx)  |       .  |  |==============>
     |          |                    |       .  |  |            |
     | External |                    |      --->| /             |
     | Optical  |                    |          |/              |
     |Transpond.|                    |    P(out)                |
     |          |                    |    ___                   |
     |          |                    |    \ /  Power Monitor    |
     |          | P(Rx)              |     V                    |
     |  +----+  | Rs    //\\         |     |     |\             |
     |  | RX |<---|-----\\//--------------------| \            |
     |  +----+  | Access Link (AL-R) |       .  |  |            |
     |          | Attenuation a(Rx)  |       .  |  |<==============
     +----------+                    |       .  |  |            |
                                     |      <---| /             |
     P(Rx) = P(out) - a(Rx)          |          |/              |
                                     |                          |
                                     |           ROADM          |
                                     +--------------------------+
```

- 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

```
    +----------+                       +--------------------------+
    | +------+ | P(Tx),P(Rx),Set(Pout) |  +-------+   +--------+  |
    | |      | | ====================> |  |       |   | Power  |  |
    | | LMP  | | P(in),P(out),Set(PTx) |  |  LMP  |   |Control |  |
    | |      | | <==================== |  |       |   | Loop   |  |
    | +------+ |                       |  +-------+   +--------+  |
    |     |    |                       |                          |
    | +------+ |                       |    P(in) = P(Tx) - a(Tx) |
    | |C.Loop| |                       |    ___                   |
    | +------+ |                       |    \ /  Power Monitor    |
    |     |    | P(Tx)                 |     V                    |
    | +------+ | Ss    //\\            |     |    |\              |
    | | TX  |>---|-----\\//---------------------->| \             |
    | +------+ | Access Link (AL-T)    |       .  |  |            |
    |  VOA(Tx) | attenuation a(Tx)     |       .  |  |==============>
    |          |                       |       .  |  |            |
    | External |                       |      --->| /             |
    | Optical  |                       |          |/              |
    |Transpond.|                       |    P(out)                |
    |          |                       |    ___                   |
    |          |                       |    \ /  Power Monitor    |
    |          | P(Rx)                 |     V                    |
    |  +----+  | Rs    //\\            |     |  VOA(out) |\       |
    |  | RX |<---|-----\\//---------------------<|-------| \      |
    |  +----+  | Access Link (AL-R)    |             .   |  |     |
    |          | attenuation a(Rx)     |             .   |  |<=======
    +----------+                       |        VOA(out) |  |     |
                                       |     <--<|-------| /      |
    P(Rx) = P(out) - a(Rx)             |                 |/       |
                                       |                          |
                                       |            ROADM         |
                                       +--------------------------+
```

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 caracteristics 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.

```
      _____________________
                           \
         Path Data Unit    |\
             (ODUk)        | \
      _____________________|  \ __________________
                           |   |                  |  >
        Tandem Data Unit   |   |                  |  |
             (ODUkT)       |   |    OCh  Layer    |   > n och IfEntries
      _____________________|   |                  |  |
                           |   |__________________|  >
              Optical      |  /|                  |  >
          Transport Unit   | / |                  |  |
              (OTUk)       |/  |    OPSn Layer    |   > m ops IfEntries
      _____________________/   |                  |  |
                               |__________________|  >
```

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
           transcievercode 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 tranceiver 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, optIfOChApplicationIdentiferNumber  }
       ::= { optIfOChSrcApplicationIdentifierTable 1 }

   OptIfOChSrcApplicationIdentifierEntry ::=
       SEQUENCE {
           optIfOChApplicationIdentiferNumber            Integer32,
           optIfOChApplicationIdentiferType              Integer32,
           optIfOChApplicationIdentiferLength            Integer32,
           optIfOChApplicationIdentifier                 DisplayString
        }


   optIfOChApplicationIdentiferNumber  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}
   optIfOChApplicationIdentiferType  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}


   optIfOChApplicationIdentiferLength  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 optIfOChApplicationIdentiferType 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: +49xxxxxxxxxx
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: 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

Copyright Notice

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

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.

```
  Overlay                                                  Overlay
  Network       +----------------------------------+       Network
+---------+     |                                  |     +---------+
|  +----+ |     |  +-----+    +-----+    +-----+   |     | +----+  |
|  |    | | UNI |  |     |    |     |    |     |   | UNI | |    |  |
| -+ EN1+-+-----+--+ CN1 +----+ CN2 +----+ CN3 +---+-----+-+ EN3+- |
|  |    | |  +--+--+     |    |     |    |     |   | +---+-|    |  |
|  +----+ |  |  |  +--+--+    +--+--+    +--+--+   | |   | +----+  |
+---------+  |  |     |          |          |      | |   +---------+
             |  |     |          |          |      | |
+---------+  |  |  +--+--+       |       +--+--+   | |   +---------+
|  +----+ |  |  |  |     |       +-------+     +-----+   | +----+  |
|  |    +-+--+  |  | CN4 +---------------+ CN5 |   |     | |    |  |
| -+ EN2+-+-----+--+     |               |     +---+-----+-+ EN4+- |
|  |    | | UNI |  +-----+               +-----+   | UNI | |    |  |
|  +----+ |     |                                  |     | +----+  |
+---------+     +----------------------------------+     +---------+
  Overlay                 Core Network                     Overlay
  Network                                                  Network

                        Legend:   EN  -  Edge Node
                                  CN  -  Core Node
```

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 singled-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.

```
       ---------------------    -----------------------------
      | Domain 1            |  |                    Domain 2 |
      |                     |  |                             |
      |        ---    ---   |  |   ---    ---     ---        |
      |       | A |--| B |--+--+--| U |--| V |---| W |       |
      |      / ---    ---   |  |   ---    ---     --- \      |
      |  ---/               |  |          /       /    \---  |
      | |Src|               |  |         /       /     |Dst| |
      |  ---\               |  |        /       /      /---  |
      |      \ ---    ---   |  |   --- /   --- /  --- /      |
      |       | C |--| D |--+--+--| X |---| Y |--| Z |       |
      |        ---    ---   |  |   ---     ---    ---        |
      |                     |  |                             |
       ---------------------    -----------------------------
```

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:

- 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].

- 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].

- 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:

```
    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
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                 IPv4 tunnel end point address                 |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |          Must Be Zero         |     Tunnel ID                 |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                       Extended Tunnel ID                      |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |          Must Be Zero         |            LSP ID             |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

The IPv4 tunnel end point address, Tunnel ID, Extended Tunnel ID and LSP ID are as defined in [RFC3209].

When the diversity identifier type is set to "IPv4 PCE Allocated 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
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |         Must Be Zero          |           Path Key            |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

The Path Key is defined in [RFC5553].

When the diversity identifier type is set to "IPv4 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 a single number that represents a summarized SRLG for the reference path against which diversity is desired. The node identified by the "IPv4 Diversity Identifier source address" field of the diversity XRO subobject assigns the PAS value.

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 counter part 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 idenifier 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:

```
    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
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                 IPv6 tunnel end point address                 |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |             IPv6 tunnel end point address (cont.)             |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |             IPv6 tunnel end point address (cont.)             |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |             IPv6 tunnel end point address (cont.)             |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |          Must Be Zero         |     Tunnel ID                 |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                       Extended Tunnel ID                      |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                   Extended Tunnel ID (cont.)                  |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                   Extended Tunnel ID (cont.)                  |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                   Extended Tunnel ID (cont.)                  |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |          Must Be Zero         |            LSP ID             |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

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:

```
    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
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |         Must Be Zero          |           Path Key            |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

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:

- When the "destination node exception" flag is set, the exclusion SHOULD be ignored for the destination node.

- 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 Resv 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

Copyright Notice

Table of Contents

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 advertized 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.

```
                            +---+    +---+
                            | P |....| P |
                            +---+    +---+
                           /              \
                      +-----+               +-----+
             +---+    | PE1 |               | PE3 |    +---+
             |CE1|----|     |               |     |----|CE2|
             +---+\   +-----+               +-----+   /+---+
                   \     |                     |     /
                    \ +-----+               +-----+ /
                     \| PE2 |               | PE4 |/
                      |     |               |     |
                      +-----+               +-----+
                            \              /
                            +---+    +---+
                            | P |....| P |
                            +---+    +---+
```

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 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:

- 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].

```
     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    |            Reserved           |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                 SRLG ID 1 (4 bytes)                           |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    ~                           ......                              ~
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                 SRLG ID n (4 bytes)                           |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

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 necesary. 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] Shiomoto, 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

Xufeng Liu
Ericsson
Vishnu Pavan Beeram
Juniper Networks
Alexander Clemm
Cisco
Igor Bryskin
Aihua Guo
ADVA Optical Networking

Expires: April 27, 2015
October 27, 2014

A Yang Data Model for Abstract TE Topologies
draft-liu-yang-abstract-te-topo-00

Status of this Memo

Copyright Notice

Abstract

This document discusses a YANG data model for Abstract TE Topologies.

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

This document defines a YANG [RFC6020] [RFC6021] module for representing and manipulating Abstract TE topologies.

2. Abstract TE Topologies

2.1. Motivation

Clients of a transport network normally have no visibility into the network’s actual Traffic-Engineering (TE) topology and resource availability information. There are numerous reasons for this, such as:

Security considerations: network operators are usually reluctant to expose the network's actual topology to its clients;

Transport network, generally speaking, is comprised of network elements that belong to a different layer network that the client devices. Also the internal network routing and traffic engineering advertisements usually contain proprietary information, which the clients cannot interpret, but discarding of which would lead to incorrect assumptions and decisions. This means that the clients cannot use actual network topology and traffic engineering information even if said information is available;

Scalability considerations: clients do not want to know any transport network information that is not related to the services provided to the clients.

On the other hand the clients need to influence to certain extent on the way the services provided to them are routed across the transport network: some services, for example, need to be as disjoint from each other as possible because they support various network failure protection schemes provisioned in the client layer network; others, on the contrary, need to be co-routed and share fate as much as possible; placement of some services needs to be optimized based on the lowest cost criteria, while other service paths need to be selected to have best optical signal quality or delay characteristics, and so forth.

Different approaches exist to allow for the clients to affect the placement of provided for them services on the transport network under conditions of no visibility into the actual transport network topology and resource availability information. For example, [GMPL-UNI] architecture allows for clients signaling their service routing policies/preferences within the service setup and modify messages and mandates the network path computers to honor said policies/preferences during the service path selection. There are also control plane based (e.g. [GMPLS-ENNI]) and SDN architectures that require the network to expose abstract TE topologies. Such topologies are decoupled from the network actual topologies and are provided on per client group/VPN/tenant basis. The abstract TE topologies are supposed to be fully comprehensible by the clients and contain sufficient information for the client path computers to select service paths according to the client policies. The service paths so selected in terms of abstract TE topology elements could be signaled or otherwise conveyed within service setup/modify requests to the transport network system responsible for the service provisioning.

2.2. Static vs Fluid Abstract TE Topologies

One problem with the abstract TE topologies exposed to the clients is their static nature. The abstract TE topologies are usually manually configured based on the transport network operator policies. This entails tedious error-prone configuration. This also does not allow for the clients to have a say as to how the abstract TE topologies exposed to them should look like, which elements (nodes, links) it should contain, what the parameters (e.g. link bandwidth, SRLGs, etc.) are, and so forth. The problem becomes especially profound in case the clients requirements with respect to the abstract TE topologies change over time and/or depend on particular week, day, time of the day, etc. It is highly desirable to have a data model understood and supported by the transport network and all its potential clients that would allow for the clients to dynamically (re-)configure the abstract TE topologies exposed to them in real time. This document introduces a data model written in YANG, that allows for the clients using NETCONF and/or RESTCONF protocols to (re-)configure abstract topologies, retrieve their data state and, thus, to automate the abstract topology manipulation.

3. Tree Structure

The structure of the groupings in this module are depicted below. Brackets enclose list keys, "rw" means configuration data, "ro" means operational state data, and "?" designates optional nodes.

```
module: abstract-te-topology
augment /nt:network-topology/nt:topology/nt:topology-types/l3t:l3-
unicast-igp-topology:
   +--rw abstract-te-topology!
augment /nt:network-topology/nt:topology/nt:node/nt:termination-
point/l3t:igp-termination-point-attributes:
   +--rw abstract-tp-attributes
      +--rw topo-ref?   leafref
      +--rw node-ref?   leafref
augment /nt:network-topology/nt:topology/nt:node/l3t:igp-node-
attributes:
   +--rw abstract-node-attributes
      +--rw schedules* [schedule-id]
      |  +--rw schedule-id          uint32
      |  +--rw start?               yang:date-and-time
```

```
         |  +--rw schedule-duration?   string
         |  +--rw repeat-interval?     string
         +--rw is-abstract?           boolean
         +--rw underlay-topology?     leafref
         +--rw connectivity-matrix* [id]
         |  +--rw id                        uint32
         |  +--rw from-tp
         |  |  +--rw topo-ref?   leafref
         |  |  +--rw node-ref?   leafref
         |  |  +--rw tp-ref?     leafref
         |  +--rw to-tp
         |  |  +--rw topo-ref?   leafref
         |  |  +--rw node-ref?   leafref
         |  |  +--rw tp-ref?     leafref
         |  +--rw is-allowed?               boolean
         |  +--rw information-source?       enumeration
         |  +--rw credibility-preference?   uint16
         +--rw ted
            +--rw te-router-id-ipv4?    inet:ipv4-address
            +--rw te-router-id-ipv6?    inet:ipv6-address
            +--rw ipv4-local-address* [ipv4-prefix]
            |  +--rw ipv4-prefix    inet:ipv4-prefix
            +--rw ipv6-local-address* [ipv6-prefix]
            |  +--rw ipv6-prefix      inet:ipv6-prefix
            |  +--rw prefix-option?   uint8
            +--rw pcc-capabilities?     pcc-capabilities
   augment /nt:network-topology/nt:topology/nt:link/l3t:igp-link-
   attributes:
      +--rw abstract-link-attributes
         +--rw schedules* [schedule-id]
         |  +--rw schedule-id          uint32
         |  +--rw start?               yang:date-and-time
         |  +--rw schedule-duration?   string
         |  +--rw repeat-interval?     string
         +--rw is-abstract?              boolean
         +--rw server-layer!
         |  +--rw dynamic?     boolean
         |  +--rw committed?   boolean
         +--rw server-path
         |  +--rw path-element* [path-element-id]
         |     +--rw path-element-id    uint32
         |     +--rw loose?             boolean
         |     +--rw (element-type)?
         |        +--:(numbered-link)
         |        |  +--rw link-ip-address?   inet:ip-address
         |        +--:(unnumbered-link)
         |        |  +--rw link-node-id?      uint32
         |        |  +--rw link-id?           uint32
         |        +--:(node)
         |        |  +--rw node-id?           uint32
         |        +--:(label)
         |           +--rw label?             uint32
         +--rw server-backup-path
         |  +--rw path-element* [path-element-id]
         |     +--rw path-element-id    uint32
         |     +--rw loose?             boolean
         |     +--rw (element-type)?
         |        +--:(numbered-link)
         |        |  +--rw link-ip-address?   inet:ip-address
         |        +--:(unnumbered-link)
         |        |  +--rw link-node-id?      uint32
         |        |  +--rw link-id?           uint32
         |        +--:(node)
         |        |  +--rw node-id?           uint32
         |        +--:(label)
         |           +--rw label?             uint32
         +--rw server-protection-type?   uint16
         +--rw server-trail-src
         |  +--rw topo-ref?   leafref
         |  +--rw node-ref?   leafref
         |  +--rw tp-ref?     leafref
         +--rw server-trail-des
         |  +--rw topo-ref?   leafref
         |  +--rw node-ref?   leafref
         |  +--rw tp-ref?     leafref
         +--rw ted
            +--rw link-index?                        uint64
            +--rw information-source?                enumeration
            +--rw credibility-preference?            uint16
            +--rw admin-status?                      enumeration
            +--rw oper-status?                        enumeration
            +--rw area-id?                            binary
            +--rw color?                              uint32
            +--rw max-link-bandwidth?                 decimal64
            +--rw max-resv-link-bandwidth?            decimal64
            +--rw unreserved-bandwidth* [priority]
            |  +--rw priority     uint8
            |  +--rw bandwidth?   decimal64
            +--rw te-default-metric?                  uint32
            +--rw link-protection-type?               enumeration
            +--rw interface-switching-capabilities* [switching-
                                                           capability]
            |  +--rw switching-capability   ted:switching-capabilities
            |  +--rw encoding?              ted:encoding-type
            |  +--rw max-lsp-bandwidth* [priority]
            |  |  +--rw priority     uint8
            |  |  +--rw bandwidth?   decimal64
            |  +--rw packet-switch-capable
            |  |  +--rw minimum-lsp-bandwidth?   decimal64
            |  |  +--rw interface-mtu?           uint16
            |  +--rw time-division-multiplex-capable
            |     +--rw minimum-lsp-bandwidth?   decimal64
            |     +--rw indication?              enumeration
            +--rw srlg
               +--rw srlg-values* [srlg-value]
                  +--rw srlg-value    uint32
   augment /l3t:igp-node-event:
      +--ro abstract-te-topology!
      +--ro abstract-node-attributes
         +--ro schedules* [schedule-id]
         |  +--ro schedule-id          uint32
         |  +--ro start?               yang:date-and-time
         |  +--ro schedule-duration?   string
         |  +--ro repeat-interval?     string
         +--ro is-abstract?          boolean
         +--ro underlay-topology?     leafref
         +--ro connectivity-matrix* [id]
         |  +--ro id                        uint32
         |  +--ro from-tp
         |  |  +--ro topo-ref?   leafref
         |  |  +--ro node-ref?   leafref
         |  |  +--ro tp-ref?     leafref
         |  +--ro to-tp
         |  |  +--ro topo-ref?   leafref
         |  |  +--ro node-ref?   leafref
         |  |  +--ro tp-ref?     leafref
         |  +--ro is-allowed?               boolean
         |  +--ro information-source?       enumeration
         |  +--ro credibility-preference?   uint16
         +--ro ted
            +--ro te-router-id-ipv4?    inet:ipv4-address
            +--ro te-router-id-ipv6?    inet:ipv6-address
            +--ro ipv4-local-address* [ipv4-prefix]
            |  +--ro ipv4-prefix    inet:ipv4-prefix
            +--ro ipv6-local-address* [ipv6-prefix]
            |  +--ro ipv6-prefix      inet:ipv6-prefix
            |  +--ro prefix-option?   uint8
            +--ro pcc-capabilities?     pcc-capabilities
   augment /l3t:igp-link-event:
      +--ro abstract-te-topology!
      +--ro abstract-link-attributes
         +--ro schedules* [schedule-id]
         |  +--ro schedule-id          uint32
         |  +--ro start?               yang:date-and-time
         |  +--ro schedule-duration?   string
         |  +--ro repeat-interval?     string
         +--ro is-abstract?              boolean
         +--ro server-layer!
         |  +--ro dynamic?     boolean
         |  +--ro committed?   boolean
         +--ro server-path
         |  +--ro path-element* [path-element-id]
         |     +--ro path-element-id    uint32
         |     +--ro loose?             boolean
         |     +--ro (element-type)?
         |        +--:(numbered-link)
         |        |  +--ro link-ip-address?   inet:ip-address
         |        +--:(unnumbered-link)
         |        |  +--ro link-node-id?      uint32
         |        |  +--ro link-id?           uint32
         |        +--:(node)
         |        |  +--ro node-id?           uint32
         |        +--:(label)
         |           +--ro label?             uint32
         +--ro server-backup-path
         |  +--ro path-element* [path-element-id]
         |     +--ro path-element-id    uint32
         |     +--ro loose?             boolean
         |     +--ro (element-type)?
         |        +--:(numbered-link)
         |        |  +--ro link-ip-address?   inet:ip-address
         |        +--:(unnumbered-link)
         |        |  +--ro link-node-id?      uint32
         |        |  +--ro link-id?           uint32
         |        +--:(node)
         |        |  +--ro node-id?           uint32
         |        +--:(label)
         |           +--ro label?             uint32
         +--ro server-protection-type?   uint16
         +--ro server-trail-src
         |  +--ro topo-ref?   leafref
         |  +--ro node-ref?   leafref
         |  +--ro tp-ref?     leafref
         +--ro server-trail-des
         |  +--ro topo-ref?   leafref
         |  +--ro node-ref?   leafref
         |  +--ro tp-ref?     leafref
         +--ro ted
            +--ro link-index?                         uint64
            +--ro information-source?                 enumeration
            +--ro credibility-preference?             uint16
            +--ro admin-status?                       enumeration
            +--ro oper-status?                        enumeration
            +--ro area-id?                            binary
            +--ro color?                              uint32
            +--ro max-link-bandwidth?                 decimal64
            +--ro max-resv-link-bandwidth?            decimal64
            +--ro unreserved-bandwidth* [priority]
            |  +--ro priority     uint8
            |  +--ro bandwidth?   decimal64
            +--ro te-default-metric?                  uint32
            +--ro link-protection-type?               enumeration
            +--ro interface-switching-capabilities* [switching-
                                                     capability]
            |  +--ro switching-capability    ted:switching-capabilities
            |  +--ro encoding?               ted:encoding-type
            |  +--ro max-lsp-bandwidth* [priority]
            |  |  +--ro priority     uint8
            |  |  +--ro bandwidth?   decimal64
            |  +--ro packet-switch-capable
            |  |  +--ro minimum-lsp-bandwidth?   decimal64
            |  |  +--ro interface-mtu?           uint16
            |  +--ro time-division-multiplex-capable
            |     +--ro minimum-lsp-bandwidth?   decimal64
            |     +--ro indication?              enumeration
            +--ro srlg
               +--ro srlg-values* [srlg-value]
                  +--ro srlg-value    uint32
```

4. Abstract TE Topology - Yang Module

```
   module abstract-te-topology {
     yang-version 1;
     namespace "urn:ietf:params:xml:ns:yang:abstract-te-topology";
     // replace with IANA namespace when assigned

     prefix "abst";

     import ietf-yang-types {
       prefix "yang";
     }

     import ietf-inet-types {
       prefix "inet";
     }

     import network-topology {
       prefix "nt";
     }

     import l3-unicast-igp-topology {
       prefix "l3t";
     }

     import ted {
       prefix "ted";
     }

     organization "TBD";
     contact "TBD";
     description "Abstract topology model";

     revision "2014-10-27" {
       description "Initial revision";
       reference "TBD";
     }

     grouping abstract-te-topology-type {
       description
         "Identifies the abstract topology type.";
       container abstract-te-topology {
         presence "indicates abstract topology";
         description
           "Its presence identifies the abstract topology type.";
       }
     }

     augment "/nt:network-topology/nt:topology/"
             + "nt:topology-types/l3t:l3-unicast-igp-topology" {
       description
         "Defines the abstract topology type.";
       uses abstract-te-topology-type;
     }

     grouping te-path-element {
       description
         "A group of attributes defining an element in a TE path
         such as TE node, TE link, TE aotomic resource or label.";
       leaf loose {
         type boolean;
         description "true if the element is loose.";
       }
       choice element-type {
         description "Attributes for various element types.";
         case numbered-link {
           leaf link-ip-address {
             type inet:ip-address;
             description "IPv4 or IPv6 address.";
           }
         }
         case unnumbered-link {
           leaf link-node-id {
             type uint32;
             description
               "Node ID of the node where the link end point resides.";
           }
           leaf link-id {
             type uint32;
             description "Identifies the link end point.";
           }
         }
         case node {
           leaf node-id {
             type uint32;
             description "Identifies the node.";
           }
         }
         case label {
           leaf label {
             type uint32;
             description "Identifies atomic TE resource or label.";
           }
         }
       }
     } // te-path-element

     grouping config-schedule-attributes {
       description
         "A list of schedules defining when a particular
          configuration takes effect.";
       list schedules {
         key "schedule-id";
         description "A list of schedule elements.";

         leaf schedule-id {
           type uint32;
           description "Identifies the schedule element.";
         }
         leaf start {
           type yang:date-and-time;
           description "Start time.";
         }
         leaf schedule-duration {
           type string {
             pattern
               ’P(\d+Y)?(\d+M)?(\d+W)?(\d+D)?T(\d+H)?(\d+M)?(\d+S)?’;
           }
           description "Schedule duration in ISO 8601 format.";
         }
         leaf repeat-interval {
           type string {
             pattern
               ’R\d*/P(\d+Y)?(\d+M)?(\d+W)?(\d+D)?T(\d+H)?(\d+M)?’
               + ’(\d+S)?’;
           }
           description "Repeat interval in ISO 8601 format.";
         }
       }
     }

     grouping abstract-node-attributes {
       description "Node attributes in an abstract topology.";
       container abstract-node-attributes {
         description "Node attributes in an abstract topology.";
         uses config-schedule-attributes;
         leaf is-abstract {
           type boolean;
           description
             "true if the node is abstract, false when the node is
             actual.";
         }
         leaf underlay-topology {
           type leafref {
             path "/nt:network-topology/nt:topology/nt:topology-id";
           }
           description
             "When an abstract node encapsulates a topology,
              this reference points to said topology.";
         }
         list connectivity-matrix {
           key "id";
           description
             "Represents node’s switching limitations, i.e. limitations
              in interconnecting network termination points (NTPs)
              across the node.";
           leaf id {
             type uint32;
             description "Identifies the connectivity-matrix entry.";
           }
           container from-tp {
             uses l3t:tp-ref;
             description
               "Reference to source NTP.";
           }
           container to-tp {
             uses l3t:tp-ref;
             description
               "Reference to destination NTP.";
           }
           leaf is-allowed {
             type boolean;
             description
               "true  - switching is allowed,
                false - switching is disallowed.";
           }
           leaf information-source {
             type enumeration {
               enum "unknown" {
                 description "The source is unknown";
               }
               enum "locally-configured" {
                 description "Configured TE link";
               }
               enum "ospfv2" {
                 description "OSPFv2";
               }
               enum "ospfv3" {
                 description "OSPFv3";
               }
               enum "isis" {
                 description "ISIS";
               }
               enum "other" {
                 description "Other source";
               }
             }
             description
               "Indicates the source of the information.";
           }
           leaf credibility-preference {
             type uint16;
             description
```

```
              "The preference value to calculate the traffic
               engineering database credibility value used for
               tie-break selection between different
               information-source values.
               Higher value is more preferable.";
           }

         }
         container ted {
           description "Includes TE node attributes.";
           uses ted:ted-node-attributes;
         }
       }
     } // abstract-node-attributes

     grouping abstract-tp-attributes {
       description
         "Termination point attributes in an abstract topology.";
       container abstract-tp-attributes {
         description
           "Termination point attributes in an abstract topology.";
         uses l3t:node-ref;
       }
     } // abstract-tp-attributes

     grouping abstract-link-attributes {
       description
         "Link attributes in an abstract topology.";
       container abstract-link-attributes {
         description "Link attributes in an abstract topology.";
         uses config-schedule-attributes;
         leaf is-abstract {
           type boolean;
           description "true if the link is abstract.";
         }
         container server-layer {
           presence
             "Indicates the server layer exists for this link.";
           description "State of the server layer of this link.";

           leaf dynamic {
             type boolean;
             description
               "true if the server layer is dynamically created.";
           }
           leaf committed {
             type boolean;
             description
               "true if the server layer is committed.";
           }
         }
         container server-path {
           description
             "The service path on the server layer topology that
              supports this link.";
           list path-element {
             key "path-element-id";
             description
               "A list of path elements describing the service path";
             leaf path-element-id {
               type uint32;
               description "To identify the element in a path.";
             }
             uses te-path-element;
           }
         } // server-path
         container server-backup-path {
           description
             "The backup service path on the server layer topology that
              supports this link.";
           list path-element {
             key "path-element-id";
             description
               "A list of path elements describing the backup service
                path";
             leaf path-element-id {
               type uint32;
               description "To identify the element in a path.";
             }
             uses te-path-element;
           }
         } // server-backup-path
         leaf server-protection-type {
           type uint16;
           description
             "Server layer protection type desired for this link";
         }
         container server-trail-src {
           uses l3t:tp-ref;
           description
             "Source termination point of the server layer trail.";
         }
         container server-trail-des {
           uses l3t:tp-ref;
           description
             "Destination termination point of the server layer
              trail.";
         }
         container ted {
           description "Includes TE link attributes.";
           uses ted:ted-link-attributes;
         }
       }
     } // abstract-link-attributes

     augment "/nt:network-topology/nt:topology/nt:node/"
             + "nt:termination-point/"
             + "l3t:igp-termination-point-attributes" {
       when "../../../topology-types/abstract-te-topology" {
         description
           "The augment is valid only for abstract topology.";
       }
       description "Augments attributes on a termination point.";
       uses abstract-tp-attributes;
     }

     augment "/nt:network-topology/nt:topology/nt:node/"
             + "l3t:igp-node-attributes" {
       when "../../topology-types/abstract-te-topology" {
         description
           "The augment is valid only for abstract topology.";
       }
       description "Augments attributes on a node.";
       uses abstract-node-attributes;
     }

     augment "/nt:network-topology/nt:topology/nt:link/"
             + "l3t:igp-link-attributes" {
       when "../../topology-types/abstract-te-topology" {
         description
           "The augment is valid only for abstract topology.";
       }
       description "Augments attributes on a link.";
       uses abstract-link-attributes;
     }

     augment "/l3t:igp-node-event" {
       description "Augments node event.";
       uses abstract-te-topology-type;
       uses abst:abstract-node-attributes;
     }

     augment "/l3t:igp-link-event" {
       description "Augments link event.";
       uses abstract-te-topology-type;
       uses abst:abstract-link-attributes;
     }
   }
```

5. Security Considerations

The protocol used for sending the TE topology data MUST support authentication and SHOULD support encryption. The data-model by itself does not create any security implications.

6. IANA Considerations

TBD

7. References

7.1. Normative References

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

[RFC6020] Bjorklund, M., "YANG - A Data Modeling Language for the Network Configuration Protocol (NETCONF)", RFC 6020, October 2010.

[RFC6021] Schoenwaelder, J., "Common YANG Data Types", RFC 6021, October 2010.

[RFC6241] Enns, R., Bjorklund, M., Schoenwaelder, J., and A. Bierman, "Network Configuration Protocol (NETCONF)", RFC 6241, June 2011.

[RFC2234] Crocker, D. and Overell, P.(Editors), "Augmented BNF for Syntax Specifications: ABNF", RFC 2234, Internet Mail Consortium and Demon Internet Ltd., November 1997.

[RFC3471] Berger, L., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Functional Description", RFC 3471, January 2003.

[RFC3811] Nadeau, T. and J. Cucchiara, "Definitions of Textual Conventions (TCs) for Multiprotocol Label Switching (MPLS) Management", RFC 3811, June 2004.

[RFC3812] Srinivasan, C., Viswanathan, A., and T. Nadeau, "Multiprotocol Label Switching (MPLS) Traffic Engineering (TE) Management Information Base (MIB)", RFC 3812, June 2004.

[RFC3813] Srinivasan, C., Viswanathan, A., and T. Nadeau, "Multiprotocol Label Switching (MPLS) Label Switching Router (LSR) Management Information Base (MIB)", RFC 3813, June 2004.

[RFC4208] Swallow, G., Drake, J., Ishimatsu, H., and Rekhter, Y., "Generalized Multiprotocol Label Switching (GMPLS) User-Network Interface (UNI): Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Support for the Overlay Model", RFC4208, October 2005.

[RFC4220] Dubuc, M., Nadeau, T., and Lang, J., " Traffic Engineering Link Management Information Base", RFC 4220, November 2005.

[RFC4801] Nadeau, T., Ed. and A. Farrel, Ed., "Definitions of Textual Conventions for Multiprotocol Label Switching (MPLS) Management", RFC 4801, February 2007.

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

7.2. Informative References

[G.8080] ITU-T Rec. G.8080/Y.1304, "Architecture for the Automatically Switched Optical Network (ASON)," November 2001 (and Revision, January 2003). For information on the availability of this document, please see http://www.itu.int.

[I-D.clemm-i2rs-yang-network-topo]

A. Clemm, "A YANG Data Model for Network Topologies", draft-clemm-i2rs-yang-network-topo-01.

[I-D.liu-yang-ted]

Xufeng Liu, "A Yang module for TED", draft-liu-yang-ted-00.

[I-D.clemm-i2rs-yang-l3-topo]

A. Clemm,"A YANG Data Model for Layer 3 Topologies", draft-clemm-i2rs-yang-l3-topo-00.

8. Acknowledgments

TBD

Authors' Addresses

Xufeng Liu
Ericsson
Email: Xufeng.liu@ericsson.com

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

Alexander Clemm
Cisco
Email: alex@cisco.com

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

Aihua Guo
ADVA Optical Networking
Email: aguo@advaoptical.com

Contributors

Gert Grammel
Juniper Networks
Email: ggrammel@juniper.net

Internet Engineering Task Force (IETF)
Internet-Draft
Intended status: Informational
Expires: April 13, 2015

A. Malis, Ed.
Huawei Technologies
R. Skoog
H. Kobrinski
Applied Communication Sciences
G. Clapp
AT&T Labs Research
V. Shukla
Verizon Communications

October 10, 2014

Requirements for Very Fast Setup of GMPLS LSPs
draft-malis-ccamp-fast-lsps-03

Abstract

Establishment and control of Label Switch Paths (LSPs) have become mainstream tools of commercial and government network providers. One of the elements of further evolving such networks is scaling their performance in terms of LSP bandwidth and traffic loads, LSP intensity (e.g., rate of LSP creation, deletion, and modification), LSP set up delay, quality of service differentiation, and different levels of resilience.

The goal of this document is to present target scaling objectives and the related protocol requirements for Generalized Multi-Protocol Label Switching (GMPLS). The document also summarizes key factors affecting current GMPLS signaling procedures in meeting these application scaling requirements.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

Generalized Multi-Protocol Label Switching (GMPLS) [RFC3945] includes an architecture and a set of control plane protocols that can be used to operate data networks ranging from packet-switch-capable networks, through those networks that use Time Division Multiplexing, to WDM networks. The Path Computation Element (PCE) architecture [RFC4655] defines functional components that can be used to compute and suggest appropriate paths in connection-oriented traffic-engineered networks. Additional wavelength switched optical networks (WSON) considerations were defined in [RFC6163].

This document refers to the same general framework and technologies, but adds requirements related to expediting LSP setup, under heavy connection churn scenarios, while achieving low blocking, under an overall distributed control plane. This document focuses on a specific problem space - high capacity and highly dynamic connection request scenarios - that may require clarification and or extensions to current GMPLS protocols and procedures. In particular, the purpose of this document is to address the potential need for protocols and procedures that enable expediting the set up of LSPs in high churn scenarios. Both single-domain and multi-domain network scenarios are considered.

This document focuses on the following two topics: 1) the driving applications and main characteristics and requirements of this problem space, and 2) the key requirements which may be novel with respect to current GMPLS protocols.

This document intends to present the objectives and related requirements for GMPLS to provide the control for networks operating with such performance requirements. While specific deployment scenarios are considered as part of the presentation of objectives, the stated requirements are aimed at ensuring the control protocols are not the limiting factor in achieving a particular network’s performance. Implementation dependencies are out of scope of this document.

It is envisioned that other documents may be needed to define how GMPLS protocols meet the requirements laid out in this document. Such future documents may define extensions, or simply clarify how existing mechanisms may be used to address the key requirements of highly dynamic networks.

2. Background

The Defense Advanced Research Projects Agency (DARPA) Core Optical Networks (CORONET) program [Chiu], is an example target environment that includes IP and optical commercial and government networks, with a focus on highly dynamic and resilient multi-terabit core networks. It anticipates the need for rapid (sub-second) setup and SONET/SDH-like restoration times for high-churn (up to tens of requests per second network-wide and holding times as short as one second) on-demand wavelength, sub-wavelength and packet services for a variety of applications (e.g., grid computing, cloud computing, data visualization, fast data transfer, etc.). This must be done while meeting stringent call blocking requirements, and while minimizing the use of resources such as time slots, switch ports, wavelength conversion, etc.

3. Motivation

The motivation for this document, and envisioned related future documents, is two-fold:

1. The anticipated need for rapid setup, while maintaining low blocking, of large bandwidth and highly churned on-demand connections (in the form of sub-wavelengths, e.g., OTN ODUx, and wavelengths, e.g., OTN OCh) for a variety of applications including grid computing, cloud computing, data visualization, and intra- and inter-datacenter communications.

2. The ability to setup circuit-like LSPs for large bandwidth flows with low setup delays provides an alternative to packet-based solutions implemented over static circuits that may require tying up more expensive and power-consuming resources (e.g., router ports). Reducing the LSP setup delay will reduce the minimum bandwidth threshold at which a GMPLS circuit approach is preferred over a layer 3 (e.g., IP) approach. Dynamic circuit and virtual circuit switching intrinsically provide guaranteed bandwidth, guaranteed low-latency and jitter, and faster restoration, all of which are very hard to provide in a packet-only networks. Again, a key element in achieving these benefits is enabling the fastest possible circuit setup times.

Future applications are expected to require setup times as fast as 100 ms in highly dynamic, national-scale network environments while meeting stringent blocking requirements and minimizing the use of resources such as switch ports, wavelength converters/regenerators, wavelength-km, and other network design parameters. Of course, the benefits of low setup delay diminish for connections with long holding times. The need for rapid setup for specific applications may override and thus get traded off, for these specific applications, against some other features currently provided in GMPLS, e.g., robustness against setup errors.

With the advent of data centers, cloud computing, video, gaming, mobile and other broadband applications, it is anticipated that connection request rates may increase, even for connections with longer holding times, either during limited time periods (such as during the restoration from a data center failure) or over the longer term, to the point where the current GMPLS procedures of path computation/selection and resource allocation may not be timely, thus leading to increased blocking or increased resource cost. Thus, extensions of GMPLS signaling and routing protocols (e.g. OSPF-TE) may also be needed to address heavy churn of connection requests (i.e., high connection request arrival rate) in networks with high traffic loads, even for connections with relatively longer holding times.

4. Driving Applications and Their Requirements

There are several emerging applications that fall under the problem space addressed here in several service areas such as provided by telecommunication carriers, government networks, enterprise networks, content providers, and cloud providers. Such applications include research and education networks/grid computing, and cloud computing. Detailing and standardizing protocols to address these applications will expedite the transition to commercial deployment.

In the target environment there are multiple Bandwidth-on-Demand service requests per second, such as might arise as cloud services proliferate. It includes dynamic services with connection setup requirements that range from seconds to milliseconds. The aggregate traffic demand, which is composed of both packet (IP) and circuit (wavelength and sub-wavelength) services, represents a five to twenty-fold increase over today’s traffic levels for the largest of any individual carrier. Thus, the aggressive requirements must be met with solutions that are scalable, cost effective, and power efficient, while providing the desired quality of service (QoS).

4.1. Key Application Requirements

There are two key performance scaling requirements in the target environment that are the main drivers behind this draft:

1. Connection request rate ranging from a few request per second for high capacity (e.g., 40 Gb/s , 100 Gb/s) wavelength-based LSPs to around 100 request per second for sub-wavelength LSPs (e.g., OTN ODU0, ODU1, and ODU2).

2. Connection setup delay of around 100 ms across a national or regional network. To meet this target, and assuming pipelined cross-connection, and worst case propagation delay and hop count, it is estimated that the maximum processing delay per hop is around 700 microseconds [Lehmen]. Optimal path selection and resource allocation may require somewhat longer processing (up to 5 milliseconds) in either the destination or source nodes and possibly tighter processing delays (around 500 microseconds) in intermediate nodes.

The model for a national network is that of the continental US with up to 100 nodes and LSPs distances up to ˜3000 km and up to 15 hops.

A connection setup delay is defined here as the time between the arrival of a connection request at an ingress edge switch - or more generally a Label Switch Router (LSR) - and the time at which information can start flowing from that ingress switch over that connection. Note that this definition is more inclusive than the LSP setup time defined in [RFC5814] and [RFC6777], which do not include PCE path computation delays.

5. Potential GMPLS Limitations

GMPLS protocols and procedures have been developed to enable automated control of Label Switched Paths (LSPs), including setup, teardown, modification, and restoration, for switching technologies extending from layer 2 and layer 3 packets, to time division multiplexing, to wavelength, and to fiber. Thus GMPLS enables substantial improvement in connection setup delays relative to manual procedures.

However, while the GMPLS protocols are geared for a wide scope of applications and robust performance, they have not specifically addressed the more aggressive characteristics envisioned here, e.g., applications requiring very fast connection setup while maintaining a high success ratio (i.e., low blocking) in a high-churn environment. Preliminary simulations and analyses of national and global scale networks, both WSON and sub-wavelength OTN [Skoog], have shown that using current GMPLS protocols and procedures does not meet the stated performance targets with respect to blocking, setup delays, and resource utilization. These simulations have also indicated limited scalability of current protocols to increasing loads and churn beyond the baseline design.

Some possible issues with existing components of GMPLS include:

1. Path selection and resource allocation in GMPLS networks is based on TE information collected via OSPF-TE LSA updates. Thus, scenarios with highly dynamic connection request activity, where the connection request arrival rate is higher than the TE update rate allowed by OSPF-TE, could lead to unacceptable blocking ratios or low resource utilization. Recall that the minimum LSA update interval is 5 seconds within which time several connections are requested in the scenarios addressed here. Stale TE information leads also, indirectly, to longer setup delays if connection attempts are re-tried. One approach to address this issue is to increase the frequency of LSA updates. Another approach is where TE information collection is incorporated into the signaling protocol which would provide a much more timely view and thus reduced blocking. Furthermore, simultaneously probing multiple paths can be another element to reduce blocking

in scenarios with highly dynamic connection requests. It should be noted that GMPLS supports distributed wavelengths allocation during the signaling phase (i.e., not just based on LSA updates) using the Label Set object and associated procedures of RSVP-TE [RFC3471]. However, in highly dynamic scenarios even the choice of route may be better made in real time rather than based on perhaps stale information. Another recent approach that can reduce the dependence of LSA updates is the use of a stateful PCE that updates an LSP data base as LSPs are set up.

2. In current GMPLS procedures, path computation, and PCC-PCE and PCC-PCC communications occur following the connection request, thus increasing overall setup delays. Although pre-computed paths are not specifically ruled out and thus can be implemented by GMPLS and stored in the PCEs or source nodes, detailed procedures need to be specified. A potential enhancement of periodical off-line downloading of multiple pre-computed paths to individual LSR nodes could, for example, significantly cut down the setup delay.

3. Current GMPLS cross-connection procedures require, as a default, a serial cross-connection processing - the cross-connection in each node must be completed before the signaling message is transmitted to the next node. This serial procedure results in cross-connection delays being accumulated in each node along the path. A procedure allowing simultaneous or pipelined cross-connections could cut this delay contribution by a factor proportional to the path hop count. Pipelined processing can be used with the RSVP-TE Path objects Suggested Label (for the forward direction) and Upstream Label (for the reverse direction). However, their successful use requires accurate resource availability information and wavelength conversion capabilities at all the nodes along the path. In heavy churned connection scenarios, the use of SL and UL objects will either mostly amount to the default serial process or require a lot of wavelength conversions. Note that this delay contribution is significant in WSON - given current optical switching delays of ˜ 10-20 ms or more; it is less significant with TDM or L2 electronic switching.

Note that GMPLS allows for signaling crankbacks when a connection setup fails. Such crankbacks increase the maximum and average setup delays. Thus, reduction of blocking rates, for example, via multiple path probing as in point 1 above, will also improve the worst case and average setup delays.

Note again that these potential GMPLS extensions should be optional as they may entail increased cost or reduced functionality and thus should only be used when needed.

6. Requirements for Very Fast Setup of GMPLS LSPs

This section lists the protocol requirements for very fast setup of GMPLS LSPs in order to adequately support the service characteristics described in the previous sections. These requirements may be the basis for future documents, some of which may be simply informational, while others may describe specific GMPLS protocol extensions. While some of these requirements may be have implications on implementations, the intent is for the requirements to apply to GMPLS protocols and their standardized mechanisms.

6.1. Protocol and Procedure Requirements

R1 Protocol extensions must be backward compatible with existing GMPLS control plane protocols. The purpose of this obvious requirement is to indicate that applications that do not need the performance addressed here and thus do not need the required protocol extensions should be able to use currently existing GMPLS protocols.

R2 Use of optional GMPLS protocol extensions for this application must be selectable by provisioning or configuration.

R3 LSP Establishment time should scale linearly based on number of traversed nodes.

R4 LSP Establishment time should be bounded by a single (worst case) per-node data path (cross-connect) establishment time and not scale linearly based on number of traversed nodes, i.e., support parallel or pipelined cross-connection establishment.

R5 LSP Establishment time shall depend on number of nodes supporting an LSP and link propagation delays and not any off (control) path transactions, e.g., PCC-PCE and PCC-PCC communications at the time of connection setup, even when PCE-based approaches are used.

R6 Must support LSP holding times as short as one second to one minute.

R7 The protocol aspects of LSP signaling must not preclude LSP request rates of tens per second.

R8 The above requirements should be met even when there are failures in connection establishment, i.e., LSPs should be established faster than when crank-back is used.

R9 These requirements are applicable even when an LSP crosses one or more administrative domains / boundaries.

R10 The above are additional requirements and do not replace existing requirements, e.g. alarm free setup and teardown, Recovery, or inter-domain confidentiality.

7. IANA Considerations

This memo includes no requests to IANA.

8. Security Considerations

Being able to support very fast setup and a high churn rate of GMPLS LSPs is not expected to adversely affect the underlying security issues associated with existing GMPLS signaling.

9. Acknowledgements

The authors would like to thank Ann Von Lehmen, Joe Gannett, and Brian Wilson of Applied Communication Sciences for their comments and assistance on this document. Lou Berger provided editorial comments on this document.

10. References

10.1. Normative References

[RFC3471] Berger, L., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Functional Description", RFC 3471, January 2003.

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

[RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, August 2006.

[RFC5814] Sun, W. and G. Zhang, "Label Switched Path (LSP) Dynamic Provisioning Performance Metrics in Generalized MPLS Networks", RFC 5814, March 2010.

[RFC6163] Lee, Y., Bernstein, G., and W. Imajuku, "Framework for GMPLS and Path Computation Element (PCE) Control of Wavelength Switched Optical Networks (WSONs)", RFC 6163, April 2011.

[RFC6777] Sun, W., Zhang, G., Gao, J., Xie, G., and R. Papneja, "Label Switched Path (LSP) Data Path Delay Metrics in Generalized MPLS and MPLS Traffic Engineering (MPLS-TE) Networks", RFC 6777, November 2012.

10.2. Informative References

[Chiu] A. Chiu, et al, "Architectures and Protocols for Capacity Efficient, Highly Dynamic and Highly Resilient Core Networks", Journal of Optical Communications and Networking vol. 4, No. 1, pp. 1-14, January 2012, <http://dx.doi.org/10.1364/JOCN.4.000001>.

[Lehmen] A. Von Lehmen, et al, "CORONET: Testbeds, Demonstration and Lessons Learned", Journal of Optical Communications and Networking vol. 7, No. 1, January 2015 (expected).

[Skoog] R. Skoog, et al, "Analysis and Implementation of a 3-Way Handshake Signaling Protocol for Highly Dynamic Transport Networks", OFC 2014, <http://www.opticsinfobase.org/abstract.cfm?URI=OFC-2014-W1K.1>.

Authors’ Addresses

Andrew G. Malis (editor)
Huawei Technologies

Email: agmalis@gmail.com

Ronald A. Skoog
Applied Communication Sciences

Email: rskoog@appcomsci.com

Haim Kobrinski
Applied Communication Sciences

Email: hkobrinski@appcomsci.com

George Clapp
AT&T Labs Research

Email: clapp@research.att.com

Vishnu Shukla
Verizon Communications

Email: vishnu.shukla@verizon.com

CCAMP Working Group
Internet-Draft
Intended Status: Standards Track
Expires: March 9, 2015

Mike Taillon
Tarek Saad, Ed.
Rakesh Gandhi, Ed.
Zafar Ali
(Cisco Systems, Inc.)
Manav Bhatia
Lizhong Jin
()
Frederic Jounay
(Orange CH)

September 5, 2014

Extensions to Resource Reservation Protocol For Fast Reroute of Traffic Engineering GMPLS LSPs

draft-tsaad-ccamp-rsvpte-bidir-lsp-fastreroute-05

Abstract

This document defines Resource Reservation Protocol - Traffic Engineering (RSVP-TE) signaling extensions to support Fast Reroute (FRR) of Packet Switched Capable (PSC) Generalized Multi-Protocol Label Switching (GMPLS) Label Switched Paths (LSPs). These signaling extensions allow the coordination of bidirectional bypass tunnel assignment protecting a common facility in both forward and reverse directions of a co-routed bidirectional LSP. In addition, these extensions enable the re-direction of bidirectional traffic and signaling onto bypass tunnels that ensure co-routedness of data and signaling paths in the forward and reverse directions after FRR to avoid RSVP soft-state timeout.

Status of this Memo

Copyright and License Notice

Table of Contents

1. Introduction

Packet Switched Capable (PSC) bidirectional Traffic Engineering (TE) tunnels are signaled using Generalized Multi-Protocol Label Switching (GMPLS) signaling procedures specified in [RFC3473]. Fast Reroute (FRR) [RFC4090] has been widely deployed in the packet TE networks today and is preferred for bidirectional TE tunnels. Using FRR also allows to leverage existing mechanisms for failure detection and restoration in the deployed networks.

FRR procedures defined in [RFC4090] describe the behavior of the Point of Local Repair (PLR) to reroute traffic and signaling onto the bypass tunnel in the event of a failure for unidirectional LSPs. These procedures are applicable to unidirectional protected LSPs signaled using either RSVP-TE [RFC3209] or GMPLS procedures [RFC3473], however don’t address issues that arise when employing FRR for bidirectional co-routed GMPLS Label Switched Paths (LSPs).

When bidirectional bypass tunnels are used to locally protect bidirectional co-routed GMPLS LSPs, the upstream and downstream PLRs may independently assign different bidirectional bypass tunnels in the forward and reverse directions. There is no mechanism in FRR procedures defined in [RFC4090] to coordinate the bidirectional bypass tunnel selection between the downstream and upstream PLRs.

When using FRR procedures with bidirectional co-routed GMPLS LSPs, it is possible in some cases (e.g. when using node protection bypass tunnels post a link failure event and when RSVP signaling is sent in-fiber and in-band with data), the RSVP signaling refreshes may stop reaching some nodes along the primary bidirectional LSP path after the PLRs complete rerouting traffic and signaling onto the bypass tunnels. This is caused by the asymmetry of paths that may be taken by the bidirectional LSP’s signaling in the forward and reverse directions after FRR reroute. In such cases, the RSVP soft-state timeout eventually causes the protected bidirectional LSP to be destroyed, and consequently impacts protected traffic flow after FRR.

This document proposes solutions to the above mentioned problems by providing mechanisms in the control plane to complement FRR procedures of [RFC4090] in order to maintain the RSVP soft-state for bidirectional co-routed protected GMPLS LSPs and achieve symmetry in the paths followed by the traffic and signaling in the forward and reverse directions post FRR. The document further extends RSVP signaling so that the bidirectional bypass tunnel selected by the upstream PLR matches the one selected by the downstream PLR node for a bidirectional co-routed LSP.

Unless otherwise specified in this document, fast reroute procedures defined in [RFC4090] are not modified for bidirectional tunnels.

2. Terminology

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

The reader is assumed to be familiar with the terminology in [RFC2205] and [RFC3209].

LSR: Label-Switch Router.

LSP: An MPLS Label-Switched Path. In this document, an LSP will always be explicitly routed.

Local Repair: Techniques used to repair LSP tunnels quickly when a node or link along the LSP’s path fails.

PLR: Point of Local Repair. The head-end LSR of a bypass tunnel or a detour LSP.

Protected LSP: An LSP is said to be protected at a given hop if it has one or multiple associated bypass tunnels originating at that hop.

Bypass Tunnel: An LSP that is used to protect a set of LSPs passing over a common facility.

NHOP Bypass Tunnel: Next-Hop Bypass Tunnel. A bypass tunnel that bypasses a single link of the protected LSP.

NNHOP Bypass Tunnel: Next-Next-Hop Bypass Tunnel. A bypass tunnel that bypasses a single node of the protected LSP.

MP: Merge Point. The LSR where one or more bypass tunnels rejoin the path of the protected LSP downstream of the potential failure. The same LSR may be both an MP and a PLR simultaneously.

Downstream PLR: A PLR that locally detects a fault and reroutes traffic in the same direction of the protected bidirectional LSP RSVP Path signaling.

Upstream PLR: A PLR that locally detects a fault and reroutes traffic in the opposite direction of the protected bidirectional LSP RSVP Path signaling.

Point of Remote Repair (PRR): An upstream PLR that triggers reroute of traffic and signaling based on procedures described in this document.

3. Fast Reroute For Unidirectional GMPLS LSPs

FRR procedures defined in [RFC4090] are applicable to unidirectional protected LSPs signaled using either RSVP-TE or GMPLS procedures and are not modified by the extensions proposed in this document. These FRR procedures also apply to bidirectional associated GMPLS LSPs where two unidirectional GMPLS LSPs are bound together by using association signaling [BID-ASSOC].

4. Bidirectional Bypass Tunnel Assignment for Bidirectional GMPLS LSPs

This section describes signaling procedures for bidirectional bypass tunnel assignment for GMPLS signaled PSC bidirectional co-routed TE LSPs.

4.1. Merge Point Labels

To correctly reroute data traffic over a node protection bypass tunnel, the downstream and upstream PLRs have to know, in advance, the downstream and upstream Merge Point (MP) labels so that data in the forward and reverse directions can be tunneled through the bypass tunnel post FRR respectively.

[RFC4090] defines procedures for the downstream PLR to obtain the protected LSP’s downstream MP label from recorded labels in the RRO of the RSVP Resv message received at the downstream PLR.

To obtain the upstream MP label, existing methods [RFC4090] to record upstream MP label are used in the RRO of the RSVP Path message. The upstream PLR can obtain the upstream MP label from the recorded label in the RRO of the received RSVP Path message.

4.2. Merge Point Addresses

To correctly assign a bidirectional bypass tunnel, the downstream and upstream PLRs have to know, in advance, the downstream and upstream Merge Point (MP) addresses. [RFC4561] defines procedures for the PLR to obtain the protected LSP’s merge point address in multi-domain routing networks where a domain is defined as an Interior Gateway Protocol (IGP) area or an Autonomous System (AS).

[RFC4561] defines procedures for the downstream PLR to obtain the protected LSP’s downstream merge point address from the recorded node-IDs in the RRO of the RSVP Resv message received at the downstream PLR.

To obtain the upstream MP address, existing methods [RFC4561] to record upstream MP node-ID are used in the RRO of the RSVP Path message. The upstream PLR can obtain the upstream MP address from the recorded node-IDs in the RRO of the received RSVP Path message.

4.3. RRO IPv4/IPv6 Subobject Flags

RRO IPv4/IPv6 subobject flags are defined in [RFC4090], Section 4.4 and are applicable to the FRR procedure for the bidirectional tunnels.

[RFC4090] defined procedure is used by the downstream PLR independently to signal the Ipv4/IPv6 subobject flags in the RRO of the RSVP Path message. Similarly, this procedure is used by the upstream PLR independently to signal the IPv4/IPv6 subobject flags in the RRO of the RSVP Resv message.

4.4. Bypass Tunnel Assignment Co-ordination

This document defines a new BYPASS_ASSIGNMENT subobject in RSVP RECORD_ROUTE object used to co-ordinate the bidirectional bypass tunnel selection between the downstream and upstream PLRs.

4.4.1. Bypass Tunnel Assignment Co-ordination Signaling Procedure

It is desirable to coordinate the bidirectional bypass tunnel selected at the downstream and upstream PLRs so that rerouted traffic and signaling flow on co-routed paths post FRR. To achieve this, a new RSVP subobject is defined for RECORD_ROUTE object (RRO) that identifies a bidirectional bypass tunnel that is assigned at a downstream PLR to protect a bidirectional LSP.

The BYPASS_ASSIGNMENT subobject is added by each downstream PLR in the RSVP Path RECORD_ROUTE message of the GMPLS signaled bidirectional primary LSP to record the downstream bidirectional bypass tunnel assignment. This subobject is sent in the RSVP Path RECORD_ROUTE message every time the downstream PLR assigns or updates the bypass tunnel assignment so the upstream PLR may reflect the assignment too. The BYPASS_ASSIGNMENT subobject is added in the RECORD_ROUTE object prior to adding the node’s IP address in the node-ID subobject. A node MUST NOT add a BYPASS_ASSIGNMENT subobject without also adding a Node-ID subobject. A node MUST NOT add a BYPASS_ASSIGNMENT subobject without also adding an IPv4 or IPv6 subobject.

The upstream PLR (downstream MP) that detects a BYPASS_ASSIGNMENT subobject whose bypass tunnel and the node-ID subobject when used as a bypass tunnel source terminates locally assigns the matching bidirectional bypass tunnel in the reverse direction, and forwards the RSVP Path message downstream. Otherwise, the bypass tunnel assignment subobject is simply forwarded downstream along in the RSVP Path message.

In the absence of BYPASS_ASSIGNMENT subobject, the upstream PLR does not assign a bypass tunnel in the reverse direction. This allows the downstream PLR to always initiate the bypass assignment and upstream PLR to simply reflect the bypass assignment.

In the case of upstream PLR receiving multiple BYPASS_ASSIGNMENT subobjects from multiple downstream PLRs, the decision of selecting a bypass tunnel in the reverse direction can be based on local policy, for example, prefer link protection versus node protection bypass tunnel, or prefer the most upstream versus least upstream node protection bypass tunnel.

Bypass assignment co-ordination procedure described above can be used for both one-to-one backup described in Section 3.1 of [RFC4090] and facility backup described in Section 3.2 of [RFC4090].

4.4.2. BYPASS_ASSIGNMENT Subobject

The BYPASS_ASSIGNMENT subobject is used to inform the MP of the bypass tunnel being used by the PLR. This can be used to coordinate the bypass tunnel used for the protected LSP by the downstream and upstream PLRs in the forward and reverse directions respectively prior or post the failure occurrence. This subobject SHOULD only be inserted into the Path message by the downstream PLR and MUST NOT be changed by downstream LSRs.

The BYPASS_ASSIGNMENT subobject in RRO 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   |      Bypass Tunnel ID         |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Type

Downstream Bypass Assignment.

Length

The Length contains the total length of the subobject in bytes, including the Type and Length fields.

Bypass Tunnel ID

The bypass tunnel identifier (16 bits).

5. Link Protection Bypass Tunnels for Bidirectional GMPLS LSPs

When a bidirectional link protection bypass tunnel is used, after a link failure, downstream PLR reroutes RSVP Path and traffic over bypass tunnel using procedures defined in [RFC4090]. Upstream PLR may reroute traffic and RSVP Resv upon detecting the link failure or upon receiving RSVP Path message over a bidirectional bypass tunnel. This allows both traffic and RSVP signaling to flow on symmetric paths in the forward and reverse directions of a bidirectional tunnel.

```
                                                <-RESV
            [R1]---[R2]----[R3]------x-----[R4]-----[R5]
               -> PATH       \              /
                              +<<-------->>+
                                    T3
                                -> PATH
                                   RESV <-


            Protected LSP:  {R1-R2-R3-R4-R5}
            R3’s Bypass T3: {R3-R4}
```

Figure 1: Flow of RSVP signaling post FRR after link failure

Consider the Traffic Engineered (TE) network shown in Figure 1. Assume every link in the network is protected with a link protection bypass tunnel (e.g. bypass tunnel T3). For the protected bidirectional co-routed LSP whose (active) head-end is on router R1 and (passive) tail-end is on router R5, each traversed router (a potential PLR) assigns a link protection bidirectional co-routed bypass tunnel. Consider a link R3-R4 on the protected LSP path fails.

5.1. Behavior Post Link Failure After FRR

The downstream PLR R3 and upstream PLR R4 independently trigger fast reroute procedures to redirect traffic onto bypass tunnels T3 in the forward and reverse directions. The downstream PLR R3 also reroutes

RSVP Path state onto the bypass tunnel T3 using procedures described in [RFC4090]. The upstream PLR R4 reroutes RSVP Resv onto the reverse bypass tunnel T3 upon receiving RSVP Path message over bypass tunnel T3.

6. Node Protection Bypass Tunnels for Bidirectional GMPLS LSPs

```
                T1
           +<<--------->>+
          /               \       <-RESV
[R1]---[R2]----[R3]--x--[R4]---[R5]---[R6]
   -> PATH       \               /
                  +<<--------->>+
                         T2


Protected LSP:  {R1-R2-R3-R4-R5-R6}
R3’s Bypass T2: {R3-R5}
R4’s Bypass T1: {R4-R2}
```

Figure 2: Flow of RSVP signaling post FRR after link failure

Consider the Traffic Engineered (TE) network shown in Figure 2. Assume every link in the network is protected with a node protection bypass tunnel. For the protected bidirectional co-routed LSP whose (active) head-end is on router R1 and (passive) tail-end is on router R6, each traversed router (a potential PLR) assigns a node protection bidirectional co-routed bypass tunnel. Consider a link R3-R4 on the protected LSP path fails.

The proposed solution introduces two phases to invoking FRR procedures by the PLR post the link failure. The first phase comprises of FRR procedures to fast reroute data traffic onto bypass tunnels in the forward and reverse directions. The second phase re-coroutes the data and signaling in the forward and reverse directions after the first phase.

6.1. Behavior Post Link Failure After FRR

The downstream PLR R3 and upstream PLR R4 independently trigger fast reroute procedures to redirect traffic onto respective bypass tunnels T2 and T1 in the forward and reverse directions. The downstream PLR R3 also reroutes RSVP Path state onto the bypass tunnel T2 using procedures described in [RFC4090]. Note, at this point, router R4 stops receiving RSVP Path refreshes for the protected bidirectional LSP while primary protected traffic continues to flow over bypass tunnels.

6.2. Behavior Post Link Failure To Re-coroute

The downstream Merge Point (MP) R5 that receives rerouted protected LSP RSVP Path message through the bypass tunnel, in addition to the regular MP processing defined in [RFC4090], gets promoted to a Point of Remote Repair (PRR role) and performs the following actions to re-coroute signaling and data traffic over the same path in both directions:

- Finds the bypass tunnel in the reverse direction that terminates on the Downstream PLR R3. Note: the Downstream PLR R3’s address is extracted from the "IPV4 tunnel sender address" in the SENDER_TEMPLATE object.

- If found, checks whether the primary LSP traffic and signaling are already rerouted over the found bypass tunnel. If not, PRR R5 activates FRR reroute procedures to direct traffic and RSVP Resv over the found bypass tunnel T2 in the reverse direction.

If downstream MP R5 receives multiple RSVP Path messages through multiple bypass tunnels (e.g. as a result of multiple failures), the PRR SHOULD identify a bypass tunnel that terminates on the farthest downstream PLR along the protected LSP path (closest to the primary bidirectional tunnel head-end) and activate the reroute procedures mentioned above.

```
                                  <- RESV
        [R1]---[R2]----[R3]--X--[R4]---[R5]---[R6]
          PATH ->        \             /
                          +<<------->>+
                         Bypass Tunnel T2
                        traffic + signaling

             Protected LSP:  {R1-R2-R3-R4-R5-R6}
             R3’s Bypass T2: {R3-R5}
```

Figure 3: Flow of RSVP signaling post FRR after re-corouted

Figure 3 describes the path taken by the traffic and signaling after completing re-coroute of data and signaling in the forward and reverse paths described earlier.

The downstream MP MAY optionally support re-corouting in data plane as follows. If the downstream MP is pre-configured with bidirectional bypass tunnel, as soon as the MP node receives the primary tunnel packets on this bypass tunnel, it MAY switch the upstream traffic on to this bypass tunnel. In order to identify the primary tunnel packets through this bypass tunnel, Penultimate Hop Popping (PHP) of the bypass tunnel MUST be disabled. The signaling procedure described above in this Section will still apply, and MP checks whether the primary tunnel traffic and signaling is already rerouted over the found bypass tunnel, if not, perform the above signaling procedure.

7. Compatibility

New RSVP subobject BYPASS_ASSIGNMENT is defined for RECORD_ROUTE in this document. Per [RFC2205], nodes not supporting this subobject will ignore the subobject but forward it without modification.

8. Security Considerations

This document introduces one new RSVP subobject that is carried in a signaling message. Thus in the event of the interception of a signaling message, slightly more information about the state of the network could be deduced than was previously the case. This is judged to be a very minor security risk as this information is already available by other means.

Otherwise, this document introduces no additional security considerations. For general discussion on MPLS and GMPLS related security issues, see the MPLS/GMPLS security framework [RFC5920].

9. IANA Considerations

A new type for the new BYPASS_ASSIGNMENT subobject for RSVP RECORD_ROUTE object is required.

10. Acknowledgements

Authors would like to thank George Swallow for his detailed and useful comments and suggestions.

11. References

11.1. Normative References

[RFC2205] Braden, R., Ed., Zhang, L., Berson, S., Herzog, S., and S. Jamin, "Resource ReSerVation Protocol (RSVP) -- Version 1 Functional Specification", RFC 2205, September 1997.

[RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, December 2001.

[RFC3473] Berger, L., Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions", RFC 3473, January 2003.

[RFC4090] Pan, P., Ed., Swallow, G., Ed., and A. Atlas, Ed., "Fast Reroute Extensions to RSVP-TE for LSP Tunnels", RFC 4090, May 2005.

[BID-ASSOC] Zhang, F., Jing, R., and Gandhi, R., "RSVP-TE Extensions for Associated Bidirectional LSPs", July 2014.

11.2. Informative References

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

[RFC4561] Vasseur, J.-P., Ed., Ali, Z., and S. Sivabalan, "Definition of a Record Route Object (RRO) Node-Id Sub-Object", RFC 4561, June 2006.

[RFC5920] Fang, L., Ed., "Security Framework for MPLS and GMPLS Networks", RFC5920, July 2010.

Authors’ Addresses

Mike Taillon
Cisco Systems, Inc.

EMail: mtaillon@cisco.com

Tarek Saad (editor)
Cisco Systems, Inc.

EMail: tsaad@cisco.com

Rakesh Gandhi (editor)
Cisco Systems, Inc.

EMail: rgandhi@cisco.com

Zafar Ali
Cisco Systems, Inc.

EMail: zali@cisco.com

Manav Bhatia
India

Email: manav@ionosnetworks.com

Lizhong Jin
Shanghai, China

Email: lizho.jin@gmail.com

Frederic Jounay
Orange CH

Email: frederic.jounay@orange.ch

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

Copyright Notice

Table of Contents

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.

```
       +-----+    +-----+     +-----+     +-----+
       |  A  +----+  B  +-----+  C  +-----+  Z  |
       +--+--+    +-----+     +-----+     +--+--+
           \                                /
            \                              /
          +--+--+       +-----+        +--+--+
          |  H  +-------+  I  +--------+  J  |
          +-----+       +-----+        +-----+
```

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.

```
             +-----+       +-----+        +-----+
             |  D  +-------+  E  +--------+  F  |
             +--+--+       +-----+        +--+--+
               /                              \
              /                                \
          +--+--+    +-----+     +-----+     +--+--+
          |  A  +----+  B  +-----+  C  +-----+  Z  |
          +--+--+    +-----+     +-----+     +--+--+
              \                                /
               \                              /
             +--+--+       +-----+        +--+--+
             |  H  +-------+  I  +--------+  J  |
             +-----+       +-----+        +-----+
```

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

```
                                 +-----+      +-----+
                                 |  F  +------+  G  +--------+
                                 +--+--+      +-----+        |
                                    |                        |
                                    |                        |
          +-----+    +-----+     +--+--+      +-----+     +--+--+
          |  A  +----+  B  +-----+  C  +--X---+  D  +-----+  E  |
          +-----+    +-----+     +-----+      +-----+     +-----+
```

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	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	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	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 phrase, the node behavior for setting up the alternative LSP can be classified into the following three categories as shown in Table 1.

```
    +---+       +---+       +---+       +---+       +---+        +---+
    | A |       | B |       | C |       | F |       | G |        | E |
    +-+-+       +-+-+       +-+-+       +-+-+       +-+-+        +-+-+
      |           |           |           |           |            |
      |   PATH    |           |           |           |            |
   C1 +----------X+ C1        |           |           |            |
      |           |   PATH    |           |           |            |
      |           +----------X+ C2        |           |            |
      |           |           |   PATH    |           |            |
      |           |           +----------X+ C3        |            |
      |           |           |           |   PATH    |            |
      |           |           |           +----------X+ C3         |
      |           |           |           |           |   PATH     |
      |           |           |           |           +-----------X+ C2
      |           |           |           |           |            |
      |           |           |           |           |            |
      |           |           |           |           |   RESV     |
      |           |           |           |        C3 +X-----------+ C2
      |           |           |           |   RESV    |            |
      |           |           |        C3 +X----------+            |
      |           |           |   RESV    |           |            |
      |           |        C2 +X----------+           |            |
      |           |   RESV    |           |           |            |
      |        C1 +X----------+           |           |            |
      |   RESV    |           |           |           |            |
   C1 +X----------+           |           |           |            |
```

Figure 4: Restoration LSP Setup Signaling Procedure

As shown in Figure 4, depending on whether the resource is re-used or not, the node behaviors differ. This deviates from normal LSP setup since some nodes do not need to re-configure the cross-connection, and thus should not be viewed as an error. Also, the judgment whether the control plane node needs to send a cross-connection setup/modification command to its corresponding data plane node(s) relies on the check whether the following two cases holds true: (1) the PATH message received include a SE reservation style; (2) the PATH message identifies a LSP that sharing the same tunnel ID as the LSP to share resource with. For the second point, the processing rules and configuration of ASSOCIATION Object defined in [RFC4872] are followed.

4.1.2. LSP Reversion

If the LSP rerouting is revertive, traffic can be reverted to the working or protecting LSP after its failure is recovered. From resource sharing perspective reversion can be divided into two types:

- 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.

```
    +---+       +---+       +---+       +---+       +---+       +---+
    | A |       | B |       | C |       | F |       | G |       | E |
    +-+-+       +-+-+       +-+-+       +-+-+       +-+-+       +-+-+
      |           |           |           |           |           |
      | PATHTEAR  |           |           |           |           |
   D1 +----------X+ D1        |           |           |           |
      |           | PATHTEAR  |           |           |           |
      |           +----------X+ D2        |           |           |
      |           |           | PATHTEAR  |           |           |
      |           |           +----------X+ D3        |           |
      |           |           |           | PATHTEAR  |           |
      |           |           |           +----------X+ D3        |
      |           |           |           |           | PATHTEAR  |
      |           |           |           |           +----------X+ D2
      |           |           |           |           |           |
```

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	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	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	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

starting restoration. Thus reversion is complete.

```
     +---+       +---+       +---+       +---+       +---+
     | A |       | B |       | C |       | D |       | E |
     +-+-+       +-+-+       +-+-+       +-+-+       +-+-+
       |           |           |           |           |
       |   PATH    |           |           |           |
    C1 +----------X+ C1        |           |           |
       |           |   PATH    |           |           |
       |           +----------X+ C2        |           |
       |           |           |   PATH    |           |
       |           |           +----------X+ C1        |
       |           |           |           |   PATH    |
       |           |           |           +----------X+ C2
       |           |           |           |           |
       |           |           |           |           |
       |           |           |           |   RESV    |
       |           |           |        C1 +X----------+ C2
       |           |           |   RESV    |           |
       |           |        C2 +X----------+           |
       |           |   RESV    |           |           |
       |        C1 +X----------+           |           |
       |   RESV    |           |           |           |
    C1 +X----------+           |           |           |
```

Figure 6: 'Make': Reversion LSP Setup follows Working LSP's Path

Figure 6 shows signaling process of reversion LSP setup for working LSP from Section 4.1.1. In this example, resource sharing between reversion and restoration LSP takes place on nodes A, B, C and E. Resource sharing between working and reversion LSP takes place on whole working LPS's path, i.e. A, B, C, D and E. Before reversion LSP is signaled, data plane configuration on nodes A, B, C and E match restoration LSP reservations. On node D data plane configuration matches working LSP reservations.

As already mentioned, MBB reversion uses make-before-break characteristics to overcome challenges related to make-while-break reversion:

- o Rollback

  If 'make' part fails, restoration LSP will still be used to carry existing traffic. Same logic applies here as for any MBB operation failure.

- o Completion guarantee

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.

- 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