CCAMP Working Group
Internet Draft
Intended status: Standard Track

Zafar Ali
Antonello Bonfanti
Cisco Systems
F. Zhang
Huawei Technologies

Expires: August 13, 2014

February 14, 2014

Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extension for Additional Signal Types in G.709 OTN

draft-ali-ccamp-additional-signal-type-g709v3-01.txt

Status of this Memo

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

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

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

This Internet-Draft will expire on August 13, 2014.

Copyright Notice

Abstract

[I-D.draft-ietf-ccamp-gmpls-signaling-g709v3] provides the

extensions to the Generalized Multi-Protocol Label Switching (GMPLS) signaling to control the full set of OTN features including ODU0, ODU4, ODU2e and ODUflex. However, it does not cover additional signal types mentioned in [G.Sup43] (ODU1e, ODU3e1, ODU3e2) or (ODU1f, ODU2f). This draft provides GMPLS signaling extension for these additional signal types.

Conventions used in this document

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

Table of Contents

1. Introduction

[I-D.draft-ietf-ccamp-gmpls-signaling-g709v3] updates the ODU-related portions of [RFC4328] to provide Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) extensions to support control for [G.709-v3]. However, it does not cover additional signal types mentioned in [G.Sup43] (ODU1e, ODU3e1, ODU3e2) or (ODU1f and ODU2f).

With the evolution and deployment of Optical Transport Network (OTN) technology, it is necessary to support additional signal types mentioned in [G.Sup43] and (ODU1f and ODU2f). [I-D.draft-khuzema-ccamp-gmpls-signaling-g709] had support for signal types mentioned in [G.Sup43] but the signal types values collides with values defined in [I-D.draft-ietf-ccamp-gmpls-signaling-g709v3]. The draft has expired and also does not support ODU1f and ODU2f signal type.

This draft provides GMPLS signaling extension to support additional signal types mentioned in [G.Sup43] and (ODU1f and ODU2f).

2. RSVP-TE extension for Additional Signal Types

[I-D.draft-ietf-ccamp-gmpls-signaling-g709v3] defines the format of Traffic Parameters in OTN-TDM SENDER_TSPEC and OTN-TDM FLOWSPEC objects. The said traffic parameters have a signal type field. This document defines the signal type for ODU1e, ODU3e1, ODU3e2, ODU1f and ODU2f, as follows:

Value	Type
23	ODU1e (10Gbps Ethernet [GSUP.43])
24	ODU1f (10Gbps Fiber Channel)
25	ODU2f (10Gbps Fiber Channel)
26	ODU3e1 (40Gbps Ethernet [GSUP.43])
27	ODU3e2 (40Gbps Ethernet [GSUP.43])

3. Security Considerations

This document does not introduce any additional security issues above those identified in [I-D.draft-ietf-ccamp-gmpls-signaling-g709v3].

4. IANA Considerations

This document defines signal type for ODU1e, ODU3e1, ODU3e2, ODU1f and ODU2f to be carried in Traffic Parameters in OTN-TDM SENDER_TSPEC and OTN-TDM FLOWSPEC objects [I-D. draft-ietf-ccamp-gmpls-signaling-g709v3].

5. Acknowledgments

The authors would like to thank Sudip Shukla for comments.

6. References

6.1. Normative References

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

[RFC4328] Papadimitriou, D., Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Extensions for G.709 Optical Transport Networks Control", RFC 4328, January 2006.

[G.709-v3] ITU-T, "Interface for the Optical Transport Network (OTN)", G.709/Y.1331 Recommendation, December 2009.

[GSUP.43] ITU-T, "Proposed revision of G.sup43 (for agreement)", December 2008.

[I-D.draft-ietf-ccamp-gmpls-signaling-g709v3] F.Zhang, G.Zhang, S.Belotti, D.Ceccarelli, K.Pithewan, "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Extensions for the evolving G.709 Optical Transport Networks Control, draft-ietf-ccamp-gmpls-signaling-g709v3, work in progress.

6.2. Informative References

[I-D.draft-khuzema-ccamp-gmpls-signaling-g709] Pithewan, K., et al, "Signaling Extensions for Generalized MPLS (GMPLS) Control of G.709 Optical Transport Networks", expired draft.

Authors' Addresses

Zafar Ali
Cisco Systems
Email: zali@cisco.com

Antonello Bonfanti
Cisco Systems
abonfant@cisco.com

Fatai Zhang
Huawei Technologies
Email: zhangfatai@huawei.com

CCAMP Working Group
Internet Draft
Intended status: Standards Track

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

Expires: August 12, 2014

February 12, 2014

Mutually Exclusive Link Group (MELG)
draft-beeram-ccamp-melg-03

Status of this Memo

Copyright Notice

Abstract

This document introduces the concept of MELG ("Mutually Exclusive Link Group") and discusses its usage in the context of mutually exclusive Virtual TE Links.

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

A Virtual TE Link (as defined in [RFC6001]) advertised into a Client Network Domain represents a potentiality to setup an LSP in the Server Network Domain to support the advertised TE link. The Virtual TE Link gets advertised like any other TE link and follows the same rules that are defined for the advertising, processing and use of regular TE links [RFC4202]. However, "mutual exclusivity" is one attribute that is specific to Virtual TE links. This document discusses the different types of mutual exclusivity (Static vs Dynamic) that come into play and explains the need to advertise this

information into the Client TEDB. It then goes onto introduce a new TE construct (MELG) to carry static mutual exclusivity information.

2. Virtual TE Link - Semantics

A Virtual TE Link (as per existing definitions) represents the potentiality to setup a server layer LSP, but there are currently no strict guidelines imposed on how the underlying server layer LSP would need to get set up. The characteristics of the underlying server-path are not necessarily pinned down until the Virtual TE Link gets actually committed. This means that some important characteristics of the Virtual TE Link like shared-risk and delay (and mutual exclusivity information) may not be known until the corresponding server layer LSP is set up. This makes resource planning (for example - pre-configuring network failure recovery schemes) in a multi-layer network that includes Virtual TE Links a very hard problem.

This document uses a slightly enhanced view of a Virtual TE Link. In the context of this document, the Virtual TE Link (even when it is uncommitted) is always aware of the key characteristics of the underlying server-path. The creation and maintenance of this Virtual TE Link is strictly driven by policy. Policy not only determines which Virtual TE Link to create (What termination points?), but it may also constrain how the corresponding underlying server layer LSP (What path?) needs to get set up. The basic idea behind this "enhanced view" is that it makes the "Virtual TE Link" get as close as it can to representing a "Real TE Link".

Also, as per this document, a Virtual TE Link remains a Virtual TE Link through-out its life-time (until it gets deleted by the user/policy). It may get committed (underlying server LSP gets set up) and uncommitted (underlying server LSP gets deleted) from time to time, but it never really loses it "Virtual" property.

3. Mutually Exclusive Virtual TE Links

Mutual Exclusivity comes into play when multiple Virtual TE Links are dependent on the usage of the same underlying server resource. Since not all of these Virtual TE Links can get committed at the same time, they are deemed to be mutually exclusive.

The existence of this "mutual exclusivity" property would need to be advertised into the Client TE Domain. This is of relevance to Client Path Computation engines; especially those that are capable of doing concurrent computations. If this information is absent, there exists

the risk of the Computation engine yielding erroneous concurrent path computation results where only a subset of the computed paths get successfully provisioned.

The "Mutual Exclusivity" property of a Virtual TE Link can be either static or dynamic in nature.

3.1. Static vs Dynamic

Static Mutual Exclusivity: This type of mutual exclusivity exists permanently within a given network configuration. It comes into play when two or more Virtual TE Links depend on the usage of the same non-shareable underlying server network domain resource. This resource gets used up in its entirety by a single Virtual TE Link when committed. Such resources exist only in the WDM layer.

Dynamic Mutual Exclusivity: This type of mutual exclusivity exists temporarily within a given network configuration. It comes into play when two or more Virtual TE Links depend on the usage of the same shareable underlying server network domain resource. Mutual Exclusivity exists when the amount of the server resource that is available for sharing is limited; it ceases to exist when sufficient amount of the resource is available for accommodating all corresponding Virtual TE Links. Such resources exist in all layers.

Because of their inherent difference, the advertisement paradigm of the TE construct required to carry static mutual exclusivity information is quite different from that of the TE construct required to carry dynamic mutual exclusivity information. Static mutual exclusivity Information can get advertised per TE-Link using a simple sub-TLV construct. There wouldn’t be any scaling issues with this approach because of the static nature of the information that gets advertised. On the contrary, advertising dynamic mutual exclusivity information per TE-Link poses serious scaling concerns and hence requires a different type of construct/paradigm.

This document introduces a new TE construct for carrying static mutual exclusivity information. The mechanisms to address dynamic mutual exclusivity are discussed in a separate document [SRcLG].

4. Static Mutual Exclusivity

Consider the network topology depicted in Figure 1a. This is a typical packet optical transport deployment scenario where the WDM layer network domain serves as a Server Network Domain providing

transport connectivity to the packet layer network Domain (Client Network Domain).

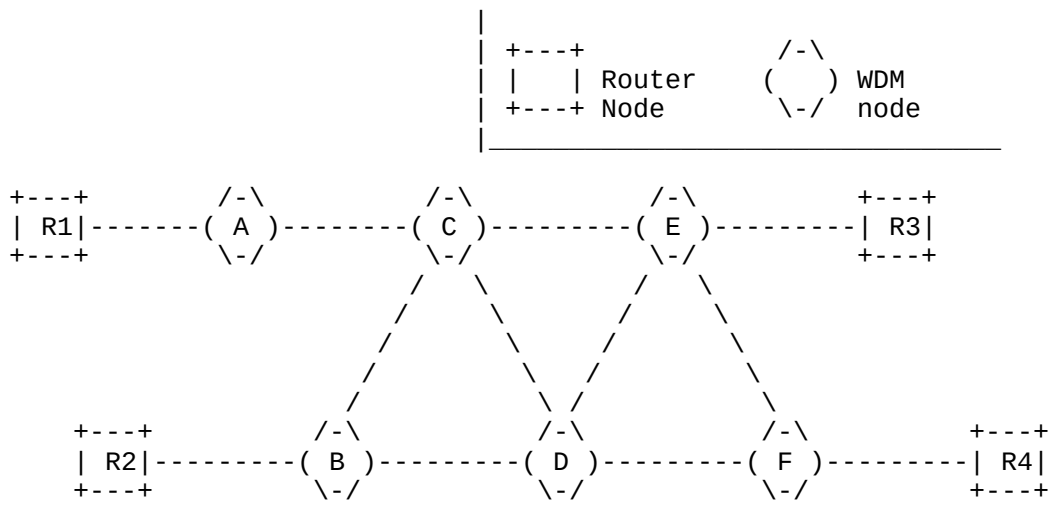


Figure 1a: Sample topology

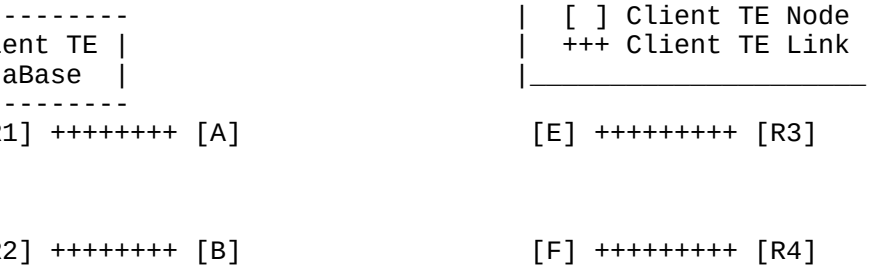


Figure 1b: Client TE Database

Nodes R1, R2, R3 and R4 are IP routers that are connected to an Optical WDM transport network. A, B, C, D, E and F are WDM nodes that constitute the Server Network Domain. The border nodes (A, B, E

and F) operate in both the server and client domains. Figure 1b depicts how the Client Network Domain TE topology looks like when there are no Client TE Links provisioned across the optical domain.

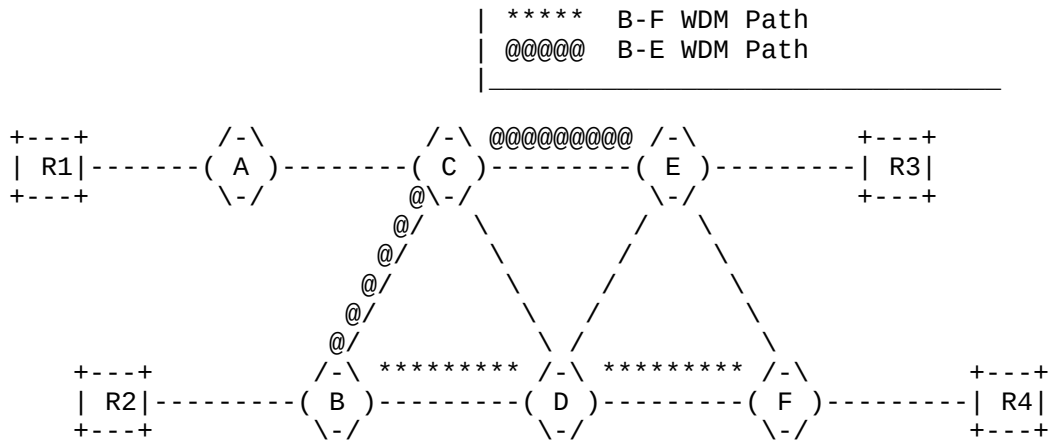


Figure 2a: Mutually Exclusive potential WDM paths

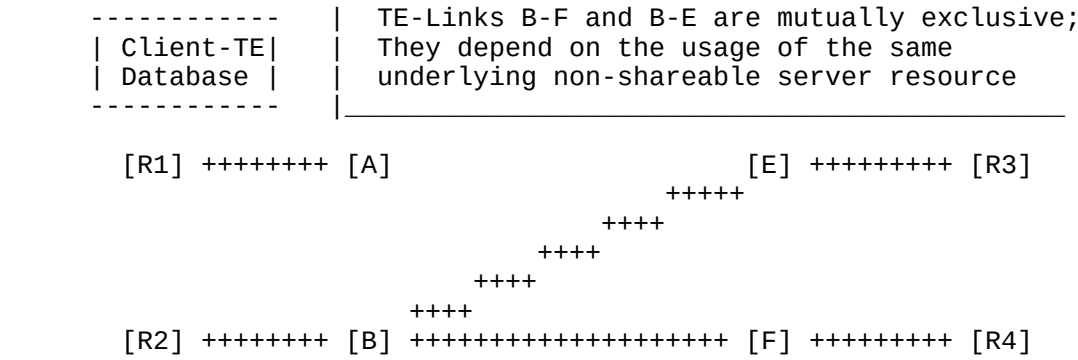


Figure 2b: Client TE Database - Mutually Exclusive Virtual TE Links

Now consider augmenting the Client TE topology by creating a couple of Virtual TE Links across the optical domain. The potential paths in the WDM network catering to these two virtual TE links are as shown in Fig 2a and the corresponding augmented Client TE topology is as illustrated in Fig 2b.

In this particular example, the potential paths in the WDM layer network supporting the Virtual TE Links require the usage of the same source transponder (on "Node B"). Because the Virtual TE Links depend on the same uncommitted network resource, only one of them could get activated at any given time. In other words they are mutually exclusive. This scenario is encountered when the potential paths depend on any common physical resource (e.g. transponder, regenerator, wavelength converter, etc.) that could be used by only one Server Network Domain LSP at a time.

This document proposes the use of "Mutually Exclusive Link Group (MELG)" for catering to this scenario.

5. Mutually Exclusive Link Group

The Mutually Exclusive Link Group (MELG) construct defined in this document has 2 purposes

- To indicate via a separate network unique number (MELG ID) an element or a situation that makes the advertised Virtual TE Link belong to one or more Mutually Exclusive Link Groups. Path computing element will be able to decide on whether two or more Virtual TE Links are mutually exclusive or not by finding an overlap of advertised MELGs (similar to deciding on whether two or more TE links share fate or not by finding common SRLGs)

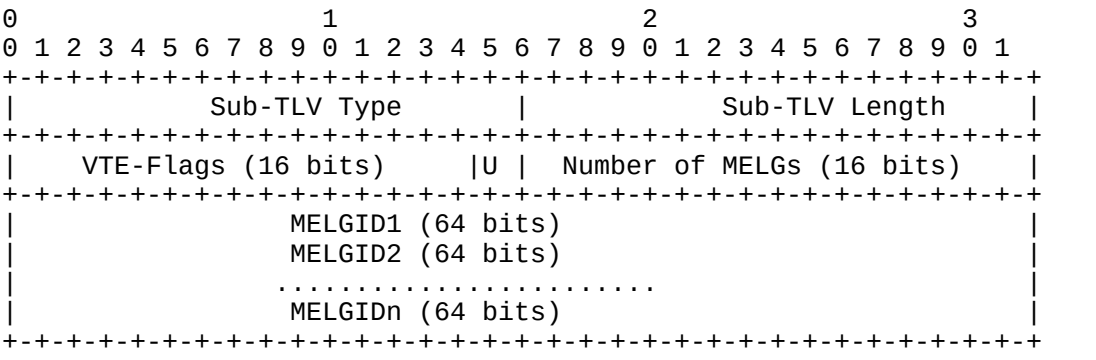
- To indicate whether the advertised Virtual TE Link is committed or not at the moment of the advertising. Such information is important for a path computation element: Committing new Virtual TE links (vs. re-using already committed ones) has a consequence of allocating more server layer resources and disabling other Virtual TE Links that have common MELGs with newly committed Virtual TE Links; Committing a new Virtual TE Link also means a longer setup time for the Client LSP and higher risk of setup-failure.

6. Protocol Extensions

6.1. OSPF

The MELG is a sub-TLV of the top level TE Link TLV. It may occur at most once within the Link TLV. The format of the MELGs sub-TLV is defined as follows:

Name: MELG
Type: TBD
Length: Variable



Number of MELGs: number of MELGS advertised for the Virtual TE Link;
VTE-Flags: Virtual TE Link specific flags;
MELGID1,MELGID2,...,MELGIDn: 64-bit network domain unique numbers associated with each of the advertised MELGs

Currently defined Virtual TE Link specific flags are:
U bit (bit 1): Uncommitted - if set, the Virtual TE Link is uncommitted at the time of the advertising (i.e. the server layer network LSP is not set up); if cleared, the Virtual TE Link is committed (i.e. the server layer LSP is fully provisioned and functioning). All other bits of the "VTE-Flags" field are reserved for future use and MUST be cleared.

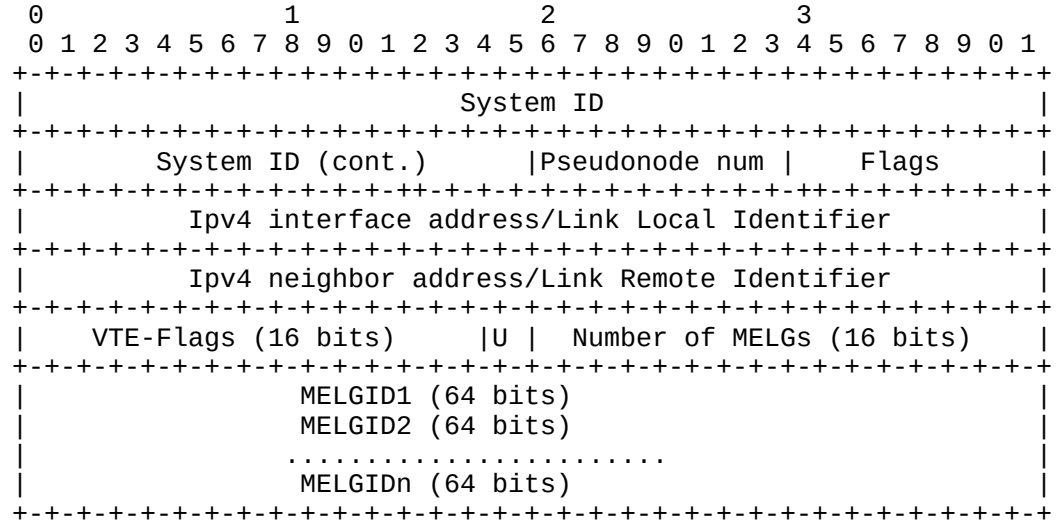
Note: A Virtual TE Link advertisement MAY include MELGs sub-TLV with zero MELGs for the purpose of communicating to the TE domain whether the Virtual TE Link is currently committed or not.

6.2. ISIS

The MELG TLV (of type TBD) contains a data structure consisting of:

6	octets of System ID
1	octet of Pseudonode Number
1	octet Flag
4	octets of IPv4 interface address or 4 octets of a Link Local Identifier
4	octets of IPv4 neighbor address or 4 octets of a Link Remote Identifier
2	octets MELG-Flags
2	octets - Number of MELGs
variable	List of MELG values, where each element in the list has 8 octets

The following illustrates encoding of the value field of the MELG TLV.



The neighbor is identified by its System ID (6 octets), plus one octet to indicate the pseudonode number if the neighbor is on a LAN interface.

The least significant bit of the Flag octet indicates whether the interface is numbered (set to 1) or unnumbered (set to 0). All other bits are reserved and should be set to 0.

The length of the TLV is 20 + 8 * (number of MELG values).

The semantics of "VTE-Flags", "Number of MELGs" and "MELGID Values" are the same as the ones defined under OSPF extensions.

The MELG TLV MAY occur more than once within the IS-IS Link State Protocol Data Units.

7. Security Considerations

TBD

8. IANA Considerations

8.1. OSPF

IANA is requested to allocate a new sub-TLV type for MELG (as defined in Section 6.1) under the top-level TE Link TLV.

8.2. ISIS

IANA is requested to allocate a new IS-IS TLV type for MELG (as defined in Section 6.2).

9. Normative References

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

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

[RFC6001] D.Papadimitriou, M.Vigoureax, K.Shiomoto, D.Brungard and JL. Le Roux, "GMPLS Protocol Extensions for Multi-Layer and Multi-Region Networks", RFC 6001, October 2010.

[SRcLG] Beeram, V., "Shared Resource Link Group", draft-beeram-ccamp-srclg, February 2014

10. Acknowledgments

Chris Bowers [cbowers@juniper.net]

Authors' Addresses

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

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

John Drake
Juniper Networks
Email: jdrake@juniper.net

Gert Grammel
Juniper Networks
Email: ggrammel@juniper.net

Wes Doonan
Email: wddlists@gmail.com

Manuel Paul
Deutsche Telekom
Email: Manuel.Paul@telekom.de

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

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

Cyril Margaria
Juniper Networks
Email: cmargaria@juniper.net

Friedrich Armbruster
Coriant GmbH

Email: friedrich.armbruster@coriant.com

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

Fatai Zhang
Huawei Technologies
Email: zhangfatai@huawei.com

CCAMP Working Group
Internet Draft
Intended status: Standards Track

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

Expires: August 14, 2014

February 14, 2014

Network Assigned Upstream-Label
draft-beeram-ccamp-network-assigned-upstream-label-02

Status of this Memo

Copyright Notice

Abstract

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

Conventions used in this document

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

Table of Contents

1. Introduction

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

scenarios. These mechanisms enable a given node to offload the task of assigning the upstream-label for a given LSP onto the network.

2. Symmetric Labels

As per [RFC3471], the upstream-label and the downstream-label for an LSP at a given hop need not be the same. The use-case discussed in this document (Section 4) pertains to Lambda Switch Capable (LSC) LSPs and it is an undocumented fact that in practice, LSC LSPs always have symmetric labels at each hop along the path of the LSP.

The protocol mechanisms discussed in this document assume "Label Symmetry" and are meant to be used only for Bidirectional LSPs that assign Symmetric Labels at each hop along the path of the LSP.

3. Unassigned Upstream Label

This document proposes the use of a special label value - "0xFFFFFFFF" - to indicate an Unassigned Label. The presence of this value in the UPSTREAM_LABEL object of a PATH message indicates that the upstream-node has not assigned an upstream label on its own and has requested the downstream-node to provide a label that it can use in both forward and reverse directions.

3.1. Processing Rules

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

In response, the downstream-node picks an appropriate symmetric label and sends it via the LABEL object in the RESV message. The upstream-node would then start using this symmetric label for both directions of the LSP. If the downstream-node cannot pick the symmetric label, it MUST issue a PATH-ERR message with a "Routing Problem/Unacceptable Label Value" indication.

The upstream-node will continue to signal the Unassigned Upstream Label in the PATH message even after it receives an appropriate symmetric label in the RESV message. This is done to make sure that

the downstream-node would pick a symmetric label if and when it needs to change the RESV label at a later point in time.

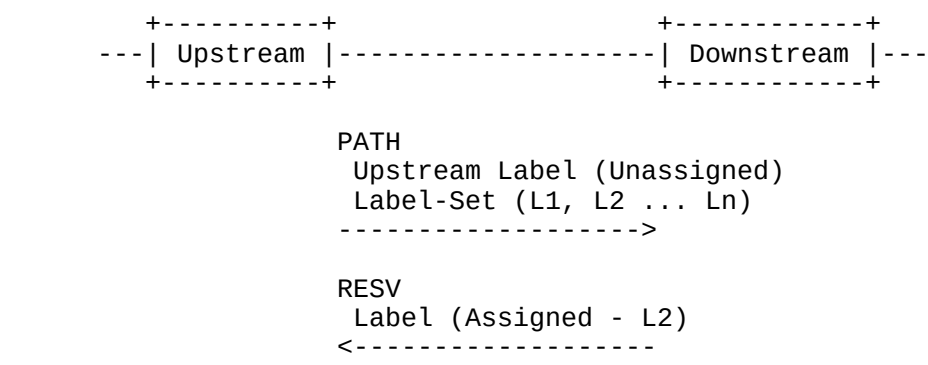


Figure 1: Unassigned UPSTREAM_LABEL

3.2. Backwards Compatibility

If the downstream-node is running an older implementation and doesn't understand the semantics of an Unassigned UPSTREAM LABEL, it will either (a) reject the special label value and generate an error or (b) accept it and treat it as a valid label.

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

4. Use-Case

4.1. Alien-Wavelength Setup

Consider the network topology depicted in Figure 2. Nodes A and B are client IP routers that are connected to an optical WDM transport network. F, H and I represent WDM nodes. The transponder sits on the router and is directly connected to the add-drop port on a WDM node.

The optical signal originating on "Router A" is tuned to a particular wavelength. On "WDM-Node F", it gets multiplexed with optical signals at other wavelengths. Depending on the implementation of this multiplexing function, it may not be

acceptable to have the router send signal into the optical network unless it is at the appropriate wavelength. In other words, having the router send signal with a wrong wavelength may adversely impact existing optical trails. If the clients do not have full visibility into the optical network, they are not in a position to pick the correct wavelength up-front.

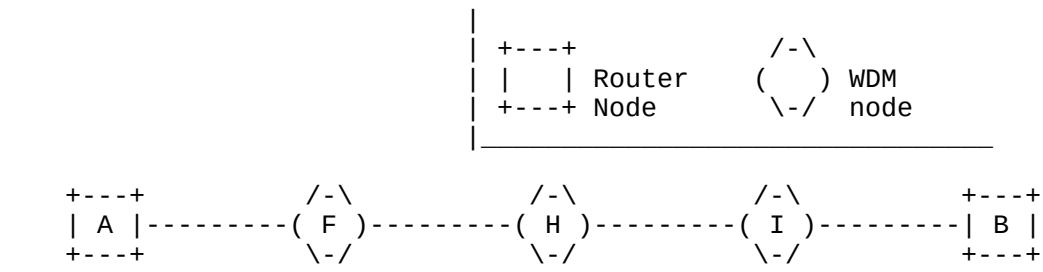


Figure 2: Sample topology

The mechanisms proposed in this document allow for the optical network to select and communicate the correct wavelength for such clients.

4.1.1. Initial Setup

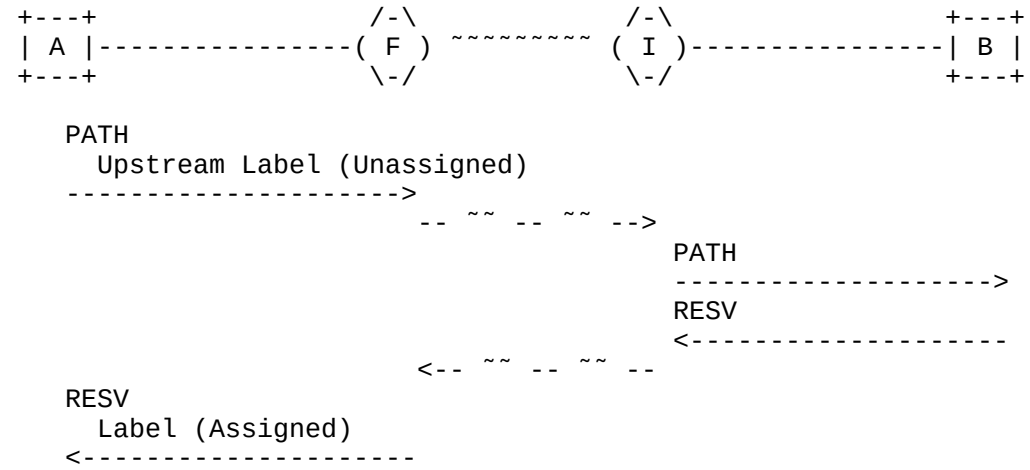


Figure 3: Alien Wavelength - Initial Setup

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

4.1.2. Wavelength Change

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

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

5. Security Considerations

TBD

6. IANA Considerations

TBD

7. Normative References

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

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

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

8. Acknowledgments

TBD

Authors' Addresses

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

John Drake
Juniper Networks
Email: jdrake@juniper.net

Gert Grammel
Juniper Networks
Email: ggrammel@juniper.net

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

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

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

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

CCAMP Working Group
Internet-Draft
Intended status: Informational
Expires: May 9, 2014

D. Ceccarelli
Ericsson
O. Gonzalez de Dios
Telefonica I+D
F. Zhang
X. Zhang
Huawei Technologies
November 5, 2013

Use cases for operating networks in the overlay model context
draft-ceccadedios-ccamp-overlay-use-cases-04

Abstract

This document defines a set of use cases for operating networks in the overlay model context through the Generalized Multiprotocol Label Switching (GMPLS) overlay interfaces.

Table of Contents

1. Introduction

The GMPLS overlay model [RFC 4208] specifies a client-server relationship between networks where client and server domains are managed as separate domains because of trustiness, scalability and operational issue. By means of procedures from the GMPLS protocol suite it is possible to build a topology in the client (overlay) network from Traffic Engineering paths in the server network. In this context, the UNI (User to Network Interface) is the demarcation point between networks. It is a boundary where policies, administrative and confidentiality issues apply that limit the exchange of information.

This GMPLS overlay model supports a wide variety of network scenarios. The packet over optical scenario is probably the most popular example where the overlay model applies.

In order to exploit the full potential of client/server network interworking in the overlay model, it may be desirable to know in advance whether is it feasible or not to connect two client network nodes [INTERCON-TE]. This requires having a certain amount of TE information of the server network in the client network. This need not be the full set of TE information available within each network, but does need to express the potential of providing TE connectivity. This subset of TE information is called TE reachability information.

The goal of this document is to define a set of solution independent use cases applicable to the overlay model. In particular it focuses on the network scenarios where the overlay model applies and analyzes the most interesting aspects of provisioning, recovery and path computation.

2. Terminology

The following terms are used within the document:

- Edge node [RFC4208]: node of the client domain belonging to the overlay network, i.e. nodes with at least one interface connected to the server domain.

- Core node [RFC4208]: node of the server domain.

- Access link: link between core node and edge node. It is the link where the UNI is usually implemented.

- Remote node: node in the client domain which has no direct access to the server domain but can reach it through an edge node

in its same administrative domain.

- Local trigger: LSP setup request issued to an edge node. It triggers the setup of a client domain FA through the server domain via a UNI interface.

- Remote trigger: LSP setup request issued to a remote node. It triggers the setup of a client domain LSP which, upon reaching an edge node, will use connectivity in the server domain dynamically provided via an UNI interface.

All the use cases listed in the sections below can be applied to any combination of, unless otherwise specified:

* Local trigger or remote signaling

* Administrative boundary or administrative plus technological boundary

* Layer transition on edge node or on core node (applicable to administrative plus technological boundary case)

With local trigger we mean the case in which a trigger for the provisioning of a service over the overlay interface is issued to one of the edge nodes belonging to the overlay network, i.e. directly connected to the UNI.

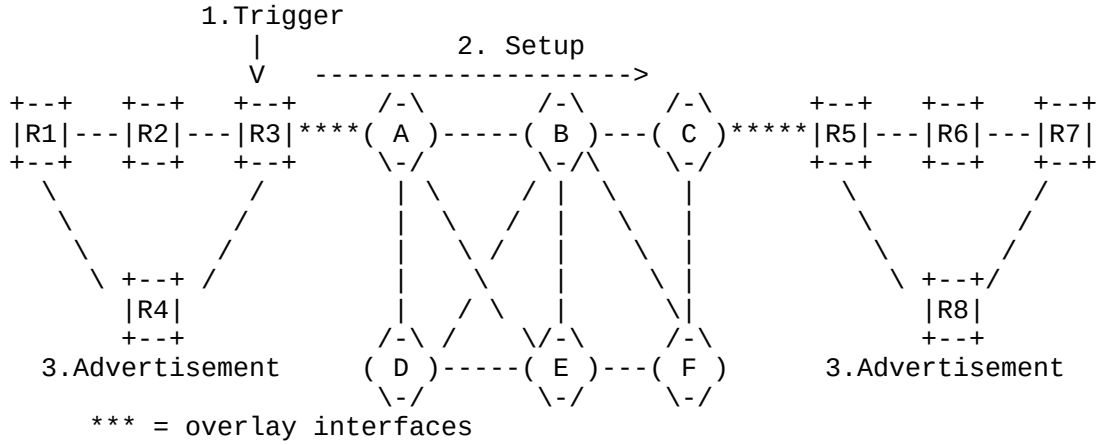


Figure 1: Local trigger

As it is possible to see in the figure above, a trigger is issued on

R3 (edge node) for starting the setup request procedure over the overlay interface (R3-A). Once the LSP in the server domain is setup and an adjacency in the packet domain between R3 and R5 is created, it can be advertised in the rest of the client domain and used by the signaling protocol (e.g. LDP) for setting up end-to-end (e.g. from R1 to R7) client domain LSPs.

On the other hand, the remote signaling consists on the utilization of a connection oriented signaling protocol in the client domain that allows issuing the end to end service setup trigger directly on the end nodes of the client domain. The signaling message, upon reaching the edge node (R3), will trigger the setup of the service in the server domain via the overlay interface.

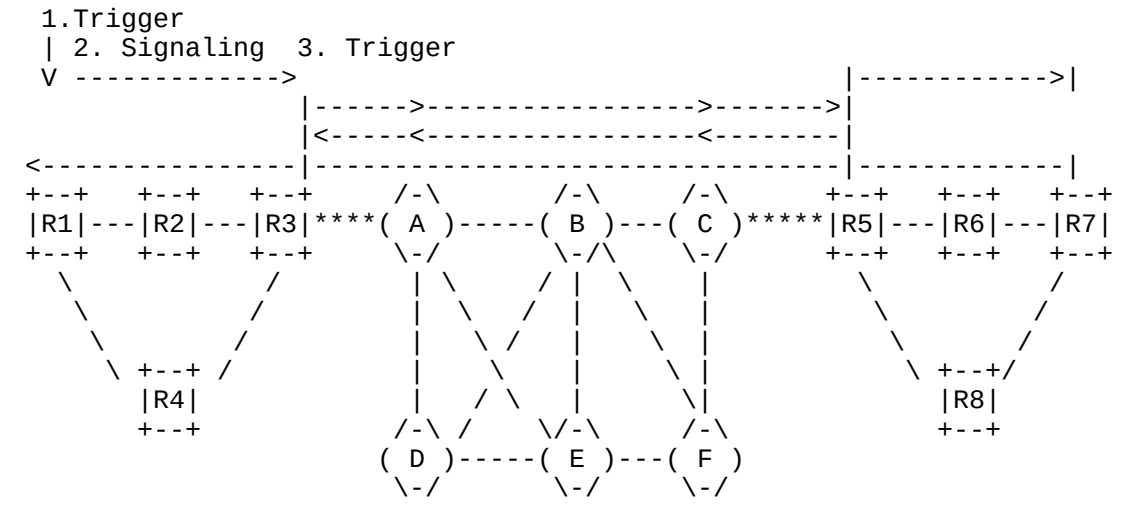


Figure 2: Remote Signaling

The utilization of the remote trigger allows for a strict control of the resources that will be used for the setup of the end to end service. In order to have a correct setup of the end to end service the trigger issued to R1 must include the overlay nodes to be used for the setup of the service in the server domain (R3 and R5). The network operator is supposed to know that the edge nodes to be used are R3 and R5.

The second bullet above speaks about administrative boundaries and administrative plus technological boundaries. Since the overlay is an administrative boundary between a client and a server domain, it is possible to configure it between a client and a server domain with

the same switching capabilities (e.g., IP over IP) or between domains with different switching capabilities (e.g., OTN over WDM). In the former case the boundary is referred to as administrative domain, while in the latter, it is referred to as both administrative and technological boundary.

In the case of boundary which is both administrative and technological a further distinction is needed and regards the node where the technological transition occurs, i.e., on the edge or on the core node.

One of the most common cases of administrative and technological boundary is the IP over WDM, where we speak about grey and colored overlay interfaces. In other words, in the case of grey interface the transponder and the domain transition are on the core node, while in the case of colored interface they are on the edge node. The physical impairments to be considered are different in the two cases (for further details please see Appendix A) but the behavior of the interface does not change and all use cases depicted below can be applied both to the grey and colored interfaces.

Generalizing what said above for the IP over WDM case, when the layer transition occurs on the edge node, the edge node is equipped with at least one interface with the switching capability of the client domain and one interface with the switching capability of the server domain. Viceversa, when layer transition occurs on the core node, it is the core node the one with at least two different interfaces with different switching capabilities.

Editor note: Actually path computation is assumed to be performed tipically at the server domain. The client domain can request the server domain for computing a path or select among a set of paths computed by the server domain and exported to the client domain as virtual/abstract topology.

3. Client domain to server domain connectivity

A further distinction criterion, which is applicable to most of the use cases below, is the degree of connectivity between the client domain and the server domain. Three scenarios are identified:

* Single homing

* Dual homing

* Multiple single homing(editor note: better name is welcome)

3.1. Single homing

In the case of single homing we consider an end to end tunnel with a single LSP in the client domain and one or more LSPs in the server domain but a single overlay interface connecting them. The scenario is shown in figure below, where an end to end circuit between R1 and R7 is built over a tunnel between R3 and R5 composed by a single LSP restorable between A and C or more (possibly restorable) LSPs between A and C.

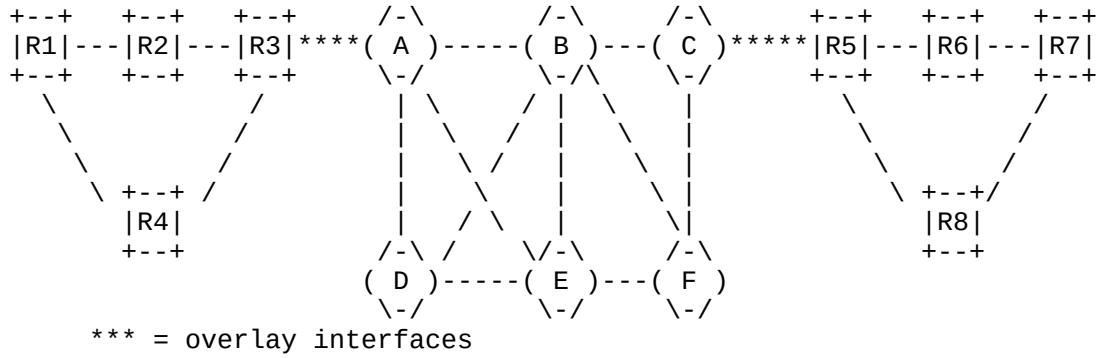


Figure 3: Single homing

Typical examples of single restorable LSP between A and C is the case of IP over WDM with single transponder on A and single transponder of C with restoration capability in the WDM domain. A common case of multiple LSPs between A and C, on the other side, it the splitting of the electical signal between a couple of transponders on A creating a 1+1 protection terminated on a couple of transponders of C.

3.2. Adjacent dual homing

The term adjacent dual homing is used to indicate two (or more) access links between the edge node and one or more core nodes. In this case we have an end to end tunnel with a single LSP in the client domain and one or more LSPs in the server domain with two or more overlay interface connecting them. The scenario is shown in figure below, where an end to end circuit between R1 and R7 is built over a tunnel between R3 and R5 composed by two LSPs between different pairs of ingress/egress nodes (A-C and D-F).

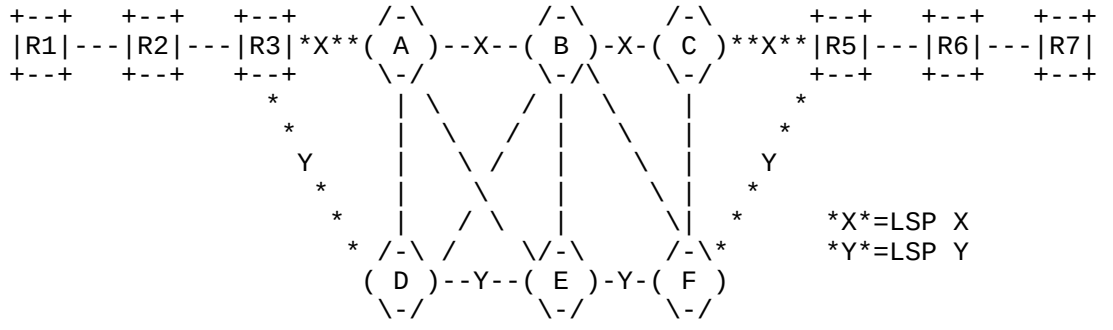


Figure 4: Adjacent dual homing

This network setup typically allows for fast client domain protection mechanisms, e.g., Fast ReRoute (FRR).

3.3. Remote dual homing

The remote dual homing scenario is based on an end to end tunnel with two (or more) LSPs in the client domain each of which relies on one (or more) LSPs in the server domain. This scenario is based on multiple independent single homing scenarios and is typically used to provide end to end diversity between two or more services. In figure below it is possible to see an end to end circuit between R1 and R7 composed by two services (A and B) which are built over two independent tunnels between R3 and R6 and between R5 and R9 respectively.

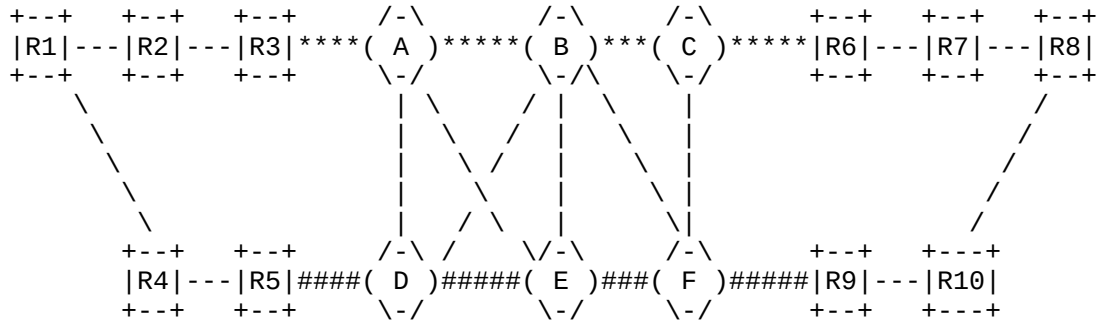


Figure 5: Remote dual homing

Typical usage of this network scenario consists on the combination of fast client domain protection mechaninsms (e.g.,1+1 protection) and server domain restoration mechanisms.

4. Use Cases

4.1. UC 1 - Provisioning

Requirement: The network operator must be able to setup an unprotected end to end service between two client domain nodes.

This use case simply consists on providing an operator with the capability of setting up a service in the client domain either by means or local trigger or remote signaling. The operator does not put any constraint over the path computation in the server domain.

4.2. UC 2 - Provisioning with optimization

Requirement: The network operator must be able to setup a service expressing which parameter must be optimized when computing the path.

This use case applies both to the local trigger and the remote signaling scenarios. In both cases the path computation function in the server domain (being it centralized or distributed) is demanded to provide a path between R3 and R5 which minimizes a given parameter (e.g. delay, jitter, TE metric).

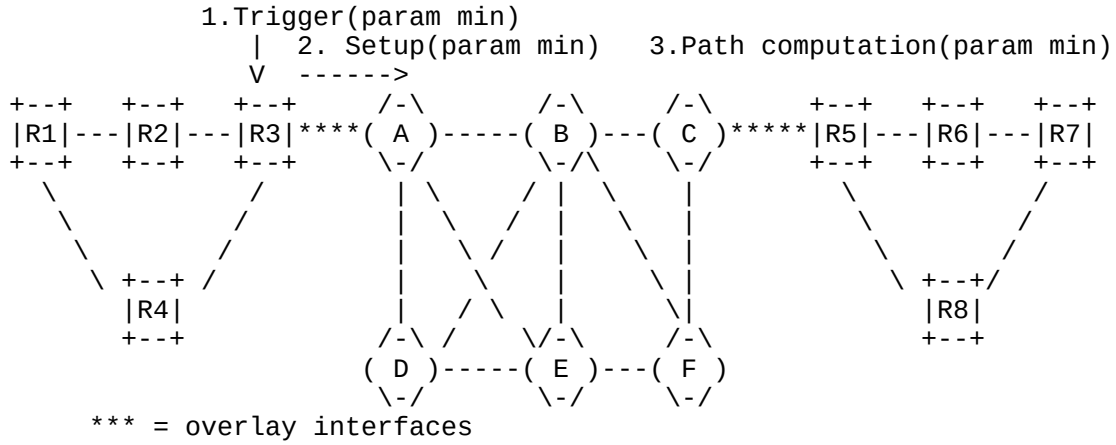


Figure 6: Provisioning with optimization

In the figure above the case of local trigger with specified parameter to be minimized is depicted, but same considerations apply to the remoe signaling (trigger on R1). In that case the parameter to be minized needs to be conveyed from R1 to R3 so that the setup request over the overlay interface can be issued taking into account the OF.

4.3. UC 3 - Provisioning with constraints

Requirement: The network operator must be able to setup a service imposing upper bounds for a set of parameters during the path computation.

This use cases is extremely similar to the provisioning with Optimization one. This time, instead of/in addition to giving the possibility of specifying which parameter needs to be optimized during the path computation, the network operator is also able to indicate and upper bound for a set of parameters which is not being minimized in the path computation.

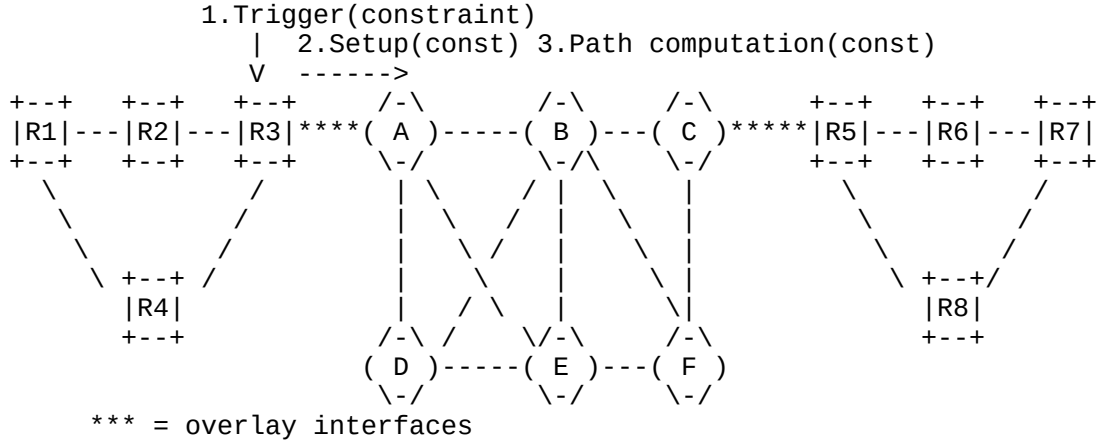


Figure 7: Provisioning with constraints

It is possible for example to ask for a path between R3 and R5 which, in addition to minimizing a given OF, does not introduce a delay higher than 10ms or where the jitter is not more than 3ms.

As per the optimization use case, when remote signaling is used (trigger on R1) a mean to convey the path computation constraints till the edge node (R3) is needed.

4.4. UC 4 - Diversity

Requirement: The network operator must be able to setup a services in the server domain in diversity with respect to server domain resources or not sharing the same fate with other server domain services. The network operator must also be able to decide whehther such diversity degree must be automatically kept by the network upon failures and optimization procedures.

This scenario is extremely common in those cases where different services in the server domain are used to provision protected services in the client domain. The services in the server domain can be computed/provisioned sequentially or in parallel but in both cases the requirement is to have them totally disjoint, so that a single failure in the server domain does not impact two or more services in the client domain which are supposed to be in a protection relationship between each other (e.g. 1+1 protection).

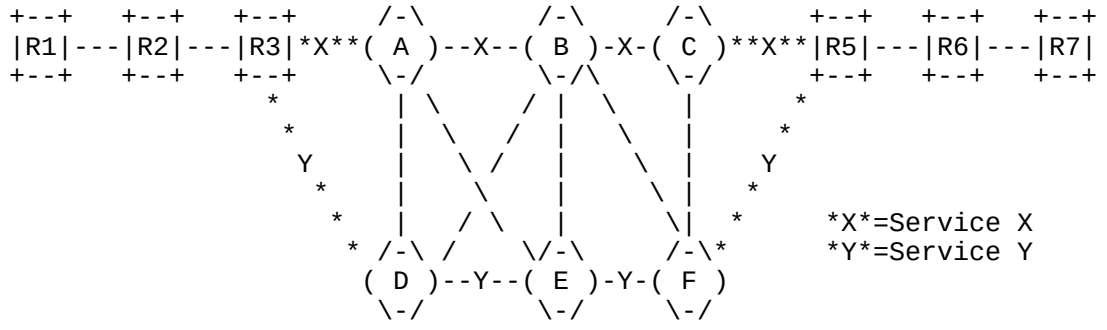


Figure 8: Diversity

In a scenario like the one depicted above, it is possible to use Service X and Service Y for the setup of a protected service in the client domain as a fault in the server domain would not impact both of them. In the case of parallel request, R3 asks the path computation in the server domain to provide two totally disjoint paths. On the other side, when sequential requests are issued, an identifier for Service X (or a set of identifiers indicating its resources) is needed so that the request for the setup of Service Y can be issued with the constraint of avoiding the resources related to such identifier.

Another case of provisioning with diversity is the one where the operator in the client domains wants the server domain PCE to exclude some resources from the path computation because of e.g. trustness

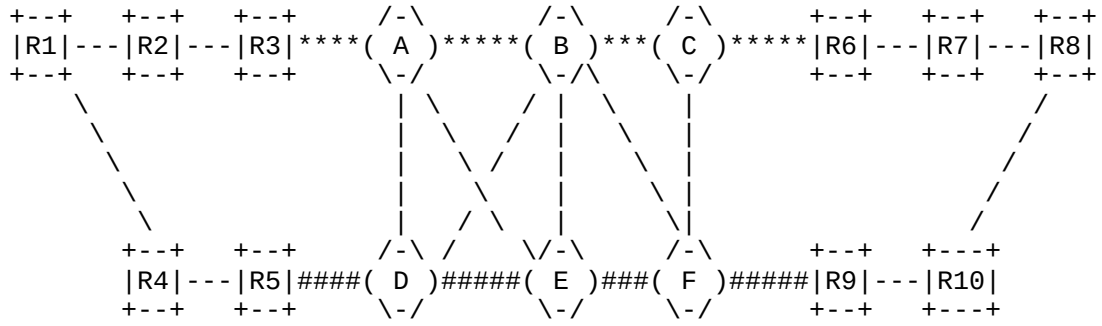
reasons. In such a case, supposing that such resources are known to the operator, it must be possible to indicate them as path computation constraint in the service setup request.

In addition to the provisioning of services with given diversity (and inclusion/exclusion) constraints, it must be possible to ask the server domain to at least keep such constraints also upon restoration or optimization procedures. It would be desirable to ask the server domain to relax constraints to be kept. The relaxation can be needed depending on resources availability, e.g., restoration of service X in partial diversity with service Y is total diverisity is not possible).

4.5. UC 5 - Concurrent provisioning

Requirement: The network operator must be able to setup a plurality of services not necessarily between the same pair of edge nodes.

Here is another case particularly interesting from a protection point of view. In the case above the same edge node was asking for different services in the server domain, but in order to have end to end diversity (i.e. from R1 to R8 in figure below), there is the need to be able to provide disjoint services between different pairs of edge nodes.



***=Service A
###=Service B

Figure 9: Concurrent provisioning

In this example Service A is provided between R3 and R6 and Service B between R5 and R9. Some sort of coordination is needed between R3 and R5 (directly between them or via R1) so that the requests to the server domain can be conveniently issued.

4.6. UC 6 - Reoptimization

Requirement: The network operator must be able to setup a plurality of services so that the overall cost of the network is minimized and not the cost of a single service.

TBD

4.7. UC 7 - Query

Requirement: The server network must be able to tell the network operator the actual parameters characterizing an existing service.

The capability of retrieving from the server domain some parameters qualifying a service can be estremely useful in different cases. One of them is the case o sequential provisioning with diversity requirements. In the case the operator wants to set-up a service in diversity from an existing one, hence it must be possible for the server domain to export some parameters univocally identifying the resources (e.g. SRLGs).

4.8. UC 8 - Availability check

Requirement: The network operator must be able to check if in the server domain there are enough resources to setup a service with given parameters.

TBD

4.9. UC 9 - P2MP services

Requirement: If allowed by the technology, the network operator must be able to setup a P2MP service with given parameters.

TBD

4.10. UC 10 - Privacy

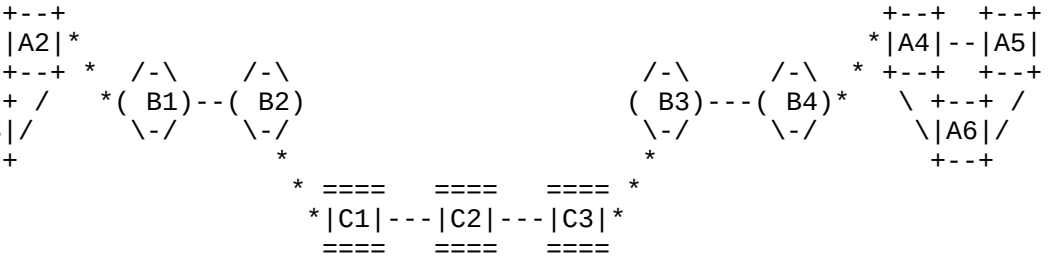
Requirement: The network operator must be able to provision different groups of users with independent addressing spaces.

This is a particularly useful functionality for those cases where the resources of the service provider are leased and shared among several other service providers or customers.

4.11. UC 12 - Stacking of overlay interfaces

Requirement: The network operator must be able manage a network with an arbitrarily high number of administrative boundaries (i.e.,>2).

Operators might want to split their overlay networks in a number of administrative domains for several reasons, among which simplifying network operations and improving scalability. In order to do so it must be possible to create a stack of overlay interfaces between the different domains as shown in figure below:



*** = overlay interfaces

Figure 10: Stacking of interfaces

Nodes "Ax" belong to a domain which is client to the domain composed by nodes "Bx". The domain composed by nodes Bx is hence server domain to the "Ax" nodes domain but client to the "Cx" nodes domain.

A pretty common deployment of this scenario consists of IP over OTN over WDM layers, where the OTN digital layer is used for the grooming of IP traffic over high bit rate lambdas. In figure 8, Node Bx can be assumed to be digital layer, which is interfacing with packet layer nodes (Ax) across overlay interface. Digital layer nodes Bx are interfacing with DWDM layer nodes Cx. If OTN (Bx) and DWDM (Cx) node belong to same IGP, then this becomes multi-layer path computation and signaling case, and it is out of scope of this document.

However, as already shown in the intro of this memo, the three different domains of the example could have the same switching capability (e.g., IP) and be kept separate just for administrative reasons.

4.12. UC 13 - Resiliency parameters

Requirement: The network operator must be able to request an LSP in the server domain with resilience parameters. The minimum set of such parameters includes 1+1 protection and restoration. Moreover, it must be possible for the operator to change the resilience level after the path is established in the network.

This functionality is interesting in a scenario like the one in Figure 6 with two concurrent paths. Let us assume service A and B are requested without any resilience requirements. If there is a failure in service A, the operator can request for protection in service B once this situation is detected.

These parameters can be used both in the case of single homing (UC1) and concurrent paths (UC6). The aim of this section is to highlight two sub-cases for every resilience case:

(1) during the provisioning the client domain can request to the server domain for resilience parameters.

(2) Once a failure occurs, the client domain has to be notified via the overlay interface thus carrying information about the situation in the server domain, so the client domain can take its own decisions.

For the different sub-use cases, the provisioning use case already highlights which is the workflow and the requirements for each scenario. This section does not include an example for each of them.

5. Security Considerations

TBD

6. IANA Considerations

TBD

7. Contributors

Diego Caviglia, Ericsson

Via E.Melen, 77 - Genova - Italy

Email: diego.caviglia@ericsson.com

Jeff Tantsura, Ericsson

300 Holger Way, San Jose, CA 95134 - USA

Email: jeff.tantsura@ericsson.com

Khuzema Pithewan, Infinera Corporation

140 Caspian CT., Sunnyvale - CA - USA

Email: kpithewan@infinera.com

Cyril Margaria, Wandl

Email: cyril.margaria@googlemail.com

John Drake, Juniper

Email: jdrake@juniper.net

Sergio Belotti, Alcatel-Lucent

Email: sergio.belotti@alcatel-lucent.com

Victor Lopez, Telefonica I+D

Email: vlopez@tid.es

Appendix A. Appendix I - Colored overlay

This use case applies to networks where the server domain is a WDM network. In those cases it is possible to either have a grey interface between client and server domains (i.e. transponder on the

border core node) or a colored interface between them (i.e. transponder on the edge node).

All the previous use cases assume the case of grey interface, but there are particular network scenarios in which it is possible to move the transponders from the core to the edge nodes and hence save on hardware cost.

The issue with this solution is that the PCE in the server domain, being either centralized or distributed, has only visibility of what is inside the server domain and hence has not all the info needed to perform the validation of a path. The edge node must provide the PCE in the server domain with a set of info needed for a correct path computation and path validation from transponder to transponder (i.e. between edge nodes) all along the server domain.

The type of information needed for this scenario can be classified into three categories:

- Feasibility: Parameters like the output power of the transponder are needed in order to state e.g. the amount of km that can be reached without regeneration.

- Compatibility: The egress transponder must be compatible with the ingress one. Parameters that influence the level of compatibility can be for example the type of FEC (Forward Error Correction) used or the modulation format (which also impacts the feasibility together with the bit rate).

- Availability: Transponders can be tunable within a range of lambdas or even locked to a single lambda. This impacts the path computation as not every path in the network might have such lambda(s) supported or available at the time the path computation is performed.

In figure below it is possible to see that the PCE is aware of all the info between A and C (i.e. within the server domain scope) but what is missing is info related to the transponders on R1 and on R2 and of the access links. (i.e. R1-A and C-R2).

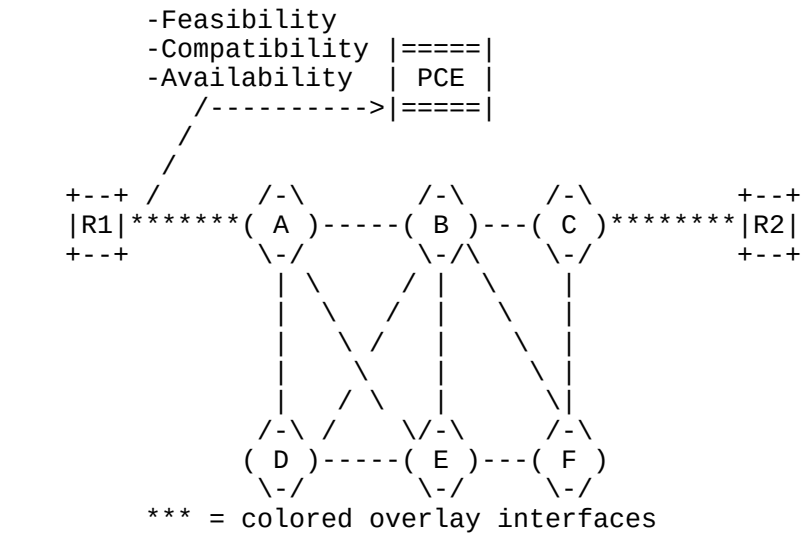


Figure 11: PCE feeding for colored UNI

There is not yet a standard set of parameters that is needed for path computation in WDM networks but an example of some of them is provided in the following list:

- Modulation format
- FEC (type or gain)
- Minimum transponder output power
- Bitrate
- Dispersion tolerance
- OSNR (minimum required)

8. References

8.1. Normative References

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

8.2. Informative References

Authors' Addresses

Daniele Ceccarelli
Ericsson
Via E. Melen 77
Genova - Erzelli
Italy

Email: daniele.ceccarelli@ericsson.com

Oscar Gonzalez de Dios
Telefonica I+D
Don Ramon de la Cruz 82-84
Madrid 28045
Spain

Email: ogondio@tid.es

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

Email: zhangfatai@huawei.com

Xian Zhang
Huawei Technologies
F3-5-B R&D Center, Huawei Base
Shenzhen 518129 P.R.China Bantian, Longgang District
Phone: +86-755-28972913

Email: zhang.xian@huawei.com

Internet Engineering Task Force
Internet-Draft
Intended status: Standards Track
Expires: August 12, 2014

D. Hiremagalur, Ed.
G. Grammel, Ed.
J. Drake, Ed.
Juniper
G. Galimberti, Ed.
Z. Ali, Ed.
Cisco
R. Kunze, Ed.
Deutsche Telekom
February 8, 2014

Extension to the Link Management Protocol (LMP/DWDM -rfc4209) for Dense Wavelength Division Multiplexing (DWDM) Optical Line Systems to manage application code of optical interface parameters in DWDM application draft-dharinigert-ccamp-g-698-2-lmp-06

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 extendsions./>

Status of This Memo

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

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

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

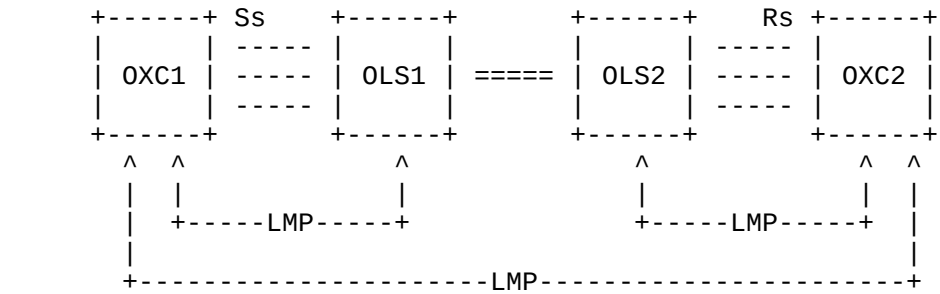
This Internet-Draft will expire on August 12, 2014.

Table of Contents

1. Introduction

This extension is based on "draft-galikunze-ccamp-g-698-2-snmp-mib-03" and "draft-kunze-g-698-2-management-control-framework-02", for the relevant interface optical parameters described in recommendations like ITU-T G.698.2 [ITU.G698.2]. The LMP Model from RFC4902 is extended to provide link property correlation between a client and an OLS device. By using LMP, the capabilities of either end of this link are exchanged 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 can agree on a common parameter window that can be supported and supervised by each device. The actual selection of a specific parameter value whithin the parameter window is outside the scope of LMP. In GMPLS the parameter selection (e.g. wavelength) is performed by RSVP-TE and Wavelength routing by IGP.

Figure 1 Extended LMP Model (from [RFC4209])



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

Figure 1: Extended LMP Model

2. Extensions to LMP-WDM Protocol

This document defines extensions to [RFC4209] to allow the Black Link (BL) parameters of G.698.2, as described in the draft draft-kunze-g-698-2-management-control-framework-02, 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]

- BL_General (sub-object Type = TBA)
- BL_ApplicationCode (sub-object Type = TBA)
- BL_VendorTransceiverClass (sub-object Type = TBA)
- BL_Ss (sub-object Type = TBA)
- BL_Rs (sub-object Type = TBA)

3. Black Link General Parameters - BL_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-04" for more details about these parameters and the [RFC6205] for the wavelength definition.

The general parameters are
1. Bit-Rate/line coding of optical tributary signals
2. Wavelength - (Tera Hertz) 4 bytes (see RFC6205 sec.3.2)
3. Number of Application Codes Supported
4. Number of Vendor Transceiver Classes Supported
5. Identifier of Application code to/in use
6. Identifier Vendor transceiver Application code to/in use

Figure 2: The format of the this sub-object (Type = TBA, Length = TBA) is as follows:

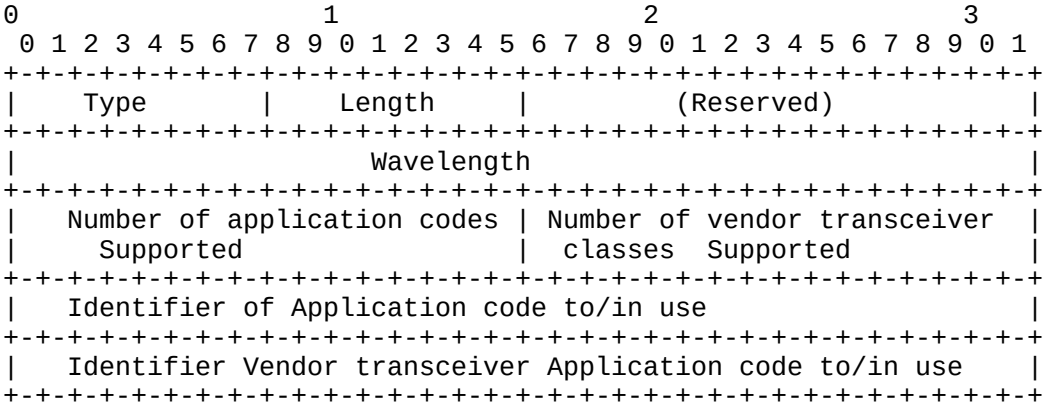


Figure 2: BL_General

4. Black Link ApplicationCode - BL_ApplicationCode

This message is to exchange the application code supported as described in [G698.2]. Please refer to the "draft-galikunze-ccamp-g-698-2-snmp-mib-04". for more details about these parameters. There can be more than one Application Code supported by the OXC/OLS. The number of application codes supported is exchanged in the "BL_General" message. (from [G698.1]/[G698.2]/[G959.1]

The parameters are

1. Single-channel application code identifire - 8 bits

2. Single-channel application codes -- 32 bytes (from [G698.1]/[G698.2]/[G959.1] - this parameter can have multiple instances as the transceiver can support multiple application codes.

Figure 3: The format of the this sub-object (Type = TBA, Length = TBA) is as follows:

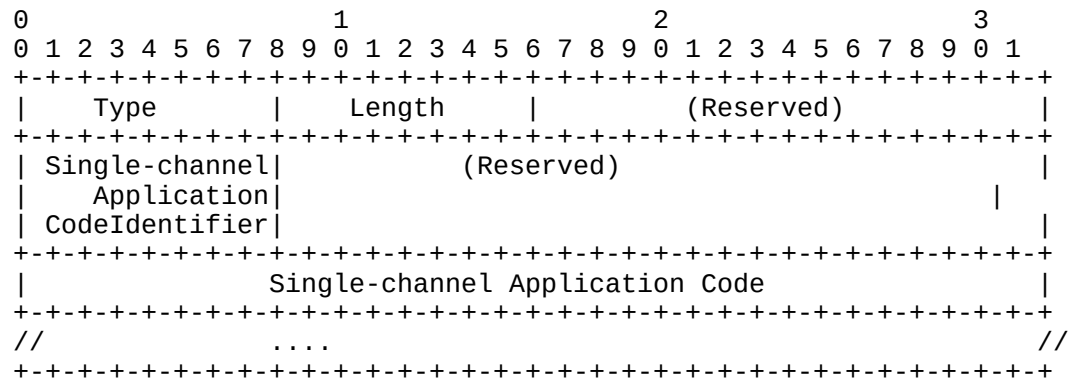


Figure 3: BL_ApplicationCode

5. Black Link Vendor Transceiver Class - BL_ApplicationCode

This message is to exchange the application code supported as described in [G698.2]. Please refer to the "draft-galikunze-ccamp-g-698-2-snmp-mib-04". for more details about these parameters. There can be more than one Vendor Transceiver Class supported by the OXC/OLS. The number of Vendor Transceiver Classes supported is exchanged in the "BL_General" message. (from [G698.1]/[G698.2]/[G959.1]

The parameters are

1. Single-channel Transceiver Class identifier - 8 bits

2. Vendor Transceiver Class -- 32 bytes
   (from [G698.1]/[G698.2]/[G959.1] - this parameter can have multiple instances as the transceiver can support multiple application codes.

Figure 4: The format of the this sub-object (Type = TBA, Length = TBA) is as follows:

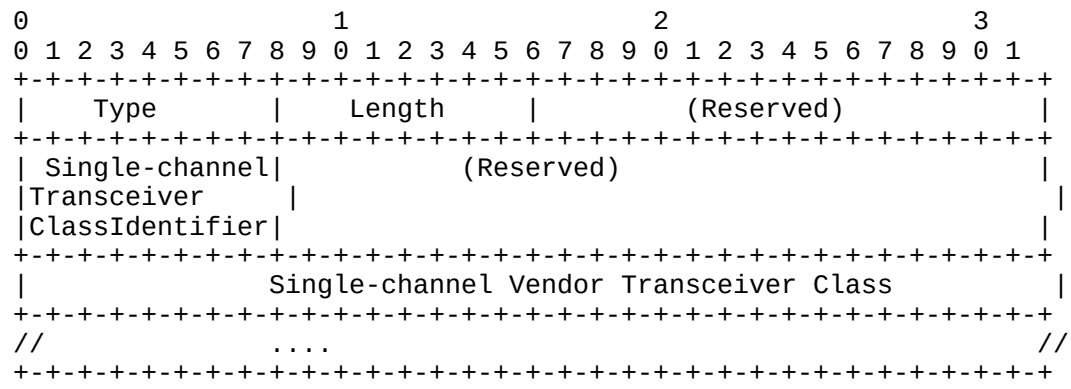


Figure 4: BL_VendorTransceiverClass

6. Black Link - BL_Ss

These are the G.698.2 parameters at the Source(Ss reference points). Please refer to "draft-galikunze-ccamp-g-698-2-snmp-mib-03" for more details about these parameters.

1. Output power

Figure 5: The format of the Black link sub-object (Type = TBA, Length = TBA) is as follows:

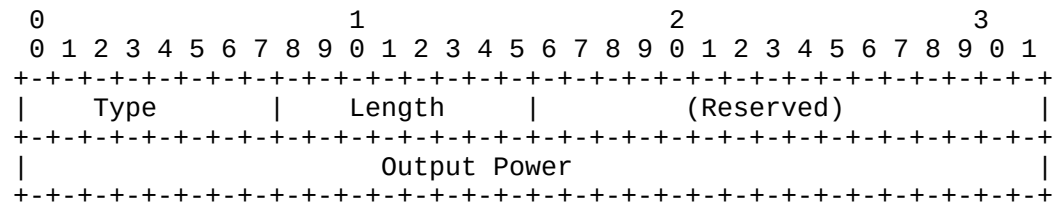


Figure 5: Black Link - BL_Ss

7. Black Link - BL_Rs

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-02" for more details about these parameters.

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

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

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

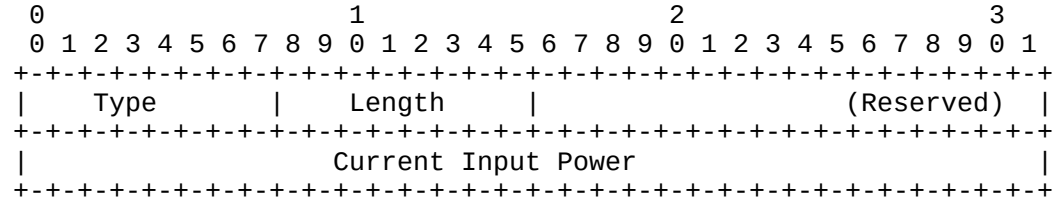


Figure 6: Black Link - BL_Rs

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"/>)

- BL_General (sub-object Type = TBA)
- BL_ApplicationCode (sub-object Type = TBA)
- BL_VendorTransceiverClass (sub-object Type = TBA)
- BL_Ss (sub-object Type = TBA)
- BL_Rs (sub-object Type = TBA)

10. References

10.1. 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, March 2003.

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

10.2. Informative References

[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.

[I-D.galimbe-kunze-g-698-2-snmp-mib]
Kunze, R. and D. Hiremagalur, "A SNMP MIB to manage black-link optical interface parameters of DWDM applications", draft-galimbe-kunze-g-698-2-snmp-mib-02 (work in progress), March 2012.

Authors' Addresses

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

Phone: +1408
Email: dharinih@juniper.net

Gert Grammel (editor)
Juniper
1194 N Mathilda Avenue
Sunnyvale - 94089 California
USA

Phone: +1408
Email: ggrammel@juniper.net

John E. Drake (editor)
Juniper
1194 N Mathilda Avenue
HW-US,Pennsylvania
USA

Phone: +1408
Email: jdrake@juniper.net

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

Network Working Group
Internet-Draft
Intended status: Informational
Expires: August 18, 2014

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

February 14, 2014

Terminology and Models for Control of Traffic Engineered Networks with Provider-Customer Relationship
draft-dios-ccamp-control-models-customer-provider-00

Abstract

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

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

Table of Contents

1. Introduction

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

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

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

1.1. Examples of Customer-Provider TE Network Domain Scenarios

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

A particular network scenario that has attracted lot of attention from the industry is the IP/MPLS/OTN/WDM over WDM. The customer network is based on multi-layer routers able to set up packet-based TE connections over wavelengths. The provider network is a WDM

optical network that provides the switching for the wavelenghts as well as restoration capabilities of the optical channels.

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

Summing up, there number of applicable scenarios is wide.

2. Terminology

2.1. Customer Domain - Provider Domain Interface

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

2.1.1. UNI in IP over Optical Networks

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

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

2.1.2. ITU-T Definition of UNI

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

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

2.1.3. OIF Definition of UNI

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

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

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

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

2.1.4. Proposed Vocabulary

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

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

2.1.4.1. Customer network

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

2.1.4.2. Provider network

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

2.1.4.3. Customer-Provider Control Plane Interface

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

2.2. Reachability

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

The document [draft-farrel-interconnected-te-info-exchange-02] provides the definition of what is reachability for client-server networks. [EDITOR's note: Text from draft-farrel-interconnected-te-info-exchange has been borrowed for this first version. Duplicated text will be deleted at later stages]

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

In the context of Traffic Engineered networks with customer and provider relationships, we can define several types of reachabiity: [draft-farrel-interconnected-te-info-exchange-02]

2.2.1. Unqualified Reachability

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

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

2.2.2. Qualified Reachability

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

2.2.3. Qualified Reachability with associated potential TE path

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

3. Control Models

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

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

3.1. Signaling Only

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

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

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

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

3.1.1. Signaling with Requirements

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

3.1.2. Signaling with Collection

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

3.2. Signaling and Reachability Model

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

3.2.1. Signalling + Basic Reachability

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

3.2.2. Signalling + Qualified Reachability

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

3.2.3. Signalling + Qualified Reachability + Potential Services

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

3.3. Other Models

3.3.1. Multi-Layer Networks / Multi-Region Networks

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

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

3.3.2. Management Model

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

4. Security Considerations

TBD

5. Contributing Authors

6. Acknowledgments

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

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.

[RFC3107] Rekhter, Y. and E. Rosen, "Carrying Label Information in BGP-4", RFC 3107, May 2001.

[RFC3717] Rajagopalan, B., Luciani, J., and D. Awduche, "IP over Optical Networks: A Framework", RFC 3717, March 2004.

[RFC4208] Swallow, G., Drake, J., Ishimatsu, H., and Y. Rekhter, "Generalized Multiprotocol Label Switching (GMPLS) User-Network Interface (UNI): Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Support for the Overlay Model", RFC 4208, October 2005.

7.2. Informative References

[draft-farrel-interconnected-te-info-exchange-02] "Farrel, A., Drake, J., Bitar, N., Swallow, G., Ceccarelli, D. draft-farrel-interconnected-te-info-exchange-02 Problem Statement and Architecture for Information Exchange Between Interconnected Traffic Engineered Networks", 2014.

Authors' Addresses

Oscar Gonzalez de Dios (editor)
Telefonica GCTO
Don Ramon de la Cruz 82-84
Madrid 28045
Spain

Phone: +34913128832
Email: ogondio@tid.es

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

Email: julien.meuric@orange.com

Daniele Ceccarelli
Ericsson
Via Calda 5
Genova
Italy

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

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: August 14, 2014

A. Farrel
J. Drake
Juniper Networks

N. Bitar
Verizon Networks

G. Swallow
Cisco Systems, Inc.

D. Ceccarelli
Ericsson

February 14, 2014

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

draft-farrel-interconnected-te-info-exchange-03.txt

Abstract

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

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

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

Status of This Memo

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

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

working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

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

Table of Contents

1. Introduction

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

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

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

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

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

1.1. What is TE Reachability?

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

TE reachability may be unqualified (there is a TE path, but no information about available resources or other constraints is supplied) which is helpful especially in determining a path to a destination that lies in an unknown domain, or may be qualified by TE attributes such as TE metrics, hop count, available bandwidth, delay, shared risk, etc.

2. Overview of Use Cases

2.1. Peer Networks

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

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

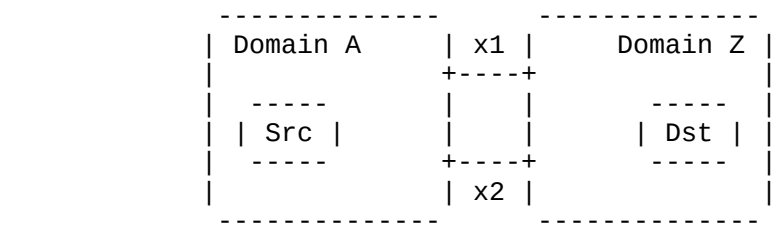


Figure 1 : Peer Networks

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

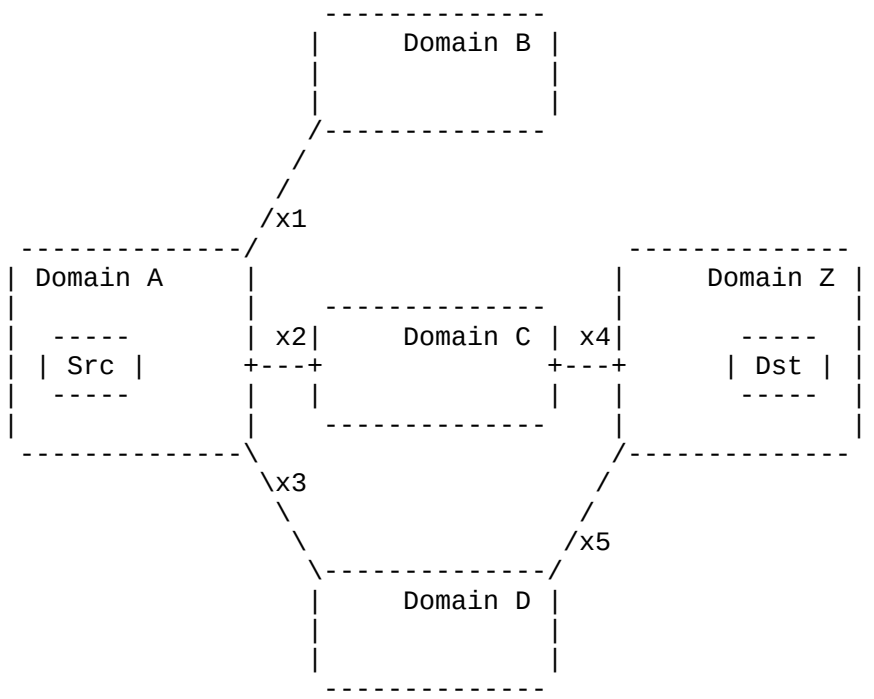


Figure 2 : Peer Networks in a Mesh

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

2.1.1. Where is the Destination?

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

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

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

2.2. Client-Server Networks

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

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

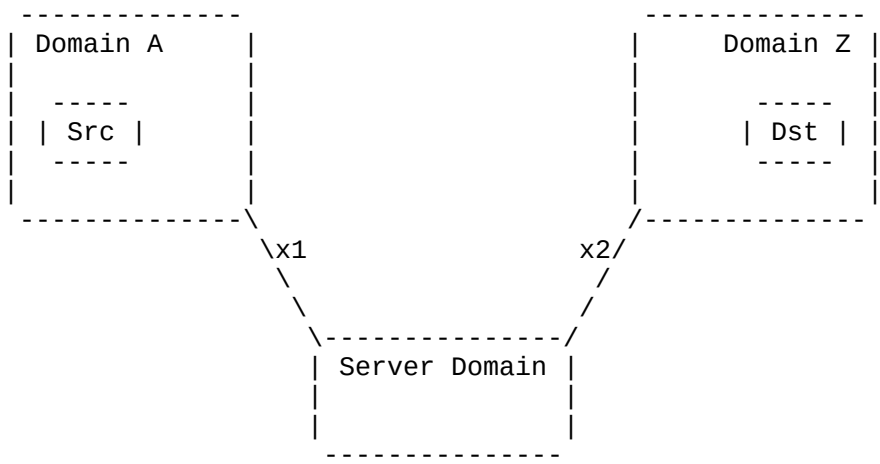


Figure 3 : Client-Server Networks

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

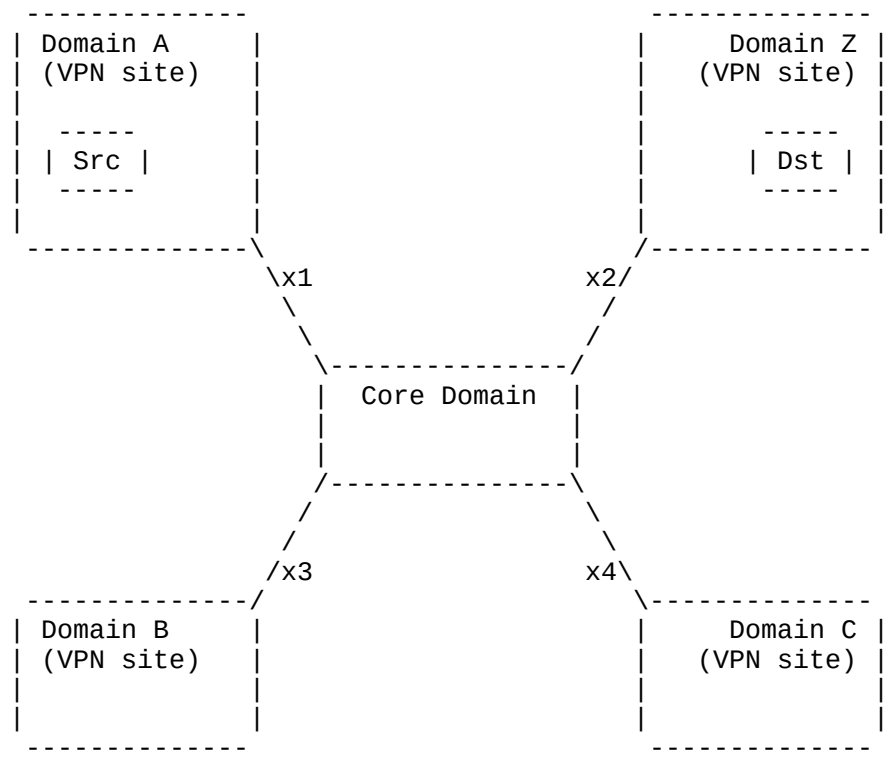


Figure 4 : A Virtual Private Network

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

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

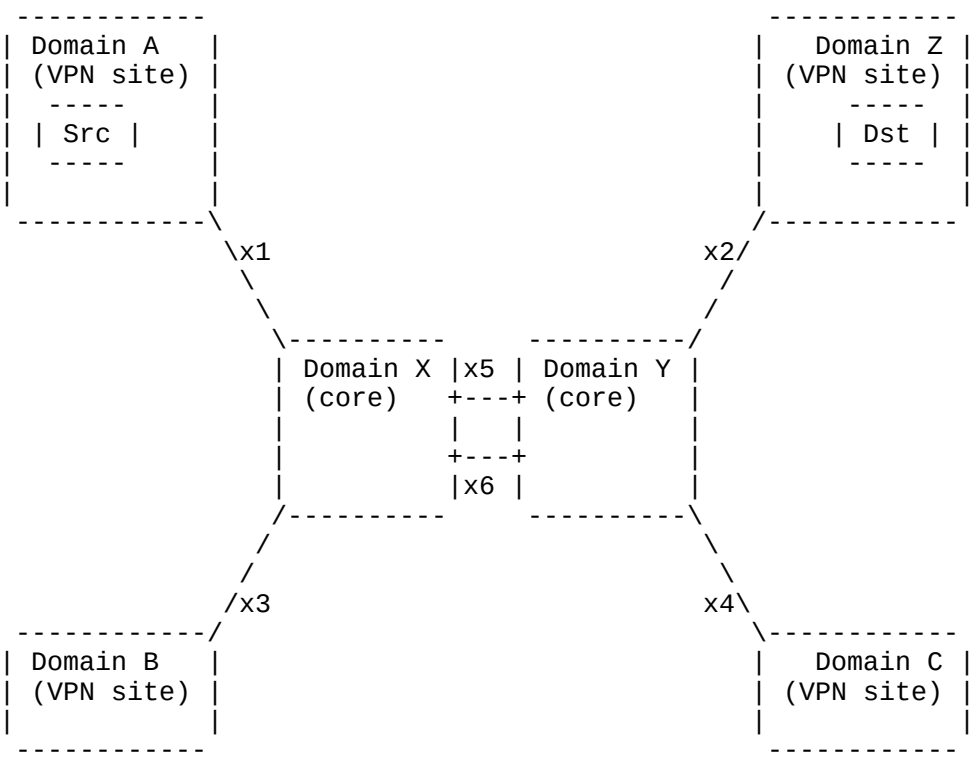


Figure 5 : A VPN Supported Over Multiple Server Domains

2.3. Dual-Homing

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

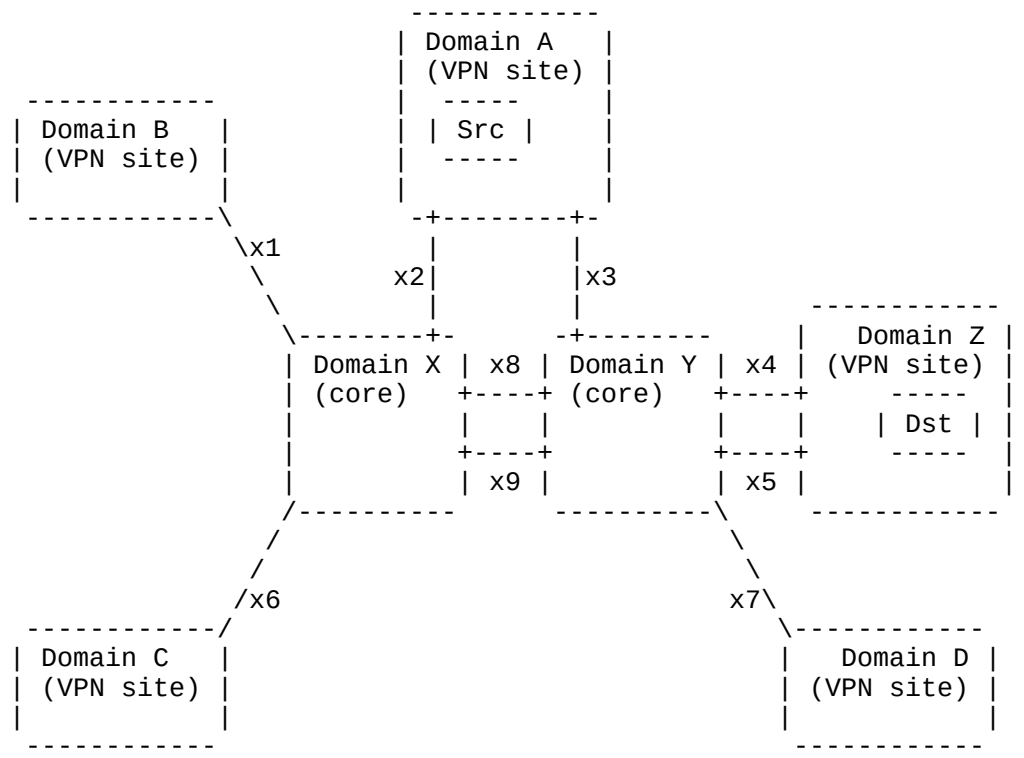


Figure 6 : Dual-Homing in a Virtual Private Network

3. Problem Statement

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

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

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

Thus, the problem is one of information collection and presentation, not about signaling. Indeed, the existing signaling mechanisms for

TE LSP establishment are likely to prove adequate [RFC4726] with the possibility of minor extensions.

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

3.1. Use of Existing Protocol Mechanisms

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

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

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

3.2. Policy and Filters

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

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

3.3. Confidentiality

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

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

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

3.4. Information Overload

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

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

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

3.5. Issues of Information Churn

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

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

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

In non-packet networks, where link resources are physical switching resources (such as timeslots or wavelengths) the capacity of an LSP may more frequently be a significant percentage of the available link resources. Furthermore, in some switching environments, it is necessary to achieve end-to-end resource continuity (such as using the same wavelength on the whole length of an LSP), so it is far more desirable to keep the TE information held at the computation points up-to-date. Fortunately, non-packet networks tend to be quite a bit smaller than packet networks, the arrival rates of non-packet LSPs are much lower, and the hold times considerably longer. Thus the information churn may be sustainable.

3.6. Issues of Aggregation

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

- Virtual node model. In this view, the domain is aggregated as if it was a single node (or router / switch). Its links to other domains are presented as real TE links, but the model assumes that any LSP entering the virtual node through a link can be routed to leave the virtual node through any other link.

- Virtual link model. In this model, the domain is reduced to a set of edge-to-edge TE links. Thus, when computing a path for an LSP that crosses the domain, a computation point can see which domain entry points can be connected to which other and with what TE attributes.

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

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

It should also be noted in the context of Section 3.5 that changes in the information within a domain may have a bearing on what aggregated data is shared with another domain. Thus, while the data shared in reduced, the aggregation algorithm (operating on the routers responsible for sharing information) may be heavily exercised.

3.7. Virtual Network Topology

The terms "virtual topology" and "virtual network topology" have become overloaded in a relatively short time. We draw on [RFC5212] and [RFC5623] for inspiration to provide a definition for use in this document. Our definition is based on the fact that a topology at the

client network layer is constructed of nodes and links. Typically, the nodes are routers in the client layer, and the links are data links. However, a layered network provides connectivity through the lower layer as LSPs, and these LSPs can provide links in the client layer. Furthermore, those LSPs may have been established in advance, or might be LSPs that could be set up if required. This leads to the definition:

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

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

- In the server layer, LSPs need to be set up either in advance in response to management instructions or in answer to dynamic requests subject to policy considerations.

- In the server layer, evaluation of available TE resources can lead to the announcement of potential connectivity (i.e., LSPs that could be set up on demand).

- In the client layer, connectivity (lower layer LSPs or potential LSPs) needs to be announced in the IGP as a normal TE link. Such links may or may not be made available to IP routing: but, they are never made available to IP until fully instantiated.

- In the client layer, requests to establish lower layer LSPs need to be made either when links supported by potential LSPs are about to be used (i.e., when a higher layer LSP is signalled to cross the link, the setup of the lower layer LSP is triggered), or when the client layer determines it needs more connectivity or capacity.

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

These activities require an architecture and protocol elements as

well as management components and policy elements.

4. Existing Work

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

4.1. Per-Domain Path Computation

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

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

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

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

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

4.2. Crankback

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

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

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

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

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

4.3. Path Computation Element

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

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

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

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

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

- Hierarchical PCE (H-PCE) [RFC6805] offers a parent PCE that is responsible for navigating a path across the domain mesh and for coordinating intra-domain computations by the child PCEs responsible for each PCE. This approach makes computing an end-to-end path across a mesh of domains far more tractable. However, it still leaves unanswered the issue of determining the location of the destination (i.e., discovering the destination domain) as described in Section 2.1.1. Furthermore, it raises the question of who operates the parent PCE especially in networks where the domains are under different administrative and commercial control.

Further issues and considerations of the use of PCE can be found in

[I-D.farrkingel-pce-questions].

4.4. GMPLS UNI and Overlay Networks

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

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

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

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

Nevertheless, in this UNI/overlay model, the edge node has limited information of precisely what LSPs could be set up across the core, and what TE services (such as diverse routes for end-to-end protection, end-to-end bandwidth, etc.) can be supported.

4.5. Layer One VPN

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

As in the UNI case, the customer edge has some control over the establishment and type of the connectivity. In the L1VPN context

three different service models have been defined classified by the semantics of information exchanged over the customer interface: Management Based, Signaling Based (a.k.a. basic), and Signaling and Routing service model (a.k.a. enhanced).

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

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

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

4.6. VNT Manager and Link Advertisement

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

The VNT Manager would, in fact, form part of the provisioning path for all server network LSPs whether they are set up ahead of client network demand or triggered by end-to-end client network LSP signaling.

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

it will be necessary to indicate which client or clients can use the link. This capability it provided in [RFC6107].

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

4.7. What Else is Needed and Why?

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

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

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

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

5. Architectural Concepts

5.1. Basic Components

This section revisits the use cases from Section 2 to present the basic architectural components that provide connectivity in the peer and client-server cases. These component models can then be used in later sections to enable discussion of a solution architecture.

5.1.1. Peer Interconnection

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

to-Network Interfaces - ENNIs) are A2-B1, B3-C1, and C3-D1.

The objective is to be able to support an end-to-end connection A1-to-D2. This connection is for TE connectivity.

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

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

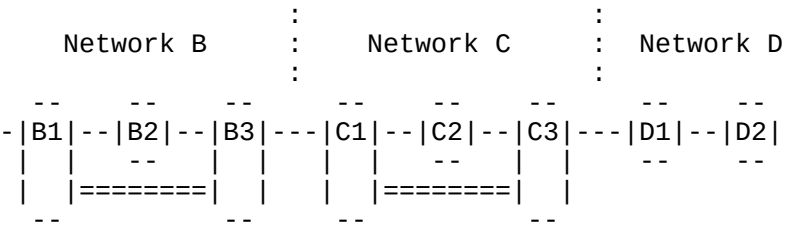


Figure 7 : Architecture for Peering

5.1.2. Client-Server Interconnection

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

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

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

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

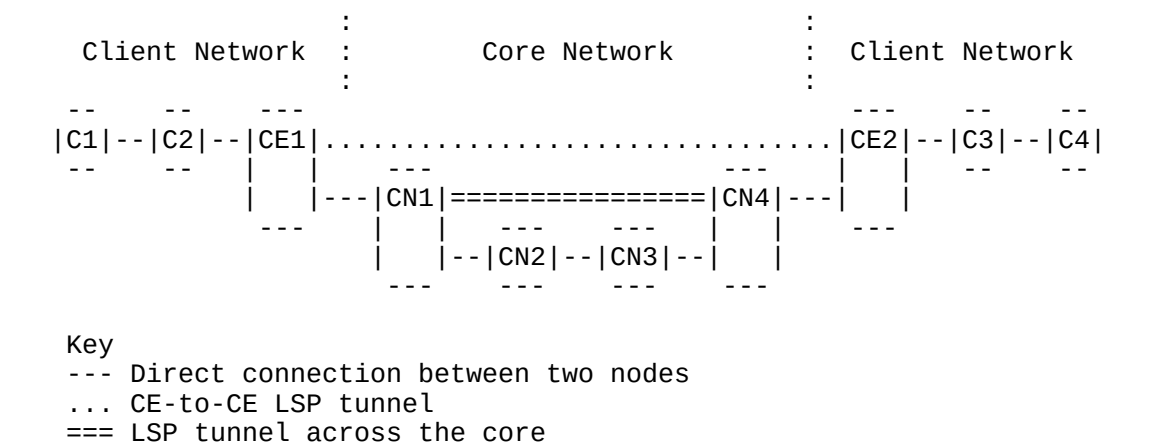


Figure 8 : Architecture for Client-Server Network

5.2. TE Reachability

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

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

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

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

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

5.3. Abstraction not Aggregation

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

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

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

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

While abstraction uses available TE information, it is subject to policy and management choices. Thus, not all potential connectivity will be advertised to each client. The filters may depend on commercial relationships, the risk of disclosing confidential information, and concerns about what use is made of the connectivity that is offered.

5.3.1. Abstract Links

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

When looking at a network such as that in Figure 8, the link from CN1

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

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

5.3.2. The Abstraction Layer Network

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

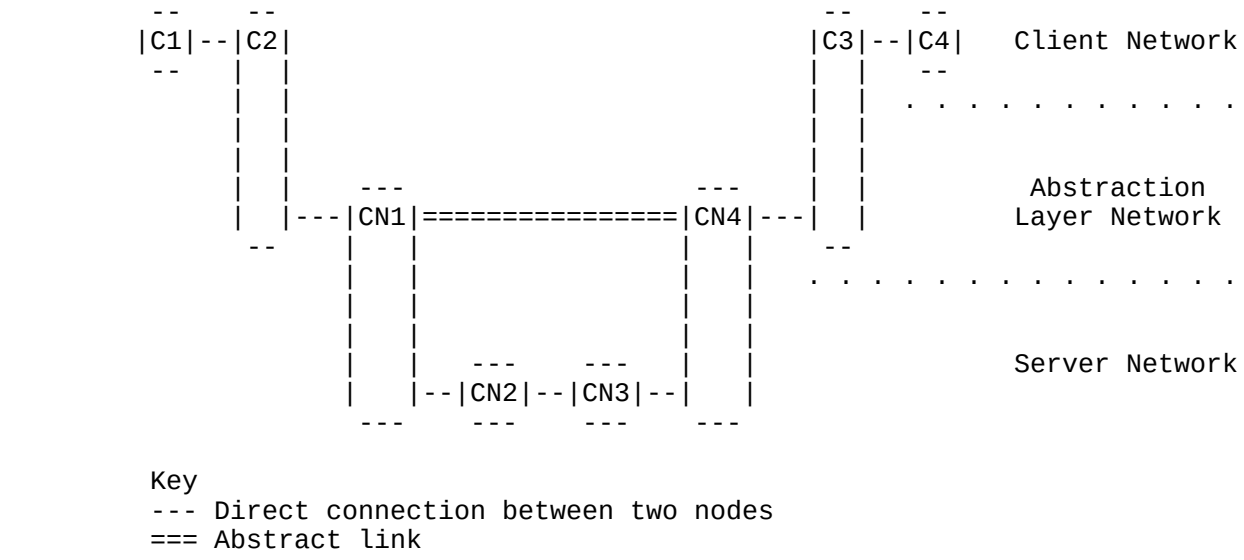


Figure 9 : Architecture for Abstraction Layer Network

The client layer network is able to operate as normal. Connectivity across the network can either be found or not found based on links

that appear in the client layer TED. If connectivity cannot be found, end-to-end LSPs cannot be set up. This failure may be reported but no dynamic action is taken by the client layer.

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

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

When the client layer needs additional connectivity it can make a request to the Abstraction Layer Network. For example, the operator of the client network may want to create a link from C2 to C3. The Abstraction Layer can see the potential path C2-CN1-CN4-C3, and asks the server layer to realise the abstract link CN1-CN4. The server layer provisions the LSP CN1-CN2-CN3-CN4 and makes the LSP available as a hierarchical LSP to turn the abstract link into a link that can be used in the client network. The Abstraction Layer can then set up an LSP C2-CN1-CN4-C3 using stitching or tunneling, and make the LSP available as a virtual link in the client network.

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

5.3.2.1. Nodes in the Abstraction Layer Network

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

Thus, as shown in Figure 10, a transit node in the Server Network (here the node is CN3) can be exposed as a node in the Abstraction Layer Network with Abstract Links connecting it to other nodes in the Abstraction Layer Network. Of course, in the network shown in Figure 10, there is little if any value in exposing CN3, but if it

had other Abstract Links to other nodes in the Abstraction Layer Network and/or direct connections to Client Network nodes, then the resulting network would be richer.

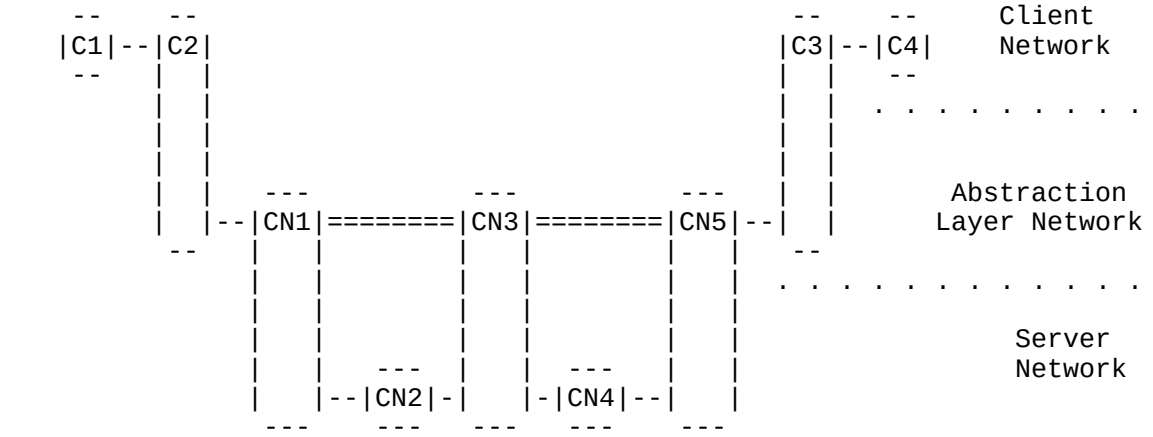


Figure 10 : Abstraction Layer Network with Additional Node

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

- Coordination between all of the external interfaces of the Abstract Node

- Management of changes in the Server Network that lead to limited capabilities to reach (cross-connect) across the Abstract Node.

5.3.3. Abstraction in Client-Server Networks

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

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

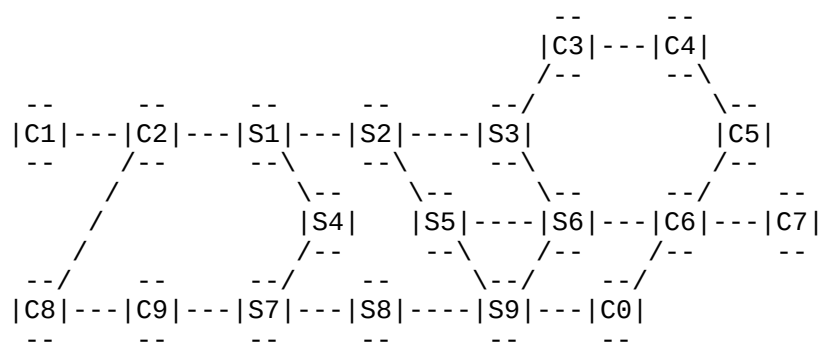


Figure 11 : An example Multi-Layer Network

If the network in Figure 11 is operated as separate client and server networks then the client layer topology will appear as shown in Figure 12. As can be clearly seen, the network is partitioned and there is no way to set up an LSP from a node on the lefthand side (say C1) to a node on the righthand side (say C7).

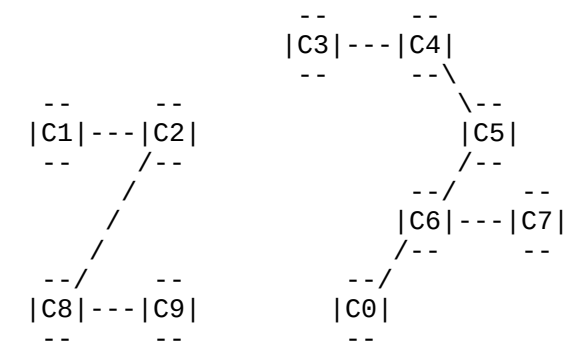


Figure 12 : Client Layer Topology Showing Partitioned Network

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

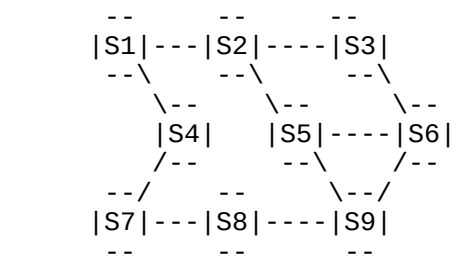


Figure 13 : Server Layer Topology

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

In our example, after consideration of which LSPs could be set up in the server layer, four abstract links are offered: S1-S3, S3-S6, S1-S9, and S7-S9. These abstract links are shown as double lines on the resulting topology of the Abstract Layer Network in Figure 14.

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

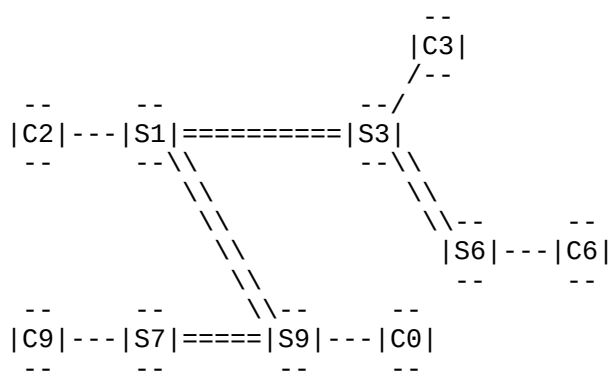


Figure 14 : Abstraction Layer Network with Abstract Links

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

results in several actions:

1. The management component for the Abstraction Layer Network asks its PCE to compute the paths necessary to make the connections. This yields C2-S1-S3-C3 and C2-S1-S9-C0.

2. The management component for the Abstraction Layer Network instructs C2 to start the signaling process for the new LSPs in the Abstraction Layer.

3. C2 signals the LSPs for setup using the explicit routes C2-S1-S3-C3 and C2-S1-S9-C0.

4. When the signaling messages reach S1 (in our example, both LSPs traverse S1) the Abstraction Layer Network may find that the necessary underlying LSPs (S1-S2-S3 and S1-S2-S5-S9) have not been established since it is not a requirement that an abstract link be backed up by a real LSP. In this case, S1 computes the paths of the underlying LSPs and signals them.

5. Once the serve layer LSPs have been established, S1 can continue to signal the Abstraction Layer LSPs either using the server layer LSPs as tunnels or as stitching segments.

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

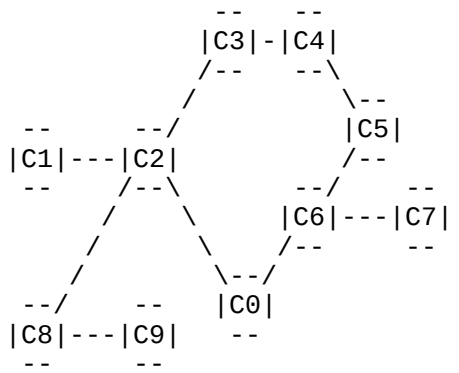


Figure 15 : Connected Client Layer Network with Additional Links

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

5.3.3.1 Macro Shared Risk Link Groups

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

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

For example, looking at the example used in Section 5.3.3 and considering the two abstract links S1-S3 and S1-S9 the is no way for the client layer to know whether the links C2-C0 and C2-C3 share fate. Clearly, if the client layer uses these links to provide a link-diverse end-to-end protection scheme, it needs to know that the links actually share a piece of network infrastructure (the server layer link S1-S2).

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

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

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

5.3.3.2 A Server with Multiple Clients

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

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

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

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

5.3.3.3 A Client with Multiple Servers

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

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

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

5.3.4. Abstraction in Peer Networks

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

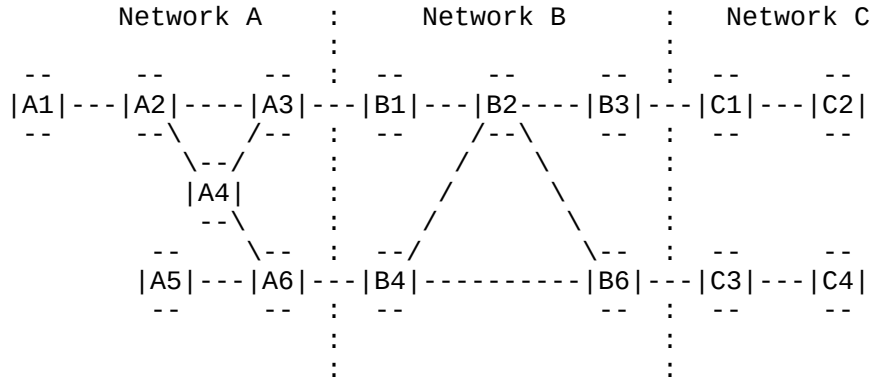


Figure 16 : A Network Comprising Three Peer Networks

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

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

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

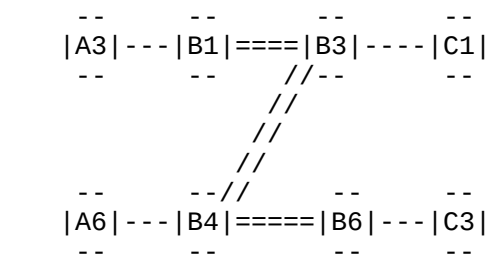


Figure 17 : Abstraction Layer Network for the Peer Network Example

Using a process similar to that described in Section 5.3.3, Network A can request connectivity to Network C and the abstract links can be instantiated as tunnels across the transit network, and edge-to-edge

LSPs can be set up to join the two networks. Furthermore, if Network C is partitioned, reachability information can be exchanged to allow Network A to select the correct edge-to-edge LSP as shown in Figure 18.

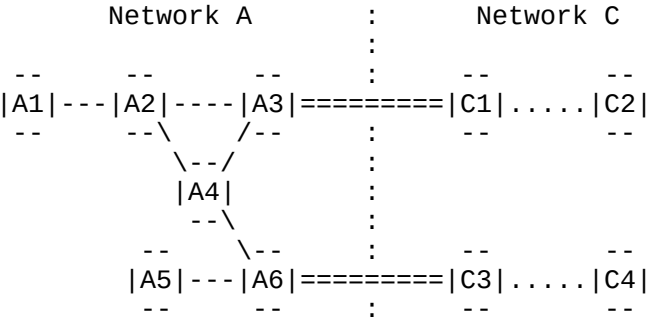


Figure 18 : Tunnel Connections to Network C with TE Reachability

Peer networking cases can be made far more complex by dual homing between network peering nodes (for example, A3 might connect to B1 and B4 in Figure 17) and by the networks themselves being arrange in a mesh (for example, A6 might connect to B4 and C1 in Figure 17). These additional complexities can be handled graceully by the Abstraction Layer Network model.

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

5.4. Considerations for Dynamic Abstraction

<TBD>

5.5. Requirements for Advertising Links and Nodes

The Abstraction Layer Network is "just another network layer". The links and nodes in the network need to be advertised so that the topology is disseminated and so that routing decisions can be made.

This requires a routing protocol running between the nodes in the Abstraction Layer Network. Note that this routing information exchange could be piggy-backed on an existing routing protocol instance, or use a new instance (or even a new protocol).

It should be noted that in some cases Abstract Link enablement is on-demand and all that is advertised in the topology for the Abstraction Layer Network is the potential for an Abstract Link to be set up. In

this case we may ponder how the routing protocol will advertise topology information over a link that is not yet established. The answer is that control plane connectivity exists in the Server Network and on the client-server edge links, and this can be used to carry the routing protocol messages for the Abstraction Layer Network. The same consideration applies to the advertisement, in the Client Network of the potential connectivity that the Abstraction Layer Network can provide.

5.6. Addressing Considerations

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

6. Building on Existing Protocols

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

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

6.1. BGP-LS

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

6.2. IGPs

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

6.3. RSVP-TE

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

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

7. Applicability to Optical Domains and Networks

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

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

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

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

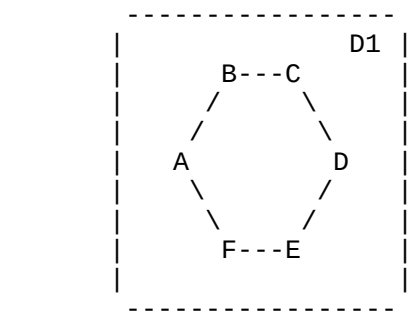


Figure 19 : A Simple Optical Domain

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

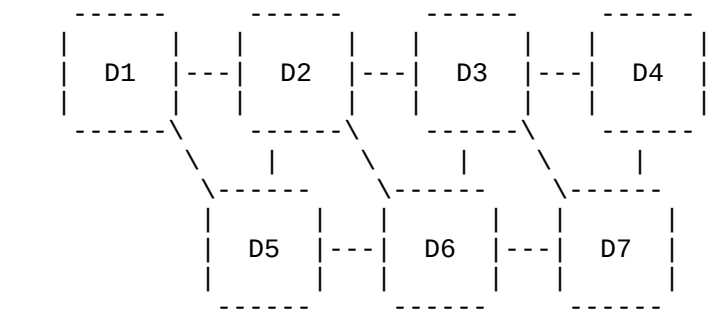


Figure 20 : A Simple Optical Domain

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

o In the first model (the Area-Based model) each domain is treated as a separate OSPF area. The end-to-end path will be specified to traverse multiple areas, and each area will be left to determine

the path across the nodes in the area. The feasibility of an end-to-end path (and, thus, the selection of the sequence of areas and their interconnections) can be derived using hierarchical PCE.

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

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

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

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

  Then, as suggested in the Area-Based model, hierarchical PCE could be used to determine the feasibility of an end-to-end path and to derive the sequence of domains and the points of interconnection to use. But, just as in that other model, the scalability of the hierarchical PCE approach must be questioned.

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

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

  This approach handles the awkwardness of representing the domains as areas or ASes by simply considering them as domains running

distinct instances of OSPF. Routing advertisements flow "upward" from the domains to the high level OSPF instance giving it a full view of the whole network and allowing end-to-end paths to be computed. Routing advertisements may also flow "downward" from the network-wide OSPF instance to any one domain so that it has visibility of the connectivity of the whole network.

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

Practically, this third model turns out to be very close to the methodology described in this document. As noted in Section 7.1 of [RFC6827], there are policy rules that can be applied to define exactly what information is exported from or imported to a low level OSPF instance. The document even notes that some forms of aggregation may be appropriate. Thus, we can apply the following simplifications to the mechanisms defined in RFC 6827:

- Zero information is imported to low level domains.

- Low level domains export only abstracted links as defined in this document and according to local abstraction policy and with appropriate removal of vendor-specific information.

- There is no need to formally define routing levels within OSPF.

- Export of abstracted links from the domains to the network-wide routing instance (the abstraction routing layer) can take place through any mechanism including BGP-LS or direct interaction between OSPF implementations.

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

links, the domain border nodes, and the abstracted (cross-domain) links.

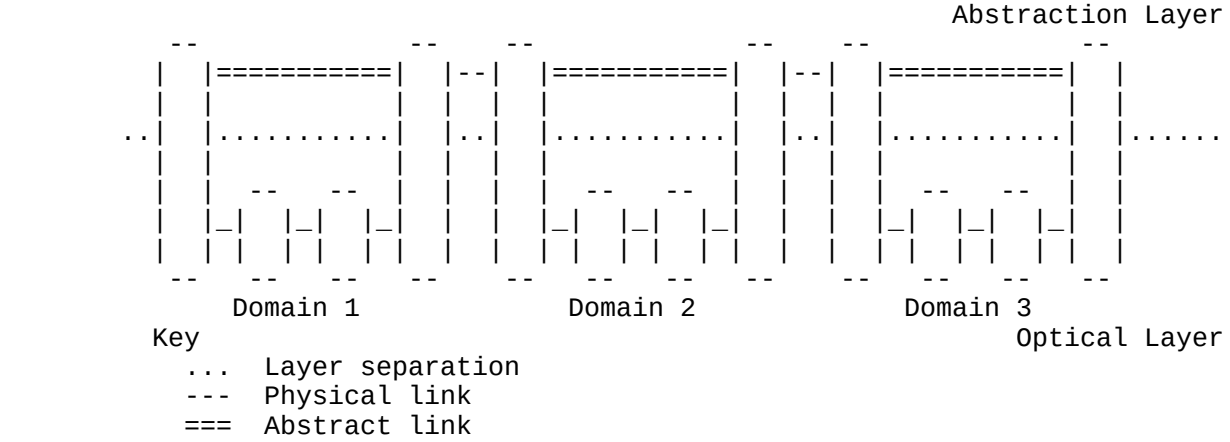


Figure 21 : The Optical Network Implemented Through the Abstraction Layer Network

8. Modeling the User-to-Network Interface

The User-to-Network Interface (UNI) is an important architectural concept in many implementations and deployments of client-server networks especially those where the client and server network have different technologies. The UNI can be seen described in [G.8080], and the GMPLS approach to the UNI is documented in [RFC4208]. Other GMPLS-related documents describe the application of GMPLS to specific UNI scenarios: for example, [RFC6005] describes how GMPLS can support a UNI that provides access to Ethernet services.

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

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

role of UNI Client-side (UNI-C) and the server edge nodes acting as the UNI Network-side (UNI-N) nodes.

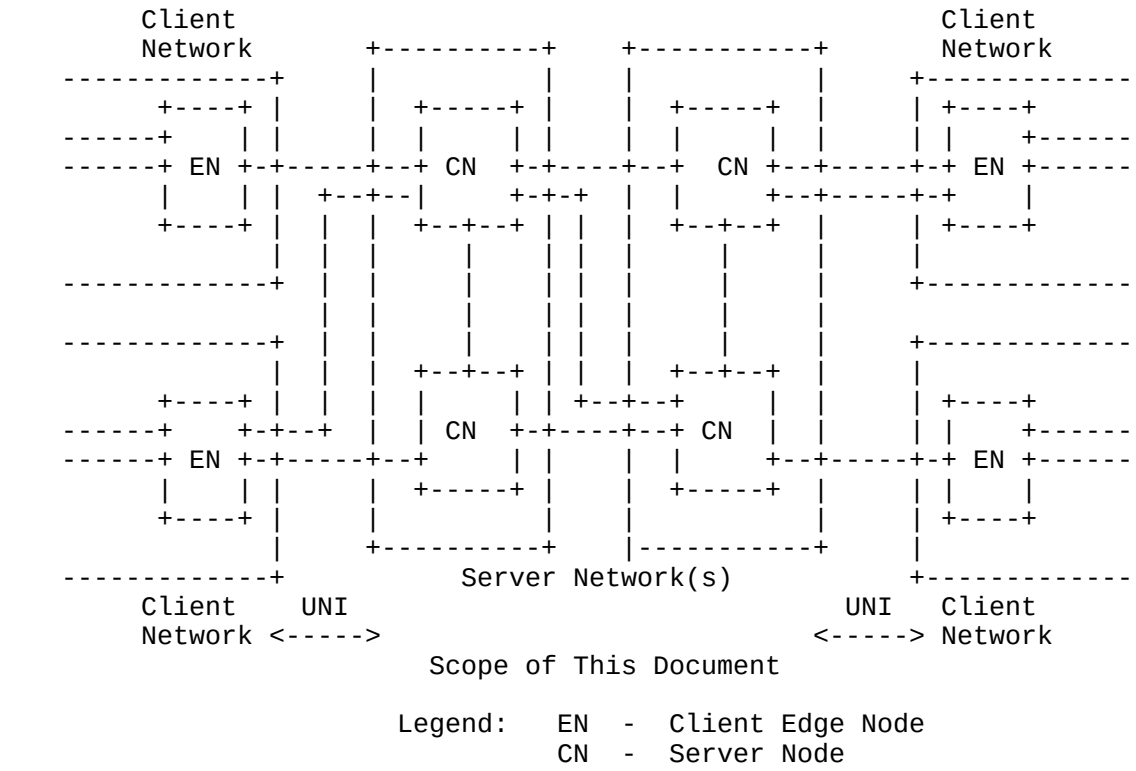


Figure 22 : Ethernet UNI Reference Model

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

9. Abstraction in L3VPN Multi-AS Environments

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

In the VPN architecture, the CE nodes are the client network edge nodes, and the PE nodes are the server network edge nodes. The Abstraction Layer Network is made up of the CE nodes, the CE-PE links, the PE nodes, and PE-PE tunnels that are the Abstract Links.

In the multi-AS or multi-operator case, the Abstraction Layer Network also includes the PEs (maybe ASBRs) at the edges of the multiple server networks, and the PE-PE (maybe inter-AS) links. This gives rise to the architecture shown in Figure 23.

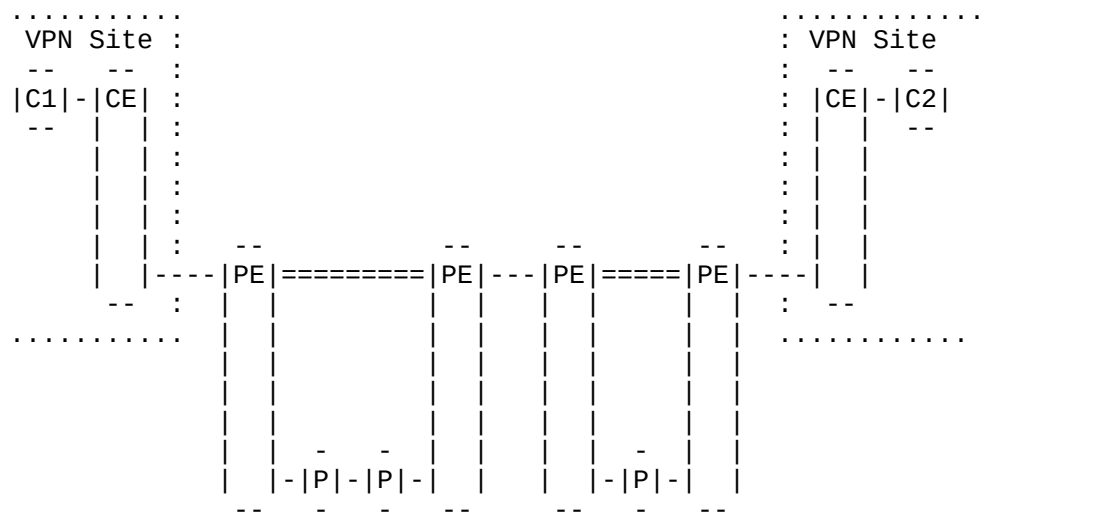


Figure 23 : The Abstraction Layer Network for a Multi-AS VPN

The policy for adding Abstract Links to the Abstraction Layer Network will be driven substantially by the needs of the VPN. Thus, when a new VPN site is added and the existing Abstraction Layer Network cannot support the required connectivity, a new Astract Link will be created out of the underlying network.

It is important to note that each VPN instance can have a separate Abstraction Layer Network. This means that the Server Network

resources can be partitioned and that traffic can be kept separate. This can be achieved even when VPN sites from different VPNs connect at the same PE. Alternatively, multiple VPNs can share the same Abstraction Layer Network if that is operationally preferable.

Lastly, just as for the UNI discussed in Section 8, the issue of dual-homing of VPN sites is a function of the Abstraction Layer Network and so is just a normal routing problem in that network.

10. Scoping Future Work

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

10.1. Not Solving the Internet

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

10.2. Working With "Related" Domains

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

10.3. Not Breaking Existing Protocols

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

10.4. Sanity and Scaling

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

11. Manageability Considerations

<TBD>

12. IANA Considerations

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

13. Security Considerations

<TBD>

14. Acknowledgements

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

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

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

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

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

15. References

15.1. Informative References

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

[I-D.beeram-ccamp-gmpls-enni] Bryskin, I., Beeram, V. P., Drake, J. et al., "Generalized Multiprotocol Label Switching (GMPLS) External Network Network Interface (E-NNI): Virtual Link Enhancements for the Overlay Model", draft-beeram-ccamp-gmpls-enni, work in progress.

[I-D.farrkingel-pce-questions]
Farrel, A., and D. King, "Unanswered Questions in the Path Computation Element Architecture", draft-farrkingel-pce-questions, work in progress.

[I-D.ietf-ccamp-xro-lsp-subobject]
Z. Ali, et al., "Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) LSP Route Diversity using Exclude Routes," draft-ali-ccamp-xro-lsp-subobject, work in progress.

[I-D.ietf-ccamp-te-metric-recording]
Z. Ali, et al., "Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) extension for recording TE Metric of a Label Switched Path," draft-ali-ccamp-te-metric-recording, work in progress.

[I-D.ietf-idr-ls-distribution]
Gredler, H., Medved, J., Previdi, S., Farrel, A., and Ray, S., "North-Bound Distribution of Link-State and TE Information using BGP", draft-ietf-idr-ls-distribution, work in progress.

[RFC2702] Awduche, D., Malcolm, J., Agogbua, J., O'Dell, M., and McManus, J., "Requirements for Traffic Engineering Over MPLS", RFC 2702, September 1999.

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

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

[RFC3630] Katz, D., Kompella, and K., Yeung, D., "Traffic Engineering (TE) Extensions to OSPF Version 2", RFC 3630, September 2003.

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

[RFC4105] Le Roux, J.-L., Vasseur, J.-P., and Boyle, J., "Requirements for Inter-Area MPLS Traffic Engineering", RFC 4105, June 2005.

[RFC4202] Kompella, K. and Y. Rekhter, "Routing Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4202, October 2005.

[RFC4206] Kompella, K. and Y. Rekhter, "Label Switched Paths (LSP) Hierarchy with Generalized Multi-Protocol Label Switching (GMPLS) Traffic Engineering (TE)", RFC 4206, October 2005.

[RFC4208] Swallow, G., Drake, J., Ishimatsu, H., and Y. Rekhter, "User-Network Interface (UNI): Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Support for the Overlay Model", RFC 4208, October 2005.

[RFC4216] Zhang, R., and Vasseur, J.-P., "MPLS Inter-Autonomous System (AS) Traffic Engineering (TE) Requirements", RFC 4216, November 2005.

[RFC4271] Rekhter, Y., Li, T., and Hares, S., "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, January 2006.

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

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

[RFC4847] T. Takeda (Ed.), "Framework and Requirements for Layer 1 Virtual Private Networks," RFC 4847, April 2007.

[RFC4920] Farrel, A., Satyanarayana, A., Iwata, A., Fujita, N., and Ash, G., "Crankback Signaling Extensions for MPLS and GMPLS RSVP-TE", RFC 4920, July 2007.

[RFC5150] Ayyangar, A., Kompella, K., Vasseur, JP., and A. Farrel, "Label Switched Path Stitching with Generalized Multiprotocol Label Switching Traffic Engineering (GMPLS TE)", RFC 5150, February 2008.

[RFC5152] Vasseur, JP., Ayyangar, A., and Zhang, R., "A Per-Domain Path Computation Method for Establishing Inter-Domain Traffic Engineering (TE) Label Switched Paths (LSPs)", RFC 5152, February 2008.

[RFC5195] Ould-Brahim, H., Fedyk, D., and Y. Rekhter, "BGP-Based Auto-Discovery for Layer-1 VPNs", RFC 5195, June 2008.

[RFC5212] Shiomoto, K., Papadimitriou, D., Le Roux, JL., Vigoureux, M., and D. Brungard, "Requirements for GMPLS-Based Multi-Region and Multi-Layer Networks (MRN/MLN)", RFC 5212, July 2008.

[RFC5251] Fedyk, D., Rekhter, Y., Papadimitriou, D., Rabbat, R., and L. Berger, "Layer 1 VPN Basic Mode", RFC 5251, July 2008.

[RFC5252] Bryskin, I. and L. Berger, "OSPF-Based Layer 1 VPN Auto-Discovery", RFC 5252, July 2008.

[RFC5305] Li, T., and Smit, H., "IS-IS Extensions for Traffic Engineering", RFC 5305, October 2008.

[RFC5440] Vasseur, JP. and Le Roux, JL., "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, March 2009.

[RFC5441] Vasseur, JP., Zhang, R., Bitar, N, and Le Roux, JL., "A Backward-Recursive PCE-Based Computation (BRPC) Procedure to Compute Shortest Constrained Inter-Domain Traffic Engineering Label Switched Paths", RFC 5441, April 2009.

[RFC5523] L. Berger, "OSPFv3-Based Layer 1 VPN Auto-Discovery", RFC 5523, April 2009.

[RFC5553] Farrel, A., Bradford, R., and JP. Vasseur, "Resource Reservation Protocol (RSVP) Extensions for Path Key Support", RFC 5553, May 2009.

[RFC5623] Oki, E., Takeda, T., Le Roux, JL., and A. Farrel, "Framework for PCE-Based Inter-Layer MPLS and GMPLS Traffic Engineering", RFC 5623, September 2009.

[RFC6005] Nerger, L., and D. Fedyk, "Generalized MPLS (GMPLS) Support for Metro Ethernet Forum and G.8011 User Network Interface

[RFC6107] Shiomoto, K., and A. Farrel, "Procedures for Dynamically Signaled Hierarchical Label Switched Paths", RFC 6107, February 2011.

[RFC6805] King, D., and A. Farrel, "The Application of the Path Computation Element Architecture to the Determination of a Sequence of Domains in MPLS and GMPLS", RFC 6805, November 2012.

[RFC6827] Malis, A., Lindem, A., and D. Papadimitriou, "Automatically Switched Optical Network (ASON) Routing for OSPFv2 Protocols", RFC 6827, January 2013.

[RFC6996] J. Mitchell, "Autonomous System (AS) Reservation for Private Use", BCP 6, RFC 6996, July 2013.

Authors' Addresses

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

John Drake
Juniper Networks
EMail: jdrake@juniper.net

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

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

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

Network Working Group
Internet Draft
Updates: 3471, 6205 (if approved)
Intended Status: Standards Track
Expires: 6 August 2014

A. Farrel
D. King
Old Dog Consulting
Y. Li
Nanjing University
F. Zhang
Huawei Technologies

6 February 2014

Generalized Labels for the Flexi-Grid in Lambda Switch Capable (LSC) Label Switching Routers

draft-farrkingel-ccamp-flexigrid-lambda-label-08.txt

Abstract

GMPLS supports the description of optical switching by identifying entries in fixed lists of switchable wavelengths (called grids) through the encoding of lambda labels. Work within the ITU-T Study Group 15 has defined a finer granularity grid, and the facility to flexibly select different widths of spectrum from the grid. This document defines a new GMPLS lambda label format to support this flexi-grid.

This document updates RFC 3471 and RFC 6205 by introducing a new label format.

Status of this Memo

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

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

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

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt

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

Table of Contents

1. Introduction

As described in [RFC3945], GMPLS extends MPLS from supporting only Packet Switching Capable (PSC) interfaces and switching, to also support four new classes of interfaces and switching that include Lambda Switch Capable (LSC).

A functional description of the extensions to MPLS signaling needed to support this new class of interface and switching is provided in

[RFC3471].

Section 3.2.1.1 of [RFC3471] states that wavelength labels "only have significance between two neighbors": global wavelength semantics are not considered. [RFC6205] defines a standard lambda label format that has a global semantic and which is compliant with both the Dense Wavelength Division Multiplexing (DWDM) grid [G.694.1] and the Coarse Wavelength Division Multiplexing (CWDM) grid [G.694.2]. The terms DWDM and CWDM are defined in [G.671].

A flexible grid network selects its data channels as arbitrarily assigned pieces of the spectrum. Mixed bitrate transmission systems can allocate their channels with different spectral bandwidths so that the channels can be optimized for the bandwidth requirements of the particular bit rate and modulation scheme of the individual channels. This technique is regarded as a promising way to improve the network utilization efficiency and fundamentally reduce the cost of the core network.

The "flexi-grid" has been developed within the ITU-T Study Group 15 to allow selection and switching of pieces of the optical spectrum chosen flexibly from a fine granularity grid of wavelengths with variable spectral bandwidth [G.694.1]. This document updates the definition of GMPLS lambda labels provided in [RFC6205] to support the flexi-grid.

This document relies on [G.694.1] for the definition of the optical data plane and does not make any updates to the work of the ITU-T in that regard.

1.1. Conventions Used in This Document

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

2. Overview of Flexi-Grid

[G.694.1] defines DWDM fixed grids. The latest version of that document extends the DWDM fixed grids to add support for flexible grids. The basis of the work is to allow a data channel to be formed from an abstract grid anchored at 193.1 THz and selected on a channel spacing of 6.25 GHz with a variable slot width measured in units of 12.5 GHz. Individual allocations may be made on this basis from anywhere in the spectrum, subject to allocations not overlapping.

[G.694.1] provides clear guidance on the support of flexible grid by implementations in Section 2 of Appendix I:

The flexible DWDM grid defined in clause 7 has a nominal central frequency granularity of 6.25 GHz and a slot width granularity of 12.5 GHz. However, devices or applications that make use of the flexible grid may not have to be capable of supporting every possible slot width or position. In other words, applications may be defined where only a subset of the possible slot widths and positions are required to be supported.

For example, an application could be defined where the nominal central frequency granularity is 12.5 GHz (by only requiring values of n that are even) and that only requires slot widths as a multiple of 25 GHz (by only requiring values of m that are even).

Some additional background on the use of GMPLS for flexible grids can be found in [FLEXFWRK].

3. Fixed Grid Lambda Label Encoding

[RFC6205] defines an encoding for a global semantic for a DWDM label based on four fields:

- Grid: used to select which grid the lambda is selected from. Values defined in [RFC6205] identify DWDM [G.694.1] and CWDM [G.694.2].

- C.S. (Channel Spacing): used to indicate the channel spacing. [RFC6205] defines values to represent spacing of 100, 50, 25 and 12.5 GHz.

- Identifier: a local-scoped integer used to distinguish different lasers (in one node) when they can transmit the same frequency lambda.

- n: a two’s-complement integer to take a positive, negative, or zero value. This value is used to compute the frequency as defined in [RFC6205] and based on [G.694.1]. The use of n is repeated here for ease of reading the rest of this document: in case of discrepancy, the definition in [RFC6205] is normative.

  Frequency (THz) = 193.1 THz + n * frequency granularity (THz)

  where the nominal central frequency granularity for the flexible grid is 0.00625 THz

4. Flexi-Grid Label Format and Values

4.1 Flexi-Grid Label Encoding

This document defines a generalized label encoding for use in flexi-grid systems. As with the other GMPLS lambda label formats defined in [RFC3471] and [RFC6205], the use of this label format is known a priori. That is, since the interpretation of all lambda labels is determined hop-by-hop, the use of this label format requires that all nodes on the path expect to use this label format.

For convenience, however, the label format is modeled on the fixed grid label defined in [RFC6205] and briefly described in Section 3.

Figure 1 shows the format of the Flexi-Grid Label. It is a 64 bit label.

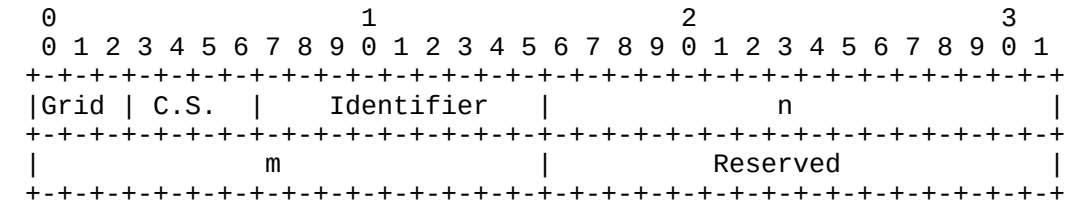


Figure 1 : The Flexi-Grid Label Encoding

This document defines a new Grid value to supplement those in [RFC6205]:

Grid	Value
ITU-T Flex	3

Within the fixed grid network, the C.S. value is used to represent the channel spacing, as the spacing between adjacent channels is constant. For the flexible grid situation, this field is used to represent the nominal central frequency granularity.

This document defines a new C.S. value to supplement those in [RFC6205]:

C.S(GHz)	Value
6.25	5

The meaning of the Identifier field is maintained from [RFC6205] (see also Section 3).

The meaning of n is maintained from [RFC6205] (see also Section 3).

The m field is used to identify the slot width according to the formula given in [G.694.1] as follows. It is a 16 bit integer value encoded in lne format.

Slot Width (GHz) = 12.5 GHz * m

The Reserved field MUST be set to zero on transmission and SHOULD be ignored on receipt.

An implementation that wishes to use the flexi-grid label encoding MUST follow the procedures of [RFC3473] and of [RFC3471] as updated by [RFC6205]. It MUST set Grid to 3 and C.S. to 5. It MUST set Identifier to indicate the local identifier of the laser in use as described in [RFC6205]. It MUST also set n according to the formula in Section 3 (inherited unchanged from [RFC6205]). Finally, the implementation MUST set m as described in the formula stated above.

4.2. Considerations of Bandwidth

There is some overlap between the concepts of bandwidth and label in many GMPLS-based systems where a label indicates a physical switching resource. This overlap is increased in a flexi-grid system where a label value indicates the slot width and so affects the bandwidth supported by an LSP. Thus the ’m’ parameter is both a property of the label (i.e., it helps define exactly what is switched) and of the bandwidth.

In GMPLS signaling [RFC3473], bandwidth is requested in the TSpec object and confirmed in the Flowspec object. The ’m’ parameter that is a parameter of the GMPLS flexi-grid label as described above, is also a parameter of the flexi-grid TSpec and Flowspec as described in [FLEXRSVP].

5. Manageability Considerations

This document introduces no new elements for management. That is, labels can continue to be used in the same way by the GMPLS protocols

and where those labels were treated as opaque quantities with local or global significance, no change is needed to the management systems.

However, this document introduces some changes to the nature of a label that may require changes to management systems. Firstly, systems that handle lambda labels as 32 bit quantities need to be updated to process the 64 bit labels described in this document even if the labels are treated as opaque quantities. Furthermore, although management systems that can handle lambda labels as defined in [RFC6205] can continue to process the fields defined in RFC 6205 as before, they have to handle new legal values of some of those fields (Grid = 3 and C.S. = 5), and they have to be aware of the new 'm' field.

6. Implementation Status

[RFC Editor Note: Please remove this entire seciton prior to publication as an RFC.]

This section records the status of known implementations of the protocol defined by this specification at the time of posting of this Internet-Draft, and is based on a proposal described in RFC 6982 [RFC6982]. The description of implementations in this section is intended to assist the IETF in its decision processes in progressing drafts to RFCs. Please note that the listing of any individual implementation here does not imply endorsement by the IETF. Furthermore, no effort has been spent to verify the information presented here that was supplied by IETF contributors. This is not intended as, and must not be construed to be, a catalog of available implementations or their features. Readers are advised to note that other implementations may exist.

According to RFC 6982, "this will allow reviewers and working groups to assign due consideration to documents that have the benefit of running code, which may serve as evidence of valuable experimentation and feedback that have made the implemented protocols more mature. It is up to the individual working groups to use this information as they see fit.

6.1. Centre Tecnologic de Telecomunicacions de Catalunya (CTTC)

Organization Responsible for the Implementation:
Centre Tecnologic de Telecomunicacions de Catalunya (CTTC)
Optical Networks and Systems Department

Implementation Name and Details:
ADRENALINE testbed

http://networks.cttc.es/experimental-testbeds/

Brief Description:
Experimental testbed implementation of GMPLS/PCE control plane.

Level of Maturity:
Implemented as extensions to a mature GMLPS/PCE control plane. It is limited to research / prototyping stages but it has been used successfully for more than the last five years.

Coverage:
Support for the 64 bit label as described version 07 of this document.
This affects mainly the implementation of RSVP-TE and PCEP protocols:
- Generalized Label Support
- Suggested Label Support
- Upstream Label Support
- ERO Label Subobjects and Explicit Label Control
It is expected that this implementation will evolve to follow the evolution of this document.

Licensing:
Proprietary

Implementation Experience:
Implementation of this document reports no issues.
General implementation experience has been reported in a number of journal papers. Contact Ramon Casellas for more information or see http://networks.cttc.es/publications/?search=GMPLS&research_area=optical-networks-systems

Contact Information:
Ramon Casellas: ramon.casellas@cttc.es

Interoperability:
No report.

7. Security Considerations

[RFC6205] notes that the definition of a new label encoding does not introduce any new security considerations to [RFC3471] and [RFC3473]. That statement applies equally to this document.

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

8. IANA Considerations

IANA maintains the "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Parameters" registry that contains several subregistries.

8.1. Grid Subregistry

IANA is requested to allocate a new entry in this subregistry as follows:

Value	Grid	Reference
3	ITU-T Flex	[This.I-D]

8.2. DWDM Channel Spacing Subregistry

IANA is requested to allocate a new entry in this subregistry as follows:

Value	Channel Spacing (GHz)	Reference
5	6.25	[This.I-D]

9. Acknowledgments

This work was supported in part by the FP-7 IDEALIST project under grant agreement number 317999.

Very many thanks to Lou Berger for discussions of labels of more than 32 bits. Many thanks to Sergio Belotti and Pietro Vittorio Grandi for their support of this work. Thanks to Gabriele Galimberti for discussion of the size of the "m" field.

Special thanks to the Vancouver 2012 Pool Party for discussions and rough consensus: Dieter Beller, Ramon Casellas, Daniele Ceccarelli, Oscar Gonzalez de Dios, Iftekhar Hussain, Cyril Margaria, Lyndon Ong, and Fatai Zhang.

The authors would like to thank Ben Niven-Jenkins for inspiring the choice of filename for this 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.

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

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

[RFC6205] Otani, T., and Li, D., "Generalized Labels for Lambda-Switch-Capable (LSC) Label Switching Routers", RFC 6205, October 2011.

[G.694.1] ITU-T Recommendation G.694.1 (revision 2), "Spectral grids for WDM applications: DWDM frequency grid", February 2012.

10.2. Informative References

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

[RFC5920] Fang, L., Ed., "Security Framework for MPLS and GMPLS Networks", RFC 5920, July 2010.

[RFC6982] Sheffer, Y. and A. Farrel, "Improving Awareness of Running Code: The Implementation Status Section", RFC 6982, July 2013.

[RFC Editor Note: This reference can be removed when Section 6 is removed]

[G.671] ITU-T Recommendation G.671, "Transmission characteristics of optical components and subsystems", 2009.

[G.694.2] ITU-T Recommendation G.694.2, "Spectral grids for WDM applications: CWDM wavelength grid", December 2003.

[FLEXFWRK] O. Gonzalez de Dios, et al., "Framework and Requirements for GMPLS based control of Flexi-grid DWDM networks", draft-ogrcetal-ccamp-flexi-grid-fwk, work in progress.

[FLEXRSVP] Zhang, F., Gonzalez de Dios, O., and D. Ceccarelli, "RSVP-TE Signaling Extensions in support of Flexible Grid", draft-zhang-ccamp-flexible-grid-rsvp-te-ext, work in progress.

Appendix A. Flexi-Grid Example

Consider a fragment of an optical LSP between node A and node B using the flexible grid. Suppose that the LSP on this hop is formed:
- using the ITU-T Flexi-Grid
- the nominal central frequency of the slot 193.05 THz
- the nominal central frequency granularity is 6.25 GHz
- the slot width is 50 GHz.

In this case the label representing the switchable quantity that is the flexi-grid quantity is encoded as described in Section 4.1 with the following parameter settings. The label can be used in signaling or in management protocols to describe the LSP.

Grid = 3 : ITU-T Flexi-Grid

C.S. = 5 : 6.25 GHz nominal central frequency granularity

Identifier = local value indicating the laser in use

n = -8 :

Frequency (THz) = 193.1 THz + n * frequency granularity (THz)

193.05 (THz) = 193.1 (THz) + n * 0.00625 (THz)

n = (193.05-193.1)/0.00625 = -8

m = 4 :

Slot Width (GHz) = 12.5 GHz * m

50 (GHz) = 12.5 (GHz) * m

m = 50 / 12.5 = 4

Authors' Addresses

Adrian Farrel
Old Dog Consulting
EMail: adrian@olddog.co.uk

Daniel King
Old Dog Consulting
EMail: daniel@olddog.co.uk

Yao Li
Nanjing University
EMail: wsliguotou@hotmail.com

Fatai Zhang
Huawei Technologies
EMail: zhangfatai@huawei.com

Contributors' Addresses

Zhang Fei
ZTE
EMail: zhang.fei3@zte.com.cn

Ramon Casellas
CTTC
EMail: ramon.casellas@cttc.es

Network Working Group
Internet Draft
Intended status: Standards Track

D. Fedyk
Hewlett-Packard
D. Beller
L. Levrau
Alcatel-Lucent
D. Ceccarelli
Ericsson
F. Zhang
Huawei Technologies
Y. Tochio
Fujitsu
X. Fu
ZTE

Expires: August 18, 2014 February 14, 2014

UNI Extensions for Diversity and Latency Support
draft-fedyk-ccamp-uni-extensions-04.txt

Status of this Memo

Abstract

This document builds on the GMPLS overlay model [RFC4208] and defines extensions to the GMPLS User-Network Interface (UNI) to support route diversity within the core network for sets of LSPs initiated by edge nodes. A particular example where route diversity within the core network is desired, are dual-homed edge nodes. The core network is typically composed of multiple network domains and those multi-domain diversity aspects that have an implication on the GMPLS UNI extensions are discussed.

The document also defines GMPLS UNI extensions to deal with latency requirements for edge node initiated LSPs.

This document uses a VPN model that is based on the same premise as L1VPN framework [RFC4847] but may also be applied to other technologies. The extensions are applicable both to VPN and non VPN environments. These extensions move the UNI from basic connectivity to enhanced mode connectivity by including additional constraints while minimizing the exchange of CE to PE information. These extensions are applicable to the overlay extension service model. Route Diversity for customer LSPs are a common requirement applicable to L1VPNs. The UNI mechanisms described in this document are L1VPN compatible and can be applied to achieve diversity for sets of customer LSPs.

The UNI extensions in support of latency constraints can also be applied to the extended overlay service model in order for the customer LSPs to meet certain latency requirements.

Table of Contents

1. Introduction

This document builds on the GMPLS overlay model [RFC4208] and defines extensions to the GMPLS User-Network Interface (UNI) to support route diversity within the core network for sets of LSPs initiated by edge nodes. In the following, the term customer edge (CE) device is used synonymously for the term edge node (EN) as in [RFC4208].

Moreover, the VPN terminology (CE and PE) [RFC4026] is used below when the core network is a VPN but is also applicable to UNI interfaces [RFC4208].

This document uses a VPN model that is based on the same premise as L1VPN framework [RFC4847] but may also be applied to other

technologies. The extensions are applicable both to VPN and non VPN environments. These extensions move the UNI from basic connectivity to enhanced mode connectivity by including additional constraints while minimizing the exchange of CE to PE information. These extensions are applicable to the overlay extension service model.

The overlay model assumes a UNI interface between the edge nodes of the respective transport domains. Route diversity for LSPs from single homed CE and dual-home CEs is a common requirement in optical transport networks. This document describes two signaling variations that may be used for supporting LSP diversity within the overlay extension service model considering dual-homing. Dual-homing is typically used to avoid a single point of failure (UNI link, PE) or if two disjoint connections are forming a protection group in the CE device, e.g., 1+1 protection. While both methods are similar in that they utilize common mechanisms in the PE network to achieve diversity, they are distinguished according to whether the CE is permitted to retrieve provider SRLG diversity information for an LSP from a PE1 and pass it on to a PE2 (SRLG information is shared with the CE), or whether a new attribute is used that allows the PE2 that receives this attribute to derive the SRLG information for an LSP based on the attribute value. Figure 1 below is depicting the scenario.

The core network is typically composed of multiple network domains (different providers, geographical separation, etc.) and some multi-domain diversity aspects have implications on the GMPLS UNI extensions defined in this document. It shall be noted that path computation can be done in two different ways for each domain: GMPLS supports distributed routing providing each node in the domain the capability to do constraint-based path computation while the utilization of the centralized path computation element (PCE) approach is another option. The GMPLS UNI extensions defined in this document are applicable to both path computation approaches and also mixed scenarios are supported where some domains utilize the distributed path computation approach while other domains are using a PCE.

The extended overlay service model can support other extensions for VPN signaling, for example, those related to latency. When requesting diverse LSPs, latency may also be an additional requirement.

2. Conventions used in this document

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

In this document, these words will appear with that interpretation only when in ALL CAPS. Lower case uses of these words are not to be interpreted as carrying RFC-2119 significance.

3. Contributors

The Authors would like to thank Eve Varma and Sergio Belotti for their review and contributions to this document.

4. LSP Diversity in the Overlay Extension Service Model

The L1VPN Framework [RFC4847] (Enhanced Mode) describes the overlay extension service model, which builds upon the UNI Overlay [RFC4208] serving as the interface between the CE edge node and the PE edge node. In this service model, a CE receives a list of CE-PE TE link addresses to which it can request a L1VPN connection (i.e., membership information) and may include additional information concerning these TE links. This document further builds on the overlay extension service model by adding shared constraint information for path diversity in the optical transport network. While the L1VPN for optical transport is an example specific VPN technology the term VPN is used generically since the extensions can apply to GMPLS UNIs and VPNs for other technologies.

Two signaling variations are outlined here that may be used for supporting LSP diversity within the overlay extension service model considering dual-homing. While both methods utilize common mechanisms in the PE network to achieve diversity, they are distinguished according to whether the CE is permitted to retrieve provider SRLG diversity information for an LSP from a PE1 and pass it on to a PE2 (SRLG information is shared with the CE or whether a new attribute is used that allows the PE2 that receives this attribute to derive the SRLG information for an LSP based on this attribute value. The selection between these methods is governed by both PE-network specific policies and approaches taken (i.e., in terms of how the provider chooses to perform routing internal to their network).

The first method (see 4.1.1) assumes that provider Shared Resource Link Group (SRLG) Identifier information is both available and shareable (policy decision) with the CE. Since SRLG IDs can then be used (passed transparently between PEs via the dual-homed CE) as signaled information on a UNI message, a mechanism supporting LSP diversity for the overlay extension service model can be provided via straightforward signaling extensions.

The second method (see 3.1.2) assumes that provider SRLG IDs are either not available or not shareable (based on provider network operator policy) with the CE. For this case, a mechanism is provided

where information signaled to the PE on UNI messages does not require shared knowledge of provider SRLG IDs to support LSP diversity for the overlay extension model.

While both methods could be implemented in the same PE network, it is likely that a GMPLS VPN CE network would use only one mechanism at a time.

4.1. LSP diversity for dual-homed customer edge (CE) devices

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

For the dual-homing case, it is possible to establish two connections from the source CE device to the same destination CE device where one connection is using one UNI link to, for example, PE1 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. This document describes how it is possible to enable such path diversity to be achieved within the provider network (which is subject to additional routing constraints). [RFC4202] defines SRLG information that can be used to allow GMPLS to provide path diversity in a GMPLS controlled transport network. As the two connections are entering the provider network at different PE devices, the PE device that receives the connection request for the second connection needs to be capable of determining the additional path computation constraints such that the path of the second LSP is disjoint with respect to the already established first connection entering the network at a different PE device. The methods described in this document allow a PE device to determine the SRLG information for a connection in the provider network that is entering the network on a different PE device.

PE SRLG information can be used directly by a CE if the CE understands the context, and the CE view is limited to its VPN context. In this case, there is a dependency on the provider information and there is a need to be able to query the SRLG in the provider network.

It may, on the other hand, be preferable to avoid this dependency and to decouple the SRLG identifier space used in the provider network from the SRLG space used in the client network. This is possible with

both methods detailed below. Even for the method where provider SRLG information is passing through the CE device (note the CE device does not need to process and decode this information) the two SRLG identifier spaces can remain fully decoupled and the operator of the client network is free to assign SRLG identifiers from the client SRLG identifier space to the CE to CE connection that is passing through the provider network.

Referring to Figure 1, the UNI signaling mechanism must support at least one of the two mechanisms described in this document for CE dual homing to achieve LSP diversity in the provider network.

The described mechanisms can also be applied to a scenario where two CE devices are connected to two different PE devices. In this case, the additional information that is exchanged across the UNI interfaces also needs to be exchanged between the two CE devices in order to achieve the desired diversity in the provider network.

This information may be configured or exchanged by some automated mechanism not described in this document.

In the dual-homing example, CE1 can locally correlate the LSP requests. For the slightly more complicated example involving CE2 and CE3, both requiring a path that shall be diverse to a connection initiated by the other CE device, CE2 and CE3 need to have a common view of the SRLG information to be signaled. In this document, we detail the required diversity information and the signaling of this diversity information; however, the means for distributing this information within the PE domain or the CE domain is out of scope.

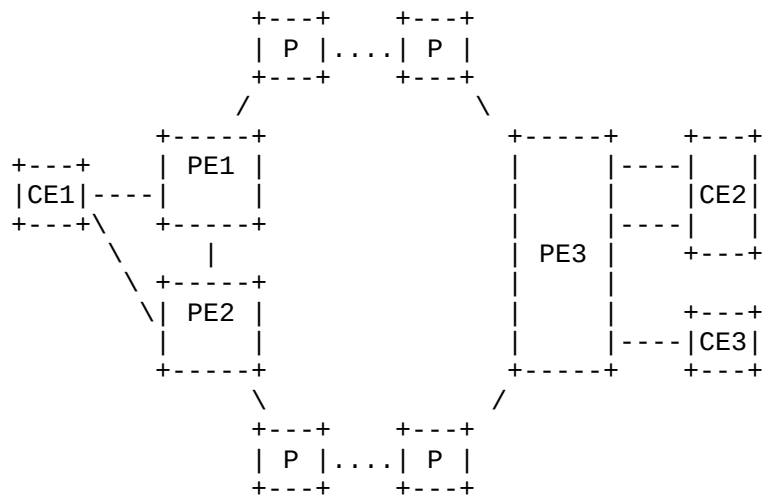


Figure 1 Overlay Reference Diagram

In an overlay model, the information exchanged between the CE and the PE is kept to a minimum.

How diversity is achieved, in terms of configuration, distribution and usage in each part of the transport networks should be kept independent and separate from how diversity is signaled at the UNI between the two transport networks.

Signaling parameters discussed in this document are:

- SRLG information (see [RFC4202])
- Path Affinity Set

4.1.1. Exchanging SRLG information between the PEs via the CE device

SRLG information is defined in [RFC4202] and if the SRLG information of an LSP is known, it can be used to calculate a path for another LSP that is SRLG diverse with respect to an existing LSP. SRLG information is an unordered list of SRLGs. SRLG information is normally not shared between the transport network and the client network; i.e., not shared with the CEs of a VPN in the VPN context. However, this becomes more challenging when a CE is dual-homed. For example, CE1 in Figure 1 may have requested an LSP1 from CE1 to CE2 via PE1 and PE3. CE1 could subsequently request an LSP2 to CE2 via PE2 and PE3 with the requirement that it should be maximally SRLG disjoint with respect to LSP1. Since PE2 does not have any information about LSP1, PE2 would need to know the SRLG information associated with LSP1. If CE1 could request the SRLG information of LSP1 from PE1, it could then transparently pass this information to PE2 as part of the LSP2 setup request, and PE2 would now be capable of calculating a path for LSP2 that is SRLG disjoint with respect to LSP1.

The exchange of SRLG information is achieved on a per VPN LSP basis using the existing RSVP-TE signaling procedures. It can be exchanged in the PATH (exclusion information) or RESV message in the original request or it can be requested by the CE at any time the path is active.

It shall be noted that SRLG information is an unordered list of SRLG identifiers and the encoding of SRLG information for RSVP signaling is already defined in [SRLG_info]. Even if SRLG information is known for several LSPs it is not possible for the CEs to derive the provider network topology from this information.

4.1.1.1. Operational Procedures

Retrieving SRLG information from a PE for an existing LSP:

When a dual-homed CE device intends to establish an LSP to the same destination CE device via another PE node, it can request the SRLG information for an already established LSP by setting the SRLG information flag in the LSP attributes sub-object of the RSVP PATH message (IANA to assign the new SRLG flag). As long as the SRLG information flag is set in the PATH message, the PE node inserts the

SRLG sub-object as defined in [SRLG_info] into the RSVP RESV message that contains the current SRLG information for the LSP. If the provider network's policy has been configured so as not to share SRLG information with the client network, the SRLG sub-object is not inserted in the RESV message even if the SRLG information flag was set in the received PATH message. Note that the SRLG information is expected to be always up-to-date.

Establishment of a new LSP with SRLG diversity constraints:

When a dual-homed CE device sends an LSP setup requests to a PE device for a new LSP that is required to be SRLG diverse with respect to an existing LSP that is entering the network via another PE device, the CE device sets the SRLG diversity flag (note: IANA to assign the new SRLG diversity flag) in the LSP attributes sub-object of the PATH message that initiates the setup of this new LSP. When the PE device receives this request it calculates a path to the given destination and uses the received SRLG information as path computation constraints.

4.1.1.2. Error Handling Procedures

When the CE device receives a RSVP PATH message with the SRLG information flag set and if the provider's network policy does not permit sharing of SRLG information, the PE device shall notify the CE device by sending a RSVP PathErr with a Notify error code (error code to be defined) "Retrieval of SRLG information not permitted". As described above, the PE device must not include the SRLG sub-object with the SRLG information for the LSP in the RSVP RESV message.

If the PE device receives a RSVP PATH message for a new LSP with the SRLG diversity flag set and SRLG information in the SRLG sub-object, the PE device tries to calculate a route to the given destination that is SRLG diverse with respect to the provided SRLG information. If no route can be found, a RSVP PathErr message with an error code (error code to be defined) "No SRLG diverse route available toward destination".

If the PE device receives a RSVP PATH message for a new LSP with the SRLG diversity flag set and SRLG information in the SRLG sub-object and if the PE device does not support the SRLG sub-object, the PE device shall send a PathErr message to the CE device, indicating an "Unknown object class".

Further error handling cases will be added in the next revision of

this document.

4.1.2. Using Path Affinity Set Extension

The Path Affinity Set (PAS) is used to signal diversity in a pure CE context by abstracting SRLG information. There are two types of diversity information in the PAS. The first type of information is a single PAS identifier. The Second part is the optional PATH information, in the form of Source and Destination addresses of an exclude path or set of paths that MAY be specified. The motive behind the PAS information is to have as little exchange of diversity information as possible between the VPN CE and PE elements.

Rather than a detailed CE or PE SRLG list, the Path Affinity Set contains an abstract SRLG identifier that associates the given path as diverse. Logically the identifier is in a VPN context and therefore only unique with respect to a particular VPN.

How the CE determines the PAS identifier is a local matter for the CE administrator. A CE may signal the PAS identifier as a diversity object in the PATH message. This identifier is a suggested identifier and may be overridden by a PE under some conditions.

For example, a PAS identifier can be used with no prior exchange of PAS information between the CE and the PE. Upon reception of the PAS identifier information the PE can infer the CEs requirements. The actual PAS identifier used will be returned in the RESV message. Optionally an empty PAS identifier allows the PE to pick the PAS identifier.

Similar to the section 4.1.1 on SRLG information, a PE can return PAS identifier as the response to a Query allowing flexibility.

A PE interprets the specific PAS identifier, for example, "123" as meaning to exclude the PE SRLG information (or equivalent) that has been allocated by LSPs associated with this Path Affinity Set identifier "123", for any LSPs associated with the resources assigned to the VPN. For example, if a Path exists for the LSP with the identifier "123", the PE would use local knowledge of the PE SRLGs associated with the "123" LSPs and exclude those SRLGs in the path request. In other words, two LSPs that need to be diverse both signal "123" and the PEs interpret this as meaning not to use shared resources. Alternatively, a PE could use the PAS identifier to select from already established LSPs. Once the path is established it becomes the "123" identifier or optionally another PAS identifier for that VPN that replaces "123".

The optional PAS Source and Destination Address tuple represents one or more source addresses and destination addresses associated with the CE Path Affinity Set identifier. These associated address tuples represent paths that use resources that should be excluded for the establishment of the current LSP. The address tuple information gives both finer grain details on the path diversity request and serves as an alternative identifier in the case when the PAS identifier is not known by the PE. The address tuples used in signaling is within a CE context and its interpretation is local to a PE that receives a Path request from a CE. The PE can use the address information to relate to PE Addresses and PE SRLG information. When a PE satisfies a connection setup for a (SRLG) diverse signaled path, the PE may optionally record the PE SRLG information for that connection in terms of PE based parameters and associate that with the CE addresses in the Path message.

Specifically for L1VPNs, Port Information table (PIT) [RFC5251] can be leveraged to translate between CE based addresses and PE based addresses. The Path Affinity Set and associated PE addresses with PE SRLG information can be distributed via the IGP in the provider transport network (or by other means such as configuration); they can be utilized by other PEs when other CE Paths are setup that would require path/connection diversity. This information is distributed on a VPN basis and contains a PAS identifier, PE addresses and SRLG information.

If diversity is not signaled, the assumption is that no diversity is required and the Provider network is free to route the LSP to optimize traffic. No Path affinity set information needs to be recorded for these LSPs. If a diversity object is included in the connection request, the PE in the Provider Network should be able to look-up the existing Provider SRLG information from the provider network and choose an LSP that is maximally diverse from other LSPs.

The mechanisms to achieve this are outside the scope of this document.

A new VPN Diverse LSP LABEL object is specified:

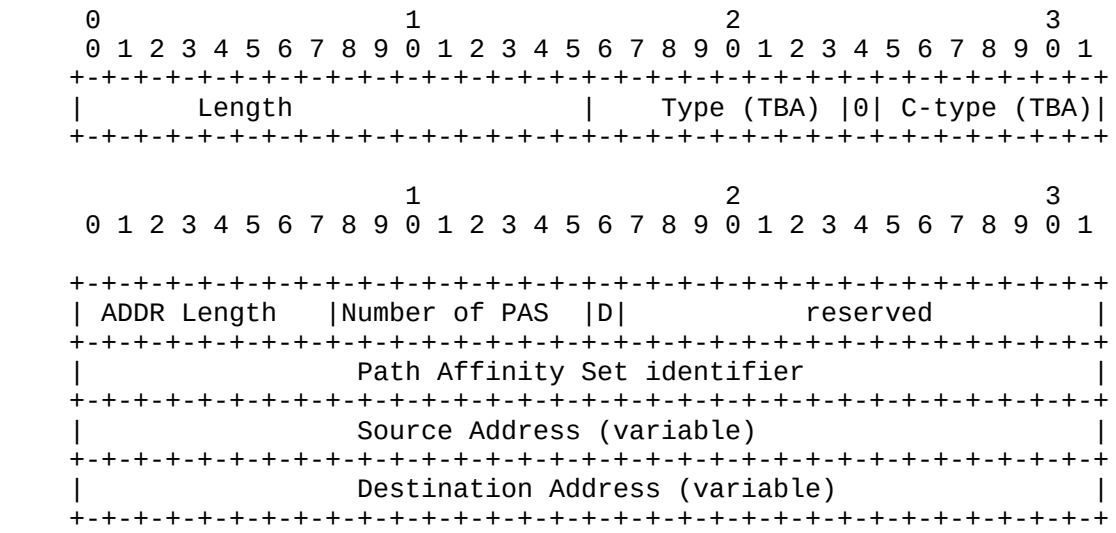


Figure 2 Diverse LSP information

1. The Address Length field (8 bits) is the number of bytes for both the source address and destination address. The address may be in any format from 1 to 32 bytes but the key point is the customers can maintain their existing addresses. A value of zero indicates there are no addresses included.

2. The Number of Path Affinity (8 bits) sets is included in the object. This is typically 1. Addition of other sets is for further study.

3. The Path affinity Set identifier (4 bytes) is a single number that represents a summarized SRLG for this path. Paths with that same Path Affinity set should be set up with diverse paths and associated with the path affinity set. A value of all zeros allows the PE to pick a PAS identifier to return. A PAS identifier of an established path may be different than the requested path identifier.

4. The diversity Bit (D) (one Bit) indicates if the diversity must be satisfied when set as a one. If a PE finds an established path with a Path Affinity set matching the signaled Path Affinity Set or the signaled Address tuple it should attempt find a diverse path.

5. The Diverse Path Source address/destination address tuple is that of an established LSP in the PE network that belongs to the same Path Affinity Set identifier. If the path for these addresses is not established or cannot be determined by the PE edge processing the PATH request then the path is established only with the Path Affinity identifier. If the path(s) for these address tuples are known by the PE the PE uses the SRLG information associated with these addresses. If in any case a diverse path cannot be setup then the Diverse bit controls whether a path is established anyway. The PE must use the PIT to translate CE Addresses into provider addresses when correlating with provider SRLG information. How SRLG information and network address tuples are distributed is for future study.

4.1.2.1. Operational Procedures

When a CE constructs a PATH message it may optionally specify and insert a Path Affinity Set in the PATH message. This Path Affinity Set may optionally include the address of an LSP that that could belong to the same Path Affinity Set. The Path Affinity Set identifier is a value (0 through 2**32-255) that is independent of the mechanism the CE or the PE use for diversity. The Path Affinity Set is a single identifier that can be used to request diversity and associate diversity.

When processing a CE PATH message in a VPN Overlay, the PE first looks up the PE based addresses in the Provider Index Table (PIT). If the Path Affinity Set is included in the PATH message, the PE must look up the SRLG information (or equivalent) in the PE network that has been allocated by LSPs associated with a Path Affinity Set and exclude those resources from the path computation for this LSP if it is a new path. The PE may alternatively choose from an existing path with a disjoint set of resources. If a path that is disjoint cannot be found, the value of the PAS diversity bit determines whether a path should be setup anyway. If the PAS diversity bit is clear, one can still attempt to setup the LSP. A PE should still attempt to minimize shared resources but that is an implementation issue, and is outside the scope of this document.

Optionally the CE may use a value of all zeros in the PAS identifier allowing the PE to select an appropriate PAS identifier. Also the PE may to override the PAS identifier allowing the PE to re-assign the identifier if required. A CE should not assume that the PAS identifier used for setup is the actual PAS identifier.

4.1.2.2. Error Handling Procedures

The PAS object must be understood by the PE device. Otherwise, the CE should not use the PAS object. Path Message processing of the PAS object SHOULD follow CTYPE 0. An Error code of IANA (TBD) indicates that the PAS object is not understood.

When a PAS identifier is not recognized by a PE it must assume this LSP defines that PAS identifier however the PE may override PAS identifier under certain conditions.

If the identifier is recognized but the Source Address-Destination address pair(s) are not recognized, this LSP must be set up using the PAS identifier only.

If the identifier is recognized and the Source Address-Destination address pair(s) are also recognized, then the PE SHOULD use the PE SRLG information associated with the LSPs identified by the address pairs to select a disjoint path.

The Following are the additional error codes:

1. Route Blocked by Exclude Route Value IANA (TBA).

4.1.2.3. Distribution of the Path Affinity Set Information

Information about SRLG is already available in the IGP TE database. A PE network can be designed to have additional opaque records for Provider paths that distribute PE paths and SRLG on a VPN basis. When a PE path is setup, the following information allows a PE to lookup the PE diversity information:

- L1 VPN Identifier 8 bytes
- Path Affinity Set Identifier
- Source PE Address
- Destination PE Address
- List of PE SRLG (variable)

The source PE address and destination PE address are the same addresses in the VPN PIT and correspond to the respective CE address identifiers.

Note that all of the information is local to the PE context and is not shared with the CE. The VPN Identifier is associated with a CE. The only value that is signaled from the CE is the Path Affinity Set and optionally the addresses of an existing LSP. The PE stores source and destination PE addresses of the LSP in their native format along with the SRLG information. This information is internal to the PE network and is always known.

PE paths may be setup on demand or they may be pre-established. When paths are pre-established, the Path Affinity Set is set to unassigned 0x0000 and is ignored. When a CE uses a pre-established path the PE may set the Path SRLG Path Affinity Set value if the CE signals one otherwise the Path Affinity Set remains unassigned 0x0000.

4.2. Multi-domain LSP Diversity Aspects for Dual-homed CE Devices

The two mechanisms described above to achieve LSP diversity for dual-homed CE devices can be applied to single-domain provider networks as well as multi-domain provider networks. This section addresses multi-domain aspects including both single provider multi-domain networks and multi-provider networks where the subdivision into multiple domains is obvious due to the organizational boundaries between different providers. Specifically, when multiple providers are involved, SRLG identifiers as well as PAS identifiers must be administrable independently for each provider network.

For the single provider multi-domain case, there are two possibilities how SRLG or PAS identifiers can be handled:

- Subdividing the identifier space into ranges assigned to domains
- Scoping the identifiers to domains

4.2.1 Subdividing Identifier Spaces into Ranges

Subdividing the identifier space into disjoint ranges and assigning the different ranges to the different domain is one possibility to apply the LSP diversity mechanisms defined in this document to a multi-domain environment. This does not require additional protocol extensions. Caution is, however, required when the identifiers are assigned. They must be selected strictly from the identifier range that has been assigned to the specific domain. From a network operations perspective, this can be an option for a single provider multi-domain network while it may be less applicable to multi-provider networks where minimal dependency is desired.

4.2.2 Scoping Identifier Spaces to Domains

[DRAFT DOMAIN SUBOBJECTS] defines new RSVP-TE domain sub-objects for the purpose of identifying domains. Domain sub-objects can be used to scope SRLG or PAS identifiers to a specific domain. With this extension, the full SRLG or PAS identifier space can be used within each domain. When a new multi-domain LSP shall be established, the diversity constraints can be signaled in the form of a sequence of a scoping domain sub-object followed by the list of SRLGs (SRLG sub-object) or the PAS sub-object, e.g.: [domain_sub-object(Dn), SRLG_sub-object(Dn)] for domain Dn.

4.2.3. Multi-domain Diversity Aspects in Case Domains Utilize a PCE

Typically, the core network is composed of multiple network domains (different providers, geographical separation, etc.) and some multi-domain diversity aspects have implications on the GMPLS UNI extensions defined in this document.

For GMPLS controlled networks, two options are defined how path computation can be done:

- Distributed path computation, i.e., each node is capable to perform constraint-based path computation
- Centralized path computation utilizing PCE as defined in [RFC4655]

The GMPLS UNI extensions defined in this document shall be applicable to both path computation approaches and also mixed scenarios shall be supported where some domains utilize the distributed path computation approach while other domains are using a PCE.

In case a network domain uses a PCE, path information for all LSPs crossing the domain can be stored in the PCE’s database and [DRAFT PATH KEY] defines a mechanism how a LSP diversity constraint can be signaled in the RSVP-TE eXclude Route Object (XRO) using a unique path key encoded in a path key sub-object. Further details can be found in [DRAFT PATH KEY].

If the scoping approach as defined in section 4.2.2 above is applied, the diversity constraint for an LSP can be signaled in the form of a sequence of a domain sub-object followed by a path key sub-object and the path key sub-object itself contains the PK-owner-ID that tells the ingress node of a domain receiving the diversity constraint which PCE instance it has to consult.

For mixed scenarios, where some domains are using the distributed path computation approach while the other domains are utilizing a PCE, the LSP diversity constraint can be signaled in the form of a

sequence of a scoping domain sub-object followed by the list of SRLGs (SRLG sub-object) or the PAS sub-object (distributed path computation) or the path key sub-object for domains using a PCE. Hence the diversity constraint for a domain Dn has the following form:

[domain_sub-object(Dn), SRLG_sub-object(Dn) | PAS_subobject(Dn) | PK_sub-object(Dn)]

5. Latency Signaling Extensions

Some network applications are sensitive to latency (sometimes also called delay) while other applications are sensitive to latency variation (sometimes also called delay variation). Specifically, real time applications typically do have certain latency requirements. It shall be noted that latency variation is typically not an issue for TDM networks including the WDM layer. For these technologies the latency is constant and there is no latency variation added. Latency variation is typically caused in packet networks or when packet based services are encapsulated into a constant bit rate server layer signal, which requires buffering of the arriving packets that may arrive in bursts. An example is an Ethernet VLAN service that is mapped into a constant bit rate server layer such as an ODUk or ODUflex OTN signal.

The GMPLS UNI as defined in [RFC4208] does not support latency as a signaling parameter that would allow a CE device to signal to the PE device that latency and/or latency variation constraints need to be met when a path is calculated for the requested LSP. The path computation function does typically calculate a route to the given destination that has the least TE metric (least cost routing). However, if a CE device requests an LSP via the UNI interface for an application that is sensitive to latency/latency variation, it should be possible to signal to the PE device that the objective function should rather take latency into account instead of the TE metric.

In order to support latency/latency variation as path computation constraint, the network has to support latency/latency variation as TE metric extension as defined in [DRAFT OSPF TE METRIC EXT] - note that [DRAFT OSPF TE METRIC EXT] is using the terms delay/delay variation instead of latency/latency variation.

A latency requirement can be added to signaling in the form of a constraint [DRAFT OBJECTIVE FUNCTION]. The constraint can take the form of:

- Minimal latency

- Maximum acceptable latency (upper bound)
- Minimal latency variation
- Maximum acceptable latency variation (upper bound), if applicable

While some systems may be able to compute routes based on delay metrics it is usual that minimizing the accumulated TE link metric (link cost) or the number of hops subject to bandwidth reservation are satisfied as the object function and delay is not considered. When considering diversity latency falls after diversity constraints have been satisfied.

Recording the latency of existing paths [DRAFT TE METRIC RECORD] to ensure they meet a maximum acceptable latency can be utilized to ensure latency constraint is met.

When a low latency path is required, the minimize latency subject to other constraints criteria should be signaled. A CE device can use the recorded latency to ensure that the maximum acceptable latency has been met.

5.1. RSVP-TE Extensions

At the UNI, the RSVP-TE extensions as defined in [DRAFT OBJECTIVE FUNCTION] SHALL be used for signaling the PE device whether a path with minimal latency is requested or whether certain latency/latency variation upper bound constraints shall be met for the end-to-end connection, i.e., from the source CE device to the destination CE device. The following objective function (OF) code point SHALL be used in the OF sub-object of the ERO to indicate that latency/latency variation constraints SHALL be taken into account when the path computation function that is invoked by the PE node that expands the route from the PE device to the destination CE device:

- OF code value 8 (to be assigned by IANA) is for the Minimum Latency Path (MLP) OF
- OF code value 9 (to be assigned by IANA) is for Minimum Latency Variation Path (MLVP) OF

Additionally, an optional OF metric-bound sub-object MAY be carried within an ERO object of the RSVP-TE Path message. The two metric-bound sub-objects defined in [DRAFT OBJECTIVE FUNCTION] that are corresponding to the two OFs above are:

- metric bound sub-object of Type T=4: Cumulative Latency

o metric bound sub-object of Type T=5: Cumulative Latency Variation

The metric-bound indicates an upper bound for the path metric that MUST NOT be exceeded for the ERO expending node to consider the computed path as acceptable. It shall be noted that the metric bound included in the RSVP-TE Path message at the UNI has end-to-end significance, which means that the upper bound metric constraint MUST be met for the path from the source CE device to the destination CE device.

5.2. Operational Procedures

The processing rules as defined in [DRAFT OBJECTIVE FUNCTION] for the OF sub-object and the optional OF metric-bound sub-object SHALL be applied at the ingress PE device when the source CE device requests an LSP (It shall be noted that [DRAFT OBJECTIVE FUNCTION] has a wider scope and may also apply to inter-domain interfaces, i.e., when the provider network is composed of multiple separate domains.).

5.3. Error Handling Procedures

The error handling rules as defined in [DRAFT OBJECTIVE FUNCTION] for the OF sub-object and the optional OF metric-bound sub-object SHALL be applied.

6. Security Considerations

Security for L1VPNs is covered in [RFC4847], [RFC5251] and [RFC5253]. In this document, the model follows a generic GMPLS VPN based on the L1VPN control plane model where CE addresses are completely distinct from the PE addresses.

The use of a private network assumes that entities outside the network cannot spoof or modify control plane communications between CE and PE. Furthermore, all entities in the private network are assumed to be trusted. Thus, no security mechanisms are required by the protocol exchanges described in this document.

However, an operator that is concerned about the security of their private control plane network may use the authentication and integrity functions available in RSVP-TE [RFC3473] or utilize IPsec ([RFC4301], [RFC4302], [RFC4835], [RFC5996], and [RFC6071]) for the point-to-point signaling between PE and CE. See [RFC5920] for a full discussion of the security options available for the GMPLS control plane.

7. IANA Considerations

TBD

8. References

8.1. Normative References

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

[RFC4202] Kompella, K., Rekhter, Y., "Routing Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4202, 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.

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

[RFC5251] Fedyk, D., Rekhter, Y., Editors "Layer 1 VPN Basic Mode", RFC 5251, July 2008.

[SRLG_info] Zhang, F., Gonzalez de Dios, O., Li, D., Margaria, C., Hartley, M., Ali, Z., "RSVP-TE Extensions for Collecting SRLG Information", draft-ietf-ccamp-rsvp-te-srlg-collect-04.txt, February 2014.

[DRAFT OBJECTIVE FUNCTION] Ali, Z., Swallow, G., Filsfils, C., Fang, L., Kumaki, K., Kunze, R., Ceccarelli, D., Zhang, X., "Resource ReserVation Protocol - Traffic Engineering (RSVP-TE) extension for signaling Objective Function and Metric Bound", draft-ali-ccamp-rc-objective-function-metric-bound-04.txt, October 2013.

[DRAFT DOMAIN SUBOBJECTS] Dhody, D., Palle, U., Kondreddy, V., Casellas, R., "Domain Subobjects for Resource ReserVation Protocol - Traffic Engineering (RSVP-TE)", draft-ietf-ccamp-rsvp-te-domain-subobjects-01.txt, January 2014.

[DRAFT PATH KEY] Zhang, X., Zhang, F., Gonzalez de Dios, O., Bryskin, I., Dhody, D., "Extensions to Resource ReSerVation Protocol-Traffic Engineering (RSVP-TE) to Support Route Exclusion Using Path Key Subobject", draft-zhang-ccamp-route-exclusion-pathkey-01.txt, February 2014

8.2. Informative References

[RFC4026] Andersson, L. and T. Madsen, "Provider Provisioned Virtual Private Network (VPN) Terminology", RFC 4026, March 2005.

[RFC6071] Frankel, S. and S. Krishnan, "IP Security (IPsec) and Internet Key Exchange (IKE) Document Roadmap", RFC 6071, February 2011.

[RFC3473] Berger, L. (editor), "Generalized MPLS Signaling - RSVP-TE Extensions", RFC 3473, January 2003.

[RFC4301] Kent, S. and K. Seo, "Security Architecture for the Internet Protocol", RFC 4301, December 2005.

[RFC4302] Kent, S., "IP Authentication Header", RFC 4302, December 2005.

[RFC5996] Kaufman, C., Hoffman, P., Nir, Y., and P. Eronen, "Internet Key Exchange Protocol Version 2 (IKEv2)", RFC 5996, September 2010.

[RFC4835] Manral, V., "Cryptographic Algorithm Implementation Requirements for Encapsulating Security Payload (ESP) and Authentication Header (AH)", RFC 4835, April 2007.

[RFC4847] Takeda, T., Editor "Framework and Requirements for Layer Virtual Private Networks", RFC 4847, April 2007.

[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.

[DRAFT TE METRIC RECORD] Ali, Z., Swallow, G., Filsfils, C., Hartley, M., Kumaki, K., Kunze, R., "Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) extension for recording TE Metric of a Label Switched Path", draft-ietf-ccamp-te-metric-recording-02.txt, July 2013.

[DRAFT OSPF TE METRIC EXT] Giacalone, S., Ward, D., Drake, J., Atlas, A., Previdi, S., "OSPF Traffic Engineering (TE) Metric Extensions", draft-ietf-ospf-te-metric-extensions-05.txt, December 2013.

Authors' Addresses

Don Fedyk
Hewlett-Packard
153 Tayor Street
Littleton, MA, 01460
Email: don.fedyk@hp.com

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

Lieven Levrau
Alcatel-Lucent

Email: Lieven.Levrau@alcatel-lucent.com

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

Fatai Zhang
Huawei Technologies
Email: zhangfatai@huawei.com

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

Xihua Fu
ZTE
Email: fu.xihua@zte.com.cn

Internet Engineering Task Force
Internet-Draft
Intended status: Standards Track
Expires: August 12, 2014

G.Galimberti, Ed.
Cisco
R.Kunze, Ed.
Deutsche Telekom
Kam Lam, Ed.
Alcatel-Lucent
D. Hiremagalur, Ed.
Juniper

February 8, 2014

An SNMP MIB extension to RFC3591 to manage optical interface parameters of DWDM applications

draft-galikunze-ccamp-g-698-2-snmp-mib-06

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 Optical parameters associated with Dense Wavelength Division Multiplexing (DWDM) interfaces. This is an extension of the RFC3591 to support the optical parameters described in ITU-T G.698.2. [ITU.G698.2]

The MIB module defined in this memo can be used for Optical Parameters monitoring and/or configuration of the endpoints of Black Links.

Status of This Memo

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

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

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

This Internet-Draft will expire on August 12, 2014.

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 Optical parameters associated with Wavelength Division Multiplexing (WDM) systems in accordance with the optical interface defined in G.698.2 [ITU.G698.2]

Black Link approach allows supporting an optical transmitter/receiver pair of one vendor to inject a DWDM channel and run it over an optical network composed of amplifiers, filters, add-drop multiplexers from a different vendor. From architectural point of view, the "Black Link" is a set of pre-configured/qualified network connections between the G.698.2 reference points S and R. The black links will be managed at the edges (i.e. 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 G.7710/Y.1701 [ITU-T G.7710] and and G.874.1 [ITU.G874.1].

The G.698.2 [ITU.G698.2] provides optical parameter values for physical layer interfaces of Dense Wavelength Division Multiplexing (DWDM) systems primarily intended for metro applications which include optical amplifiers. Applications are defined 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 specify the details of the optical link, e.g. the maximum fibre length, explicitly. The Recommendation currently includes unidirectional DWDM applications at 2.5 and 10 Gbit/s (with 100 GHz 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] plus few optical parameter not included in the application code definition.

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] G.798 [ITU.G798], G.874 [ITU.G874], parameters specified G.7710/Y.1701 [ITU-T G.7710] allows the different vendors and operator to retrieve, provision and exchange information related to Optical blak links in a standardized way.

This facilitates interworking in case of using optical interfaces from different vendors at the end of the link.

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

2. The Internet-Standard Management Framework

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

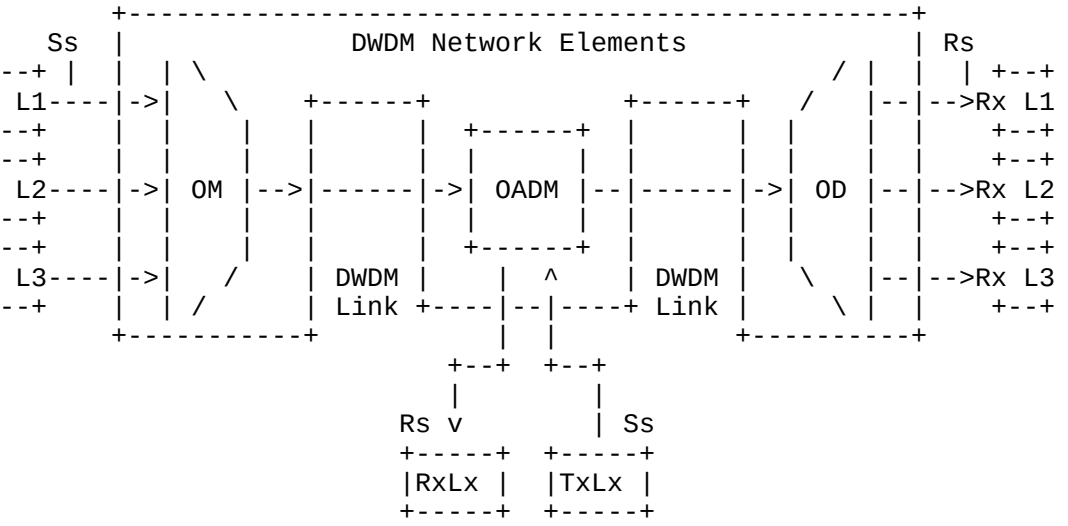
Managed objects are accessed via a virtual information store, termed the Management Information Base or MIB. MIB objects are generally accessed through the Simple Network Management Protocol (SNMP). Objects in the MIB are defined using the mechanisms defined in the Structure of Management Information (SMI). This memo specifies a MIB module that is compliant to the SMIv2, which is described in STD 58, RFC 2578 [RFC2578], STD 58, RFC 2579 [RFC2579] and STD 58, RFC 2580 [RFC2580].

3. Conventions

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119] 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 = reference point at the DWDM network element tributary output
Rs = reference point at the DWDM network element tributary input
Lx = Lambda x
OM = Optical Mux
OD = Optical Demux
OADM = Optical Add Drop Mux

from Fig. 5.1/G.698.2

Figure 1: Linear Black Link

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

4.1. Optical Parameters Description

The black links are managed at the edges, i.e. at the transmitters (Tx) and receivers (Rx) attached to the S and R reference points respectively. The parameters that could be managed at the black link

edges 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).

4.1.1. Rs-Ss Configuration

The Rs-Ss configuration table allows configuration of Wavelength, 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).

Wavelength Value (see G.694.1 Table 1):
: This parameter indicates the wavelength value that Ss and Rs will be set to work (in THz) se in particular Section 6/G.694.1 (G, S).

Number of Vendor Transceiver Class Supported
: This parameter indicates the number of Vendor Transceiver codes supported by this interface (G).

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
Wavelength Value	G,S	G.694.1 S.6
Vendor Transceiver Class	G	N.A.
Number of Vendor Transceiver Class Supported	G	N.A.
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:
This is the application code that is defined in G.698.2.

4.1.3. Table of Vendor Application Codes

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

Vendor Transceiver Class Identifier::
The Identifier for the vendor transceiver class.

Vendor Transceiver Class:
Other than specifying all the Transceiver parameter, it might be convenient for the vendors to summarize a set of parameters in a single proprietary parameter: the Class of transceiver. The Transceiver classification will be based on the Vendor Name and the main TX and RX parameters (i.e. Trunk Mode, Framing, Bit rate, Trunk Type, Channel Band, Channel Grid, Modulation Format, Channel Modulation Format, FEC Coding, Electrical Signal Framing at Tx, Minimum maximum Chromatic Dispersion (CD) at Rx, Maximum Polarization Mode Dispersion (PMD) at Rx, Maximum differential

group delay at Rx, Loopbacks, TDC, Pre-FEC BER, Q-factor, Q-margin,etc.). If this parameter is used, the MIB parameters specifying the Transceiver characteristics may not be significant and the vendor will be responsible to specify the Class contents and values. The Vendor can publish the parameters of its Classes or declare to be compatible with published Classes.(G) Optional for compliance. (not mentioned in G.698)

4.2. 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 Optical Channel (OCh) layer. There is a one to n relationship between the OPS and OCh layers.

EDITOR NOTE: Reason for changing from OChr to OCh: Work on revised G.872 in the SG15 December 2011 meeting agreed to remove OChr from the architecture and to update G.709 to account for this architectural change. The meeting also agreed to consent the revised text of G.872 and G.709 at the September 2012 SG15 meeting.

Figure 2 In the following figures, opticalChannel and opticalPhysicalSection are abbreviated as OCh and ops respectively.

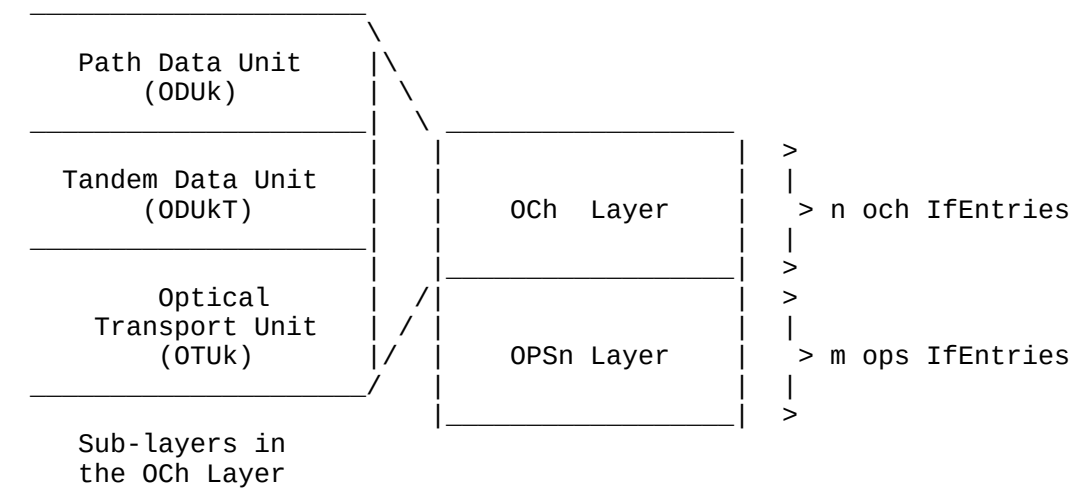


Figure 2: OTN Layers for OPS and OCh

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.

4.2.1. 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.

4.2.2. Use of ifTable for OCh Layer

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

4.2.3. 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 3

HigherLayer	LowerLayer
0	j
0	k
j	i
k	i
i	0

Figure 3: 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"
    ::= { 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 {
         optIfOChWavelengthn                          Unsigned32,
         optIfOChInterfaceVendorTransceiverClass      DisplayString,
         optIfOChNumberVendorClassesSupported         Unsigned32,
         optIfOChInterfaceApplicationCode             DisplayString,
         optIfOChNumberApplicationCodesSupported      Unsigned32,
         optIfOChOutputPower                          Integer32,
         optIfOChInputPower                           Integer32
     }

 optIfOChWavelengthn  OBJECT-TYPE
     SYNTAX  Unsigned32
     MAX-ACCESS  read-write
     STATUS  current
     DESCRIPTION
         " This parameter indicate minimum wavelength spectrum - n, in
          a definite wavelength Band (L, C and S) as represented in
          [RFC6205] by the formula -
          Wavelength (nm ) = 1471nm + n* optIfOChMiminumChannelSpacing
                                                (converted to nm)
          Eg - optIfOChMiminumChannelSpacing in nm
          'Wavelength (nm ) = 1471nm + n* 20nm  (20nm is the spacing
           for CWDM)'
         "
     ::= { optIfOChSsRsConfigEntry  1 }
```

```
 optIfOChInterfaceVendorTransceiverClass  OBJECT-TYPE
     SYNTAX DisplayString
     MAX-ACCESS  read-write
     STATUS  current
     DESCRIPTION
        "As defined in G.698
         Vendors can summarize a set of parameters in a
         single proprietary parameter: the Class of transceiver.  The
         Transceiver classification will be based on the Vendor Name
         and the main TX and RX parameters (i.e.  Trunk Mode, Framing,
         Bit rate, Trunk Type etc).
         This defines the tranceiver class that is/should be used by
         this interface. The optIfOChSrcVendorTranscieverClassTable
         has all the vendor classes supported by this interface."

     ::= { optIfOChSsRsConfigEntry  2 }

 optIfOChNumberVendorClassesSupported  OBJECT-TYPE
     SYNTAX Unsigned32
     MAX-ACCESS  read-only
     STATUS  current
     DESCRIPTION
        " Number of Vedor classes supported by this interface."
     ::= { optIfOChSsRsConfigEntry  3 }

 optIfOChInterfaceApplicationCode  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/should be used by this interface. The
          optIfOChSrcApplicationCodeTable  has all the application
          codes supported by this interface. "
     ::= { optIfOChSsRsConfigEntry  4 }

 optIfOChNumberApplicationCodesSupported  OBJECT-TYPE
     SYNTAX Unsigned32
     MAX-ACCESS  read-only
     STATUS  current
     DESCRIPTION
        " Number of Application codes supported by this interface."
     ::= { optIfOChSsRsConfigEntry  5 }

 optIfOChOutputPower  OBJECT-TYPE
     SYNTAX  Integer32
     UNITS   "0.01dbm"
```

```
    MAX-ACCESS  read-write
    STATUS  current
    DESCRIPTION
        " The output power for this interface in .01 dbm "
    ::= { optIfOChSsRsConfigEntry  6 }

optIfOChInputPower  OBJECT-TYPE
    SYNTAX  Integer32
    UNITS   "0.01dbm"
    MAX-ACCESS  read-only
    STATUS  current
    DESCRIPTION
        " The input power for this interface in .01 dbm "
    ::= { optIfOChSsRsConfigEntry  7 }

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

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

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

OptIfOChSrcApplicationCodeEntry ::=
    SEQUENCE {
        optIfOChApplicationCodeIdentifer           Integer32,
        optIfOChApplicationCode                    DisplayString
     }


optIfOChApplicationCodeIdentifer  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.
       "
     ::= { optIfOChSrcApplicationCodeEntry  1}

 optIfOChApplicationCode  OBJECT-TYPE
     SYNTAX DisplayString
     MAX-ACCESS  read-only
     STATUS  current
     DESCRIPTION
         " The application code supported by this interface DWDM
           link."
     ::= { optIfOChSrcApplicationCodeEntry  2}

 -- Table of Vendor Transceiver class supported by the interface
 --  OptIfOChSrcVendorTranscieverClassEntry

 optIfOChSrcVendorTranscieverClassTable  OBJECT-TYPE
     SYNTAX  SEQUENCE OF OptIfOChSrcVendorTranscieverClassEntry
     MAX-ACCESS  not-accessible
     STATUS  current
     DESCRIPTION
         "A  table of OCh Src (Ss) tranceiver classes supported by
          this interface."
     ::= { optIfOChSsRsGroup 3 }

 optIfOChSrcVendorTranscieverClassEntry OBJECT-TYPE
     SYNTAX      OptIfOChSrcVendorTranscieverClassEntry
     MAX-ACCESS  not-accessible
     STATUS  current
     DESCRIPTION
         "A conceptual row that contains the tranceiver classes
          supported by this interface."
     INDEX  { ifIndex, optIfOChTranscieverClassIdentifer  }
     ::= { optIfOChSrcVendorTranscieverClassTable 1 }

 OptIfOChSrcVendorTranscieverClassEntry ::=
     SEQUENCE {
         optIfOChTranscieverClassIdentifer           Integer32,
         optIfOChTranscieverClass                    DisplayString
      }

 optIfOChTranscieverClassIdentifer  OBJECT-TYPE
     SYNTAX  Integer32 (1..255)
     MAX-ACCESS  not-accessible
     STATUS  current
     DESCRIPTION
```

```
       " The number/identifer of the application code supported at this
         interface. The interface can support more than one
         application codes.
       "
     ::= { optIfOChSrcVendorTranscieverClassEntry  1}

 optIfOChTranscieverClass  OBJECT-TYPE
     SYNTAX DisplayString
     MAX-ACCESS  read-only
     STATUS  current
     DESCRIPTION
         " Vendor tranceiver class supported by this interface."
     ::= { optIfOChSrcVendorTranscieverClassEntry  2}


 -- Notifications

 -- Wavelength Change Notification
 optIfOChWavelengthChange NOTIFICATION-TYPE
     OBJECTS { optIfOChWavelengthn }
     STATUS  current
     DESCRIPTION
         "Notification of a change in the wavelength."
     ::= { 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
  Cisco
  32 HaMelacha St., (HaSharon Bldg)
  SOUTH NETANYA, HAMERKAZ 42504
  ISRAEL
  email ogerstel@cisco.com

12. References

12.1. Normative References

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

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

[RFC2578] McCloghrie, K., Ed., Perkins, D., Ed., and J. Schoenwaelder, Ed., "Structure of Management Information Version 2 (SMIv2)", STD 58, RFC 2578, April 1999.

[RFC2579] McCloghrie, K., Ed., Perkins, D., Ed., and J. Schoenwaelder, Ed., "Textual Conventions for SMIv2", STD 58, RFC 2579, April 1999.

[RFC2580] McCloghrie, K., Perkins, D., and J. Schoenwaelder, "Conformance Statements for SMIv2", STD 58, RFC 2580, April 1999.

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

[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.

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 Philips,12
20052 - Monza
Italy

Phone: +390392091462
Email: ggalimbe@cisco.com

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

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

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

CCAMP Working Group
Internet-Draft
Intended status: Informational
Expires: July 19, 2014

Rakesh Gandhi
Zafar Ali
Gabriele Maria Galimberti
Cisco Systems, Inc.
Xian Zhang
Huawei

January 15, 2014

RSVP-TE Signaling For GMPLS Restoration LSP
draft-gandhi-ccamp-gmpls-restoration-lsp-02

Abstract

In transport networks, there are requirements where Generalized Multi-Protocol Label Switching (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. This draft describes Resource reSerVation Protocol - Traffic Engineering (RSVP-TE) signaling for GMPLS end-to-end recovery when using restoration LSP.

Status of this Memo

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

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

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

Table of Contents

1. Introduction

Generalized Multi-Protocol Label Switching (GMPLS) extends MPLS to include support for different switching technologies [RFC3471] [RFC3473]. These switching technologies provide several protection schemes [RFC4426][RFC4427] (e.g., 1+1, 1:N and M:N). GMPLS RSVP-TE signaling has been extended to support various recovery schemes to establish Label Switched Paths (LSPs) [RFC4872][RFC4873], typically working LSP and protecting LSP. [RFC4427] Section 7 specifies various schemes for GMPLS restoration.

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. Consequently, revertive recovery mode is usually preferred by recovery schemes used in transport networks.

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

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

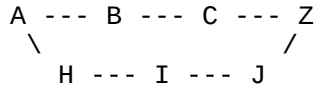


Figure 1: An example of 1+R recovery scheme

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

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

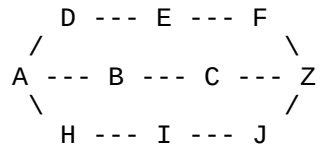


Figure 2: An example of 1+1+R recovery scheme

[RFC4872] Section 14 defines PROTECTION object for GMPLS recovery signaling. The PROTECTION object is used to identify primary and secondary LSPs using S bit and protecting and working LSPs using P bit. [RFC4872] and [RFC6689] define the usage of ASSOCIATION object for further associating GMPLS working and protecting LSPs. However, these existing methods do not specify how to identify restoration LSP when working/protecting LSPs are not torn down.

This draft describes procedures for identifying the restoration LSP for GMPLS end-to-end recovery where working and protecting LSP resources are kept reserved after the failure.

2. Conventions used in this document

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

3. Restoration LSP Signaling

3.1. Signaling Procedure

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

The procedure for signaling the PROTECTION object is specified in [RFC4872] and is changed by this document. Restoration LSP being used as a working LSP is signaled with P bit cleared and as a protecting LSP is signaled with P bit set.

When using a GMPLS recovery mode, where the restoration LSP is promoted to be the new working LSP, restoration LSP RSVP Path message MUST be refreshed by using the ASSOCIATION_OBJECT.LSP_ID to contain the LSP ID of the protecting LSP if known or LSP ID of itself if protecting LSP is not known as defined in [RFC6689].

When using a GMPLS recovery mode, where the restoration LSP is promoted to be the new protecting LSP, restoration LSP RSVP Path message MUST be refreshed by using the ASSOCIATIN_OBJECT.LSP_ID to contain the LSP ID of the working LSP if known or LSP ID of itself if working LSP is not known as defined in [RFC6689].

4. IANA Considerations

This document makes no request for IANA action.

5. Security Considerations

This document introduces no additional security considerations. For a general discussion on MPLS and GMPLS related security issues, see the MPLS/GMPLS security framework [RFC5920]. In addition, the considerations specified in [RFC4872] will apply.

6. Acknowledgement

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

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.

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

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

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

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

[RFC4872] Lang, J., Rekhter, Y., and D. Papadimitriou, "RSVP-TE Extensions in Support of End-to-End Generalized Multi-Protocol Label Switching (GMPLS) Recovery", RFC 4872, May 2007.

[RFC4873] Berger, L., Bryskin, I., Papadimitriou, D., and A. Farrel,

"GMPLS Segment Recovery", RFC 4873, May 2007.

[RFC6689] Berger, L, "Usage of the RSVP ASSOCIATION Object", RFC 6689, July 2012.

7.2. Informative References

[RFC4426] Lang, J., Rajagopalan B., and D.Papadimitriou, Editors, "Generalized Multiprotocol Label Switching (GMPLS) Recovery Functional Specification", RFC 4426, March 2006.

[RFC4427] Mannie, E., Ed. and D. Papadimitriou, Ed., "Recovery (Protection and Restoration) Terminology for Generalized Multi-Protocol Label Switching, RFC 4427, March 2006.

[RFC5920] Fang, L., "Security Framework for MPLS and GMPLS Networks", RFC 5920, July 2010.

Authors' Addresses

Rakesh Gandhi
Cisco Systems, Inc.

Email: rgandhi@cisco.com

Zafar Ali
Cisco Systems, Inc.

Email: zali@cisco.com

Gabriele Maria Galimberti
Cisco Systems, Inc.

Email: ggalimbe@cisco.com

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

Email: zhang.xian@huawei.com

CCAMP Working Group
Internet Draft
Intended status: Standards Track
Expires: August 13, 2014

Matt Hartley
Zafar Ali
Cisco Systems
O. Gonzalez de Dios
Telefonica Global CTO
C. Margaria
Coriant R&D GmbH
Xian Zhang
Huawei
February 14, 2014

RSVP-TE Extensions for RRO Editing
draft-hartley-ccamp-rro-editing-01.txt

Status of this Memo

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

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

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

This Internet-Draft will expire on August 13, 2014.

Copyright Notice

Abstract

This document provides extensions for the Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) to allow the communication of changes made by a node to the information provided by other nodes in a ROUTE_RECORD Object (RRO) in Path and Resv messages, or to indicate that it has itself provided incomplete information.

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].

Table of Contents

1. Introduction

The signaling process of a Label-Switched Path (LSP) may require gathering information of the actual path traversed by the LSP. The procedure for collecting this information includes the hop-by-hop construction of a Record Route Object (RRO) in the Path and Resv messages, containing information about the path traversed by the LSP ([RFC-3209], [RFC-3473], [RFC-4873], [RFC-5420], [RFC-5553], [DRAFT-SRLG], [DRAFT-METRIC]). There are cases, described in this document, in which one or more nodes on the path of an LSP may require that the data contained in the RRO in the Path and/or Resv be removed or

summarized. However, it is important for the ingress or egress nodes to know which RRO subobjects have been edited by intermediate nodes. This document addresses this requirement.

1.1. Use Cases

1.1.1. Overlay and Multi-domain Networks

In the GMPLS overlay model there is a client-server relationship [RFC4208]. The GMPLS User-Network Interface (UNI) is the reference point where policies may be applied. In this case, policy at the server network boundary may require that some or all information related to the server network be edited, summarized or removed when communicating with the client nodes. Similar policy requirements exist for inter-domain LSPs and in E-NNI use case.

1.1.2. RRO Reduction

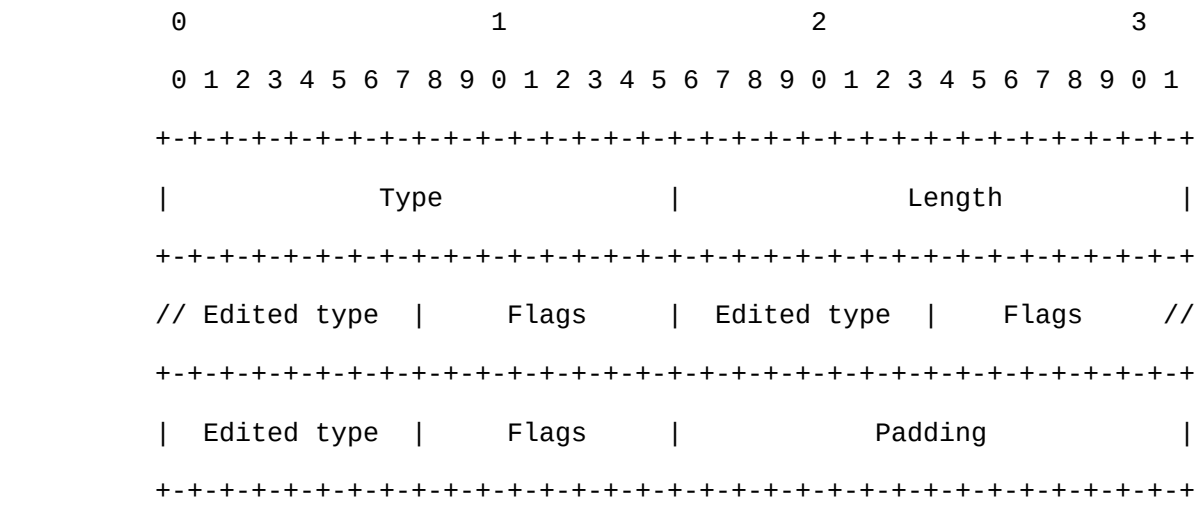
If an LSP with many hops is signaled and a great deal of information is collected at each hop, it is possible that the RRO may grow to the point where it reaches its maximum possible size or is too large to fit in the Path or Resv message. In this case a node may summarize or remove information from the RRO to reduce its size, rather than dropping it entirely as specified by [RFC-3209].

2. RSVP-TE Signaling Extensions

This section describes the signaling extensions required to address the aforementioned requirements. Specifically, the requirements are addressed by defining a new LSP_ATTRIBUTES TLV that can be used to reference what information in RRO has been edited.

2.1. RRO-edit LSP_ATTRIBUTES TLV

A new LSP_ATTRIBUTES TLV is defined in order to indicate that RRO sub-object(s) of a specified type have been edited.



The sub-object fields are defined as follows:

Type (2 bytes): The sub-object type, to be assigned by IANA (suggested value: 3).

Length (2 bytes): the total length of the TLV, in bytes. It MUST be a multiple of 4, and at least 8.

The following fields are repeated for each edited type:

Edited type (1 byte): the type of the RRO sub-object to which the immediately following flags in this sub-object apply.

Flags (1 byte): the flags that apply to the preceding Edited Type, numbered from 0 as the most significant bit in the field. Three flags are defined by this document:

- Bit position 0: P (Partial) bit. When set, this bit indicates that the data contained in RRO sub-objects of the immediately preceding type is incomplete. This may be because some information was withheld by a node (i.e. never placed into the RRO) or because information provided by one node has been removed by another.
- Bit position 1: S (Summary) bit. When set, this bit indicates that the data contained in the specified RRO sub-object has been summarized.

. Bit position 2: R (Removed) bit. When set, this bit indicates that the specified RRO sub-object has been removed entirely.

The remaining bits of the Flags field are undefined. They MUST be set to 0 on transmission and MUST be ignored when received.

Padding: This field is present only if an odd number of edited type/flags pairs is present in the TLV. It is used to ensure the TLV length is always a multiple of 4 bytes.

2.2. RRO-edit TLV Processing Rules

The processing rules in this section apply to the processing of both Path and Resv RROs.

The RRO-edit TLV provides information on the changes made to RRO sub-objects. It MAY be present in the LSP_ATTRIBUTES object in a Path or Resv message. It MUST NOT be added to the LSP_REQUIRED_ATTRIBUTES object.

The LSP_ATTRIBUTES object SHOULD contain no more than one RRO-edit TLV. If a received LSP_ATTRIBUTES object contains multiple RRO-edit TLVs, the second and subsequent RRO-edit TLVs MUST be ignored.

The RRO-edit TLVs contains pairs of RRO subobject types and flags relating to that type. Any RRO subobject type MAY be present in the RRO-edit TLV. Each RRO subobject type SHOULD appear only once; if a RRO subobject type occurs more than once then only the first occurrence is meaningful, and subsequent occurrences MUST be ignored.

Normal RRO processing involves a node simply adding data related to the local hop to the RRO received from the prior node to RRO, and placing the new RRO in the message to be transmitted. In this case the transmitted RRO contains all data that was present in the received RRO and no further processing is required.

If a node edits the data in the received RRO such that the same data is not present in the transmitted RRO, or if it is supplying incomplete or summarized data on its own behalf, then the following rules apply at the processing node.

. The node MAY choose not to add or amend the RRO-edit TLV if its local policy prevents this.
. For each RRO subobject type that the processing node has edited, a RRO-edit type/flags pair SHOULD be added to the RRO-edit TLV if it does not already exist. If a RRO-edit type/flags

pair for the edited subobject type is already present in the RRO-edit TLV, the node SHOULD set additional flags in that subobject if appropriate.

- The node SHOULD set the appropriate P/S/R bits for the RRO subobject in the RRO-edit TLV to indicate the changes that have been made to RRO subobjects of that type.
- A node SHOULD NOT insert a RRO-edit type/flags pair with all flags set to zero.
- A node SHOULD NOT unset any P/S/R bit that is set in a received RRO-edit TLV.
- A node SHOULD NOT remove any RRO-edit type/flags pair from the RRO-edit TLV.
- A RRO-edit TLV with no RRO-edit type/flags pairs (i.e. one of length 4) is considered invalid. It MUST be ignored on receipt and MUST NOT be added to a LSP_ATTRIBUTES object.
- Unassigned flag bits are considered reserved. They MUST be set to zero.
- The RRO-edit TLV length MUST be a multiple of 4. If an odd number of RRO-subobject/flags pairs is present on transmission, a 16-bit Padding field MUST be added to the TLV. If an even number of RRO-subobject/flags pairs is present on transmission, the Padding MUST NOT be added. If present, the Padding bytes MUST be set to zero on transmission and MUST be ignored on receipt.
- Any set flag whose meaning is either unassigned or not understood SHOULD be ignored, and MUST be included unchanged in the transmitted RRO-edit TLV.
- A RRO-edit type/flags pair with an unknown RRO subobject type SHOULD be ignored and MUST passed unchanged in the transmitted RRO-edit TLV.

3. Security Considerations

There are no new security considerations introduced by this document.

4. IANA Considerations

4.1. LSP_ATTRIBUTES Object

IANA has made the following assignments in the "Attributes TLV Space" section of the "RSVP-TE PARAMETERS" registry located at http://www.iana.org/assignments/rsvp-te-parameters/rsvp-te-parameters.xml.

This document introduces a new LSP_ATTRIBUTES sub-object:

Type	Name	Reference
TBD (suggested value: 3)	RRO-edited TLV	This I-D

This TLV is allowed on LSP_ATTRIBUTES, and not allowed on LSP_REQUIRED_ATTRIBUTES.

5. Acknowledgments

The authors would like to thank Lou Berger for suggesting the core idea described in this draft. The authors would also like to thank George Swallow for his input.

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.

[RFC5420] Farrel, A., Ed., 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.

6.2. Informative References

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

[RFC4873] Berger, L., Bryskin, I., Papadimitriou, D., and A. Farrel, "GMPLS Segment Recovery", RFC 4873, May 2007.

[RFC5553] Farrel, A., Ed., Bradford, R., Vasseur, JP., "Resource Reservation Protocol (RSVP) Extensions for Path Key Support", RFC 5553, May 2009.

[DRAFT-SRLG] Zhang, F., Li, D., Gonzalez de Dios, O., Margaria, C., Hartley, M., "RSVP-TE Extensions for Collecting SRLG Information", draft-ietf-ccamp-rsvp-te-srlg-collect-03, October 2013.

[DRAFT-METRIC] Ali, Z., Swallow, G., Filsfils, C., Hartley, M., Kumaki, K., Kunze, R., "Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) extension for recording TE Metric of a Label Switched Path", draft-ietf-ccamp-te-metric-recording-02, July 2013.

Author's Addresses

Matt Hartley
Cisco Systems
Email: mhartley@cisco.com

Zafar Ali
Cisco Systems
Email: zali@cisco.com

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

Cyril Margaria
Coriant R&D GmBH
Email: cyril.margaria@gmail.com

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

Disclaimer of Validity

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: August 18, 2014

O. Gonzalez de Dios, Ed.
Telefonica I+D
R. Casellas, Ed.
CTTC
F. Zhang
Huawei
X. Fu
ZTE
D. Ceccarelli
Ericsson
I. Hussain
Infinera

February 14, 2014

Framework and Requirements for GMPLS based control of Flexi-grid DWDM networks
draft-ietf-ccamp-flexi-grid-fwk-01

Abstract

This document defines a framework and the associated control plane requirements for the GMPLS based control of flexi-grid DWDM networks. To allow efficient allocation of optical spectral bandwidth for high bit-rate systems, the International Telecommunication Union Telecommunication Standardization Sector (ITU-T) has extended the recommendations [G.694.1] and [G.872] to include the concept of flexible grid. A new DWDM grid has been developed within the ITU-T Study Group 15 by defining a set of nominal central frequencies, channel spacings and the concept of "frequency slot". In such environment, a data plane connection is switched based on allocated, variable-sized frequency ranges within the optical spectrum.

This Internet-Draft will expire on August 18, 2014.

Table of Contents

1. Requirements Language

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

2. Introduction

The term "Flexible grid" (flexi-grid for short) as defined by the International Telecommunication Union Telecommunication Standardization Sector (ITU-T) Study Group 15 in the latest version of [G.694.1], refers to the updated set of nominal central frequencies (a frequency grid), channel spacing and optical spectrum management/allocation considerations that have been defined in order to allow an efficient and flexible allocation and configuration of optical spectral bandwidth for high bit-rate systems.

A key concept of flexi-grid is the "frequency slot"; a variable-sized optical frequency range that can be allocated to a data connection. As detailed later in the document, a frequency slot is characterized by its nominal central frequency and its slot width which, as per [G.694.1], is constrained to be a multiple of a given slot width granularity.

Compared to a traditional fixed grid network, which uses fixed size optical spectrum frequency ranges or "frequency slots" with typical channel separations of 50 GHz, a flexible grid network can select its media channels with a more flexible choice of slot widths, allocating as much optical spectrum as required, allowing high bit rate signals (e.g., 400G, 1T or higher) that do not fit in the fixed grid.

From a networking perspective, a flexible grid network is assumed to be a layered network [G.872][G.800] in which the media layer is the server layer and the optical signal layer is the client layer. In the media layer, switching is based on a frequency slot, and the size of a media channel is given by the properties of the associated

frequency slot. In this layered network, the media channel transports an Optical Tributary Signal.

A Wavelength Switched Optical Network (WSON), addressed in [RFC6163], is a term commonly used to refer to the application/deployment of a Generalized Multi-Protocol Label Switching (GMPLS)-based control plane for the control (provisioning/recovery, etc) of a fixed grid WDM network in which media (spectrum) and signal are jointly considered

This document defines the framework for a GMPLS-based control of flexi-grid enabled DWDM networks (in the scope defined by ITU-T layered Optical Transport Networks [G.872]), as well as a set of associated control plane requirements. An important design consideration relates to the decoupling of the management of the optical spectrum resource and the client signals to be transported.

3. Acronyms

EFS: Effective Frequency Slot

FS: Frequency Slot

NCF: Nominal Central Frequency

OCh: Optical Channel

OCh-P: Optical Channel Payload

OTS: Optical Tributary Signal

OCC: Optical Channel Carrier

SWG: Slot Width Granularity

4. Flexi-grid Networks

4.1. Flexi-grid in the context of OTN

[G.872] describes from a network level the functional architecture of Optical Transport Networks (OTN). The OTN is decomposed into independent layer networks with client/layer relationships among them. A simplified view of the OTN layers is shown in Figure 1.

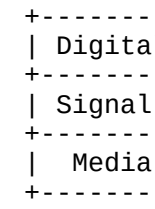


Figure 1: Generic OTN overview

In the OTN layering context, the media layer is the server layer of the optical signal layer. The optical signal is guided to its destination by the media layer by means of a network media channel. In the media layer, switching is based on a frequency slot, and the size of a media channel is given by the properties of the associated frequency slot.

In this scope, this document uses the term flexi-grid enabled DWDM network to refer to a network in which switching is based on frequency slots defined using the flexible grid, and covers mainly the Media Layer as well as the required adaptations from the Signal layer. The present document is thus focused on the control and management of the media layer.

4.2. Terminology

This section presents the definition of the terms used in flexi-grid networks. These terms are included in the ITU-T recommendations [G.694.1], [G.872]), [G.870], [G.8080] and [G.959.1-2013].

Where appropriate, this documents also uses terminology and lexicography from [RFC4397].

4.2.1. Frequency Slots

This subsection is focused on the frequency slot related terms.

- Frequency Slot [G.694.1]: The frequency range allocated to a slot within the flexible grid and unavailable to other slots. A frequency slot is defined by its nominal central frequency and its slot width.

Nominal Central Frequency: each of the allowed frequencies as per the definition of flexible DWDM grid in [G.694.1]. The set of nominal central frequencies can be built using the following expression f = 193.1 THz + n x 0.00625 THz, where 193.1 THz is ITU-T ''anchor frequency'' for transmission over the C band, n is a positive or negative integer including 0.

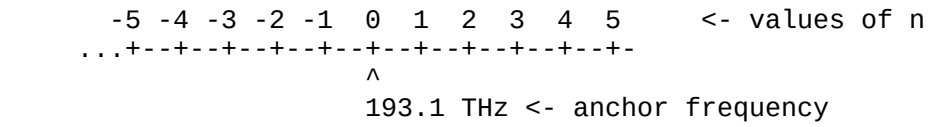


Figure 2: Anchor frequency and set of nominal central frequencies

Nominal Central Frequency Granularity: It is the spacing between allowed nominal central frequencies and it is set to 6.25 GHz (note: sometimes referred to as 0.00625 THz).

Slot Width Granularity: 12.5 GHz, as defined in [G.694.1].

Slot Width: The slot width determines the "amount" of optical spectrum regardless of its actual "position" in the frequency axis. A slot width is constrained to be m x SWG (that is, m x 12.5 GHz), where m is an integer greater than or equal to 1.

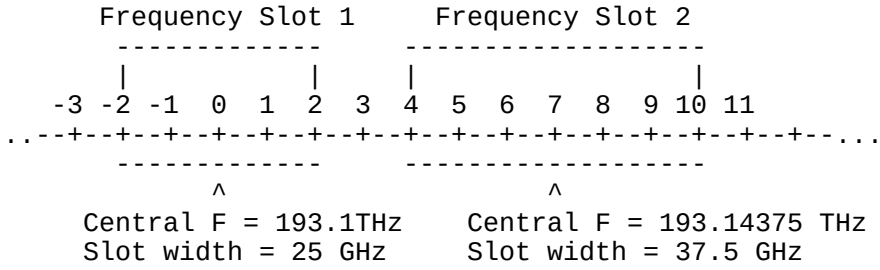


Figure 3: Example Frequency slots

o The symbol '+' represents the allowed nominal central frequencies, the '--' represents the nominal central frequency granularity, and the '^' represents the slot nominal central frequency. The number on the top of the '+' symbol represents the 'n' in the frequency calculation formula. The nominal central frequency is 193.1 THz when n equals zero.

Effective Frequency Slot: the effective frequency slot of a media channel is the common part of the frequency slots along the media channel through a particular path through the optical network. It is a logical construct derived from the (intersection of) frequency slots allocated to each device in the path. The effective frequency slot is an attribute of a media channel and, being a frequency slot, it is described by its nominal central frequency and slot width, according to the already described rules.

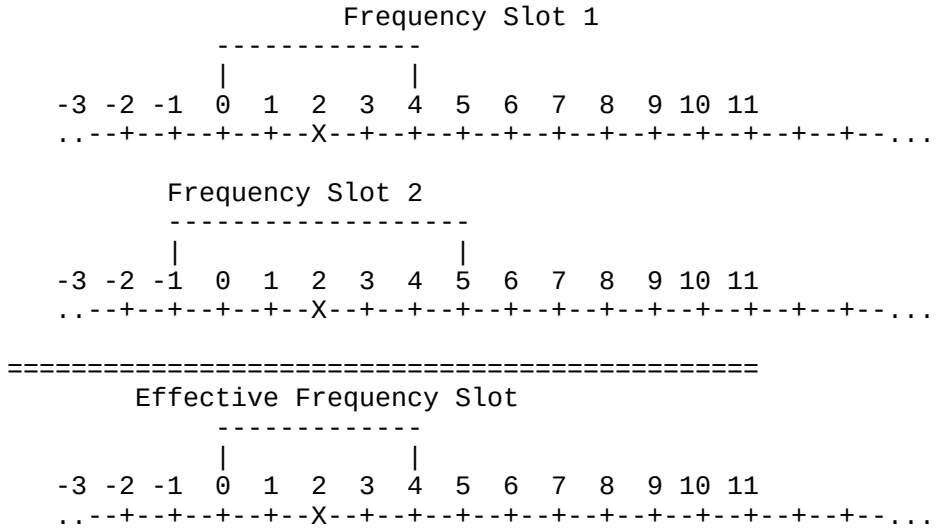


Figure 4: Effective Frequency Slot

4.2.2. Media Channels

Media Channel: a media association that represents both the topology (i.e., path through the media) and the resource (frequency slot) that it occupies. As a topological construct, it represents a (effective) frequency slot supported by a concatenation of media elements (fibers, amplifiers, filters, switching matrices...). This term is used to identify the end-to-end physical layer entity with its corresponding (one or more) frequency slots local at each link filters.

Network Media Channel: It is a media channel that transports an Optical Tributary Signal [Editor's note: this definition goes beyond current G.870 definition, which is still tightened to a particular case of OTS, the OCh-P]

4.2.3. Media Layer Elements

Media Element: a media element only directs the optical signal or affects the properties of an optical signal, it does not modify the properties of the information that has been modulated to produce the optical signal [G.870]. Examples of media elements include fibers, amplifiers, filters and switching matrices.

Media Channel Matrixes: the media channel matrix provides flexible connectivity for the media channels. That is, it represents a point

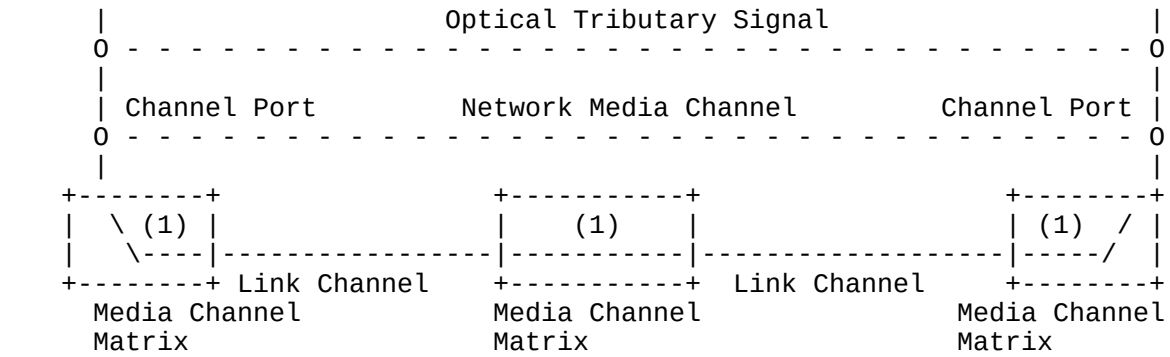
of flexibility where relationships between the media ports at the edge of a media channel matrix may be created and broken. The relationship between these ports is called a matrix channel. (Network) Media Channels are switched in a Media Channel Matrix.

4.2.4. Optical Tributary Signals

Optical Tributary Signal [G.959.1-2013]: The optical signal that is placed within a network media channel for transport across the optical network. This may consist of a single modulated optical carrier or a group of modulated optical carriers or subcarriers. One particular example of Optical Tributary Signal is an Optical Channel Payload (OCh-P) [G.872].

4.3. Flexi-grid layered network model

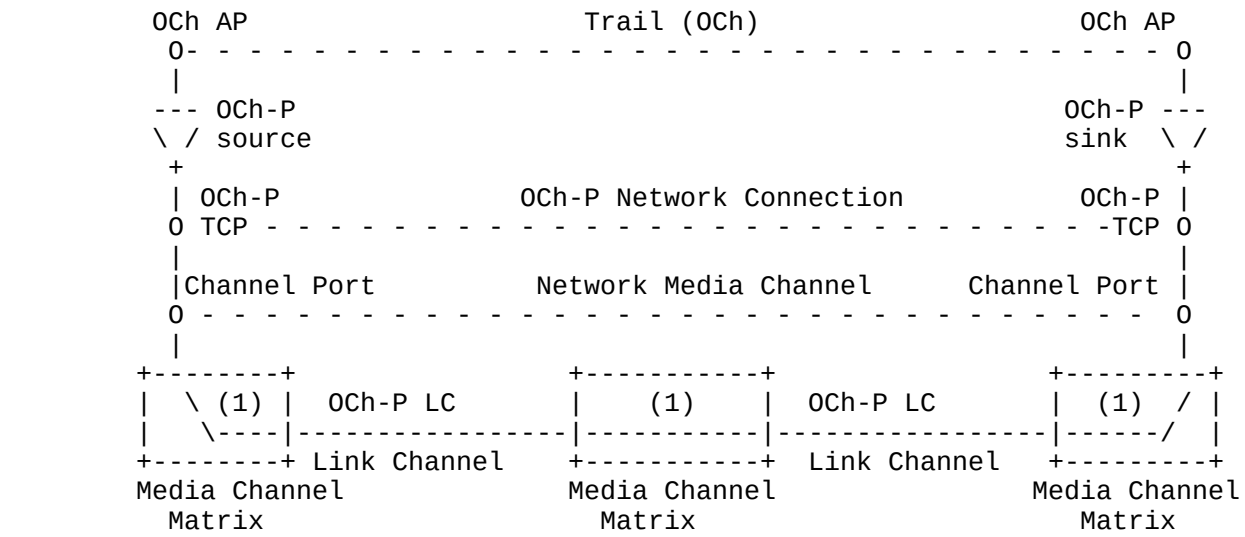
In the OTN layered network, the network media channel transports a single Optical Tributary Signal (see Figure 5)



(1) - Matrix Channel

Figure 5: Simplified Layered Network Model

A particular example of Optical Tributary Signal is the OCh-P. Figure Figure 6 shows the example of the layered network model particularized for the OCH-P case, as defined in G.805.



(1) - Matrix Channel

Figure 6: Layered Network Model according to G.805

By definition a network media channel only supports a single Optical Tributary signal. How several Optical Tributary signals are bound together is out of the scope of the present document and is a matter of the signal layer.

4.3.1. Hierarchy in the Media Layer

In summary, the concept of frequency slot is a logical abstraction that represents a frequency range while the media layer represents the underlying media support. Media Channels are media associations, characterized by their (effective) frequency slot, respectively; and media channels are switched in media channel matrixes. From the control and management perspective, a media channel can be logically splited in other media channels.

In Figure 7 , a Media Channel has been configured and dimensioned to support two network media channels, each of them carrying one optical tributary signal.

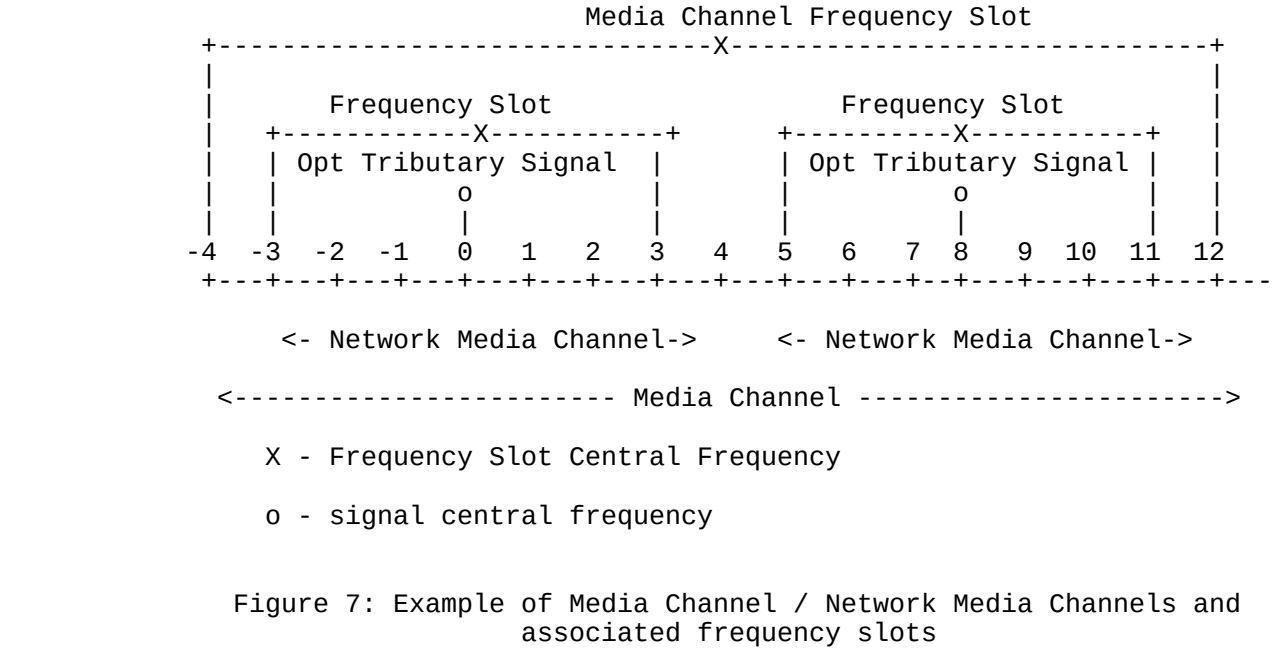


Figure 7: Example of Media Channel / Network Media Channels and associated frequency slots

4.3.2. DWDM flexi-grid enabled network element models

Similar to fixed grid networks, a flexible grid network is also constructed from subsystems that include Wavelength Division Multiplexing (WDM) links, tunable transmitters and receivers, i.e, media elements including media layer switching elements (media matrices), as well as electro-optical network elements, all of them with flexible grid characteristics.

As stated in [G.694.1] the flexible DWDM grid defined in Clause 7 has a nominal central frequency granularity of 6.25 GHz and a slot width granularity of 12.5 GHz. However, devices or applications that make use of the flexible grid may not be capable of supporting every possible slot width or position. In other words, applications may be defined where only a subset of the possible slot widths and positions are required to be supported. For example, an application could be defined where the nominal central frequency granularity is 12.5 GHz (by only requiring values of n that are even) and that only requires slot widths as a multiple of 25 GHz (by only requiring values of m that are even).

5. GMPLS applicability

The goal of this section is to provide an insight of the application of GMPLS to control flexi-grid networks, while specific requirements are covered in the next section. The present framework is aimed at controlling the media layer within the Optical Transport Network (OTN) hierarchy and the required adaptations of the signal layer. This document also defines the term SSON (Spectrum-Switched Optical Network) to refer to a Flexi-grid enabled DWDM network that is controlled by a GMPLS/PCE control plane.

This section provides a mapping of the ITU-T G.872 architectural aspects to GMPLS/Control plane terms, and considers the relationship between the architectural concept/construct of media channel and its control plane representations (e.g. as a TE link).

5.1. General considerations

The GMPLS control of the media layer deals with the establishment of media channels, which are switched in media channel matrixes. GMPLS labels locally represent the media channel and its associated frequency slot. Network media channels are considered a particular case of media channels when the end points are transceivers (that is, source and destination of an Optical Tributary Signal)

5.2. Considerations on TE Links

From a theoretical / abstract point of view, a fiber can be modeled has having a frequency slot that ranges from (-inf, +inf). This representation helps understand the relationship between frequency slots / ranges.

The frequency slot is a local concept that applies locally to a component / element. When applied to a media channel, we are referring to its effective frequency slot as defined in [G.872].

The association of a filter, a fiber and a filter is a media channel in its most basic form, which from the control plane perspective may modeled as a (physical) TE-link with a contiguous optical spectrum at start of day. A means to represent this is that the portion of spectrum available at time t0 depends on which filters are placed at the ends of the fiber and how they have been configured. Once filters are placed we have the one hop media channel. In practical terms, associating a fiber with the terminating filters determines the usable optical spectrum.

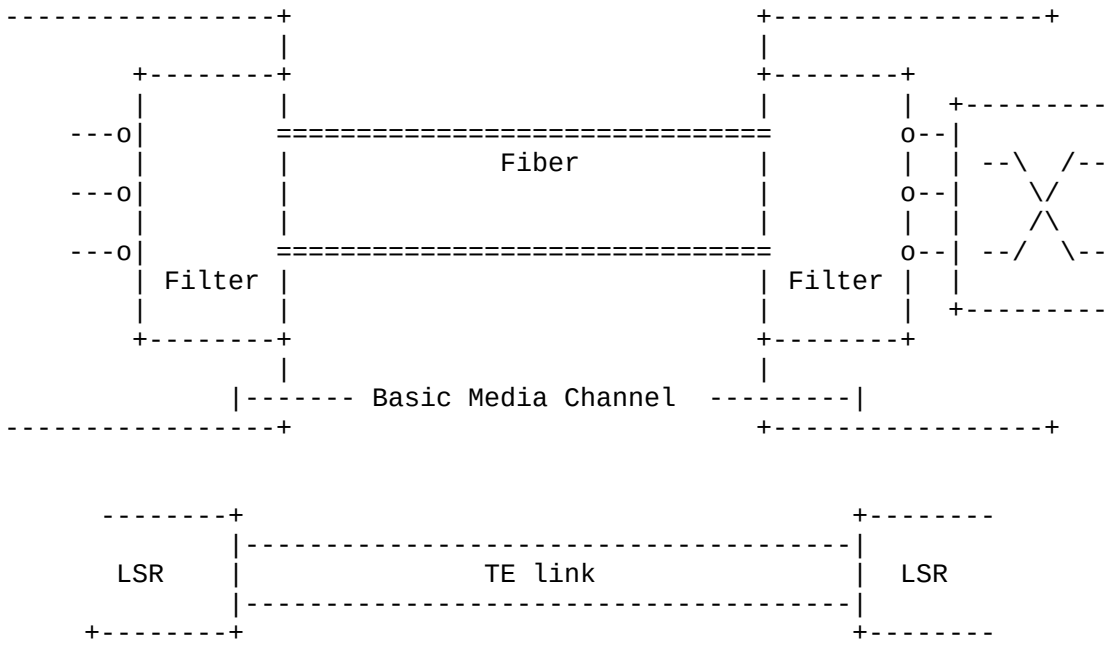


Figure 8: (Basic) Media channel and TE link

Additionally, when a cross-connect for a specific frequency slot is considered, the underlying media support is still a media channel, augmented, so to speak, with a bigger association of media elements and a resulting effective slot. When this media channel is the result of the association of basic media channels and media layer matrix cross-connects, this architectural construct can be represented as / corresponds to a Label Switched Path (LSP) from a control plane perspective. In other words, It is possible to "concatenate" several media channels (e.g. Patch on intermediate nodes) to create a single media channel.

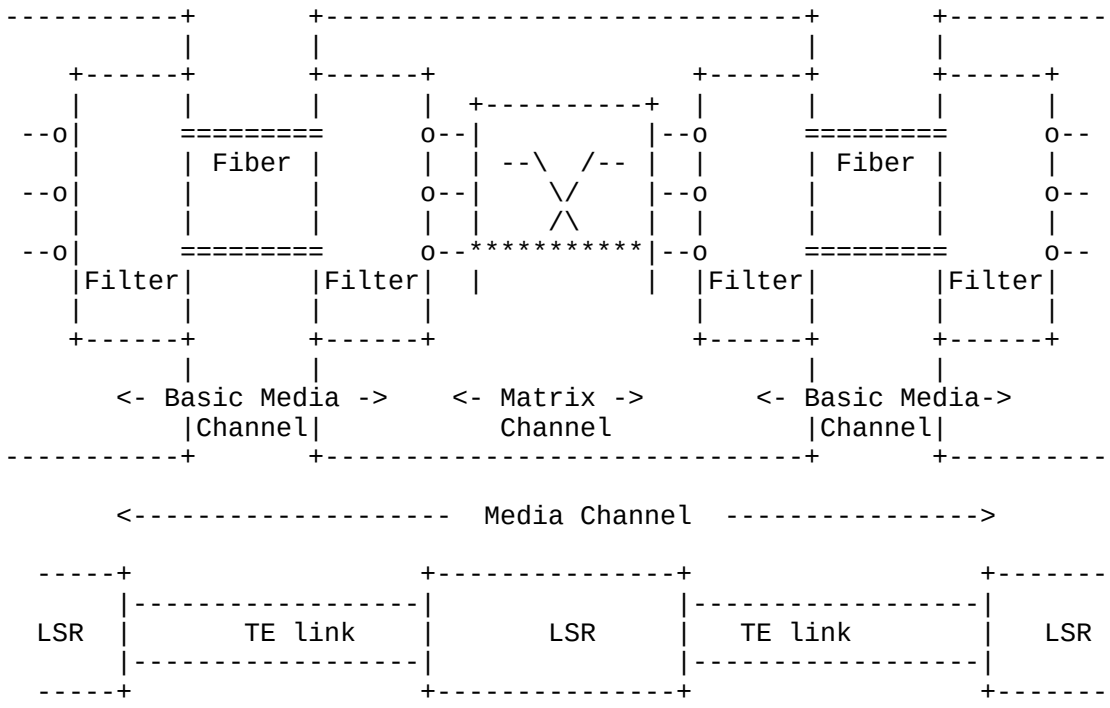


Figure 9: Extended Media Channel

Additionally, if appropriate, it can also be represented as a TE link or Forwarding Adjacency (FA), augmenting the control plane network model.

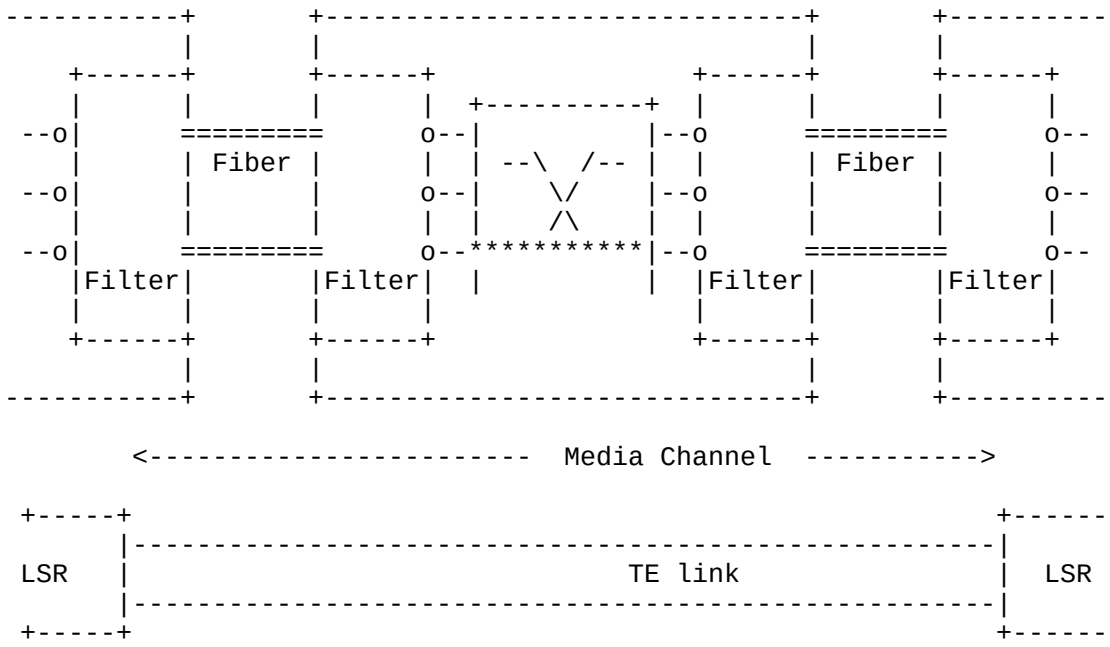


Figure 10: Extended Media Channel / TE Link / FA

5.3. Considerations on Labeled Switched Path (LSP) in Flexi-grid

The flexi-grid LSP is seen as a control plane representation of a media channel. Since network media channels are media channels, an LSP may also be the control plane representation of a network media channel, in a particular context. From a control plane perspective, the main difference (regardless of the actual effective frequency slot which may be dimensioned arbitrarily) is that the LSP that represents a network media channel also includes the endpoints (transceivers) , including the cross-connects at the ingress / egress nodes. The ports towards the client can still be represented as interfaces from the control plane perspective.

Figure 11 describes an LSP routed along 3 nodes. The LSP is terminated before the optical matrix of the ingress and egress nodes and can represent a Media Channel. This case does NOT (and cannot) represent a network media channel as it does not include (and cannot include) the transceivers.

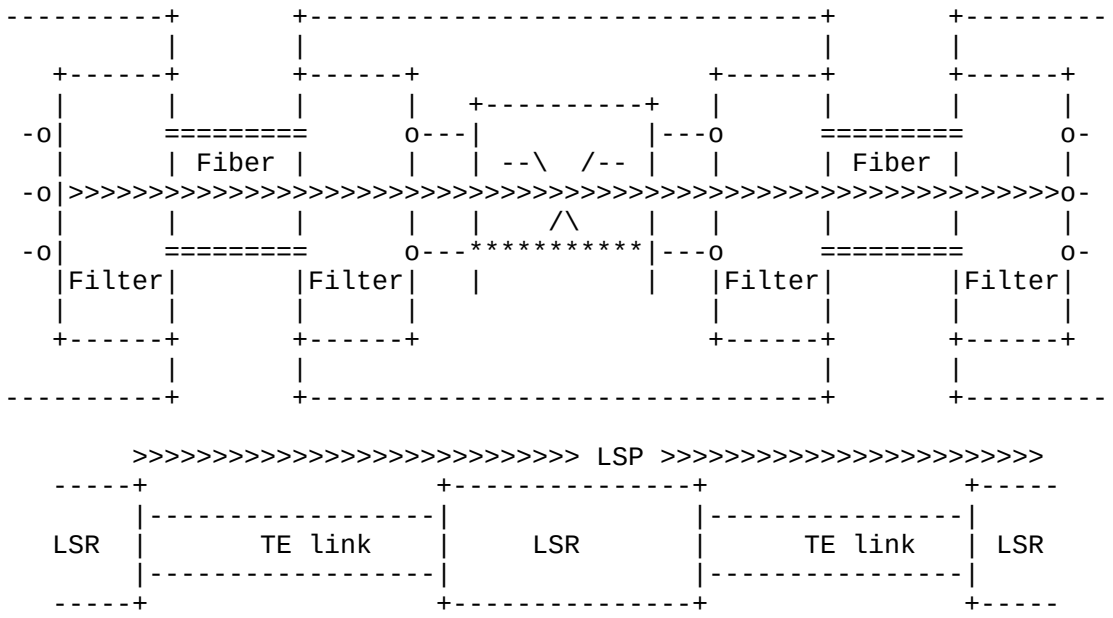


Figure 11: Flex-grid LSP representing a media channel that starts at the filter of the outgoing interface of the ingress LSR and ends at the filter of the incoming interface of the egress LSR

In Figure 12 a Network Media Channel is represented as terminated at the DWDM side of the transponder, this is commonly named as OCh-trail connection.

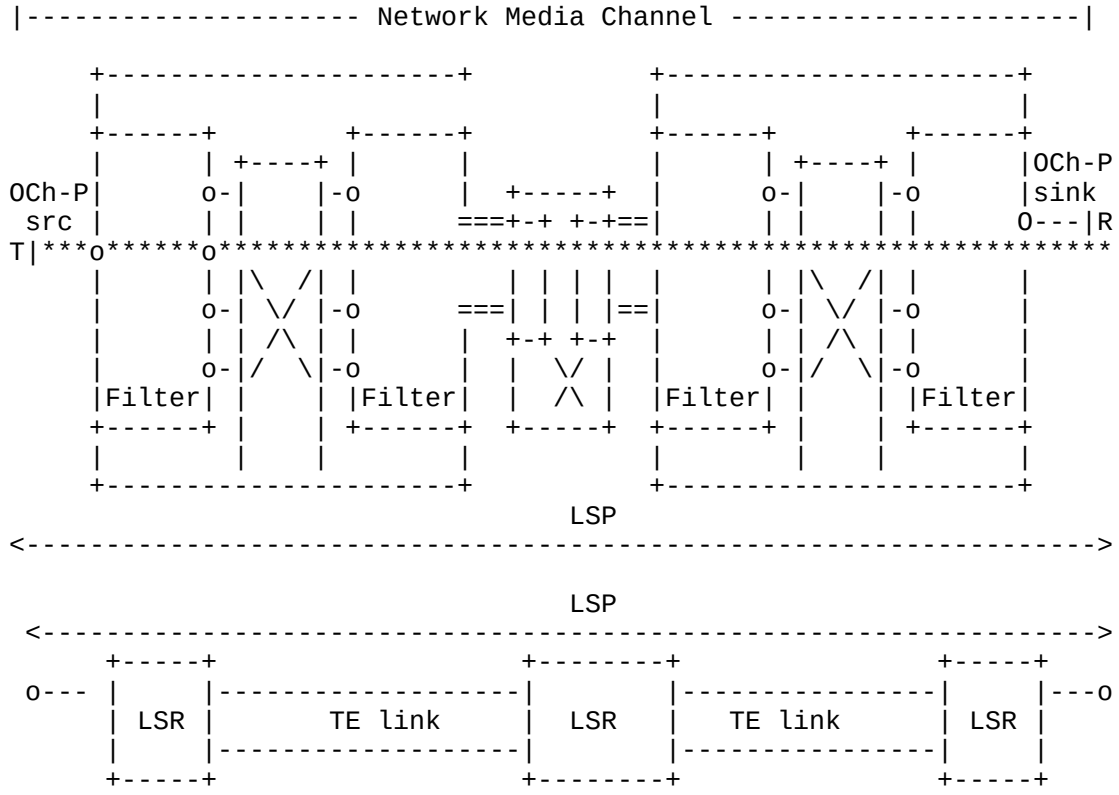


Figure 12: LSP representing a network media channel (OCh-Trail)

In a third case, a Network Media Channel terminated on the Filter ports of the Ingress and Egress nodes. This is named in G.872 as OCh-NC (we need to discuss the implications, if any, once modeled at the control plane level of models B and C).

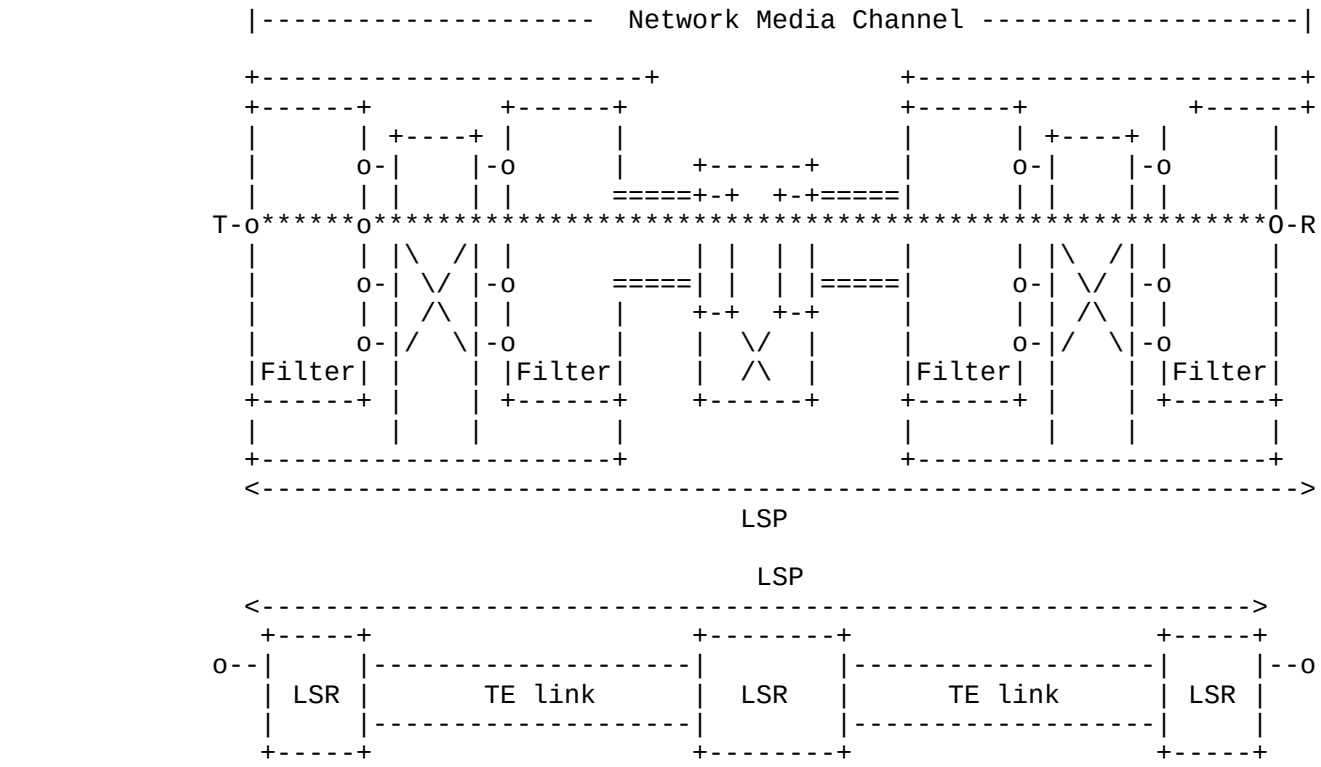


Figure 13: LSP representing a network media channel (OCh-P NC)

[Note: not clear the difference, from a control plane perspective, of figs Figure 12 and Figure 13.]

Applying the notion of hierarchy at the media layer, by using the LSP as a FA, the media channel created can support multiple (sub) media channels. [Editot note : a specific behavior related to Hierarchies will be verified at a later point in time].

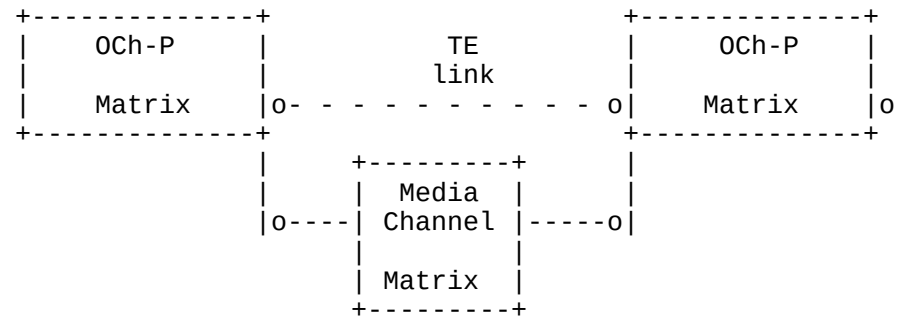


Figure 14: MRN/MLN topology view with TE link / FA

Note that there is only one media layer switch matrix (one implementation is FlexGrid ROADM) in SSON, while "signal layer LSP is mainly for the purpose of management and control of individual optical signal". Signal layer LSPs (OChs) with the same attributions (such as source and destination) could be grouped into one media-layer LSP (media channel), which has advantages in spectral efficiency (reduce guard band between adjacent OChs in one FSC) and LSP management. However, assuming some network elements indeed perform signal layer switch in SSON, there must be enough guard band between adjacent OChs in one media channel, in order to compensate filter concatenation effect and other effects caused by signal layer switching elements. In such condition, the separation of signal layer from media layer cannot bring any benefit in spectral efficiency and in other aspects, but make the network switch and control more complex. If two OChs must switch to different ports, it is better to carry them by diferent FSCs and the media layer switch is enough in this scenario.

5.4. Control Plane modeling of Network elements

Optical transmitters/receivers may have different tunability constraints, and media channel matrixes may have switching restrictions. Additionally, a key feature of their implementation is their highly asymmetric switching capability which is described in [RFC6163] in detail. Media matrices include line side ports which are connected to DWDM links and tributary side input/output ports which can be connected to transmitters/receivers.

A set of common constraints can be defined:

o The minimum and maximum slot width.

- Granularity: the optical hardware may not be able to select parameters with the lowest granularity (e.g. 6.25 GHz for nominal central frequencies or 12.5 GHz for slot width granularity).

- Available frequency ranges: the set or union of frequency ranges that are not allocated (i.e. available). The relative grouping and distribution of available frequency ranges in a fiber is usually referred to as ''fragmentation''.

- Available slot width ranges: the set or union of slot width ranges supported by media matrices. It includes the following information.

  * Slot width threshold: the minimum and maximum Slot Width supported by the media matrix. For example, the slot width can be from 50GHz to 200GHz.

  * Step granularity: the minimum step by which the optical filter bandwidth of the media matrix can be increased or decreased. This parameter is typically equal to slot width granularity (i.e. 12.5GHz) or integer multiples of 12.5GHz.

[Editor's note: different configurations such as C/CD/CDC will be added later. This section should state specifics to media channel matrices, ROADM models need to be moved to an appendix].

5.5. Media Layer Resource Allocation considerations

A media channel has an associated effective frequency slot. From the perspective of network control and management, this effective slot is seen as the "usable" frequency slot end to end. The establishment of an LSP related the establishment of the media channel and effective frequency slot.

In this context, when used unqualified, the frequency slot is a local term, which applies at each hop. An effective frequency slot applies at the media chall (LSP) level

A "service" request is characterized as a minimum, by its required effective slot width. This does not preclude that the request may add additional constraints such as imposing also the nominal central frequency. A given frequency slot is requested for the media channel say, with the Path message. Regardless of the actual encoding, the Path message sender descriptor sender_tspec shall specify a minimum frequency slot width that needs to be fulfilled.

In order to allocate a proper effective frequency slot for a LSP, the signaling should specify its required slot width.

An effective frequency slot must equally be described in terms of a central nominal frequency and its slot width (in terms of usable spectrum of the effective frequency slot). That is, one must be able to obtain an end-to-end equivalent n and m parameters. We refer to this as the "effective frequency slot of the media channel/LSP must be valid".

In GMPLS the requested effective frequency slot is represented to the TSpec and the effective frequency slot is mapped to the FlowSpec.

The switched element corresponds in GMPLS to the 'label'. As in flexi-grid the switched element is a frequency slot, the label represents a frequency slot. Consequently, the label in flexi-grid must convey the necessary information to obtain the frequency slot characteristics (i.e, center and width, the n and m parameters). The frequency slot is locally identified by the label

The local frequency slot may change at each hop, typically given hardware constraints (e.g. a given node cannot support the finest granularity). Locally n and m may change. As long as a given downstream node allocates enough optical spectrum, m can be different along the path. This covers the issue where concrete media matrices can have different slot width granularities. Such "local" m will appear in the allocated label that encodes the frequency slot as well as the flow descriptor flowspec.

Different modes are considered: RSA with explicit label control, and for R+DSA, the GMPLS signaling procedure is similar to the one described in section 4.1.3 of [RFC6163] except that the label set should specify the available nominal central frequencies that meet the slot width requirement of the LSP. The intermediate nodes can collect the acceptable central frequencies that meet the slot width requirement hop by hop. The tail-end node also needs to know the slot width of a LSP to assign the proper frequency resource. Compared with [RFC6163], except identifying the resource (i.e., fixed wavelength for WSON and frequency resource for flexible grids), the other signaling requirements (e.g., unidirectional or bidirectional, with or without converters) are the same as WSON described in the section 6.1 of [RFC6163].

Regarding how a GMPLS control plane can assign n and m, different cases can apply:

a) n and m can both change. It is the effective slot what matters. Some entity needs to make sure the effective frequency slot remains valid.

b) m can change; n needs to be the same along the path. This ensures that the nominal central frequency stays the same.

c) n and m need to be the same.

d)n can change, m needs to be the same.

In consequence, an entity such as a PCE can make sure that the n and m stay the same along the path. Any constraint (including frequency slot and width granularities) is taken into account during path computation. alternatively, A PCE (or a source node) can compute a path and the actual frequency slot assignment is done, for example, with a distributed (signaling) procedure:

Each downstream node ensures that m is >= requested_m.

Since a downstream node cannot foresee what an upstream node will allocate in turn, a way we can ensure that the effective frequency slot is valid is then by ensuring that the same "n" is allocated. By forcing the same n, we avoid cases where the effective frequency slot of the media channel is invalid (that is, the resulting frequency slot cannot be described by its n and m parameters).

Maybe this is a too hard restriction, since a node (or even a centralized/combined RSA entity) can make sure that the resulting end to end (effective) frequency slot is valid, even if n is different locally. That means, the effective (end to end) frequency slot that characterizes the media channel is one and determined by its n and m, but are logical, in the sense that they are the result of the intersection of local (filters) freq slots which may have different freq. slots

For Figure Figure 15 the effective slot is valid by ensuring that the minimum m is greater than the requested m. The effective slot (intersection) is the lowest m (bottleneck).

For Figure Figure 16 the effective slot is valid by ensuring that it is valid at each hop in the upstream direction. The intersection needs to be computed. Invalid slots could result otherwise.

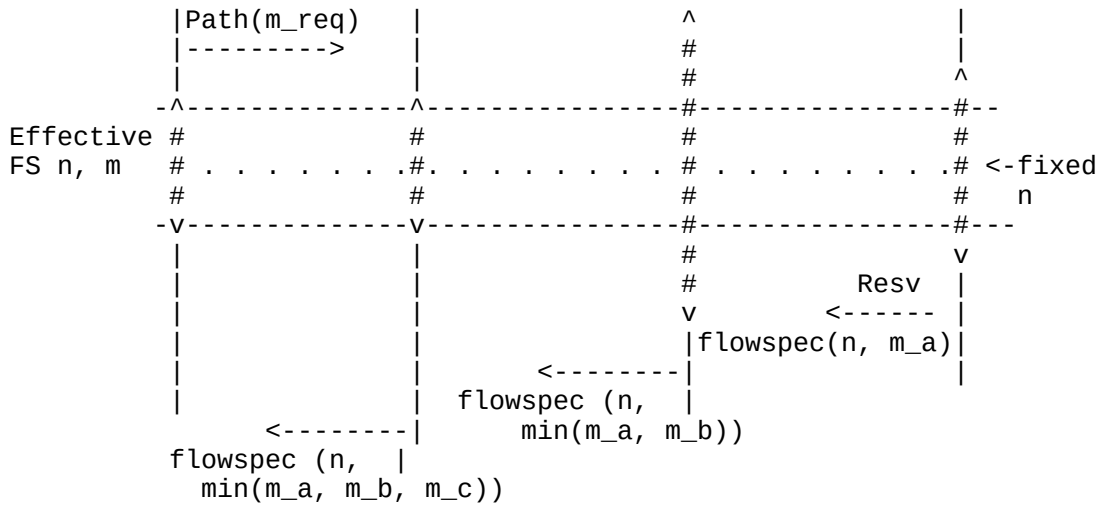


Figure 15: Distributed allocation with different m and same n

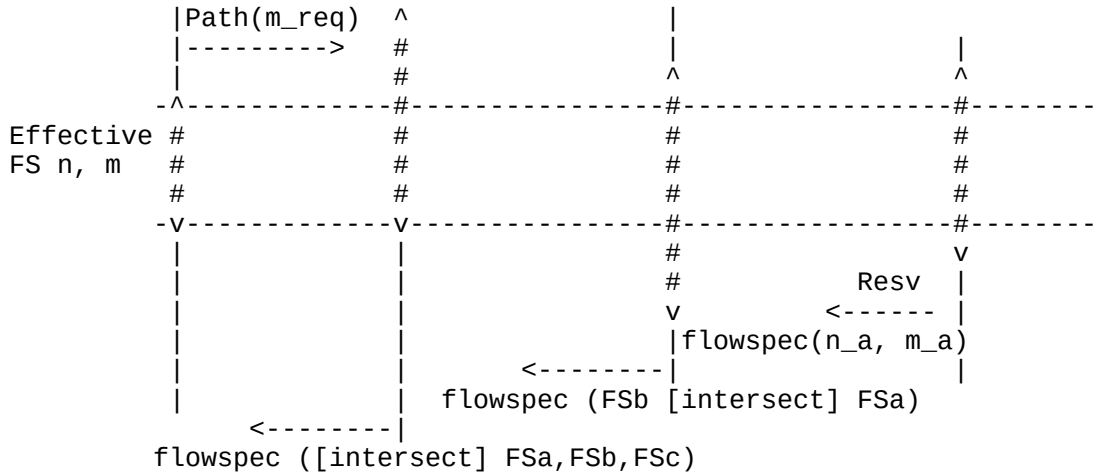


Figure 16: Distributed allocation with different m and different n

Note, when a media channel is bound to one OCh-P (i.e is a Network media channel), the EFS must be the one of the Och-P. The media channel setup by the LSP may contains the EFS of the network media channel EFS. This is an endpoint property, the egress and ingress SHOULD constrain the EFS to Och-P EFS .

5.6. Neighbor Discovery and Link Property Correlation

Potential interworking problems between fixed-grid DWDM and flexible-grid DWDM nodes, may appear. Additionally, even two flexible-grid optical nodes may have different grid properties, leading to link property conflict.

Devices or applications that make use of the flexible-grid may not be able to support every possible slot width. In other words, applications may be defined where different grid granularity can be supported. Taking node F as an example, an application could be defined where the nominal central frequency granularity is 12.5 GHz requiring slot widths being multiple of 25 GHz. Therefore the link between two optical nodes with different grid granularity must be configured to align with the larger of both granularities. Besides, different nodes may have different slot width tuning ranges.

In summary, in a DWDM Link between two nodes, at least the following properties should be negotiated:

Grid capability (channel spacing) - Between fixed-grid and flexible-grid nodes.

Grid granularity - Between two flexible-grid nodes.

Slot width tuning range - Between two flexible-grid nodes.

5.7. Path Computation / Routing and Spectrum Assignment (RSA)

Much like in WSON, in which if there is no (available) wavelength converters in an optical network, an LSP is subject to the ''wavelength continuity constraint'' (see section 4 of [RFC6163]), if the capability of shifting or converting an allocated frequency slot, the LSP is subject to the Optical ''Spectrum Continuity Constraint''.

Because of the limited availability of wavelength/spectrum converters (sparse translucent optical network) the wavelength/spectrum continuity constraint should always be considered. When available, information regarding spectrum conversion capabilities at the optical nodes may be used by RSA (Routing and Spectrum Assignment) mechanisms.

The RSA process determines a route and frequency slot for a LSP. Hence, when a route is computed the spectrum assignment process (SA) should determine the central frequency and slot width based on the slot width and available central frequencies information of the transmitter and receiver, and the available frequency ranges

information and available slot width ranges of the links that the route traverses.

5.7.1. Architectural Approaches to RSA

Similar to RWA for fixed grids, different ways of performing RSA in conjunction with the control plane can be considered. The approaches included in this document are provided for reference purposes only; other possible options could also be deployed.

5.7.1.1. Combined RSA (R&SA)

In this case, a computation entity performs both routing and frequency slot assignment. The computation entity should have the detailed network information, e.g. connectivity topology constructed by nodes/links information, available frequency ranges on each link, node capabilities, etc.

The computation entity could reside either on a PCE or the ingress node.

5.7.1.2. Separated RSA (R+SA)

In this case, routing computation and frequency slot assignment are performed by different entities. The first entity computes the routes and provides them to the second entity; the second entity assigns the frequency slot.

The first entity should get the connectivity topology to compute the proper routes; the second entity should get the available frequency ranges of the links and nodes' capabilities information to assign the spectrum.

5.7.1.3. Routing and Distributed SA (R+DSA)

In this case, one entity computes the route but the frequency slot assignment is performed hop-by-hop in a distributed way along the route. The available central frequencies which meet the spectrum continuity constraint should be collected hop by hop along the route. This procedure can be implemented by the GMPLS signaling protocol.

5.8. Routing / Topology dissemination

In the case of combined RSA architecture, the computation entity needs to get the detailed network information, i.e. connectivity topology, node capabilities and available frequency ranges of the links. Route computation is performed based on the connectivity topology and node capabilities; spectrum assignment is performed

based on the available frequency ranges of the links. The computation entity may get the detailed network information by the GMPLS routing protocol. Compared with [RFC6163], except wavelength-specific availability information, the connectivity topology and node capabilities are the same as WSON, which can be advertised by GMPLS routing protocol (refer to section 6.2 of [RFC6163]. This section analyses the necessary changes on link information brought by flexible grids.

5.8.1. Available Frequency Ranges/slots of DWDM Links

In the case of flexible grids, channel central frequencies span from 193.1 THz towards both ends of the C band spectrum with 6.25 GHz granularity. Different LSPs could make use of different slot widths on the same link. Hence, the available frequency ranges should be advertised.

5.8.2. Available Slot Width Ranges of DWDM Links

The available slot width ranges needs to be advertised, in combination with the Available frequency ranges, in order to verify whether a LSP with a given slot width can be set up or not; this is is constrained by the available slot width ranges of the media matrix Depending on the availability of the slot width ranges, it is possible to allocate more spectrum than strictly needed by the LSP.

5.8.3. Spectrum Management

[Editors' note: the part on the hierarchy of the optical spectrum could be confusing, we can discuss it]. The total available spectrum on a fiber could be described as a resource that can be divided by a media device into a set of Frequency Slots. In terms of managing spectrum, it is necessary to be able to speak about different granularities of managed spectrum. For example, a part of the spectrum could be assigned to a third party to manage. This need to partition creates the impression that spectrum is a hierarchy in view of Management and Control Plane. The hierarchy is created within a management system, and it is an access right hierarchy only. It is a management hierarchy without any actual resource hierarchy within fiber. The end of fiber is a link end and presents a fiber port which represents all of spectrum available on the fiber. Each spectrum allocation appears as Link Channel Port (i.e., frequency slot port) within fiber.

5.8.4. Information Model

Fixed DM grids can also be described via suitable choices of slots in a flexible DWDM grid. However, devices or applications that make use of the flexible grid may not be capable of supporting every possible slot width or central frequency position. Following is the definition of information model, not intended to limit any IGP encoding implementation. For example, information required for routing/path selection may be the set of available nominal central frequencies from which a frequency slot of the required width can be allocated. A convenient encoding for this information (may be as a frequency slot or sets of contiguous slices) is further study in IGP encoding document.

```
<Available Spectrum in Fiber for frequency slot> ::=
    <Available Frequency Range-List>
    <Available Central Frequency Granularity >
    <Available Slot Width Granularity>
    <Minimal Slot Width>
    <Maximal Slot Width>

<Available Frequency Range-List> ::=
    <Available Frequency Range >[< Available Frequency Range-List>]

<Available Frequency Range >::=
  <Start Spectrum Position><End Spectrum Position> |
  <Sets of contiguous slices>

<Available Central Frequency Granularity> ::= n A&#151; 6.25GHz,
  where n is positive integer, such as 6.25GHz, 12.5GHz, 25GHz, 50GHz
  or 100GHz

<Available Slot Width Granularity> ::= m A&#151; 12.5GHz,
  where m is positive integer

<Minimal Slot Width> ::= j x 12.5GHz,
  j is a positive integer

<Maximal Slot Width> ::= k x 12.5GHz,
    k is a positive integer (k >= j)
```

Figure 17: Routing Information model

6. Control Plane Requirements

The GMPLS based control plane of a flexi-grid networks provides aditional requirements to GMPLS. In this section the features to be covered by GMPLS signaling for flexi-grid are identified. [Editor's note: Only discussed requirements are included at this stage. Routing requirements will come in the next version]

6.1. Support for Media Channels

The control plane SHALL be able to support Media Channels, characterized by a single frequency slot. The representation of the Media Channel in the GMPLS Control plane is the so-called flexi-grid LSP. Since network media channels are media channels, an LSP may also be the control plane representation of a network media channel. Consequently, the control plane SHALL be able to support Network Media Channels.

The signaling procedure SHALL be able to configure the nominal central frequency (n) of a flexi-grid LSP.

The control plane protocols SHALL allow flexible range of values for the frequency slot width (m) parameter. Specifically, the control plane SHALL allow setting up a media channel with frequency slot width (m) ranging from a minimum of m=1 (12.5GHz) to a maximum of the entire C-band with a slot width granularity of 12.5GHz.

The signaling procedure of the GMPLS control plane SHALL be able to configure the minimum width (m) of a flexi-grid LSP. In adition, the control plane SHALL be able to configure local frecuency slots,

The control plane architecture SHOULD allow for the support of L-band and S-band

The signalling process of the control plane SHALL allow to collect the local frequency slot asigned at each link along the path

6.2. Support for Media Channel Resizing

The control plane SHALL allow resizing (grow or shrink) the frequency slot width of a media channel/network media channel. The resizing MAY imply resizing the local frequency slots along the path of the flexi-grid LSP.

6.3. Support for Logical Associations of multiple media channels

A set of media channels can be used to transport signals that have a logical association between them. The control plane architecture SHOULD allow multiple media channels to be logically associated. The control plane SHOULD allow the co-routing of a set of media channels logically associated

7. Security Considerations

TBD

8. Contributing Authors

Qilei Wang
ZTE
Ruanjian Avenue, Nanjing, China
wang.qilei@zte.com.cn

Malcolm Betts
ZTE
malcolm.betts@zte.com.cn

Xian Zhang
Huawei
zhang.xian@huawei.com

Cyril Margaria
Nokia Siemens Networks
St Martin Strasse 76, Munich, 81541, Germany
+49 89 5159 16934
cyril.margaria@nsn.com

Sergio Belotti
Alcatel Lucent
Optics CTO
Via Trento 30 20059 Vimercate (Milano) Italy
+39 039 6863033
sergio.belotti@alcatel-lucent.com

Yao Li
Nanjing University
wsliguotou@hotmail.com

Fei Zhang
ZTE
Zijinghua Road, Nanjing, China
zhang.fei3@zte.com.cn

Lei Wang
ZTE
East Huayuan Road, Haidian district, Beijing, China
wang.lei131@zte.com.cn

Guoying Zhang
China Academy of Telecom Research
No.52 Huayuan Bei Road, Beijing, China
zhangguoying@ritt.cn

Takehiro Tsuritani
KDDI R&D Laboratories Inc.
2-1-15 Ohara, Fujimino, Saitama, Japan
tsuri@kddilabs.jp

Lei Liu
KDDI R&D Laboratories Inc.
2-1-15 Ohara, Fujimino, Saitama, Japan
le-liu@kddilabs.jp

Eve Varma
Alcatel-Lucent
+1 732 239 7656
eve.varma@alcatel-lucent.com

Young Lee
Huawei

Jianrui Han
Huawei

Sharfuddin Syed
Infinera

Rajan Rao
Infinera

Marco Sosa
Infinera

Biao Lu
Infinera

Abinder Dhillon
Infinera

Felipe Jimenez Arribas
TelefA^3nica I+D

Andrew G. Malis
Verizon

Adrian Farrel
Old Dog Consulting

Daniel King
Old Dog Consulting

Huub van Helvoort

9. Acknowledgments

The authors would like to thank Pete Anslow for his insights and clarifications. This work was supported in part by the FP-7 IDEALIST project under grant agreement number 317999.

10. References

10.1. Normative References

[G.694.1] International Telecomunications Union, "ITU-T Recommendation G.694.1, Spectral grids for WDM applications: DWDM frequency grid", November 2012.

[G.709] International Telecomunications Union, "ITU-T Recommendation G.709, Interfaces for the Optical Transport Network (OTN).", March 2009.

[G.800] International Telecomunications Union, "ITU-T Recommendation G.800, Unified functional architecture of transport networks.", February 2012.

[G.805] International Telecomunications Union, "ITU-T Recommendation G.805, Generic functional architecture of transport networks.", March 2000.

[G.8080] International Telecomunications Union, "ITU-T Recommendation G.8080/Y.1304, Architecture for the automatically switched optical network", 2012.

[G.870] International Telecomunications Union, "ITU-T Recommendation G.870/Y.1352, Terms and definitions for optical transport networks", November 2012.

[G.872] International Telecomunications Union, "ITU-T Recommendation G.872, Architecture of optical transport networks, draft v0.16 2012/09 (for discussion)", 2012.

[G.959.1-2013] International Telecomunications Union, "Update of ITU-T Recommendation G.959.1, Optical transport network physical layer interfaces (to appear in July 2013)", 2013.

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

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

[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.

[RFC5150] Ayyangar, A., Kompella, K., Vasseur, JP., and A. Farrel, "Label Switched Path Stitching with Generalized Multiprotocol Label Switching Traffic Engineering (GMPLS TE)", RFC 5150, February 2008.

[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.

10.2. Informative References

[RFC4397] Bryskin, I. and A. Farrel, "A Lexicography for the Interpretation of Generalized Multiprotocol Label Switching (GMPLS) Terminology within the Context of the ITU-T's Automatically Switched Optical Network (ASON) Architecture", RFC 4397, February 2006.

Authors' Addresses

Oscar Gonzalez de Dios (editor)
Telefonica I+D
Don Ramon de la Cruz 82-84
Madrid 28045
Spain

Phone: +34913128832
Email: ogondio@tid.es

Ramon Casellas (editor)
CTTC
Av. Carl Friedrich Gauss n.7
Castelldefels Barcelona
Spain

Phone: +34 93 645 29 00
Email: ramon.casellas@cttc.es

Fatai Zhang
Huawei
Huawei Base, Bantian, Longgang District
Shenzhen 518129
China

Phone: +86-755-28972912
Email: zhangfatai@huawei.com

Xihua Fu
ZTE
Ruanjian Avenue
Nanjing
China

Email: fu.xihua@zte.com.cn

Daniele Ceccarelli
Ericsson
Via Calda 5
Genova
Italy

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

Iftekhar Hussain
Infinera
140 Caspian Ct.
Sunnyvale 94089
USA

Phone: 408-572-5233
Email: ihussain@infinera.com

CCAMP Working Group
Internet Draft
Intended status: Standard Track
Expires: August 2, 2014

Zafar Ali
George Swallow
Clarence Filsfils
Matt Hartley
Gabriele Maria Galimberti
Cisco Systems
Ori Gerstel
SDN Solutions Ltd.
Kenji Kumaki
KDDI Corporation
Ruediger Kunze
Deutsche Telekom AG
Julien Meuric
France Telecom Orange
February 3, 2014

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

draft-ietf-ccamp-lsp-diversity-03.txt

Status of this Memo

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

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

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

This Internet-Draft will expire on August 2, 2014.

Copyright Notice

Abstract

RFC 4874 specifies methods by which path exclusions may be communicated during RSVP-TE signaling in networks where precise explicit paths are not computed by the LSP source node. This document specifies signaling for additional route exclusion subobjects 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 is a well-known requirement from Service Providers. Such diversity ensures Label-Switched Paths (LSPs) may 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 can be computed locally. However, there are scenarios when path computations are performed by remote nodes, thus there is a need for relevant diversity requirements 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 (server layer) 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).

[RFC4874] 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 signaling the diverse LSP, in which case the addresses 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, however the node has incomplete or no path information, e.g. due to policy. In other words, the scenario in which the reference path is known by the source / requesting node but the properties required to construct an XRO object are not fully known. Inter-domain and GMPLS overlay networks can present such restrictions.

This document defines procedures that may be used to exclude the path taken by a particular LSP, or the paths taken by all LSPs belonging to a single tunnel. The diversity requirements considered in this document do not require that the paths in question belong to the same tunnel or share the same source or destination node.

If mutually diverse paths are desired for two LSPs belonging to different tunnels, it is recommended that they be signaled with XRO LSP subobjects referencing each other. The processing rules specified in this document cover this case.

The means by which the node calculating or expanding the route of the signaled LSP discovers the route of the path(s) from which the signaled LSP requires diversity are beyond the scope of this document.

This document addresses only the exclusion of point-to-point paths. Exclusion of point-to-multipoint paths is beyond the scope of this document.

2. RSVP-TE signaling extensions

This section describes the signaling extensions required to address the aforementioned requirements. Specifically, this document defines a new LSP subobject to be signaled in the EXCLUDE_ROUTE object (XRO) and/ or Explicit Exclusion Route Subobject (EXRS) defined in [RFC4874]. Inclusion of the LSP subobject in any other RSVP object is not defined.

2.1. Terminology

In this document, the following terminology is adopted:

Excluded path: the path from which diversity is required.

Diverse LSP: the LSP being signaled with XRO/ EXRS containing the path subobject referencing the excluded path(s).

Processing node: the node performing a path-calculation involving exclusion specified in an XRO or EXRS.

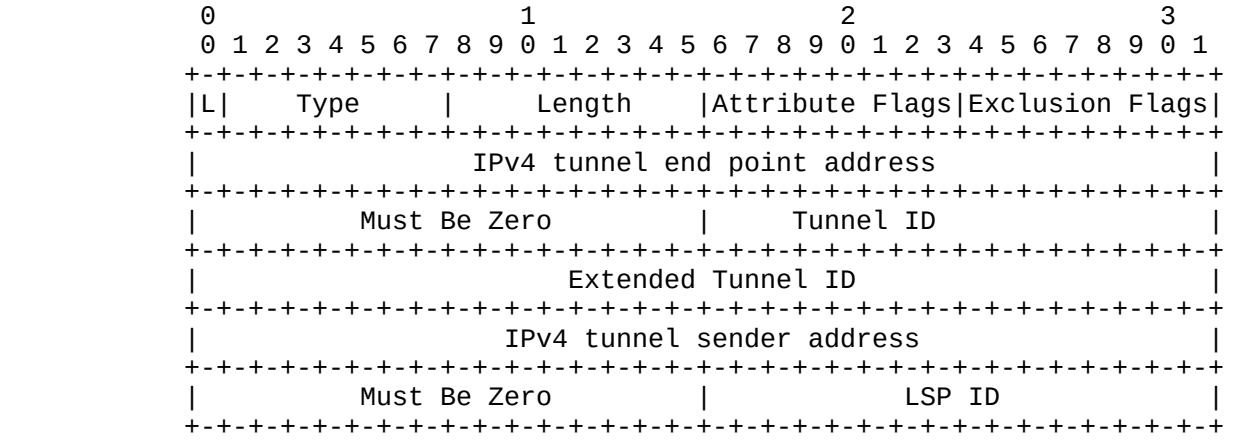
Destination node: in the context of an XRO, this is the destination of the LSP being signaled. In the context of an EXRS, the destination node is the last explicit node to which the loose hop is expanded.

Penultimate node: in the context of an XRO, this is the penultimate hop of the LSP being signaled. In the context of an EXRS, the penultimate node is the penultimate node of the loose hop undergoing expansion.

2.2. Path XRO Subobjects

New IPv4 and IPv6 Point-to-Point (P2P) Path XRO subobjects are defined by this document as follows.

2.2.1. IPv4 Point-to-Point Path subobject



L:

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

0 indicates that the attribute specified MUST be excluded.

1 indicates that the attribute specified SHOULD be avoided.

Type:

IPv4 Point-to-Point Path subobject (to be assigned by IANA; suggested value: 36).

Length:

The length contains the total length of the subobject in bytes, including the type and length fields. The length is always 24.

Attribute Flags:

The Attribute 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 = LSP ID to be ignored

Indicates tunnel level exclusion. Specifically, this flag is used to indicate that the lsp-id field of the subobject is to be ignored and the exclusion applies to any LSP matching the rest of the supplied FEC.

0x02 = Destination node exception

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

0x04 = Processing node exception

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

0x08 = 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.

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

Exclusion Flags

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

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 LSP 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 LSP subobject.

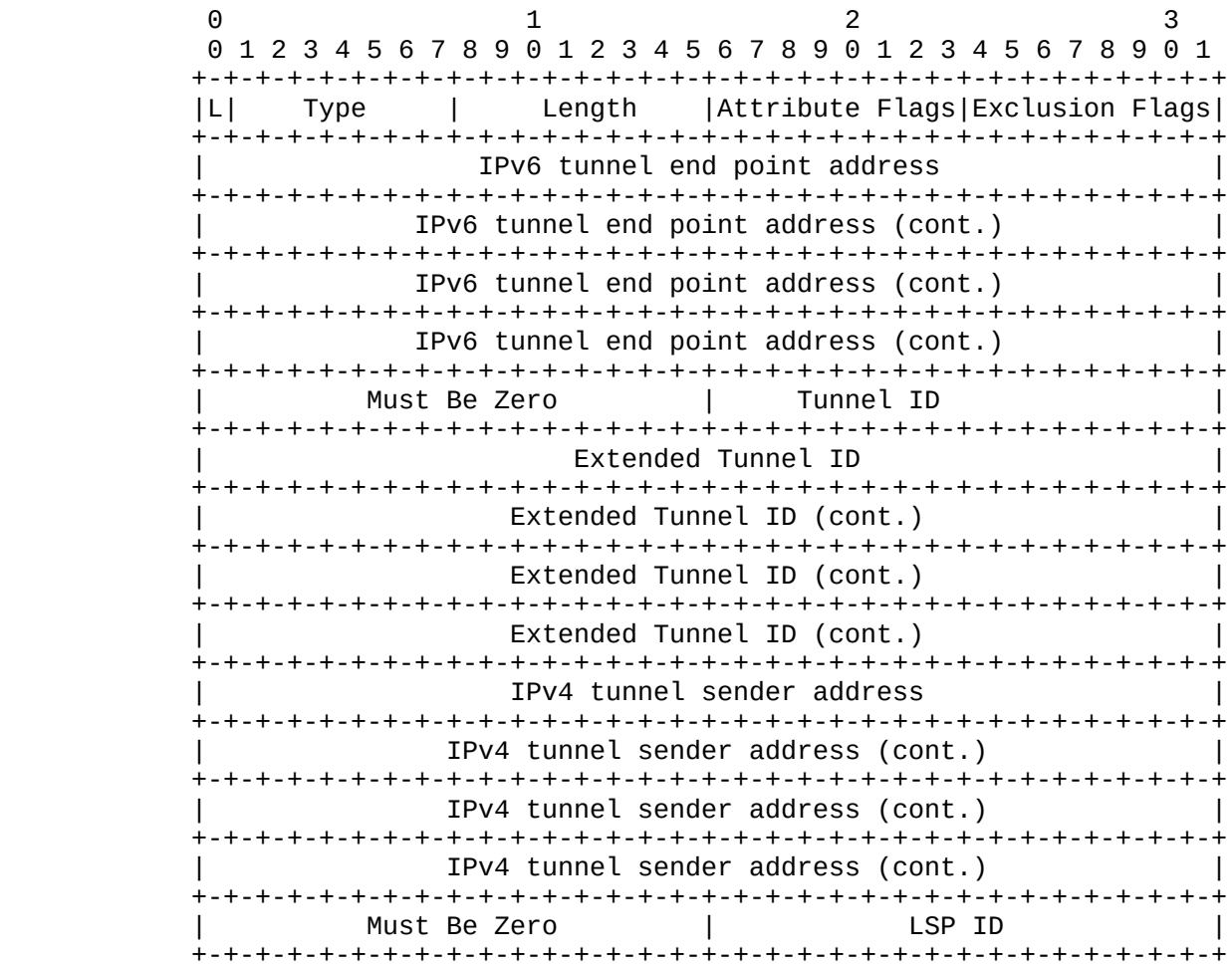
(Note: the meaning of this flag may be modified by the value of the Attribute-flags.)

0x04 = Link exclusion

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

The remaining fields are as defined in [RFC3209].

2.2.2. IPv6 Point-to-Point Path subobject



L

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

0 indicates that the attribute specified MUST be excluded.

1 indicates that the attribute specified SHOULD be avoided.

Type

IPv6 Point-to-Point Path 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 always 48.

The Attribute Flags and Exclusion Flags are as defined for the IPv4 Point-to-Point LSP XRO subobject.

The remaining fields are as defined in [RFC3209].

2.3. Processing rules for the Path XRO subobjects

XRO processing as described in [RFC4874] is unchanged.

If the processing node is the destination for the LSP being signaled, it SHOULD NOT process a Path XRO subobject.

If the L-flag is not set, the processing node follows the following procedure:

- 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 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 excluded path referenced in the LSP subobject is unknown to the processing node, the processing node SHOULD ignore the LSP subobject in the XRO 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 of XRO path unknown" (value to be assigned by IANA, suggested value: 13) for the signaled LSP.

If the L-flag is set, the processing node follows the procedure below:

- 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 the excluded path referenced in the LSP subobject is unknown to the processing node, the processing node SHOULD ignore the LSP subobject in the XRO and SHOULD proceed with the signaling request. After sending the Resv for signaled LSP, the processing node SHOULD return a PathErr message with the error code "Notify Error" (25) and error sub-code "Route of XRO path unknown" for the signaled LSP.

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

- an excluded path referenced in the diverse LSP’s XRO subobject becomes known to the processing node (e.g. when the excluded path is signaled), 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, 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). A source node receiving a PathErr message with this error code and sub-code combination MAY try to reoptimize the diverse tunnel to the new 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 NOT 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 following rules apply whether or not the L-flag is set:

- An XRO object MAY contain multiple path subobjects.

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

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

  - o When the "LSP ID to be ignored" flag is set, the processing node MUST calculate a path based on exclusions from the paths of all known LSPs matching the tunnel-id, source, destination and extended tunnel-id specified in the subobject (i.e., tunnel level exclusion). 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).

  - o When the "destination node exception" flag is not set, the exclusion flags SHOULD also be respected for the destination node.

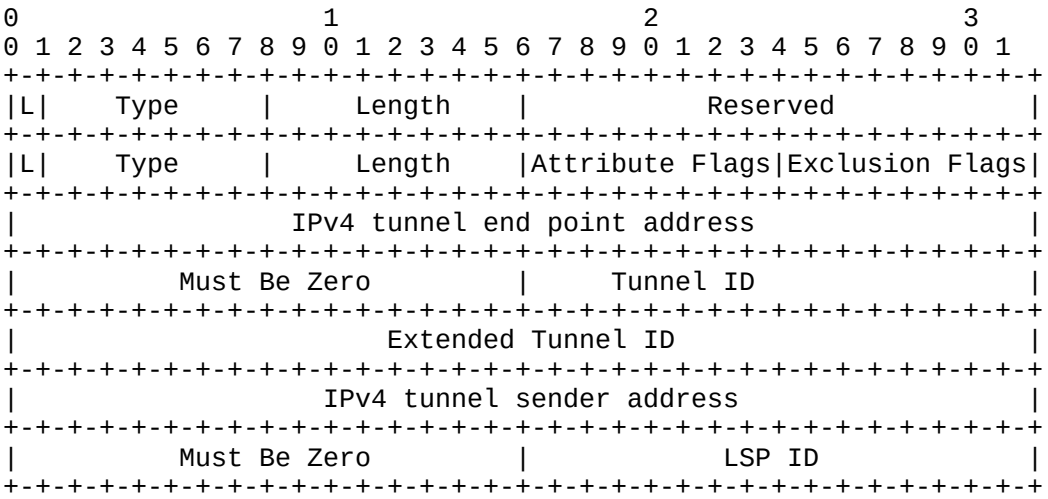
- When the "processing node exception" flag is not set, the exclusion flags SHOULD also be respected for the processing node.

- When the "penultimate node exception" flag is not set, the exclusion flags SHOULD also be respected for the penultimate node.

2.4. Path 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 an IPv4 Point-to-Point (P2P) Path subobject as specified in section 2.2.1. In this case, the EXRS format would be as follows:



The meanings of respective fields in EXRS header are as defined in [RFC4874]. The meanings of respective fields in IPv4 P2P Path subobject are as defined earlier in this document.

The processing rules for the EXRS object are unchanged from [RFC4874]. When the EXRS contains one or more Path subobject(s),

the processing rules specified in Section 2.3 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 Type	Subobject Description
To be assigned by IANA (suggested value: 36)	IPv4 P2P Path subobject
To be assigned by IANA (suggested value: 37)	IPv6 P2P Path subobject

4.2. New EXRS subobject types

The IPv4 and IPv6 P2P Path 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 path 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.

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.

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

[RFC5920] Fang, L., Ed., "Security Framework for MPLS and GMPLS Networks", RFC 5920, July 2010.

Authors' Addresses

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

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

Gabriele Maria Galimberti
Cisco Systems
ggalimbe@cisco.com

Ori Gerstel
SDN Solutions Ltd.
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
Ruediger.Kunze@telekom.de

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

George Swallow
Cisco Systems
swallow@cisco.com

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: August 18, 2014

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

M. Hartley
Z. Ali
Cisco

February 14, 2014

RSVP-TE Extensions for Collecting SRLG Information
draft-ietf-ccamp-rsvp-te-srlg-collect-04

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 LSP.

Status of This Memo

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

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

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

This Internet-Draft will expire on August 18, 2014.

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 may constitute a 'shared risk link group' (SRLG) if they share a resource whose failure may 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 an automatic mechanism to collect the SRLG for the TE link formed by a LSP. Note that how to use the collected SRLG information is out of scope of this document

2. RSVP-TE Requirements

2.1. SRLG Collection Indication

The head nodes of the LSP must be capable of indicating whether the SRLG information of the LSP should be collected during the signaling procedure of setting up an LSP. SRLG information should not be collected without an explicit request for it being made by the head node.

2.2. SRLG Collection

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

2.3. 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 must be capable of updating the SRLG information of the LSP. This means that that the signaling procedure must be capable of updating the new SRLG information.

3. RSVP-TE Extensions (Encoding)

3.1. SRLG Collection Flag

In order to indicate nodes that SRLG collection is desired, this document defines a new flag in the Attribute Flags TLV, which is carried in an LSP_REQUIRED_ATTRIBUTES or LSP_ATTRIBUTE Object:

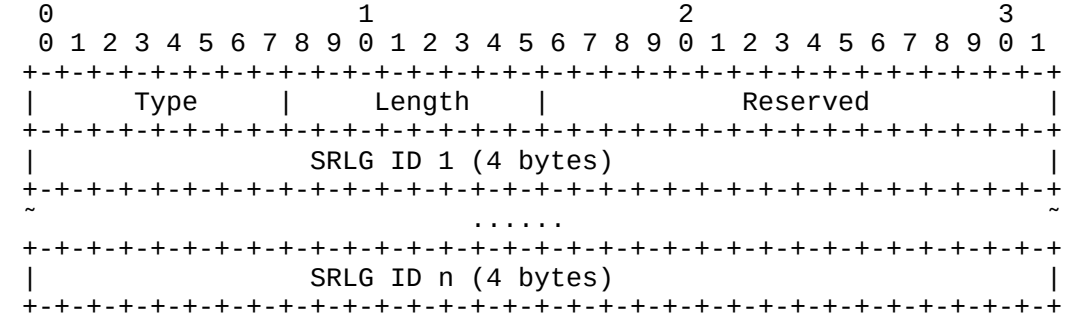
- Bit Number (to be assigned by IANA, recommended bit 12): SRLG Collection flag

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 head and tail node along the setup of the LSP.

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

3.2. SRLG sub-object

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



Type

The type of the sub-object, to be assigned by IANA, which is recommended 34.

Length

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

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

4. Signaling Procedures

4.1. SRLG Collection

Typically, the head node gets the route information of an LSP by adding a RRO which contains the sender’s IP addresses in the Path message. If a head node also desires SRLG recording, it sets the SRLG Collection Flag in the Attribute Flags TLV which can be carried either in an LSP_REQUIRED_ATTRIBUTES Object if the collection is mandatory, or in an LSP_ATTRIBUTES Object if 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 is set, if local policy determines that the SRLG information should not be provided to the endpoints, it MUST return a PathErr message with

Error Code 2 (policy) and Error subcode "SRLG Recording Rejected" (value to be assigned by IANA, suggest value 108) to reject the Path message.

When a node receives a Path message which carries an LSP_ATTRIBUTES Object and the SRLG Collection Flag is set, if local policy determines that the SRLG information should not 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 the SRLG sub-object(s) in the Path RRO.

If local policy permits the recording of the SRLG information, the processing node SHOULD add an SRLG sub-object to the RRO to carry the local SRLG information. It then forwards the Path message to the next node in the downstream direction.

Following the steps described above, the intermediate nodes of the LSP can collect the SRLG information in the RRO during the forwarding of the Path message hop by hop. When the Path message arrives at the tail node, the tail node can get the SRLG information from the RRO.

Before the Resv message is sent to the upstream node, the tail node adds the SRLG subobject with the SRLG value(s) associated with the local hop to the Resv RRO in a similar manner to that specified above for the addition of Path RRO sub-objects by midpoint nodes.

When a node receives a Resv message for an LSP for which SRLG Collection is specified, if local policy determines that the SRLG information should not be provided to the endpoints, if the SRLG-recording request was in a LSP_REQUIRED_ATTRIBUTES object, then a ResvErr with Error code 2 (policy) and Error subcode "SRLG Recording Rejected" (value to be assigned by IANA, suggest value 108) MUST be sent. If the request was in a LSP_ATTRIBUTES object, then a ResvErr SHOULD NOT be generated, but SRLG information must not be added in the RRO. Otherwise, if local policy allows to provide the SRLG information, it MUST add an SRLG sub-object to the RRO to carry the SRLG information in the upstream direction. When the Resv message arrives at the head node, the head node can get the SRLG information from the RRO in the same way as the tail node.

Note that a link’s SRLG information for the upstream direction cannot be assumed to be the same as that in the downstream.

- 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.

- 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.

4.2. SRLG Update

When the SRLG information of a link is changed, the LSPs using that link should 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 need not send an update.

5. Manageability Considerations

5.1. Policy Configuration

In a border node of inter-domain or inter-layer network, the following SRLG processing policy should be capable of being configured:

- 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 or removed entirely.

5.2. Coherent SRLG IDs

In a multi-layer multi-domain scenario, SRLG ids may 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 must 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.

6. Security Considerations

This document does not introduce any additional security issues above those identified in [RFC5920][RFC3209][RFC3473]

7. IANA Considerations

7.1. RSVP Attribute Bit Flags

IANA has created a registry and manages the space of attributes bit flags of Attribute Flags TLV, as described in section 11.3 of [RFC5420], in the "Attributes TLV Space" section of the "Resource Reservation Protocol-Traffic Engineering (RSVP-TE) Parameters" registry located in https://www.iana.org/assignments/rsvp-te-parameters/rsvp-te-parameters.xhtml. It is requested that IANA makes assignments from the Attribute Bit Flags.

This document introduces a new Attribute Bit Flag:

- Bit number: TBD (10)
- Defining RFC: this I-D
- Name of bit: SRLG Collection Flag
- The meaning of the SRLG Collection Flag is defined in this I-D.

7.2. ROUTE_RECORD Object

IANA has made the following assignments in the "Class Names, Class Numbers, and Class Types" section of the "RSVP PARAMETERS" registry located at http://www.iana.org/assignments/rsvp-parameters. We request that IANA make assignments from the ROUTE_RECORD RFC 3209 [RFC3209] portions of this registry.

This document introduces a new RRO sub-object:

Type	Name	Reference
TBD (34)	SRLG sub-object	This I-D

7.3. Policy Control Failure Error subcodes

IANA has made the following 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. We request that IANA make assignments from the Policy Control Failure Sub-Codes registry.

This document introduces a new Policy Control Failure Error sub-code:

- Error sub-code: TBD (108)
- Defining RFC: this I-D
- Name of error sub-code: SRLG Recording Rejected
- The meaning of the SRLG Recording Rejected error sub-code is defined in this I-D

8. Acknowledgements

The authors would like to thank Igor Bryskin, Ramon Casellas, Lou Berger and Alan Davey for their useful comments and improvements to the document.

9. Normative References

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

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

[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.

[RFC5150] Ayyangar, A., Kompella, K., Vasseur, JP., and A. Farrel, "Label Switched Path Stitching with Generalized Multiprotocol Label Switching Traffic Engineering (GMPLS TE)", RFC 5150, February 2008.

[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
Don Ramon de la Cruz
Madrid 28006
Spain

Phone: +34 913328832
Email: ogondio@tid.es

Dan Li
Huawei
F3-5-B RD Center
Bantian, Longgang District, Shenzhen 518129
P.R.China

Email: danli@huawei.com

Cyril Margaria
SabenerStr. 44
Munich 81547
Germany

Phone: +49 89 5159 16934
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: Informational
Expires: August 2014

CCAMP
Y.Li
ZTE
G. Zhang
CATR
X.Fu
ZTE
R. Casellas
CTTC
Y Wang
CATR

February 14, 2014

Link Management Protocol Extensions for Grid Property Negotiation
draft-li-ccamp-grid-property-lmp-03.txt

Status of this Memo

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

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

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

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt

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

This Internet-Draft will expire on August 10, 2014.

Copyright Notice

Abstract

The recent updated version of ITU-T [G.694.1] has introduced the flexible-grid DWDM technique, which provides a new tool that operators can implement to provide a higher degree of network optimization than is possible with fixed-grid systems. This document describes the extensions to the Link Management Protocol (LMP) to negotiate link grid property between the adjacent DWDM nodes before the link is brought up.

Table of Contents

1. Introduction

The recent updated version of ITU-T [G.694.1] has introduced the flexible-grid DWDM technique, which provides a new tool that operators can implement to provide a higher degree of network optimization than is possible with fixed-grid systems. A flexible-grid network supports allocating a variable-sized spectral slot to a channel. Flexible-grid DWDM transmission systems can allocate their channels with different spectral bandwidths/slot widths so that they can be optimized for the bandwidth requirements of the particular bit rate and modulation scheme of the individual channels. This technique is regarded to be a promising way to improve the spectrum utilization efficiency and can be used in the beyond 100Gb/s transport systems.

Fixed-grid DWDM system is regarded as a special case of Flexi-grid DWDM. It is expected that fixed-grid optical nodes will be gradually replaced by flexible nodes and interworking between fixed-grid DWDM and flexible-grid DWDM nodes will be needed as the network evolves. Additionally, even two flexible-grid optical nodes may have different grid properties based on the filtering component characteristics, thus need to negotiate on the specific parameters to be used during neighbor discovery process [draft-ietf-ccamp-flexi-grid-fwk-00]. This document describes the extensions to the Link Management Protocol (LMP) to negotiate a link grid property between two adjacent Flexi-grid nodes before the link is brought up.

1.1. Conventions Used in This Document

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

2. Terminology

For the flexible-grid DWDM, the spectral resource is called frequency slot which is represented by the central frequency and the slot width. The defined nominal central frequency and the slot width can be referred to [FLEX-FWK].

In this contribution, some other definitions are listed below:

Central frequency granularity: It is the granularity of the allowed central frequencies and is set to the multiple of 6.25 GHz.

Slot width granularity: It is the granularity of the allowed slot width, and is set to the multiple of 12.5 GHz.

Tuning range: It describes the supported spectrum slot range of the switching nodes or interfaces. It is represented by the supported minimal slot width and the maximum slot width.

Channel spacing: It is used in traditional fixed-grid network to identify spectrum spacing between two adjacent channels.

3. Requirements for Grid Property Negotiation

3.1. Flexi-fixed Grid Nodes Interworking

Figure 1 shows an example of interworking between flexible and fixed-grid nodes. Node A, B, D and E support flexible-grid. All these nodes can support frequency slots with a central frequency granularity of 6.25 GHz and slot width granularity of 12.5 GHz. Given the flexibility in flexible-grid nodes, it is possible to configure the nodes in such a way that the central frequencies and slot width parameters are backwards compatible with the fixed DWDM grids (adjacent flexible frequency slots with channel spacing of 8*6.25 and slot width of 4*12.5 GHz is equivalent to fixed DWDM grids with channel spacing of 50 GHz).

As node C can only support the fixed-grid DWDM property with channel spacing of 50 GHz, to establish a LSP through node B, C, D, the links between B to C and C to D must set to align with the fixed-grid values. This link grid property must be negotiated before establishing the LSP.

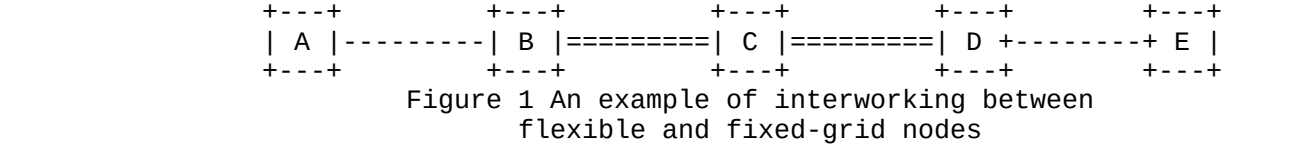
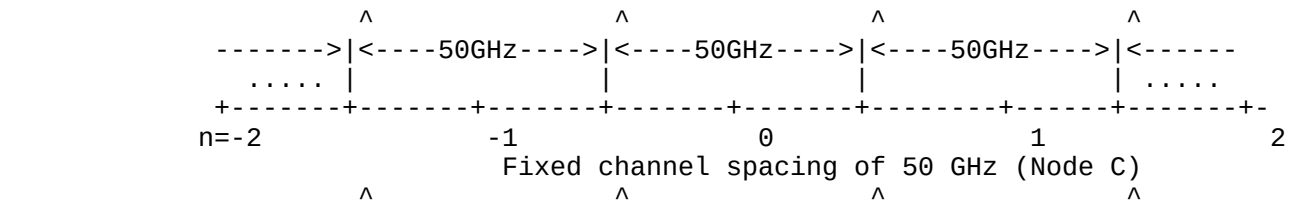
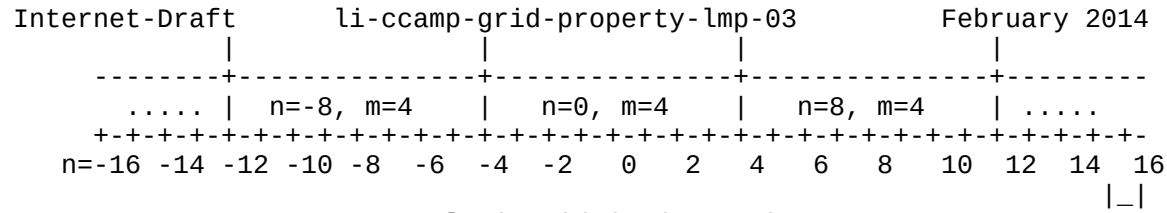


Figure 1 An example of interworking between flexible and fixed-grid nodes





Central frequency granularity=6.25 GHz
Slot width granularity=12.5 GHz

Figure 2 Representation of fixed channel spacing and flexi-grid spectrum slot

3.2. Flexible-Grid Capability Negotiation

The updated version of ITU-T [G.694.1] has defined the flexible-grid with a central frequency granularity of 6.25 GHz and a slot width granularity of 12.5 GHz. However, devices or applications that make use of the flexible-grid may not be able to support every possible slot width. In other words, applications may be defined where different grid granularity can be supported. Taking node G as an example, an application could be defined where the central frequency granularity is 12.5 GHz requiring slot widths being multiple of 25 GHz. Therefore the link between two optical nodes with different grid granularity must be configured to align with the larger of both granularities. Besides, different nodes may have different slot width tuning ranges. For example, in figure 3, node F can only support slot width with tuning change from 12.5 to 100 GHz, while node G supports tuning range from 25 GHz to 200 GHz. The link property of slot width tuning range for the link between F and G should be chosen as the range intersection, resulting in a range from 25 GHz to 100 GHz.

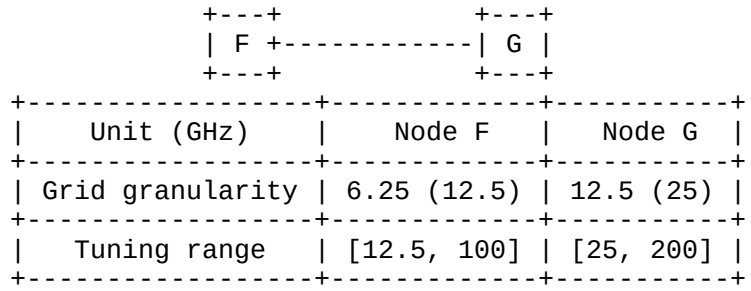


Figure 3 An example of flexible-grid capability negotiation

3.3. Summary

In summary, in a DWDM Link between two nodes, the following properties can be negotiated:

- o Grid capability: flexible grid or fixed grid DWDM.

- o Central frequency granularity: a multiplier of 6.25 GHz.

- o Slot width granularity: a multiplier of 12.5 GHz.

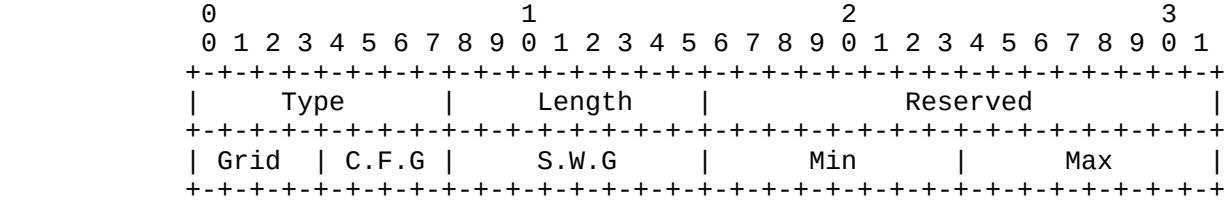
- o Slot width tuning range: two multipliers of 12.5GHz, each indicate the minimal and maximal slot width supported by a port respectively.

4. LMP extensions

4.1. Grid Property Subobject

According to [RFC4204], the LinkSummary message is used to verify the consistency of the link property on both sides of the link before it is brought up. The LinkSummary message contains negotiable and non-negotiable DATA_LINK objects, carrying a series of variable-length data items called subobjects, which illustrate the detailed link properties. The subobjects are defined in Section 12.12.1 in [RFC4204].

To solve the problems stated in section 3, this draft extends the LMP protocol by introducing a new DATA_LINK subobject called "Grid property", allowing the grid property correlation between adjacent nodes. The encoding format of this new subobject is as follows:



Type=TBD, Grid property type.

Grid:

The value is used to represent which grid the node/interface supports. Values defined in [RFC6205] identify DWDM [G.694.1] and CWDM [G.694.2]. The value defined in [I-D.farrkingel-ccamp-flexigrid-lambda-label] identifies flexible DWDM.

Grid	Value

Reserved	0
ITU-T DWDM	1
ITU-T CWDM	2
Flexible DWDM	3
Future use	4-16

C.F.G (central frequency granularity):

For a fixed-grid node/interface, the C.F.G value is used to represent the channel spacing, as the spacing between adjacent channels is constant. For a flexible-grid node/interface, this field should be used to represent the central frequency granularity which is the multiple of 6.25 GHz.

C.F.G (GHz)	Value
Reserved	0
100	1
50	2
25	3
12.5	4
6.25	5
Future use	6-15

S.W.G (Slot Width Granularity):

It is a positive integer value which indicates the slot width granularity which is the multiple of 12.5 GHz.

Min & Max:

Min & Max indicate the slot width tuning range the interface supports (as defined in section 2). For example, for slot width

tuning range from 25 GHz to 100 GHz (with regard to a node with slot width granularity of 12.5 GHz), the values of Min and Max should be 2 and 8 respectively. For fixed-grid nodes, these two fields are meaningless and should be set to zero.

5. Messages Exchange Procedure

5.1. Flexi-fixed Grid Nodes Messages Exchange

To demonstrate the procedure of grid property correlation, the model shown in Figure 1 is reused. Node B starts sending messages.

- o After inspecting its own node/interface property, node B sends node C a LinkSummary message including the MESSAGE ID, TE_LINK ID and DATA_LINK objects. The setting and negotiating of MESSAGE ID and TE_link ID can be referenced to [RFC4204]. As node B supports flexible-grid property, the Grid and C.S. values in the grid property subobject are set to be 3 and 5 respectively. The slot width tuning range is from 12.5 GHz to 200 GHz. Meanwhile, the N bit of the DATA_LINK object is set to 1, indicating that the property is negotiable.

- o When node C receives the LinkSummary message from B, it checks the Grid, C.S., Min and Max values in the grid property subobject. Node C can only support fixed-grid DWDM and realizes that the flexible-grid property is not acceptable for the link. Since the receiving N bit in the DATA_LINK object is set, indicating that the Grid property of B is negotiable, node C responds to B with a LinkSummaryNack containing a new Error_code object and state that the property needs further negotiation. Meanwhile, an accepted grid property subobject (Grid=2, C.S.=2, fixed DWDM with channel spacing of 50 GHz) is carried in LinkSummaryNack message. At this moment, the N bit in the DATA_LINK object is set to 0, indicating that the grid property subobject is non-negotiable.

- o As the channel spacing and slot width of node B can be configured to be any integral multiples of 6.25 GHz and 12.5 GHz respectively, node B supports the fixed DWDM values announced by node C. Consequently, node B will resend the LinkSummary message carrying the grid property subobject with values of Grid=2 and C.S.=2.

- o Once received the LinkSummary message from node B, node C replies with a LinkSummaryACK message. After the message exchange, the link between node B and C is brought up with a fixed channel spacing of 50 GHz.

In the above mentioned grid property correlation scenario, the node supporting a flexible-grid is the one that starts sending LMP

messages. The procedure where the initiator is the fixed-grid node is as follows:

- o After inspecting its own interface property, Node C sends B a LinkSummary message containing a grid property subobject with Grid=2, C.S.=2. The N bit in the DATA_LINK object is set to 0, indicating that it is non-negotiable.

- o As the channel spacing and slot width of node B can be configured to be any integral multiples of 6.25 GHz and 12.5 GHz respectively, node B is able to support the fixed DWDM parameters. Then, node B will make appropriate configuration and reply node C the LinkSummaryACK message.

- o After the message exchange, the link between node B and C is brought up with a fixed channel spacing of 50 GHz.

5.2. Flexible Nodes Messages Exchange

To demonstrate the procedure of grid property correlation between to flexi-grid capable nodes, the model shown in figure 3 is reused. The procedure of grid property correlation (negotiating the grid granularity and slot width tuning range) is similar to the scenarios mentioned above.

- o The Grid, C.S., Min and Max values in the grid property subobject sent from node F to G are set to be 3,5,1,8 respectively. Meanwhile, the N bit of the DATA_LINK object is set to 1, indicating that the grid property is negotiable.

- o When node G has received the LinkSummary message from F, it will analyze the Grid, C.S., Min and Max values in the Grid property subobject. But node G can only support grid granularity of 12.5 GHz and a slotwdith tuning range from 25 GHz to 200 GHz. Considering the property of node F, node G then will respond F a LinkSummaryNack containing a new Error_code object and state that the property need further negotiation. Meanwhile, an accepted grid property subobject (Grid=3, C.S.=4, Min=1, Max=4, the slot width tuning range is set to the intersection of Node F and G) is carried in LinkSummaryNack message. Meanwhile, the N bit in the DATA_LINK object is set to 1, indicating that the grid property subobject is non-negotiable.

- o As the channel spacing and slot width of node F can be configured to be any integral multiples of 6.25 GHz and 12.5 GHz respectively, node F can support the lager granularity. The suggested slot width tuning range is acceptable for node F. In consequence, node F will resend the LinkSummary message carrying the grid subobject with values of Grid=3, C.S.=4, Min=1 and Max=4.

- o Once received the LinkSummary message from node F, node G replies with a LinkSummaryACK message. After the message exchange, the link between node F and G is brought up supporting central frequency granularity of 12.5 GHz and slot width tuning range from 25 GHz to 100 GHz.

From the perspective of the control plane, once the links have been brought up, wavelength constraint information can be advertised and the wavelength label can be assigned hop-by-hop when establishing a LSP based on the link grid property.

6. Security Considerations

TBD.

7. IANA Considerations

TBD.

8. References

8.1. Normative references

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

[G.694.2] International Telecommunications Union, "Spectral grids for WDM applications: CWDM wavelength grid", Recommendation G.694.2, December 2003.

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

[RFC4204] Lang, J., "Link Management Protocol (LMP)", RFC 4204, October 2005.

[RFC6205] Otani, T. and D. Li, "Generalized Labels for Lambda-Switch-Capable (LSC) Label Switching Routers", RFC 6205, March 2011.

8.2. Informative References

[I-D.farrkingel-ccamp-flexigrid-lambda-label]
Farrel, A., King, D., Li, Y., Zhang, F., "Generalized Labels for the Flexi-Grid in Lambda-Switch-Capable (LSC) Label Switching Routers", draft-farrkingel-ccamp-flexigrid-lambda-label-08 (work in progress), February 2014.

[FLEX-FWK]
Dios, O., Casellas, R., Zhang, F., Fu, X., Ceccarelli, D., and I. Hussain, "Framework for GMPLS based control of Flexi-grid DWDM networks", draft-ietf-ccamp-flexi-grid-fwk-00 (work in progress), October 2013.

9. Authors' Address

Yao Li (editor)

ZTE

Email: li.yao3@zte.com.cn

Guoying Zhang (editor)

China Academy of Telecom Research, MIIT

Email: zhangguoying@catr.cn

Xihua Fu (editor)

ZTE

Email: fu.xihua@zte.com.cn

Ramon Casellas

CTTC

Email: ramon.casellas@cttc.es

Yu Wang

China Academy of Telecom Research, MIIT

Email: wangyu@catr.cn

10. Contributors' Address

Wenjuan He (editor)

ZTE

Email: he.wenjuan1@zte.com.cn

Network Working Group
Internet Draft
Intended status: Standards Track

H. Long, M.Ye
Huawei Technologies Co., Ltd
G. Mirsky
Ericsson
A. Alessandro
Telecom Italia S.p.A
H. Shah
Ciena

Expires: August 2014

February 13, 2014

OSPF Routing Extension for Links with Variable Discrete Bandwidth
draft-long-ccamp-ospf-availability-extension-02.txt

Abstract

Packet switching network may contain links with variable discrete bandwidth, e.g., copper, radio, etc. The bandwidth of such link may change discretely in reaction to changing external environment. Availability is typically used for describing such links during network planning. This document describes an extension for OSPF routing for route computation in a Packet Switched Network (PSN) which contains links with variable discrete bandwidth by introducing an optional availability sub-TLV.

Table of Contents

Conventions used in this document

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

The following acronyms are used in this draft:

OSPF Open Shortest Path First

PSN Packet Switched Network

SNR Signal-to-noise Ratio

LSP Label Switched Path

ISCD Interface Switching Capacity Descriptor

PE Provider Edge

LSA Link State Advertisement

1. Introduction

Some data communication technologies allow seamless change of maximum physical bandwidth through a set of known discrete values. For example, in mobile backhaul network, microwave links are very popular for providing connection of last hops. In case of heavy rain, to maintain the link connectivity, the microwave link may lower the modulation level since demodulating lower modulation level need lower signal-to-noise ratio (SNR). This is called adaptive modulation technology [EN 302 217]. However, lower modulation level also means lower link bandwidth. When link bandwidth reduced because of modulation down-shifting, high priority traffic can be maintained, while lower priority traffic is dropped. Similarly the cooper links may change their effective link bandwidth due to external interference.

The parameter, availability [G.827, F.1703, P.530], is often used to describe the link capacity during network planning. Assigning different availability classes to different types of service over such kind of links provides more efficient planning of link capacity. To set up an LSP across these links, availability information is required for the nodes to verify bandwidth satisfaction and make bandwidth reservation. The availability information should be inherited from the availability requirements of the services expected to be carried on the LSP, voice service usually needs "five nines" availability, while non-real time services may adequately perform at four or three nines availability.

For the route computation, the availability information should be provided along with bandwidth resource information. In this document, an extension on Interface Switching Capacity Descriptor (ISCD) [RFC4202] for availability information is defined to support in routing signaling. The extension reuses the reserved field in the ISCD and also introduces an optional availability sub-TLV.

If there is a hop that cannot support the availability sub-TLV, the availability sub-TLV is ignored.

2. Overview

A node which has link(s) with variable bandwidth attached SHOULD contain a <bandwidth, availability> information list in its OSPF TE LSA messages. The list provides the information that how much

bandwidth a link can support for a specified availability. This information is used for path calculation by the PE node(s).

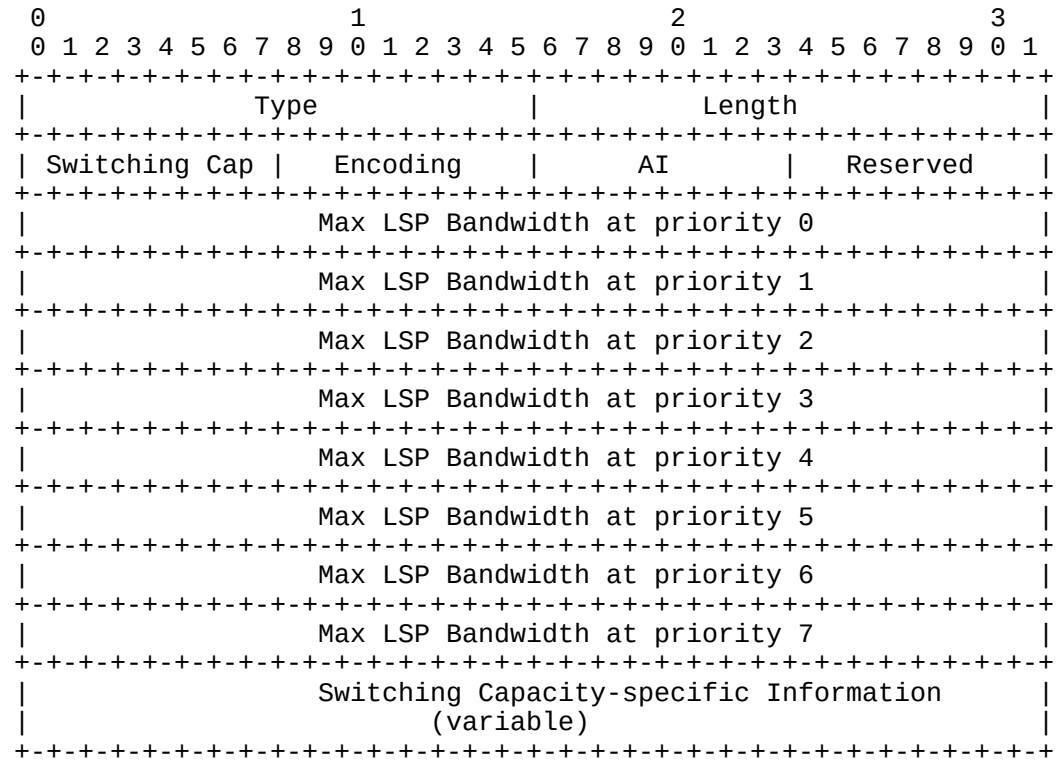
To setup an label switching path (LSP), a PE node may collect link information which is spread in OSPF TE LSA message by network nodes to get know about the network topology, and calculate out an LSP route based on the network topology, and send the calculated LSP route to signaling to initiate a PATH/RESV message for setting up the LSP.

Availability information is required to carry in the signaling message to better utilize the link bandwidth. The signaling extension for availability can be found in [ASTE].

3. Extension to OSPF Routing Protocol

3.1. Interface Switching Capacity Descriptor

The Interface Switching Capacity Descriptor (ISCD) sub-TLV [RFC 4203] has the following format:



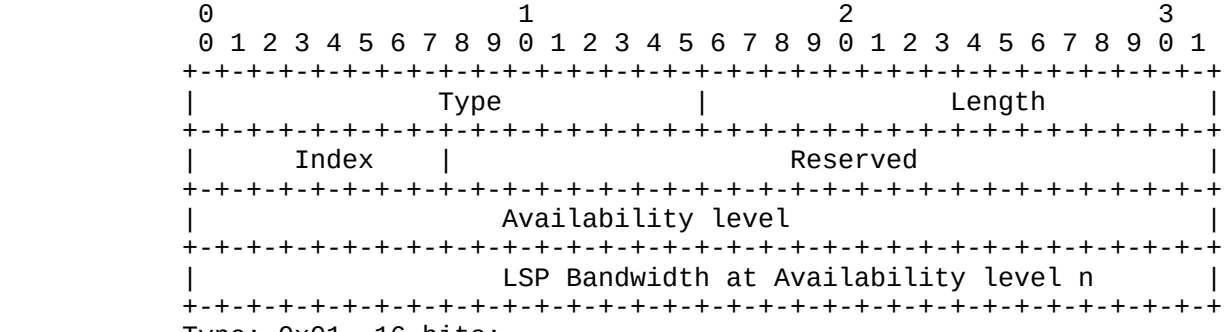
A new AI filed is defined in this document.

AI: ISCD Availability sub-TLV index, 8 bits

This new field is the index of availability sub-TLV for this ISCD sub-TLV.

3.2. ISCD Availability sub-TLV

The ISCD availability sub-TLV has the following format:



Type: 0x01, 16 bits;

Length: 16 bits;

Index: 8 bits

This field is the index of this availability sub-TLV, referred by the AI field of the ISCD sub-TLV.

Availability level: 32 bits

This field is a 32-bit IEEE floating point number which describes the availability guarantee of the switching capacity in the ISCD object which has the AI value equal to Index of this sub-TLV. The value must be less than 1.

LSP Bandwidth at Availability level n: 32 bits

This field is a 32-bit IEEE floating point number which describes the LSP Bandwidth at a certain Availability level which was described in the Availability filed.

3.3. Signaling Process

A node which has link(s) with variable bandwidth attached SHOULD contain one or more ISCD Availability sub-TLVs in its OSPF TE LSA messages. Each ISCD Availability sub-TLV provides the information that how much bandwidth a link can support for a specified availability. This information is used for path calculation by the PE node(s).

A node who doesn't support availability sub-TLV should ignore ISCD availability sub-TLV.

4. Security Considerations

This document does not introduce new security considerations to the existing OSPF protocol.

5. IANA Considerations

This document introduces an Availability sub-TLV of the ISCD sub-TLV of the TE Link TLV in the TE Opaque LSA for OSPF v2. This document proposes a suggested value for the Availability sub-TLV; it is recommended that the suggested value be granted by IANA. Initial values are as follows:

Type	Length	Format	Description
0	-	Reserved	Reserved value
0x01	8	see Section 3.2	Availability sub-TLV

6. References

6.1. Normative References

[RFC2210] Wroclawski, J., "The Use of RSVP with IETF Integrated Services", RFC 2210, September 1997.

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

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

[RFC4202] Kompella, K. and Rekhter, Y. (Editors), "Routing Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4202, October 2005.

[RFC4203] Kompella, K., Ed., and Y. Rekhter, Ed., "OSPF Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4203, October 2005.

[G.827] ITU-T Recommendation, "Availability performance parameters and objectives for end-to-end international constant bit-rate digital paths", September, 2003.

[F.1703] ITU-R Recommendation, "Availability objectives for real digital fixed wireless links used in 27 500 km hypothetical reference paths and connections", January, 2005.

[P.530] ITU-R Recommendation," Propagation data and prediction methods required for the design of terrestrial line-of-sight systems", February, 2012

[EN 302 217] ETSI standard, "Fixed Radio Systems; Characteristics and requirements for point-to-point equipment and antennas", April, 2009

[ASTE] H., Long, M., Ye, Mirsky, G., Alessandro, A., Shah, H., "RSVP-TE Signaling Extension for Links with Variable Discrete Bandwidth", Work in Progress, December, 2013

6.2. Informative References

[MCOS] Minei, I., Gan, D., Kompella, K., and X. Li, "Extensions for Differentiated Services-aware Traffic Engineered LSPs", Work in Progress, June 2006.

7. Acknowledgments

Authors' Addresses

Hao Long
Huawei Technologies Co., Ltd.
No.1899, Xiyuan Avenue, Hi-tech Western District
Chengdu 611731, P.R.China

Phone: +86-18615778750
Email: longhao@huawei.com

Min Ye (editor)
Huawei Technologies Co., Ltd.
No.1899, Xiyuan Avenue, Hi-tech Western District
Chengdu 611731, P.R.China

Email: amy.yemin@huawei.com

Greg Mirsky (editor)
Ericsson

Email: gregory.mirsky@ericsson.com

Alessandro D'Alessandro
Telecom Italia S.p.A

Email: alessandro.dalessandro@telecomitalia.it

Himanshu Shah
Ciena Corp.
3939 North First Street
San Jose, CA 95134
US

Emial: hshah@ciena.com

Network Working Group
Internet Draft
Intended status: Standards Track

H. Long, M. Ye
Huawei Technologies Co., Ltd
G. Mirsky
Ericsson
A. Alessandro
Telecom Italia S.p.A
H. Shah
Ciena

Expires: August 2014

February 13, 2014

RSVP-TE Signaling Extension for Links with Variable Discrete Bandwidth

draft-long-ccamp-rsvp-te-bandwidth-availability-03.txt

Abstract

Packet switching network may contain links with variable bandwidth, e.g., copper, radio, etc. The bandwidth of such link is sensitive to external environment. Availability is typically used for describing the link during network planning. This document describes an extension for RSVP-TE signaling for setting up a label switching path (LSP) in a Packet Switched Network (PSN) network which contains links with discretely variable bandwidth by introducing an optional availability field in RSVP-TE signaling.

Status of this Memo

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

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

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

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt

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

This Internet-Draft will expire on August 13, 2014.

Table of Contents

Conventions used in this document

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

The following acronyms are used in this draft:

RSVP-TE Resource Reservation Protocol-Traffic Engineering

LSP Label Switched Path

PSN Packet Switched Network

SNR Signal-to-noise Ratio

TLV Type Length Value

PE Provider Edge

LSA Link State Advertisement

1. Introduction

The RSVP-TE specification [RFC3209] and GMPLS extensions [RFC3473] specify the signaling message including the bandwidth request for setting up a label switching path in a PSN network.

Some data communication technologies allow seamless change of maximum physical bandwidth through a set of known discrete values. For example, in mobile backhaul network, microwave links are very popular for providing connection of last hops. In case of heavy rain, to maintain the link connectivity, the microwave link may lower the modulation level since demodulating lower modulation level need lower signal-to-noise ratio (SNR). This is called adaptive modulation technology [EN 302 217]. However, lower modulation level also means lower link bandwidth. When link bandwidth reduced because of modulation down-shifting, high priority traffic can be maintained, while lower priority traffic is dropped. Similarly the cooper links may change their link bandwidth due to external interference.

The parameter, availability [G.827, F.1703, P.530], is often used to describe the link capacity during network planning. A more detailed example on the bandwidth availability can be found in Appendix A. Assigning different availability classes to different types of service over such kind of links provides more efficient planning of link capacity. To set up a LSP across these links, availability information is required for the nodes to verify bandwidth satisfaction and make bandwidth reservation. The availability information should be inherited from the availability requirements of the services expected to be carried on the LSP, voice service usually needs "five nines" availability, while non-real time services may adequately perform at four or three nines availability. Since different service types may need different availabilities guarantee, multiple <availability, bandwidth> pairs may be required when signaling.

If the availability requirement is not specified in the signaling message, the bandwidth will be reserved as the highest availability. For example, the bandwidth with 99.999% availability of a link is 100Mbps; the bandwidth with 99.99% availability is 200Mbps. When a

video application requests for 120Mbps without availability requirement, the system will compare 120Mbps with 100Mbps, therefore cannot set up the LSP path. But in fact, video application doesn't need 99.999% availability, 99.99% availability is enough. In this case, the LSP could be set up if availability is specified in the signaling message.

To fulfill LSP setup by signaling in these scenarios, this document specifies a new availability sub-TLV as the sub-TLV of Ethernet bandwidth profiles [RFC6003]. Multiple bandwidth profiles with different availability can be carried in the SENDER_TSPEC object.

2. Overview

A PSN tunnel may span one or more links in a network. To setup a label switching path (LSP), a PE node may collect link information which is spread in routing message, e.g., OSPF TE LSA message, by network nodes to get to know about the network topology, and calculate out an LSP route based on the network topology, and send the calculated LSP route to signaling to initiate a PATH/RESV message for setting up the LSP.

In case that there is(are) link(s) with variable discrete bandwidth in a network, a <bandwidth, availability> requirement list should be specified for an LSP. Each <bandwidth, availability> pair in the list means that listed bandwidth with specified availability is required. The list could be inherited from the results of service planning for the LSP.

A node which has link(s) with variable discrete bandwidth attached SHOULD contain a <bandwidth, availability> information list in its OSPF TE LSA messages. The list provides the information that how much bandwidth a link can support for a specified availability. This information is used for path calculation by the PE node(s). The routing extension for availability can be found in [ARTE].

When a PE node initiates a PATH/RESV signaling to set up an LSP, the PATH message SHOULD carry the <bandwidth, availability> requirement list as bandwidth request. Intermediate node(s) will allocate the bandwidth resource for each availability requirement from the remaining bandwidth with corresponding availability. An error message may be returned if any <bandwidth, availability> request cannot be satisfied.

3. Extension to RSVP-TE Signaling

The RSVP-TE signaling extension in this document is based on RFC6003: a new sub-TLV for Ethernet Bandwidth Profile TLV is defined.

3.1.1. Availability sub-TLV

The Ethernet Bandwidth Profile TLV in RFC6003 has the following format. A new filed is defined in this document as shown in Figure 1.

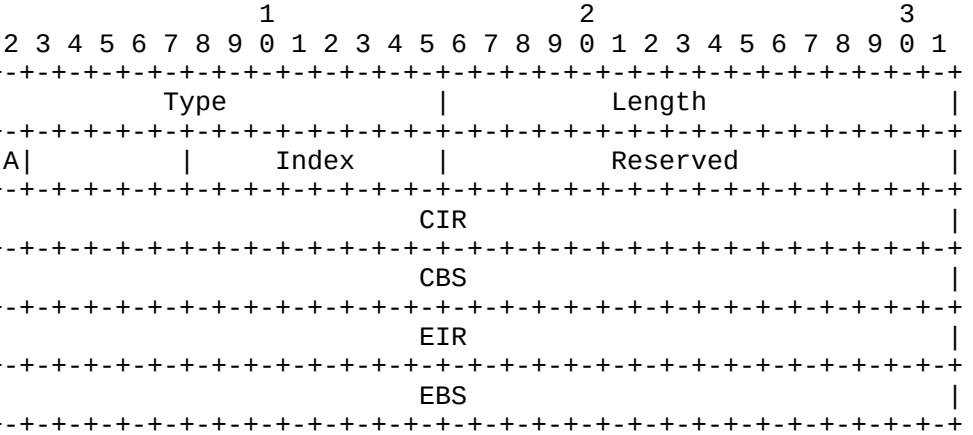


Figure 1: A new "AF" filed in Ethernet Bandwidth Profile TLV

A new filed is defined in this document:

AF filed (bit 2): Availability Field (AF)

If the AF filed is set to 1, Availability sub-TLV MUST be included in the Bandwidth Profile TLV. If the AF field is set to value 0, then an Availability sub-TLV SHOULD NOT be included. The availability sub-TLV has the following format:

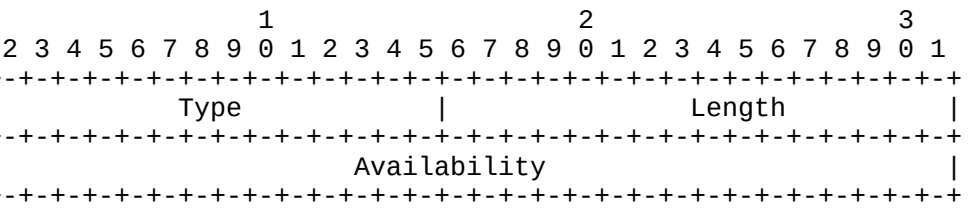


Figure 2: Availability sub-TLV

Type (2 octets): TBD

Length (2 octets): 4

Availability (4 octets): a 32-bit floating number describes availability requirement for this bandwidth request. The value must be less than 1.

As the Ethernet Bandwidth Profile TLV can be carried for one or more times in the Ethernet SENDER_TSPEC object, the availability sub-TLV can also be present for one or more times.

3.2. FLOWSPEC Object

The FLOWSPEC object (Class-Num = 9, Class-Type = TBD) has the same format as the Ethernet SENDER_TSPEC object.

3.3. Signaling Process

The source node initiates PATH messages including one or more Bandwidth Profile TLVs with different availability value in the SENDER_TSPEC object. Each Bandwidth Profile TLV specifies the portion of bandwidth request with referred availability requirement.

The destination node checks whether it can satisfy the bandwidth requirements by comparing each bandwidth requirement inside the SENDER_TSPEC objects with the remaining link sub-bandwidth resource with respective availability guarantee when received the PATH message.

- If all <bandwidth, availability> requirements can be satisfied, it should reserve the bandwidth resource from each remaining sub-bandwidth portion to set up this LSP. Optionally, the higher availability bandwidth can be allocated to lower availability request when the lower availability bandwidth cannot satisfy the request.

- If at least one <bandwidth, availability> requirement cannot be satisfied, it should generate PathErr message with the error code "Admission Control Error" and the error value "Requested Bandwidth Unavailable" (see [RFC2205]).

If two LSP request for the bandwidth with the same availability requirement, a way to resolve the contention is comparing the node ID, the node with the higher node ID will win the contention. More details can be found in [RFC3473].

If a node does not support the Availability sub-TLV, then it MUST ignore the sub-TLV and only use the bandwidth request in the

Ethernet Bandwidth Profile TLV. The [RFC6003] states that a node that does not support a flag should ignore it. Thus a legacy implementation will ignore the Availability Flag.

4. Security Considerations

This document does not introduce new security considerations to the existing RSVP-TE signaling protocol.

5. IANA Considerations

IANA maintains registries and sub-registries for RSVP-TE used by GMPLS. IANA is requested to make allocations from these registries as set out in the following sections.

5.1 Ethernet Bandwidth Profile TLV

IANA maintains a registry of GMPLS parameters called "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Parameters".

IANA has created a new sub-registry called "Ethernet Bandwidth Profiles" to contain bit flags carried in the Ethernet Bandwidth Profile TLV of the Ethernet SENDER_TSPEC object.

Bits are to be allocated by IETF Standards Action. Bits are numbered from bit 0 as the low order bit. A new bit field is as follow:

Bit	Hex	Description	Reference
2	0x03	Availability Field (AF)	[This ID]

Sub-TLV types for Ethernet Bandwidth Profiles are to be allocated by IETF Standard Action. Initial values are as follows:

Type	Length	Format	Description
0	-	Reserved	Reserved value
TBD	4	see Section 3.1	Availability sub-TLV

6. References

6.1. Normative References

[RFC2210] Wroclawski, J., "The Use of RSVP with IETF Integrated Services", RFC 2210, September 1997.

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

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

[RFC6003] Papadimitriou, D. "Ethernet Traffic Parameters", RFC 6003, October 2010.

[G.827] ITU-T Recommendation, "Availability performance parameters and objectives for end-to-end international constant bit-rate digital paths", September, 2003.

[F.1703] ITU-R Recommendation, "Availability objectives for real digital fixed wireless links used in 27 500 km hypothetical reference paths and connections", January, 2005.

[P.530] ITU-R Recommendation," Propagation data and prediction methods required for the design of terrestrial line-of-sight systems", February, 2012

[EN 302 217] ETSI standard, "Fixed Radio Systems; Characteristics and requirements for point-to-point equipment and antennas", April, 2009

[ARTE] H., Long, M., Ye, Mirsky, G., Alessandro, A., Shah, H., "OSPF Routing Extension for Links with Variable Discrete Bandwidth", Work in Progress, December, 2013

6.2. Informative References

[MCOS] Minei, I., Gan, D., Kompella, K., and X. Li, "Extensions for Differentiated Services-aware Traffic Engineered LSPs", Work in Progress, June 2006.

7. Acknowledgments

The authors would like to thank Khuzema Pithewan, Lou Berger, Yuji Tochio, Dieter Beller, and Autumn Liu for their comments on the document.

Appendix A

Presuming that a link has three discrete bandwidth levels:

The link bandwidth under modulation level 1, e.g., QPSK, is 100M;

The link bandwidth under modulation level 2, e.g., 16QAM, is 200M;

The link bandwidth under modulation level 3, e.g., 256QAM, is 400M.

In sunny day, the modulation level 3 can be used to achieve 400M link bandwidth.

A light rain with X mm/h rate triggers the system to change the modulation level from level 3 to level 2, with bandwidth changing from 400M to 200M. The probability of X mm/h rain in the local area is 53 minutes in a year. Then the dropped 200M bandwidth has 99.99% availability.

A heavy rain with Y(Y>X) mm/h rate triggers the system to change the modulation level from level 2 to level 1, with bandwidth changing from 200M to 100M. The probability of Y mm/h rain in the local area is 26 minutes in a year. Then the dropped 100M bandwidth has 99.995% availability.

For the 100M bandwidth of the modulation level 1, only the extreme weather condition can cause the whole system unavailable, which only happens for 5 minutes in a year. So the 100M bandwidth of the modulation level 1 owns the availability of 999.99%.

In a word, the maximum bandwidth is 400Mbps. According to the weather condition, the sub-bandwidth and its availability are shown as follows:

Sub-bandwidth(Mbps)	Availability
200	99.99%
100	99.995%

100	99.999%

Authors' Addresses

Hao Long
Huawei Technologies Co., Ltd.
No.1899, Xiyuan Avenue, Hi-tech Western District
Chengdu 611731, P.R.China

Phone: +86-18615778750
Email: longhao@huawei.com

Min Ye (editor)
Huawei Technologies Co., Ltd.
No.1899, Xiyuan Avenue, Hi-tech Western District
Chengdu 611731, P.R.China

Email: amy.yemin@huawei.com

Greg Mirsky (editor)
Ericsson

Email: gregory.mirsky@ericsson.com

Alessandro D'Alessandro
Telecom Italia S.p.A

Email: alessandro.dalessandro@telecomitalia.it

Himanshu Shah
Ciena Corp.
3939 North First Street
San Jose, CA 95134
US

Emial: hshah@ciena.com

Routing Working Group
Internet-Draft
Intended status: Informational
Expires: August 15, 2014

M. Jethanandani
Ciena Corporation

February 11, 2014

Analysis of LMP Security According to KARP Design Guide draft-mahesh-karp-lmp-analysis-01.txt

Abstract

This document analyzes Link Management Protocol (LMP) according to guidelines set forth in section 4.2 of KARP Design Guidelines (RFC 6518).

Status of This Memo

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

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

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

This Internet-Draft will expire on August 15, 2014.

Copyright Notice

Table of Contents

1. Introduction

In March 2006, the Internet Architecture Board (IAB) described an attack on core routing infrastructure as an ideal attack that would inflict the greatest amount of damage, in their Report from the IAB workshop on Unwanted Traffic March 9-10, 2006 [RFC4948], and suggested steps to tighten the infrastructure against the attack. Four main steps were identified for that tightening:

1. Create secure mechanisms and practices for operating routers.

2. Clean up the Internet Routing Registry (IRR) repository, and securing both the database and the access, so that it can be used for routing verifications.

3. Create specifications for cryptographic validation of routing message content.

4. Secure the routing protocols' packets on the wire.

In order to secure the routing protocols this document performs an initial analysis of the current state of LMP according to the requirements of KARP Design Guidelines [RFC6518]. This draft builds on several previous analysis efforts into routing security:

- Issues with existing Cryptographic Protection Methods for Routing Protocols [RFC6039] an analysis of cryptographic issues with routing protocols.

- Analysis of OSPF Security According to KARP Design Guide [RFC6863].

- Analysis of BGP, LDP, PCEP, and MSDP Issues According to KARP Design Guide [RFC6952] which is a analysis of the four routing protocols.

Link Management Protocol (LMP) [RFC4204] is used to manage Traffic Engineering (TE) links. According to the document, LMP can be subject to a number of attacks. Some examples include:

- an adversary may spoof control packets

- an adversary may modify the control packet in transit

- an adversary may replay control packets

- an adversary may study a number of control packets and try to break the key using cryptographic tools.

Section 2 looks at the current security state of LMP. Section 3 suggest an optimal security state and section 4 does an analysis of the gap between the existing and the optimal security state of the protocol and suggest some areas where we need to improve.

1.1. Abbreviations

LMP - Link Management Protocol

TE - Traffic Engineering

2. Current Assessment of LMP

This section looks at LMP procedure, the underlying transport layer and security assessment associated with LMP.

2.1. LMP Procedure

The two core procedures of LMP procedure are control channel management and link property correlation. Control channel management is used to establish and maintain control channels between adjacent nodes. This is done using a Config message exchange and a fast keep-alive mechanism between the nodes. Link property correlation is used

to synchronize the TE link properties and verify the TE link configuration.

Two additional procedures include link connectivity verification and fault management. Link connectivity verification is used for data plane discovery, Interface_Id exchange, and physical connectivity verification. This is done by sending Test messages over the data channel and the TestStatus messages coming back over the control plane. The LMP link connectivity verification procedure is coordinated using the BeginVerify message exchanged over the control channel.

The LMP fault management procedure is based on a ChannelStatus message exchange. The ChannelStatus message is sent unsolicited and is used to notify an LMP neighbor about the status of one or more data channels. ChannelStatusAck is used to acknowledge receipt of the ChannelStatus message. Similarly, a ChannelStatusResponse message is used to acknowledge receipt of a ChannelStatusRequest message.

2.2. Transport Layer

Except for Test messages, all LMP packets use UDP to communicate with its piers over a LMP port number. Multiple "LMP adjacencies" may be formed and be active between two nodes. LMP messages are transmitted reliably using Message_Ids and retransmissions.

Unlike TCP which can use TCP-AO [RFC5925] for message authentication, UDP does not have any of authenticating packets.

2.3. Message Integrity and Node Authentication

LMP [RFC4204] recommends the use of IPSec for authentication. That document also states that there is currently no requirement that LMP headers or payload be encrypted. It also states that LMP endpoint identity does not need to be protected.

To authenticate LMP, the document further states that manual keying mode be supported. However, it notes that manual keying cannot effectively support replay protection and automatic re-keying. It therefore recommends that manual keying should only be used for diagnostic purposes and only use automatic re-keying for replay protection and automatic re-keying.

2.4. Replay Attack

MESSAGE_ID and MESSAGE_ID_ACK objects are included in the LMP messages to support reliable message delivery. The Message_Id field of the MESSAGE_ID object contains a generator selected value. This value is supposed to be monotonically increasing. A value is considered to be used when it has been sent in an LMP message with the same CC_Id or LMP adjacency. The Message_Id field of the MESSAGE_ID_ACK contains the Message_Id field of the message being acknowledged.

Unacknowledged messages sent with the MESSAGE_ID object are to be retransmitted until the message is acknowledged or until a retry limit is reached. The Message_Id field is 32 bit wide and may wrap.

The 32-bit Message_Id number space is not large enough to guarantee that the Message_Id number will not wrap around within a reasonable long period. Therefore, the system is susceptible to a replay attack.

In addition, LMP does not provide for a generation of a unique monotonically increasing sequence numbers across a failure or a restart.

2.5. Out-of-order Protection

LMP states that nodes processing incoming messages are supposed to check to see if the newly received message is out of order messages, and if so, they are to be ignored and dropped silently.

Specifically, if the message is a Config message, and the Message_Id value is less than the largest Message_Id value previously received from the sender for the CC_Id, then the message is supposed to be treated as being out-of-order. If the message is a LinkSummary message and the Message_Id value is less than the largest Message_Id value previously received from the sender of the TE link, then the message is supposed to be treated as being out-of-order. Similarly, if the message is a ChannelStatus message and the Message_Id value is less than the largest Message_id value previously received from he sender of the specific TE link, then the receiver is supposed to check for the Message_Id value previously received from the state of each data channel included in the ChannelStatus message. If the Message_Id value associated with at least one of the data channels included in the message, the message is not supposed to be treated as out-of-order. All other messages are not supposed to be treated as out-of-order.

3. Security Requirements for LMP

LMP [RFC4204] states that the following requirements should be applied to secure the protocol.

- LMP security must be able to provide authentication, integrity and replay protection.
- Confidentiality is not needed for LMP traffic.
- The protection of identity of the LMP end-points is not commonly required.
- The security mechanism should provide for a well defined key management scheme. The key management scheme should be scalable and should provide for automatic key rollover.
- The algorithm used for authentication must be cryptographically sound and it should provide for algorithm agility.

4. Gap Analysis for LMP

This section outlines the differences between the current state of LMP and the desired state as outlined in sections 4.1 and 4.2 of KARP Design Guidelines [RFC6518].

4.1. Replay Protection

As outlined above, LMP protocol is subject to replay attacks. Solutions to replay protection include:

1. Maintaining Message_Id numbers in stable memory
2. Introducing the data from a local time clock into the generation of Message_Id numbers after a restart
3. Introducing the timing information from a Network Recovered Clock into the generation of Message_Id numbers after a restart.

In addition, a handshake is defined for a receiver to get the latest value of a Message_Id number. Therefore, this solution is effective in addressing the issues caused by the rollback of Message_Id numbers across a system restart or failure. However, when a router uses the approach to generating Message_Id numbers with the time information from NTP, an attacker may try to deceive the router to generate a Message_Id number which is less than the Message_Id numbers it used to have, by sending replayed or foiled NTP information.

5. IANA Requirements

This document makes no IANA requests, and the RFC Editor may consider deleting this section on publication of this document as a RFC.

6. Security Consideration

This document is all about security considerations for LMP.

7. Acknowledgements

8. References

8.1. Normative References

[RFC4204] Lang, J., "Link Management Protocol (LMP)", RFC 4204, October 2005.

[RFC6518] Lebovitz, G. and M. Bhatia, "Keying and Authentication for Routing Protocols (KARP) Design Guidelines", RFC 6518, February 2012.

8.2. Informative References

[RFC4948] Andersson, L., Davies, E., and L. Zhang, "Report from the IAB workshop on Unwanted Traffic March 9-10, 2006", RFC 4948, August 2007.

[RFC5925] Touch, J., Mankin, A., and R. Bonica, "The TCP Authentication Option", RFC 5925, June 2010.

[RFC6039] Manral, V., Bhatia, M., Jaeggli, J., and R. White, "Issues with Existing Cryptographic Protection Methods for Routing Protocols", RFC 6039, October 2010.

[RFC6863] Hartman, S. and D. Zhang, "Analysis of OSPF Security According to the Keying and Authentication for Routing Protocols (KARP) Design Guide", RFC 6863, March 2013.

[RFC6952] Jethanandani, M., Patel, K., and L. Zheng, "Analysis of BGP, LDP, PCEP, and MSDP Issues According to the Keying and Authentication for Routing Protocols (KARP) Design Guide", RFC 6952, May 2013.

Author's Address

Mahesh Jethanandani
Ciena Corporation
3939 North 1st Street
San Jose, CA 95134
USA

Phone: +1 (408) 904-2160
Email: mjethanandani@gmail.com

Internet Engineering Task Force (IETF)
Internet-Draft
Intended status: Informational
Expires: August 17, 2014

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

February 13, 2014

Requirements for Very Fast Setup of GMPLS LSPs
draft-malis-ccamp-fast-lsps-01

Abstract

The Defense Advanced Research Projects Agency (DARPA) Core Optical Networks (CORONET) program has laid out a vision for the next evolution of 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 one second to one minute holding times) 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 and wavelength-km.

This document discusses the requirements for extensions to Generalized Multi-Protocol Label Switching (GMPLS) signaling for expediting the control of Label Switched Paths (LSPs), including sub-wavelengths (e.g., OTN ODUs) and full wavelengths, in order to satisfy application requirements laid out in this program.

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/.

Table of Contents

1. Introduction

The Defense Advanced Research Projects Agency (DARPA) Core Optical Networks (CORONET) program [Chiu] has laid out a vision for the next evolution of IP and optical commercial and government networks, with a focus on highly dynamic and resilient multi-terabit core networks. The program anticipates an environment where 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 are two to three orders of magnitude faster than possible with current connection setup

protocols. 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. It is the desired goal of the program to achieve transition of these advances to commercial and government networks in the next few years. 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).

Thus, CORONET anticipates the need for rapid (sub-second) setup and restoration times for high-churn (up to tens of requests per second network-wide and one second to one minute holding times) 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 and wavelength-km.

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.

However, while the current GMPLS constituent 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 low connection setup times while maintaining a high success ratio (i.e., low blocking) in a high-churn environment. For example, in Internet2, a network which provides CORONET-like high bandwidth circuit services for the Research & Education community, a circuit is currently established, on average, roughly at a rate of one per hour. In contrast, the CORONET vision is a churn rate of up to tens of circuits per second, over four orders of magnitude greater.

Furthermore, 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. The purpose of this draft is to determine the requirements to augment the GMPLS framework to allow specific applications, or users, to rapidly set up connections over GMPLS networks with minimal delays and a high probability of success.

Preliminary simulations and analyses of national and global scale networks, both WSON and sub-wavelength OTN, have shown that using current GMPLS protocols and procedures does not meet the CORONET 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 of the factors affecting these results in a highly dynamic network include:

1. Stale TE information when the connection request rate exceeds TE information update rate based on OSPF-TE LS updates. This leads to increased blocking and indirectly to longer setup delays.

2. Real-time path computation and PCE communication, i.e., following connection request, thus increasing setup delays.

3. Cross-connection procedures resulting in accumulating cross-connection delays when cross-connection must be completed before the Resv signaling message is propagated upstream. This contribution may be significant in WSON but less so with TDM or L2 switching.

4. Crankbacks.

2. Scope and Motivation

[RFC6163] provides the framework, basic elements, and terminology of wavelength switched optical networks (WSON) and wavelength-based LSPs. These basic elements generally apply to other GMPLS technologies as well, e.g., spectral switching (SSON), sub-wavelength TDM, and L2 LSPs. This draft refers to the same general framework and technologies, but addresses an extension of the general problem space addressed in [RFC6163]. Specifically, this draft addresses the requirements of expediting LSP setup, under heavy connection churn scenarios, while achieving low blocking, under an overall distributed control plane. Once there is agreement on the requirements, further drafts will describe the procedures and signaling contents required to meet the requirements (potentially more than one if separate standard track drafts are found necessary for wavelength and sub-wavelength LSPs). Both single-domain and multi-domain network scenarios are addressed. 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.

The motivation for GMPLS extensions as described here is thus two-fold:

1. The anticipated need for rapid setup while maintaining low blocking, on-demand, of large bandwidth 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 performance of current GMPLS protocols and procedures in networks with the above characteristics.

The ability to setup circuit-like LSPs for large bandwidth flows and 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 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 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 maximum frequency of TE information updates is not sufficient to provide adequate path computation and resource allocation, as network conditions and resource attributes may be changing faster than can be reflected in OSPF-TE updates.

Thus, GMPLS and routing protocol traffic engineering (e.g. OSPF-TE) extensions are also 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.

3. Requirements for Very Fast Setup of GMPLS LSPs

This section lists the requirements for very fast setup of GMPLS LSPs in order to provide the services described in the previous sections. They will be the basis for future standards-track drafts to satisfy these requirements. Some of these requirements may be implementation-dependent to some extent, but they may also have LSP signaling protocol dependencies as well. Protocols that satisfy these requirements can be further compared based on other important factors such as resource efficiency, and implementation complexity.

The requirements are divided in two general categories - control plane requirements and network requirements. Note that network requirements essentially reflect DARPA CORONET program requirements, but anticipate cloud and other emerging application requirements. The networks considered in the CORONET program are primarily long haul national and global networks. 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.

3.1. Control Plane Requirements

R1 Protocol extensions must be backward compatible with existing GMPLS control plane protocols.

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

R3 Must support the use of PCE for path computation, and in particular the PCE-based approach for multi-domain LSPs in [RFC5441].

3.2. Network Requirements

R4 Must have an LSP setup time less than or equal to 100 ms for intra-continental LSPs, and less than or equal to 250 ms for transcontinental LSPs, including PCE path computation delays.

R5 Must support LSP holding times of one second to one minute.

R6 While there are implementation-dependent aspects of supporting high LSP setup rates, the protocol aspects of LSP signaling must

not preclude LSP request rates of tens per second. A possible example of a protocol aspect is the ability to update the IGP TE database to accurately reflect resource availability at all times. Note that LSP request rates may be dependent on LSP bandwidth, where very high bandwidth LSPs (such as for an entire wavelength) could be less frequent than lower-rate LSPs (such as an ODUx connection).

R7 Must support restoration for all cases of single node or link failures.

R8 At most one blocked LSP setup request per 1000 requests. LSP setup blocking depends on network variables (topology, available resources) and on the setup protocol. The choice of selected protocol is primarily determined by the level of resource utilization.

4. IANA Considerations

This memo includes no request to IANA.

5. 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, and potentially could improve GMPLS' resistance against denial of service attacks that attempt to deny service through the use of a high frequency of GMPLS LSP setup requests.

6. 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.

7. References

7.1. Normative References

[RFC5441] Vasseur, JP., Zhang, R., Bitar, N., and JL. Le Roux, "A Backward-Recursive PCE-Based Computation (BRPC) Procedure to Compute Shortest Constrained Inter-Domain Traffic Engineering Label Switched Paths", RFC 5441, April 2009.

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

[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.

7.2. 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>.

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

John E. Drake
Juniper

Email: jdrake@juniper.net

Vishnu Shukla
Verizon Communications

Email: vishnu.shukla@verizon.com

CCAMP
Internet-Draft
Intended status: Standards Track
Expires: August 16, 2014

G. Martinelli, Ed.
Cisco
D. Siracusa, Ed.
CREATE-NET
X. Zhang, Ed.
Huawei Technologies
G. Galimberti
Cisco
A. Zanardi
CREATE-NET

February 12, 2014

Information Encoding for WSON with Impairments Validation
draft-martinelli-ccamp-wson-iv-encode-03

Abstract

Impairment-Aware (IA) Routing and Wavelength Assignment (RWA) function might be required in Wavelength Switched Optical Networks (WSON) that already support RWA. This document defines proper encoding to support this operation. It goes in addition to the available impairment-free WSON encoding and it is fully compatible with it.

As the information model, the encoding is independent from control plane architectures and protocol implementations. Its definitions can be used in related protocol extensions.

Status of This Memo

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

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

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

This Internet-Draft will expire on August 16, 2014.

Table of Contents

1. Introduction

In case of WSON where optical impairments play a significant role, the framework document [RFC6566] defines related control plane architectural options for Impairment Aware Routing and Wavelength Assignment (IA-RWA). This document provides a suitable encoding for the related WSON impairment information model as defined [I-D.martinelli-ccamp-wson-iv-info].

This document directly refers to ITU recommendations [ITU.G680] and [ITU.G697] as already detailed in the information model.

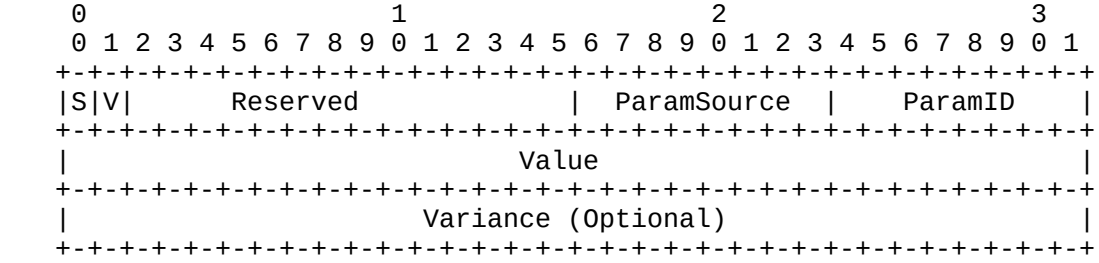
1.1. Requirements Language

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

2. Encoding

2.1. Optical Parameter

The OPTICAL_PARAM is defined as a sub TLV object.



The following flag is defined:

S. Standard bit.
S=1 indentifies a set of parameters standardized by ITU; while S=0 identifies a non-standardized set of parameters.

V. Variance bit.
V=0 only parameter value, V=1 parameter value and variance.

With the flag S=1 the following parameters are defined:

ParamSource. Where this parameter is defined. Currently only [ITU.G697] has defined this with value 1.

ParamID. Parameter identifier according to the source. [ITU.G697] table V.3 defines the following identifiers:

1. Total Power (dBm)

2. Channel Power (dBm)

3. Reserved (Defined in [ITU.G697] but not used)

4. Reserved (Defined in [ITU.G697] but not used)

5. OSNR (db)

6. Q Factor (a pure number)

7. PMD (ps)

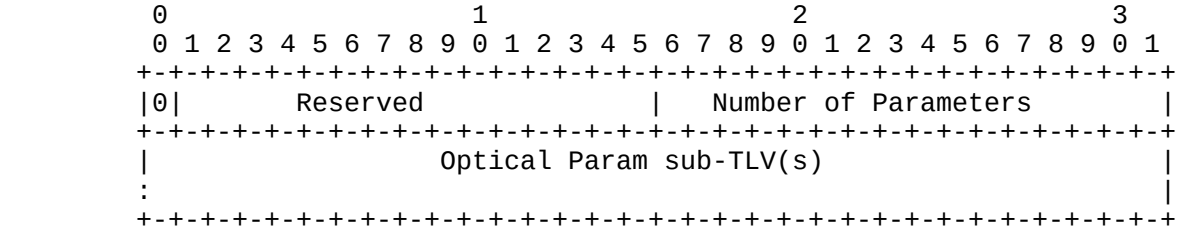
8. Residual Chromatic Dispersion (ps/nm)

Value. Value for the parameter. As defined by [ITU.G697], it is a 32 bit IEEE floating point number.

Variance. Variance for the parameter, a 32 bit IEEE floating point number.

2.2. Impairment Vector

This sub-TLV is a list of optical parameters and they MAY have a wavelength dependency information.

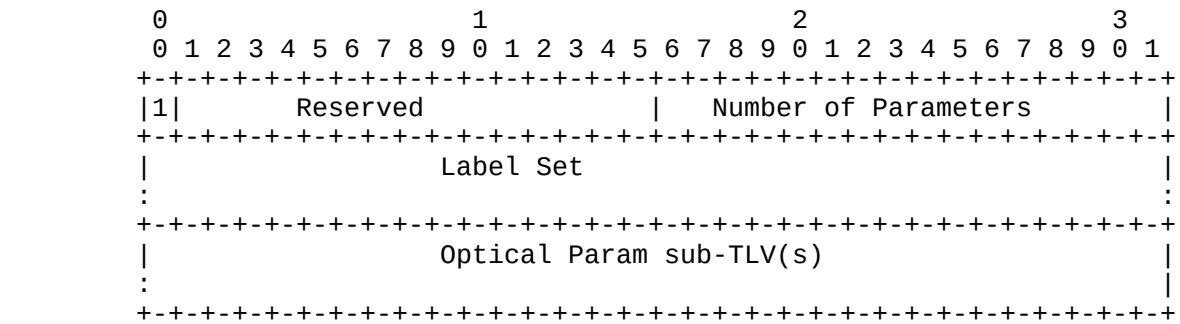


Where:

W = 0. Wavelength Dependency flag. There is no wavelength dependency.

Number of Parameters contained in this vector.

Optical Param sub-TLV(s) present a list of Object as defined in Section 2.1.



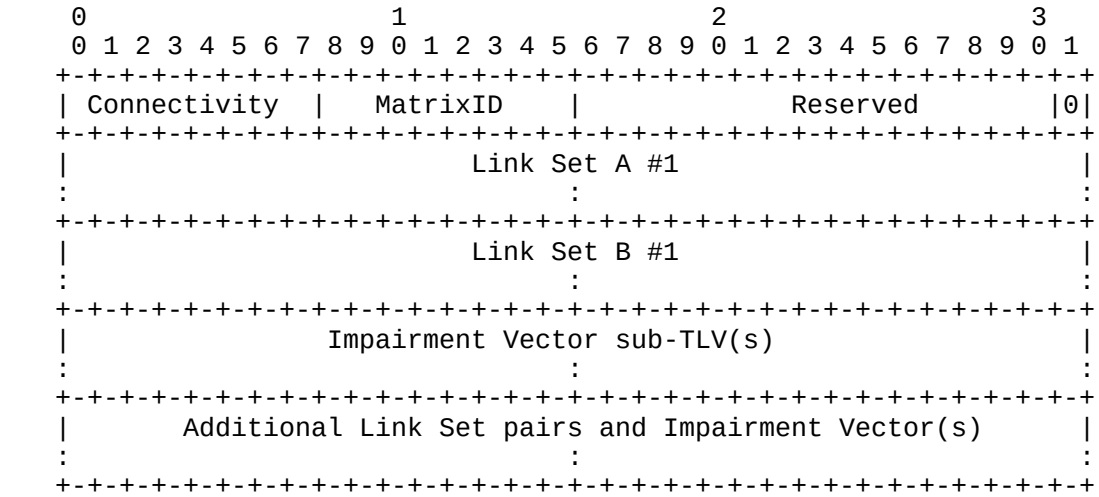
Where:

W = 1. Wavelength Dependency flag. There is wavelength dependency.

The Label Set object is defined in [I-D.ietf-ccamp-general-constraint-encode] Section 2.1. Likely an inclusive range will be the only option required by the Action defined in the Label Set.

2.3. Impairment Matrix

As defined by the [I-D.martinelli-ccamp-wson-iv-info], the impairment matrix follows the same structure as the connectivity matrix.



Where:

Connectivity: value MUST be 2 for the impairment matrix (Values 0 and 1 are already defined by [I-D.ietf-ccamp-general-constraint-encode].

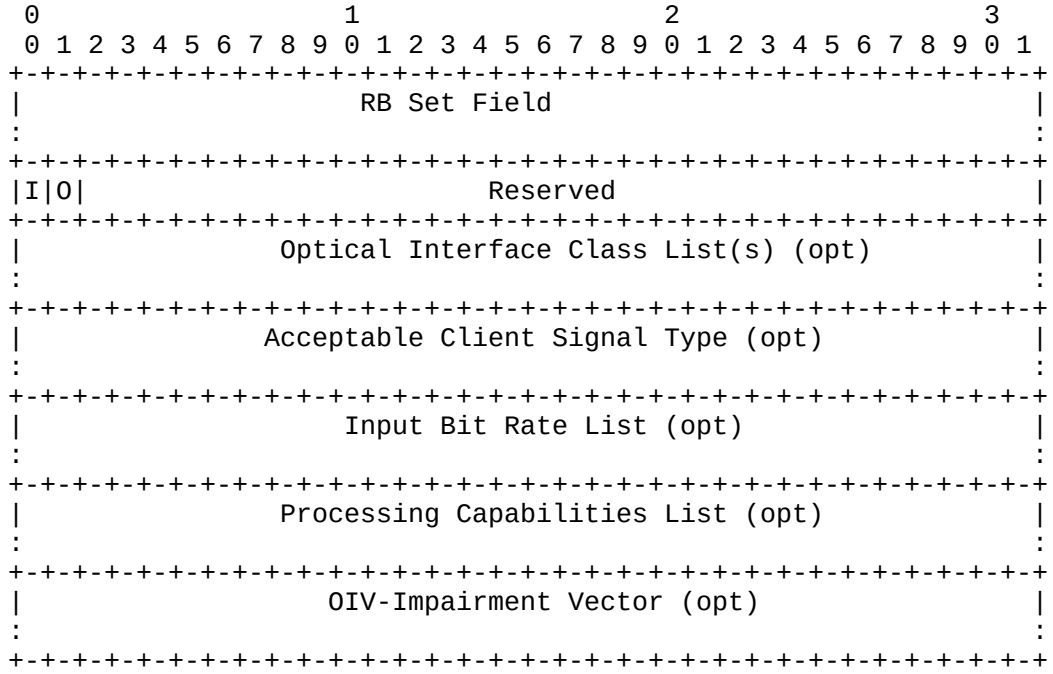
MatrixID: matrix identifier, the scope of this integer number is shared with [I-D.ietf-ccamp-rwa-info].

N: Node scope flag. With this flag set there's no Link Set information but only a list of optical parameters TLVs that apply to the whole optical node.

The usage of multiple matrixes with connectivity type equal to 2 (Impairment Matrix) MIGHT be used to grup optical parameters by connectivity. For example, if a subset of parameters apply to the whole node, a unique matrix with flag N=1 is used. At the same some another subset of parameters applies only to some LinkSet pairs, a specific Impairment Matrix will be added.

2.4. Resource Block Information

As defined by [I-D.martinelli-ccamp-wson-iv-info], the concept of resource block is extended to support the description of the impairments related to that block. The encoding follows the same structure as the one defined in [I-D.ietf-ccamp-rwa-wson-encode], with the addition of an optional Impairment Vector sub-object:



The Impairment Vector is defined within Section 2.2. All the other fields are defined within [I-D.ietf-ccamp-rwa-wson-encode].

3. Acknowledgements

TBD

4. Contributing Authors

This document was the collective work of several authors. The text and content of this document was contributed by the editors and the co-authors listed below (the contact information for the editors appears in appropriate section and is not repeated below):

Moustafa Kattan
Cisco
DUBAI,   500321
UNITED ARAB EMIRATES

Email: mkattan@cisco.com

Young Lee
Huawei
1700 Alma Drive, Suite 100
Plano, TX 75075
USA

Phone: +1 972 509 5599 x2240
Fax: +1 469 229 5397
Email: ylee@huawei.com

Fatai Zhang
Huawei
F3-5-B R&D Center, Huawei Base
Bantian, Longgang District
P.R. China

Phone: +86-755-28972912
Email: zhangfatai@huawei.com

Federico Pederzolli
CREATE-NET
via alla Cascata 56/D, Povo
Trento 38123
Italy

Email: federico.pederzolli@create-net.org

5. IANA Considerations

This document does not contain any IANA request.

6. Security Considerations

This document defines an protocol-neutral encoding for an information model describing impairments in optical networks and it does not introduce any security issues. If such a encoding is put into use

within a network it will by its nature contain details of the physical characteristics of an optical network. Such information would need to be protected from intentional or unintentional disclosure.

7. References

7.1. Normative References

[ITU.G680] International Telecommunications Union, "Physical transfer functions of optical network elements", ITU-T Recommendation G.680, July 2007.

[ITU.G697] International Telecommunications Union, "Optical monitoring for dense wavelength division multiplexing systems", ITU-T Recommendation G.697, February 2012.

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

7.2. Informative References

[I-D.ietf-ccamp-general-constraint-encode] Bernstein, G., Lee, Y., Li, D., and W. Imajuku, "General Network Element Constraint Encoding for GMPLS Controlled Networks", draft-ietf-ccamp-general-constraint-encode-13 (work in progress), November 2013.

[I-D.ietf-ccamp-rwa-info] Lee, Y., Bernstein, G., Li, D., and W. Imajuku, "Routing and Wavelength Assignment Information Model for Wavelength Switched Optical Networks", draft-ietf-ccamp-rwa-info-19 (work in progress), November 2013.

[I-D.ietf-ccamp-rwa-wson-encode] Bernstein, G., Lee, Y., Li, D., and W. Imajuku, "Routing and Wavelength Assignment Information Encoding for Wavelength Switched Optical Networks", draft-ietf-ccamp-rwa-wson-encode-23 (work in progress), November 2013.

[I-D.martinelli-ccamp-wson-iv-info] Martinelli, G., Zhang, X., Galimberti, G., Zanardi, A., and D. Siracusa, "Information Model for Wavelength Switched Optical Networks (WSONs) with Impairments Validation", draft-martinelli-ccamp-wson-iv-info-03 (work in progress), February 2014.

[RFC6566] Lee, Y., Bernstein, G., Li, D., and G. Martinelli, "A Framework for the Control of Wavelength Switched Optical Networks (WSONs) with Impairments", RFC 6566, March 2012.

Authors' Addresses

Giovanni Martinelli (editor)
Cisco
via Philips 12
Monza 20900
Italy

Phone: +39 039 2092044
Email: giomarti@cisco.com

Domenico Siracusa (editor)
CREATE-NET
via alla Cascata 56/D, Povo
Trento 38123
Italy

Email: domenico.siracusa@create-net.org

Xian Zhang (editor)
Huawei Technologies
F3-5-B R&D Center, Huawei Base
Bantian, Longgang District
Shenzen 518129
P.R. China

Phone: +86 755 28972913
Email: zhang.xian@huawei.com

Gabriele M. Galimberti
Cisco
Via Philips,12
Monza 20900
Italy

Phone: +39 039 2091462
Email: ggalimbe@cisco.com

Andrea Zanardi
CREATE-NET
via alla Cascata 56/D, Povo
Trento 38123
Italy

Email: andrea.zanardi@create-net.org

CCAMP
Internet-Draft
Intended status: Informational
Expires: August 16, 2014

G. Martinelli, Ed.
Cisco
X. Zhang, Ed.
Huawei Technologies
G. Galimberti
Cisco
A. Zanardi
D. Siracusa
CREATE-NET

February 12, 2014

Information Model for Wavelength Switched Optical Networks (WSONs) with Impairments Validation
draft-martinelli-ccamp-wson-iv-info-03

Abstract

This document defines an information model to support Impairment-Aware (IA) Routing and Wavelength Assignment (RWA) function. This operation might be required in Wavelength Switched Optical Networks (WSON) that already support RWA and the information model defined here goes in addition and it is fully compatible with the already defined information model for impairment-free RWA process in WSON.

This information model shall support all control plane architectural options defined for WSON with impairment validation.

Status of This Memo

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

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

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

This Internet-Draft will expire on August 16, 2014.

Table of Contents

1. Introduction

In the context of Wavelength Switched Optical Network (WSON), [RFC6163] describes the basic framework for a GMPLS and PCE-based Routing and Wavelength Assignment (RWA) control plane. The associated information model [I-D.ietf-ccamp-rwa-info] defines all information/parameters required by an RWA process.

There are cases of WSON where optical impairments plays a significant role and are considered as important constraints. The framework document [RFC6566] defines problem scope and related control plane architectural options for the Impairment Aware Routing and Wavelength Assignment (IA-RWA) operation. Options include different combinations of Impairment Validation (IV) and RWA functions in term of different combination of control plane functions (i.e., PCE, Routing, Signaling).

This document provides an information model for the impairment aware case to allow the impairment validation function implemented in the control plane or enabled by control plane available information. This model goes in addition to [I-D.ietf-ccamp-rwa-info] and it shall support any control plane architectural option described by the framework document (see sections 4.2 and 4.3 of [RFC6566]) where a set of control plane combinations of control plane functions vs. IV function is provided.

2. Definitions, Applicability and Properties

This section provides some concepts to help understand concepts used along the document and to make a clear sepration about what coming from data plane definitions (ITU-T G recomandations) and are taken as input for this Information Model. The first sub-section provides raw definitions while the Applicability sections reuses the defined concepts to scope this document.

2.1. Definitions

- Computational Model / Optical Computational Model.
  Defined by ITU standard documents. In this context we looks for models that are able to compute optical impairments for a give lightpath.

- Information Model.
  It is defined by IETF (this draft) and provide the set of information required by the Computational Model to be applied.

- Level of Approximation.

This concept refer to the Computational Model as it may compute optical impairment with a certain level of uncertainty. This level is generally not measured but [RFC6566] make a rough classification about it.

- o Feasible Path.
  It is the output of the CSPF with RWA-IV capability. It's a path that satisfies the constraints in particular the optical impairment contraints. The path, instantiated through wavelength, may actually work or not work depending of the level of approximation.

- o Existing Service Disruption.
  A known effect to optical network designers is the cross-interaction among adjacent (specrum) wavelengths, e,g,,a wavelength may exeperience some increased BER due to the setting up of an adjacent wavelength. Solving this problem is a typical optical network design activity. Just as an example a simple method is adding optical margings (e.g., additional OSNR), other complex and detailed methods exist.

2.2. Applicability

This document targets at Scenario C defined in [RFC6566] section 4.1.1. as approximate impairment estimation. The Approximate concept refer to the fact that this Information Model cover information mainly provided by the [ITU.G680] Computational Model.

Computational models having no approximation, referred as IV-Detailed in the [RFC6566], currently does not exist in term of ITU-T recomandation. They generally refer to non-linear optical impairment and they are usually vendor specific.

The current information model does not speculate about mathematical formula used to fill up information model parameters hence, it does not preclude changing the computational model. At the same time authors does not belive this Information Model is exhaustive and if necessary further documents will cover additional models as long as they become available.

The result of RWA-IV process implementing this Information Model will result in a path (a wavelength in the data plane) that have better chance to be feasible than if it was computed without any IV function. The Existing Service Disruption, as per the definition above, would still be a problem left to network designers: this model does not replace by any means the optical network design phase. The Information Model targets, the GMPLS context with the releated relationship between data plane(s) and control plane.

2.3. Properties

An information model may have several attributes or properties that need to be defined for each optical parameter made available to the control plane. The properties will help to determine how the control plane can deal with a specific impairment parameter, depending on architectural options chosen within the overall impairment framework [RFC6566]. In some case, properties value will help to identify the level of approximation supported by the IV process.

- Time Dependency
  This identifies how an impairment parameter may vary with time. There could be cases where there is no time dependency, while in other cases there may be need of re-evaluation after a certain time. In this category, variations in impairments due to environmental factors such as those discussed in [G.sup47] are considered. In some cases, an impairment parameter that has time dependency may be considered as a constant for approximation. In this information model, we do neglect this property.

- Wavelength Dependency
  This property identifies if an impairment parameter can be considered as constant over all the wavelength spectrum of interest or not. Also in this case a detailed impairment evaluation might lead to consider the exact value while an approximation IV might take a constant value for all wavelengths. In this information model, we consider both case: dependency / no dependency on a specific wavelength. This property appears directly in the information model definitions and related encoding.

- Linearity
  As impairments are representation of physical effects, there are some that have a linear behavior while other are non-linear. Linear approximation is in scope of Scenario C of [RFC6566]. During the impairment validation process, this property implies that the optical effect (or quantity) satisfies the superposition principle, thus a final result can be calculated by the sum of each component. The linearity implies the additivity of optical quantities considered during an impairment validation process. The non-linear effects in general does not satisfy this property. The information model presented in this document however, easily allow introduction of non-linear optical effects with a linear approximated contribution to the linear ones.

- Multi-Channel
  There are cases where a channel’s impairments take different values depending on the aside wavelengths already in place, this

is mostly due to non-linear impairments. The result would be a dependency among different LSPs sharing the same path. This information model do not cosider this kind of property.

The following table summarize the above considerations where in the first column reports the list of properties to be considered for each optical parameter, while the second column states if this property is taken into account or not by this information model.

Property	Info Model Awareness
Time Dependency	no
Wavelength Dependency	yes
Linearity	yes
Multi-channel	no

Table 1: Optical Impairment Properties

3. ITU-T List of Optical Parameters

[EDITOR NOTE: To better integrate material coming from ITU WD06-31 October 2013 and future liasons]

As stated by Section 2.2 this Information Model does not intend to be exaustive and targets an approximate computational model although not precluding future evolutions towards more detailed impairments estimation methods.

On the same line, ITU SG15/Q6 provides a list of optitical parameters with following observations:

(a) the problem of calculating the non-linear impairments in a multi-vendor environment is not solved. The transfer functions works only for the so called [ITU.G680] "Situation 1".

(b) The generated list of parameters is not definitive or exaustive.

In particular, [ITU.G680] contains many parameters that would be required to estimate linear impairments and [ITU.G697] contains information on which parameters can be monitored in an optical network.

[ITU.G671] contains some additional parameters defintions required by here above recomandation.

The list of optical parameters starts from [ITU.G680] Section 9 which provides the optical computational models for the following:

P1 OSNR. Section 9.1

P2 Optical Power. As per Section 9.1, required by Optical Computation Model for OSNR calculation.

P3 Chromatic Dispersion (CD). Section 9.2

P4 Polarization Mode Dispersion (PMD). Section 9.3

P5 Polarization Dependent Loss (PDL). Section 9.3

In addition to the above, the following list of parameters has been mentioned by ITU SG15/Q6.

P6 Channel Frequency Range [ITU.G671].

P7 Ripple

P8 Channel Signal-Spontaneous noise figure. This is considered within OSNR computational model above.

P9 Differential Group Delay [ITU.G671]. Required for PMD above.

P10 Reflectance.

P11 Isolation.

P12 Channel extintion.

P13 Non-Linear Coefficient (for a fibre segment). Needed for non-linear impairment

4. Background from WSON-RWA Information Model

In this section we report terms already defined for the WSON-RWA (impairment free) as in [I-D.ietf-ccamp-rwa-info] and [I-D.ietf-ccamp-general-constraint-encode]. The purpose is to provide essential information that will be reused or extended for the impairment case.

In particular [I-D.ietf-ccamp-rwa-info] defines the connectivity matrix as the following:

ConnectivityMatrix ::= <MatrixID> <ConnType> <Matrix>

According to [I-D.ietf-ccamp-general-constraint-encode], this definition is further detailed as:

```
ConnectivityMatrix ::=
      <MatrixID> <ConnType> ((<LinkSet> <LinkSet>) ...)
```

This second formula highlights how the connectivity matrix is built by pairs of LinkSet objects identifying the internal connectivity capability due to internal optical node constraint(s). It's essentially binary information and tell if a wavelength or a set of wavelengths can go from an input port to an output port.

As an additional note, connectivity matrix belongs to node information and is purely static. Dynamic information related to the actual usage of the connections is available through specific extension to link information.

Furthermore [I-D.ietf-ccamp-rwa-info] define the resource block as follow:

```
 ResourceBlockInfo ::= <ResourceBlockSet> [<InputConstraints>]
   [<ProcessingCapabilities>] [<OutputConstraints>]
```

Which is an efficient way to model constrains of a WSON node.

5. Optical Impairment Information Model

The idea behind this information model is to categorize the impairment parameters into three types and extend the information model already defined for impairment-free WSONs. The three categories are:

- Node Information. The concept of connectivity matrix is reused and extended to introduce an impairment matrix, which represents the impairments suffered on the internal path between two ports. In addition, the concept of Resource Block is also reused and extended to provide an efficient modelization of per-port impairment.

- Link Information representing impairment information related to a specific link or hop.

- Path Information representing the impairment information related to the whole path.

All the above three categories will make use of a generic container, the Impairment Vector, to transport optical impairment information.

This information model however will allow however to add additional parameters beyond the one defined by [ITU.G680] in order to support additional computational models. This mechanism could eventually applicable to both linear and non-linear parameters.

This information model makes the assumption that the each optical node in the network is able to provide the control plane protocols with its own parameter values however, no assumption is made on how the optical node gets those value information (e.g. internally computed, provisioned by a network management system, etc.). To this extent, the information model intentionally ignores all internal detailed parameters that are used by the formulas of the Optical Computational Model (i.e., "transfer function") and simply provides the object containers to carry results of the formulas.

5.1. The Optical Impairment Vector

Optical Impairment Vector (OIV) is defined as a list of optical parameters to be associated to a WSON node or a WSON link. It is defined as:

<OIV> ::= ([<LabelSet>] <OPTICAL_PARAM>) ...

The optional LabelSet object enables wavelength dependency property as per Table 1. LabelSet has its definition in [I-D.ietf-ccamp-general-constraint-encode].

OPTICAL_PARAM. This object represents an optical parameter. The Impairment vector can contain a set of parameters as identified by [ITU.G697] since those parameters match the terms of the linear impairments computational models provided by [ITU.G680]. This information model does not speculate about the set of parameters (since defined elsewhere, e.g. ITU-T), however it does not preclude extentions by adding new parameters.

5.2. Node Information

5.2.1. Impairment Matrix

Impairment matrix describes a list of the optical parameters that applies to a network element as a whole or ingress/egress port pairs of a network element. Wavelength dependency property of optical paramters is also considered.

```
ImpairmentMatrix ::=  <MatrixID> <ConnType>
      ((<LinkSet> <LinkSet> <OIV>) ...)
```

Where:

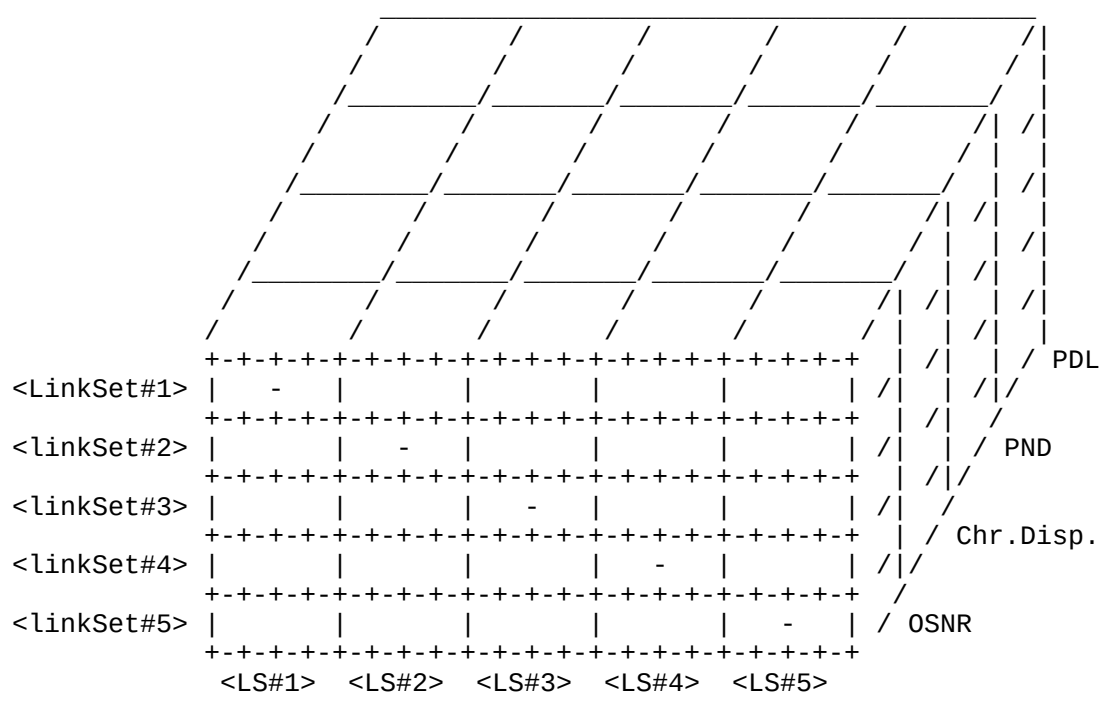
MatrixID. This ID is a unique identifier for the matrix. It shall be unique in scope among connectivity matrices defined in [I-D.ietf-ccamp-rwa-info] and impairment matrices defined here.

ConnType. This number identifies the type of matrix and it shall be unique in scope with other values defined by impairment-free WSON documents.

LinkSet. Same object definition and usage as [I-D.ietf-ccamp-general-constraint-encode]. The pairs of LinkSet identify one or more internal node constrain.

OIV. The Optical Impairment Vector defined above.

The model can be represented as a multidimensional matrix shown in the following picture



The connectivity matrix from [I-D.ietf-ccamp-general-constraint-encode] is only a two dimensional matrix, containing only binary information, through the LinkSet pairs. In this model, a third dimension is added by generalizing the binary information through the Optical Impairment Vector associated with each LinkSet pair. Optical parameters in the picture are reported just as examples while details go into specific encoding draft [I-D.martinelli-ccamp-wson-iv-encode].

This representation shows the most general case however, the total amount of information transported by control plane protocols can be greatly reduced by proper encoding when the same set of values apply to all LinkSet pairs.

[EDITOR NODE: first run of the information model does looks for generality not for optimizing the quantity of information. We'll deal with optimization in a further step.]

5.2.2. Impariment Resource Block Information

This information model reuse the definition of Resource Block Information adding the associated impairment vector.

```
ResourceBlockInfo ::= <ResourceBlockSet> [<InputConstraints>]
  [<ProcessingCapabilities>] [<OutputConstraints>] [<OIV>]
```

The object ResourceBlockInfo is than used as specified within [I-D.ietf-ccamp-rwa-info].

5.3. Link Information

For the list of optical parameters associated to the link, the same approach used for the node-specific impairment information can be applied. The link-specific impairment information is extended from [I-D.ietf-ccamp-rwa-info] as the following:

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

DynamicLinkInfo is already defined in [I-D.ietf-ccamp-rwa-info] while OIV is the Optical Impairment Vector is defined in the previous section.

5.4. Path Information

There are cases where the optical impariments can only be described as a contrains on the overall end to end path. In such case, the optical impariment and/or parameter, cannot be derived (using a simple function) from the set of node / link contributions.

An equivalent case is the option reported by [RFC6566] on IV-Candidate paths where, the control plane knows a list of optically feasible paths so a new path setup can be selected among that list. Independent from the protocols and functions combination (i.e. RWA vs. Routing vs. PCE), the IV-Candidates imply a path property stating that a path is optically feasible.

```
<PathInfo> ::=  <OIV>
```

[EDITOR NOTE: section to be completed, especially to evaluate protocol implications. Likely resemble to RSVP ADSPEC].

6. Encoding Considerations

Details about encoding will be defined in a separate document [I-D.martinelli-ccamp-wson-iv-encode] however worth remembering that, within [ITU.G697] Appending V, ITU already provides a guideline for encoding some optical parameters.

In particular [ITU.G697] indicates that each parameter shall be represented by a 32 bit floating point number.

Values for optical parameters are provided by optical node and it could provide by direct measurement or from some internal computation starting from indirect measurement. In such cases could be useful to un understand the variance associated with the value of the optical parmater hence, the encoding shall provide the possibility to include a variance as well.

This kind of information will enable IA-RWA process to make some additional considerations on wavelength feasibility. [RFC6566] Section 4.1.3 reports some considerations regarding this degree of confidence during the impairment validation process.

7. Control Plane Architectures

This section briefly describes how the defintions contained in this information model will match the architectural options described by [RFC6566].

The first assumption is that the WSON GMPLS extentions are available and operational. To such extent, the WSON-RWA will provide the following information through its path computation (and RWA process):

- The wavelengths connectivity, considering also the connectivity constraints limited by reconfigurable optics, and wavelengths availability.
- The interface compatibility at the physical level.
- The Optical-Elettro-Optical (OEO) availability within the network (and related physical interface compatibility). As already stated by the framework this information it’s very important for impairment validation:
  - A. If the IV functions fail (path optically infeasible), the path computation function may use an available OEO point to find a

feasible path. In normally operated networks OEO are mainly uses to support optically unfeasible path than mere wavelength conversion.

B. The OEO points reset the optical impairment information since a new light is generated.

7.1. IV-Centralized

Centralized IV process is performed by a single entity (e.g., a PCE). Given sufficient impairment information, it can either be used to provide a list of paths between two nodes, which are valid in terms of optical impairments. Alternatively, it can help validate whether a particular selected path and wavelength is feasiable or not. This requires distribution of impairment information to the entity performing the IV process.

[EDITOR NOTE: to be completed]

7.2. IV-Distributed

For the distributed IV process, common computational models are needed together with the information model defined in this document. Computational models for the optical impairments are defined by ITU standard body. The currently available computation models are reported in [ITU.G680] and only cover the linear impairment case. This does not require the distribution of impairment information since they can be collected hop-by-hop using a control plane signaling protocol.

[EDITOR NOTE: to be completed]

8. Acknowledgements

Authors would like to thank ITU SG15/Q6 and in particular Pete Anslow for providing text and information to CCAMP through join meetings and liasons.

9. Contributing Authors

This document was the collective work of several authors. The text and content of this document was contributed by the editors and the co-authors listed below (the contact information for the editors appears in appropriate section and is not repeated below):

Moustafa Kattan
Cisco
DUBAI, 500321
UNITED ARAB EMIRATES

Email: mkattan@cisco.com

Young Lee
Huawei
1700 Alma Drive, Suite 100
Plano, TX 75075
USA

Phone: +1 972 509 5599 x2240
Fax: +1 469 229 5397
Email: ylee@huawei.com

Greg M. Bernstein
Grotto Networking
Fremont, CA
USA

Phone: +1 510 573 2237
Email: gregb@grotto-networking.com

Fatai Zhang
Huawei
F3-5-B R&D Center, Huawei Base
Bantian, Longgang District
P.R. China

Phone: +86-755-28972912
Email: zhangfatai@huawei.com

Federico Pederzolli
CREATE-NET
via alla Cascata 56/D, Povo
Trento 38123
Italy

Email: federico.pederzolli@create-net.org

10. IANA Considerations

This document does not contain any IANA requirement.

11. Security Considerations

This document defines an information model for impairments in optical networks. If such a model is put into use within a network it will by its nature contain details of the physical characteristics of an optical network. Such information would need to be protected from intentional or unintentional disclosure.

12. References

12.1. Normative References

[ITU.G671]
International Telecommunications Union, "Transmission characteristics of optical components and subsystems", ITU-T Recommendation G.671, February 2012.

[ITU.G680]
International Telecommunications Union, "Physical transfer functions of optical network elements", ITU-T Recommendation G.680, July 2007.

[ITU.G697]
International Telecommunications Union, "Optical monitoring for dense wavelength division multiplexing systems", ITU-T Recommendation G.697, February 2012.

12.2. Informative References

[I-D.ietf-ccamp-general-constraint-encode]
Bernstein, G., Lee, Y., Li, D., and W. Imajuku, "General Network Element Constraint Encoding for GMPLS Controlled Networks", draft-ietf-ccamp-general-constraint-encode-13 (work in progress), November 2013.

[I-D.ietf-ccamp-rwa-info]
Lee, Y., Bernstein, G., Li, D., and W. Imajuku, "Routing and Wavelength Assignment Information Model for Wavelength Switched Optical Networks", draft-ietf-ccamp-rwa-info-19 (work in progress), November 2013.

[I-D.martinelli-ccamp-wson-iv-encode]
Martinelli, G., Zanardi, A., Zhang, X., Galimberti, G., and D. Siracusa, "Information Encoding for WSON with Impairments Validation", draft-martinelli-ccamp-wson-iv-encode-02 (work in progress), July 2013.

[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.

[RFC6566] Lee, Y., Bernstein, G., Li, D., and G. Martinelli, "A Framework for the Control of Wavelength Switched Optical Networks (WSONs) with Impairments", RFC 6566, March 2012.

Appendix A. ITU-T Liason Tracking

[EDITOR NOTE: appendix reserved to track liason to/from ITU related to this draft]

Authors' Addresses

Giovanni Martinelli (editor)
Cisco
via Philips 12
Monza 20900
Italy

Phone: +39 039 2092044
Email: giomarti@cisco.com

Xian Zhang (editor)
Huawei Technologies
F3-5-B R&D Center, Huawei Base
Bantian, Longgang District
Shenzen 518129
P.R. China

Phone: +86 755 28972465
Email: zhang.xian@huawei.com

Gabriele M. Galimberti
Cisco
Via Philips,12
Monza 20900
Italy

Phone: +39 039 2091462
Email: ggalimbe@cisco.com

Andrea Zanardi
CREATE-NET
via alla Cascata 56/D, Povo
Trento 38123
Italy

Email: andrea.zanardi@create-net.org

Domenico Siracusa
CREATE-NET
via alla Cascata 56/D, Povo
Trento 38123
Italy

Email: domenico.siracusa@create-net.org

INTERNET-DRAFT
Intended Status: Standard Track

Expires: April 20, 2014

Khuzema Pithewan
Rajan Rao
Infinera

October 17, 2013

OSPF-TE extensions for MLNMRN based on OTN
draft-rao-ccamp-mlnmrn-otn-ospfte-ext-03.txt

Abstract

This document specifies OSPF extensions for multi-layer/multi-region where one of the regions is multi-layer e.g. OTN, SONET/SDH.

Status of this Memo

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

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

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

The list of current Internet-Drafts can be accessed at http://www.ietf.org/1id-abstracts.html

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

Copyright and License Notice

Table of Contents

1 Introduction

In order to do end-to-end path computation, where a path may involve more than one region and part of single routing domain, TE Links connecting the two regions need to have bandwidth capacity advertised for the switch that connects the two regions. This document specifies the OSPF extensions that are required if any of the region is a multi-layer network. The specification is based on the requirement as specified in RFC 5212. As per the said RFC, ISCD characterizes the information associated to one or more network layers. Same RFC also says that the information about the adjustment capabilities of the nodes in the network allow the path computation process to select an end-to-end multi-layer or multi- region path that includes links with different switching capabilities joined by LSRs that can adapt (i.e., adjust) the signal between the links. By inference, information about the adjustment capabilities should be able to identify a layer in ISCD, if ISCD specifies more than one layer.

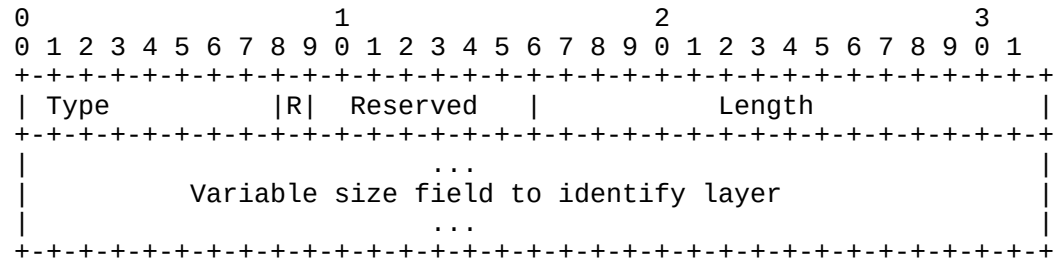
RFC6001 specifies how to advertise adjustment capabilities between two switching regions. IACD definition has provision to extend it for a specific technology through Adjustment Capability Specific information (ACSI) field, if required. ACSI field can be used to identify a layer in the multi-layer ISCD.

While OTN multi-layer technology is a primary driver for this extension, the extensions in this document does cover specifications for multi-layer technologies in general. To make sure the extensions are extensible to other multi-layer technologies as well, this document covers SDH/SONET as well.

2 Layer Identification

Multi-region path computation requires to identify a layer in the multi-layer region. This mandates layer identification along with identification of technology in the region. The technology identification is done via Switching capability and Encoding type.

IACD needs to be extended to be able to carry layer identification. the layer Identification is OPTIONAL and used only when interface supports layer multiplexing and hence creating a need to identify a layer. A new Layer ID Sub-TLV has been defined to carry layer identification.



Type : Type field is used to identify a particular structure of variable size field, which is specific to the particular Switching Capability and Encoding type combination

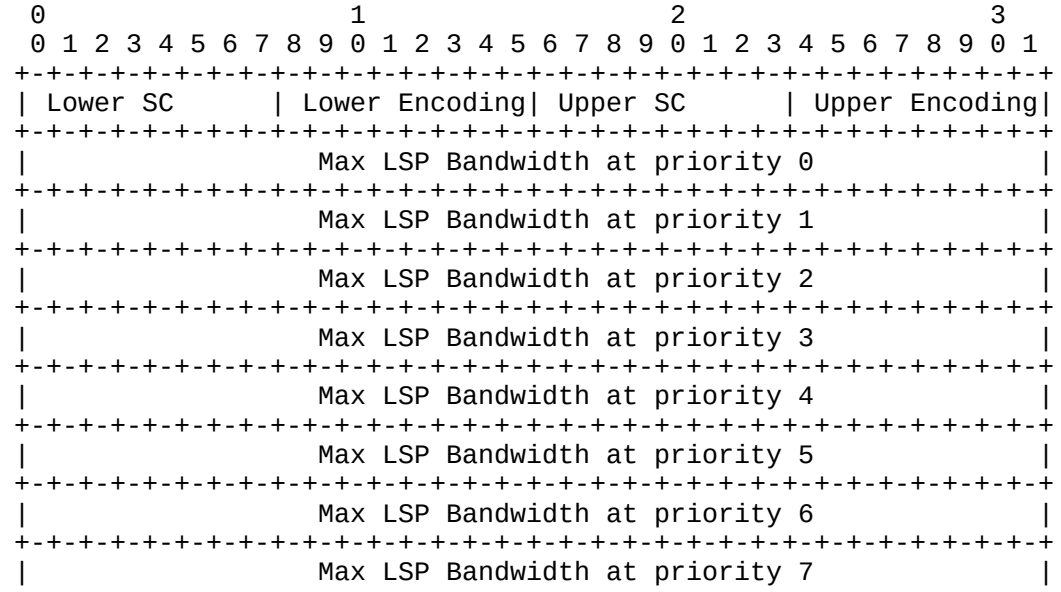
R : This bit is used to make sense whether the Layer ID is for Lower region or upper region. 1 means upper region and 0 means lower.

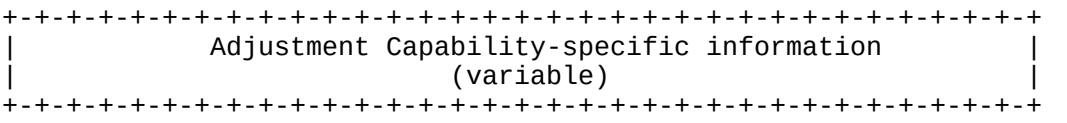
IACD can have at-most 2 Layer ID TLVs, if both the regions are multi-layer.

Next two sections specifies Layer ID for two multi-layer technologies namely, OTN and SONET/SDH

3 OTN Layer ID

RFC6001 defines IACD sub-TLV as follows. Please refer to the RFC for definition of individual fields of the sub-TLV.

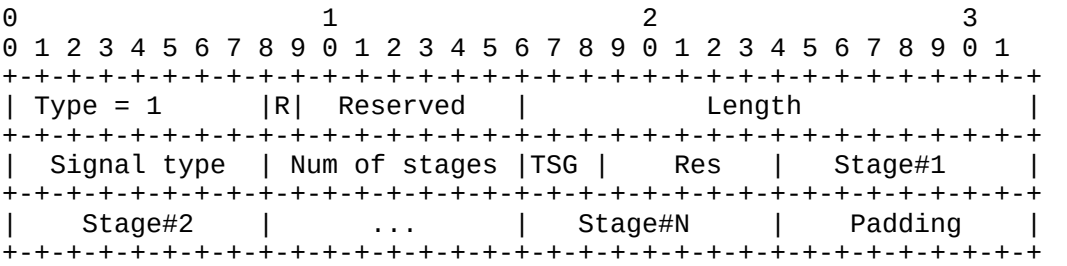




[GMPLS-OTN-OSPF] defines attributes that identifies a layer in multi-layer OTN ISCD. These attributes are part of Bandwidth sub-TLV in Switch capability specific information of ISCD. These attributes are reproduced here for completeness sake.

* Signal Type: Layer for which bandwidth is being advertised.
* Hierarchy : also called as multiplexing branch that specifies all the layers between server layer and signal type.
* TSG : Time Slot Granularity

Adjustment Capability-specific information abbreviated as ACSI henceforth for OTN G.709v3 carries LayerID Sub-TLV which is defined as follows



This LayerID sub-TLV is applicable only when one of the regions is OTN, which means either lower or upper SC and Encoding type MUST have Switch Cap as OTN-TDM and encoding type as G.709 ODUk.

R bit is used to make sense whether the Layer ID is for Lower region or upper region. 1 means upper region and 0 means lower.

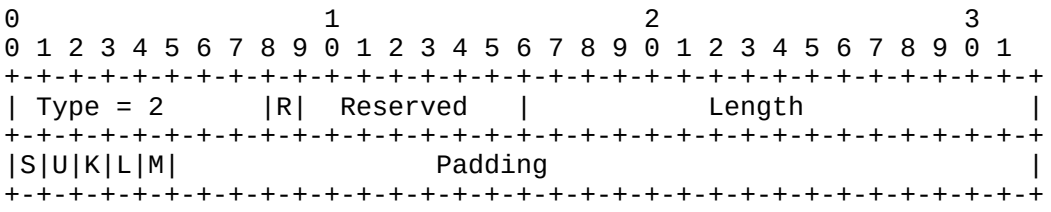
The 8 priorities of the BW as defined in main IACD structure, is adjustment capability between the two regions where one of the region is identifies by LayerID sub-TLV.

Absence of this sub-TLV for OTN means that the OTN ISCD doesn’t support multiplexing.

4 SONET/SDH Layer Identification

G.707 defines the structure of SDH multiplexing hierarchy and RFC 4606 defines generalized label structure needed to fully specify SONET/SDH multiplexing hierarchy. This Label structure also referred as SUKLM structure identifies all the layers of the multiplexing hierarchy along with time slots. For the purpose of this draft, only layer identification is needed, hence each layer can be identified by a bit. Bit value 1 signifies presence of the layer and 0, its absence. 5 Bits, each representing one layer is sufficient to fully identify the SONET/SDH multiplexing hierarchy.

Layer ID sub TLV for SONET/SDH is defined as follows



SUKLM bits signifies the presence of SONET/SDH layers and these bits together fully specifies the multiplexing hierarchy. Refer to Section 3 of RFC 4606 for full specification of SUKLM bits.

Absence of sub-TLV means that the SONET/SDH ISCD doesn’t support multiplexing and needs only transparent mapping to other Interface.

5 Procedure

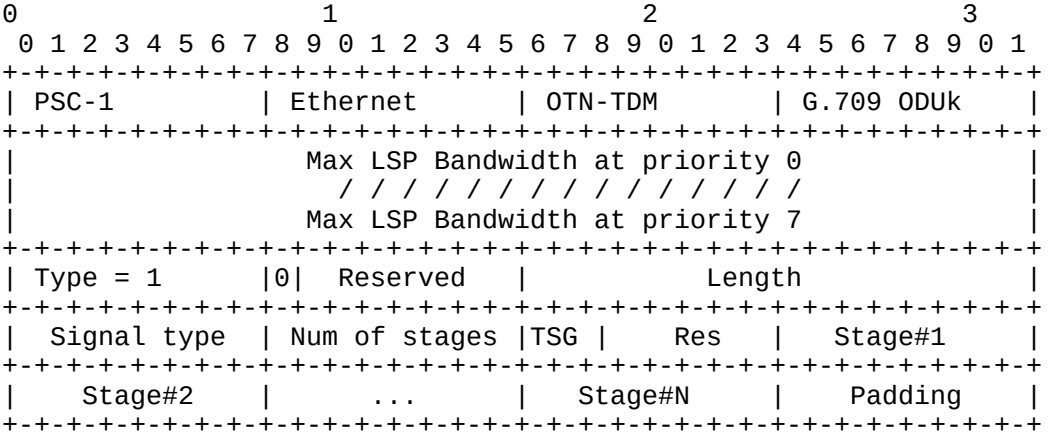
A node advertising IACD for the bandwidth between regions where one or both of them are hierarchical i.e. OTN or SONET/SDH, MUST include the Layer ID sub-TLV as part of ACSI as defined above.

For multi-region path computation, the path computing node MUST look at the LayerID sub-TLV (in ACSI part of IACD) if lower/upper {SC,Enc] is {OTN-TDM,G.709ODUk} or {TDM,SONET/SDH} to identify the layer for correct layer for BW check.

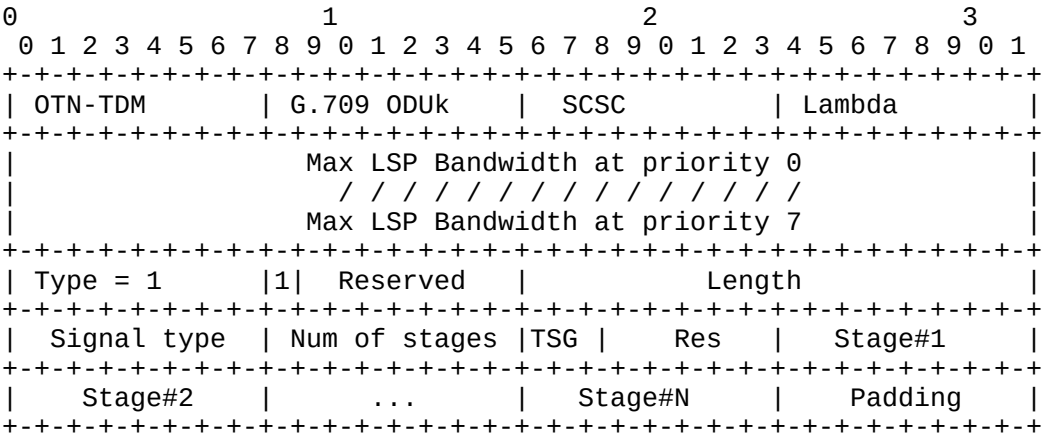
6 Examples

This section exemplifies TLV values for various technology region combinations, where one of the region is OTN

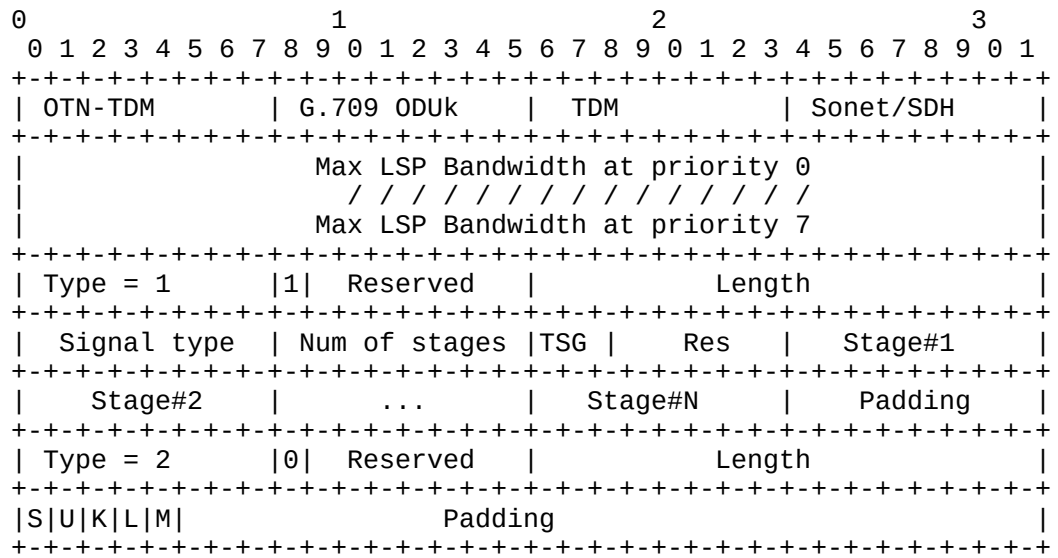
6.1. Ethernet and OTN When upper region is Ethernet and lower region is OTN



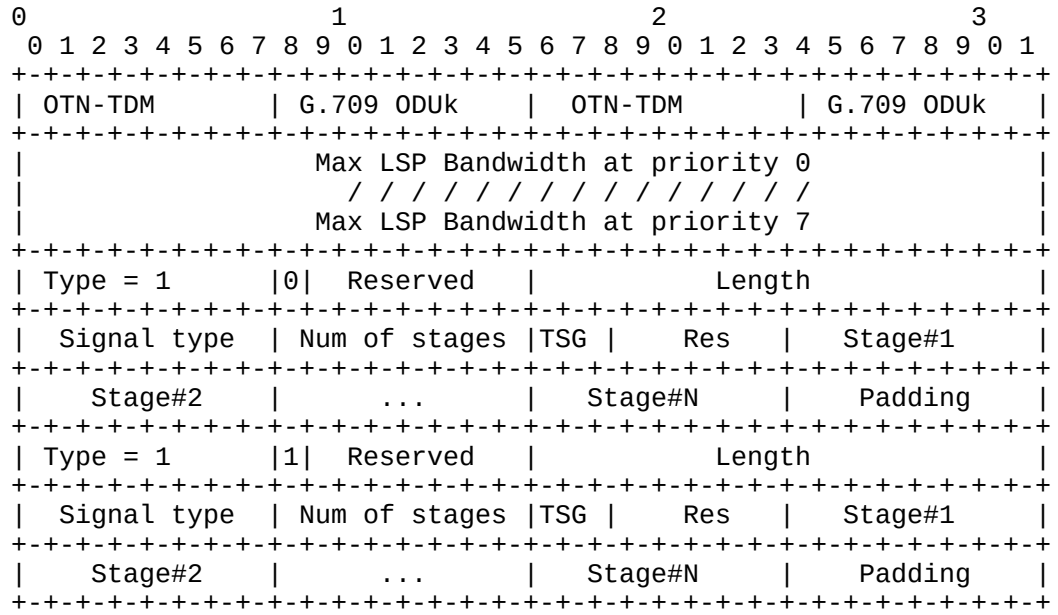
6.2. OTN and FlexGrid



6.3. OTN and SONET/SDH



6.4. OTN and OTN



7 IANA Considerations

TBD

8 Security Considerations

TBD

9 References

[RFC5212] K. Shiomoto, Papadimitriou, D.,JL. Le Roux, Vigoureux, M., Brungard, D., "Requirements for GMPLS-Based Multi-Layer and Multi-Region Networks (MLN/ MRN)",RFC 5212, July 2008.

[RFC6001] Papadimitriou, D., Vigoureux, M., Shiomoto, K., Brungard, D., and JL. Le Roux, "Generalized MPLS (GMPLS) Protocol Extensions for Multi-Layer and Multi-Region Networks (MLN/MRN)", RFC 6001, October 2010.

[RFC4606] E. Mannie, Perceval, D. Papadimitriou, "Generalized Multi-Protocol Label Switching (GMPLS) Extensions for Synchronous Optical Network (SONET) and Synchronous Digital Hierarchy (SDH) Control", RFC 4606, Aug 2006

[GMPLS-OTN-OSPF] Traffic Engineering Extensions to OSPF for Generalized MPLS (GMPLS)Control of Evolving G.709 OTN Networks

10. Authors' Addresses

Khuzema Pithewan
Infinera
140 Caspian Ct., Sunnyvale, CA 94089
Email: kpithewan@infinera.com

Rajan Rao
Infinera
140 Caspian Ct., Sunnyvale, CA 94089
Email: rrao@infinera.com

CCAMP Working Group
Internet-Draft
Intended Status: Standards Track
Expires: August 7, 2014

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

February 3, 2014

Extensions to Resource Reservation Protocol For Fast Reroute of Bidirectional Co-routed Traffic Engineering LSPs

draft-tsaad-ccamp-rsvpte-bidir-lsp-fastreroute-03

Abstract

This document defines Resource Reservation Protocol - Traffic Engineering (RSVP-TE) signaling extensions to support Fast Reroute (FRR) of bidirectional co-routed Traffic Engineering (TE) LSPs. 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. In addition, the RSVP-TE signaling extensions allow the coordination of bypass tunnel assignment protecting a common facility in both forward and reverse directions prior to or post failure occurrence.

Status of this Memo

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

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

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

Copyright and License Notice

Table of Contents

1. Introduction

Co-routed bidirectional tunnels are signaled using GMPLS signaling procedures specified in [RFC3473] and [RFC3471]. Existing 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, and don't address issues that arise when employing FRR for bidirectional co-routed Label Switched Paths (LSPs).

When using current FRR procedures with bidirectional co-routed LSPs, it is possible in some cases (e.g. when using node-protecting 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.

When co-routed bidirectional bypass tunnels are used to locally protect bidirectional LSPs, the upstream and downstream PLRs may independently assign different bidirectional bypass tunnels in the forward and reverse directions. Currently, there is no means to coordinate the bypass tunnel selection between the downstream and upstream PLRs. In case of mismatch and after FRR, data traffic and signaling may flow over asymmetric paths in the forward and reverse directions which may be undesirable for certain applications.

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

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.

Facility Bypass: A local repair method in which a bypass tunnel is used to protect one or more protected LSPs that traverse the PLR, the resource being protected, and the Merge Point in that order.

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.

CSPF: Constraint-based Shortest Path First.

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. Link Failure With Node-protection Bypass Tunnels

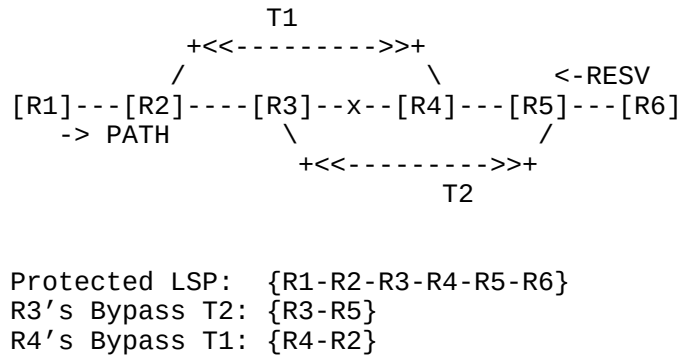


Figure 1: Flow of RSVP signaling post FRR after failure

Consider the Traffic Engineered (TE) network shown in Figure 1. 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 cases where they go over asymmetric paths (i.e. non co-routed) in the forward and reverse directions after the first phase.

3.1. Behavior Before Local Repair

To correctly reroute data traffic over a node-protection 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.

3.1.1. Downstream Merge Point Label Discovery

[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.

3.1.2. Upstream Merge Point Label Discovery

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

3.2. 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.

3.3. 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 signaling (RSVP Resv) over the found bypass tunnel T3 in the reverse

- If PRR R5 is unable to successfully find a bypass tunnel that terminates on the downstream PLR, it may send an immediate RSVP Notify message back to the head-end. The head-end may tear and re-setup the protected LSP immediately.

If 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.

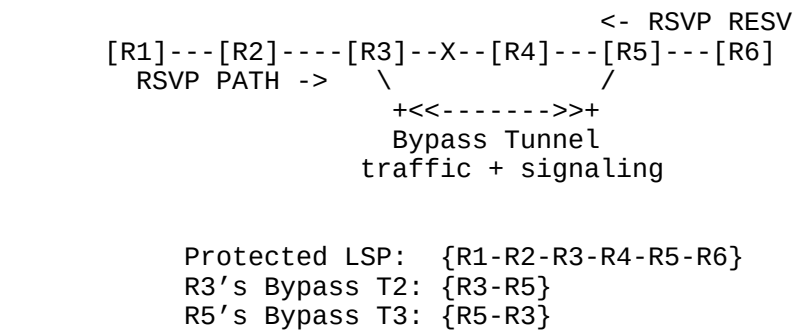


Figure 2: Flow of RSVP signaling post FRR after re-corouted

Figure 2 describes the path taken by traffic and signaling after completing re-coroute of data and signaling in the forward and reverse paths described earlier.

The MP MAY optionally support handling in data plane as follows. If the MP is preconfigured 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.

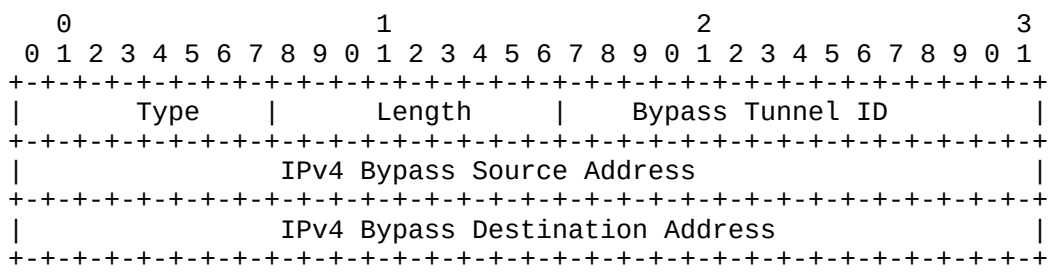
4. Bypass Tunnel Assignment Coordination

This document defines a new subobject in RSVP RECORD_ROUTE object, DOWNSTREAM_BYPASS_ASSIGNMENT, to extend RSVP-TE for fast-reroute signaling.

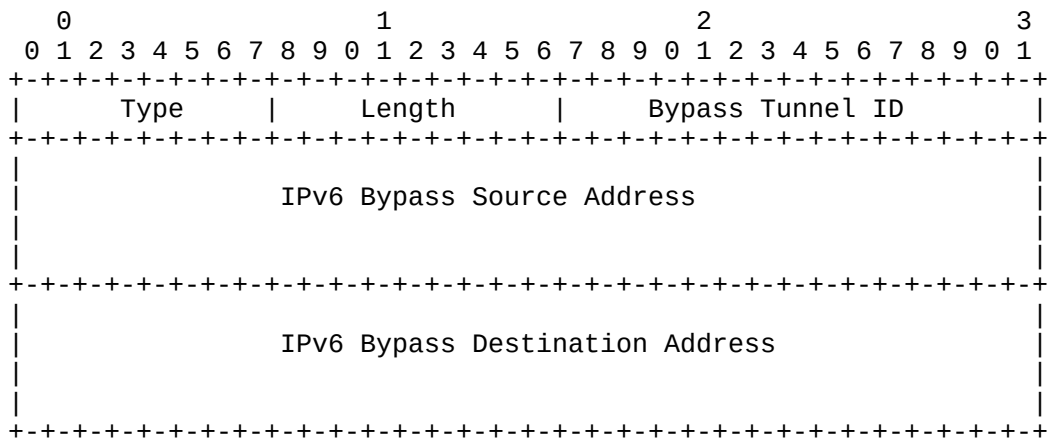
4.1. DOWNSTREAM_BYPASS_ASSIGNMENT Subobject

The DOWNSTREAM_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 MUST only be inserted into the Path message by the downstream PLR and MUST NOT be changed by downstream LSRs. The DOWNSTREAM_BYPASS_ASSIGNMENT subobject has the following format:

The IPv4 DOWNSTREAM_BYPASS_ASSIGNMENT subobject has the following format:



The IPv6 DOWNSTREAM_BYPASS_ASSIGNMENT subobject has the following format:



Type

Downstream Bypass Assignment

Length

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

Bypass Source Address

The bypass tunnel source IPV4 or IPV6 address.

Bypass Destination Address

The bypass tunnel destination IPV4 or IPV6 address.

Bypass Tunnel ID

The bypass tunnel identifier.

4.2. Bypass Tunnel Assignment Signaling Procedure

In cases where bidirectional bypass tunnels are used for FRR Local Repair for a bidirectional co-routed LSP, it is desirable to coordinate the bypass tunnel selected at the downstream and upstream PLRs so that rerouted traffic and signaling flows on symmetrical 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 DOWNSTREAM_BYPASS_ASSIGNMENT subobject is added by each downstream PLR in the RSVP Path RECORD_ROUTE message of the 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 DOWNSTREAM_BYPASS_ASSIGNMENT subobject is added in the RECORD_ROUTE object prior to adding the node’s IP address. A node MUST NOT add a DOWNSTREAM_BYPASS_ASSIGNMENT subobject without also adding an IPv4 or IPv6 subobject.

The upstream PLR (downstream MP) that detects a DOWNSTREAM_BYPASS_ASSIGNMENT subobject whose bypass tunnel destination matching its own address 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 DOWNSTREAM_BYPASS_ASSIGNMENT subobject, the downstream MP can independently assign a bypass tunnel in the reverse direction. In the case of downstream MP receiving multiple DOWNSTREAM_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. Note, the bypass tunnel selection will be corrected for co-routeness after FRR based on the PRR behavior after failure.

5. Compatibility

New RSVP subobject DOWNSTREAM_BYPASS_ASSIGNMENT is defined for RECORD_ROUTE in this document. Per [RFC2205], nodes not supporting this subobject will ignore but forward it, unexamined and unmodified, in all messages resulting from this message.

6. Security Considerations

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

7. IANA Considerations

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

8. Acknowledgements

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

9. References

9.1. 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.

9.1. Informative References

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

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

[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
Cisco Systems, Inc.

EMail: tsaad@cisco.com

Rakesh Gandhi
Cisco Systems, Inc.

EMail: rgandhi@cisco.com

Zafar Ali
Cisco Systems, Inc.

EMail: zali@cisco.com

Manav Bhatia
Alcatel-Lucent
India

Email: manav.bhatia@alcatel-lucent.com

Lizhong Jin
Shanghai, China

Email: lizho.jin@gmail.com

Frederic Jounay
Orange CH

Email: frederic.jounay@orange.ch

Network Working Group
Internet-Draft
Intended status: Informational

Xian Zhang
Haomian Zheng
Huawei
Ramon Casellas
CTTC
O. Gonzalez de Dios
Telefonica
D. Ceccarelli
Ericsson

Expires: August 14, 2014 February 14, 2014

GMPLS OSPF-TE Extensions in support of Flexible Grid

draft-zhang-ccamp-flexible-grid-ospf-ext-04.txt

Abstract

This memo describes the OSPF-TE extensions in support of GMPLS control of networks that include devices that use the new flexible optical grid.

Status of this Memo

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

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

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

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt.

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

This Internet-Draft will expire on August 14, 2014.

Copyright Notice

Table of Contents

1. Introduction

[G.694.1] defines the Dense Wavelength Division Multiplexing (DWDM) frequency grids for Wavelength Division Multiplexing (WDM)

applications. A frequency grid is a reference set of frequencies used to denote allowed nominal central frequencies that may be used for defining applications. The channel spacing is the frequency spacing between two allowed nominal central frequencies. All of the wavelengths on a fiber should use different central frequencies and occupy a fixed bandwidth of frequency.

Fixed grid channel spacing is selected from 12.5 GHz, 25 GHz, 50 GHz, 100 GHz and integer multiples of 100 GHz. But [G.694.1] also defines "flexible grids", also known as "flexi-grid". The terms "frequency slot" (i.e., the frequency range allocated to a specific channel and unavailable to other channels within a flexible grid) and "slot width" (i.e., the full width of a frequency slot in a flexible grid) are used to define a flexible grid.

[FLEX-FWK] defines a framework and the associated control plane requirements for the GMPLS based control of flexi-grid DWDM networks.

[RFC6163] provides a framework for GMPLS and Path Computation Element (PCE) control of Wavelength Switched Optical Networks (WSONs), and [WSON-OSPF] defines the requirements and OSPF-TE extensions in support of GMPLS control of a WSON.

[FLEX-SIG] describes requirements and protocol extensions for signaling to set up LSPs in networks that support the flexi-grid, and this document complements [FLEX-SIG] by describing the requirement and extensions for OSPF-TE routing in a flexi-grid network.

2. Terminology

For terminology related to flexi-grid, please consult [FLEX-FWK] and [G.694.1].

2.1. Conventions Used in this Document

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

3. Requirements for Flexi-grid Routing

The architecture for establishing LSPs in a Spectrum Switched optical Network (SSON) is described in [FLEX-FWK].

A flexi-LSP occupies a specific frequency slot, i.e. a range of frequencies. The process of computing a route and the allocation of

a frequency slot is referred to as RSA (Routing and Spectrum Assignment). [FLEX-FWK] describes three types of architectural approaches to RSA: combined RSA; separated RSA; and distributed SA. The first two approaches among them could be called "centralized SA" because both routing and spectrum (frequency slot) assignment are performed by centralized entity before the signaling procedure.

In the case of centralized SA, the assigned frequency slot is specified in the Path message during LSP setup. In the case of distributed SA, the slot width of the flexi-grid LSP is specified in the Path message, allowing the involved network elements to select the frequency slot to be used.

If the capability of switching or converting the whole optical spectrum allocated to an optical spectrum LSP is not available at nodes along the path of the LSP, the LSP is subject to the Optical "Spectrum Continuity Constraint", as described in [FLEX-FWK].

The remainder of this section states the additional extensions on the routing protocols in a flexi-grid network. That is, the additional information that must be collected and passed between nodes in the network by the routing protocols in order to enable correct path computation and signaling in support of LSPs within the network.

3.1. Available Frequency Ranges

In the case of flexi-grids, the central frequency steps from 193.1 THz with 6.25 GHz granularity. The calculation method of central frequency and the frequency slot width of flexi-LSP are defined in [G.694.1].

On a DWDM link, the frequency slots must not overlap with each other. However, the border frequencies of two frequency slots may be the same frequency, i.e., the highest frequency of a frequency slot may be the lowest frequency of the next frequency slot.

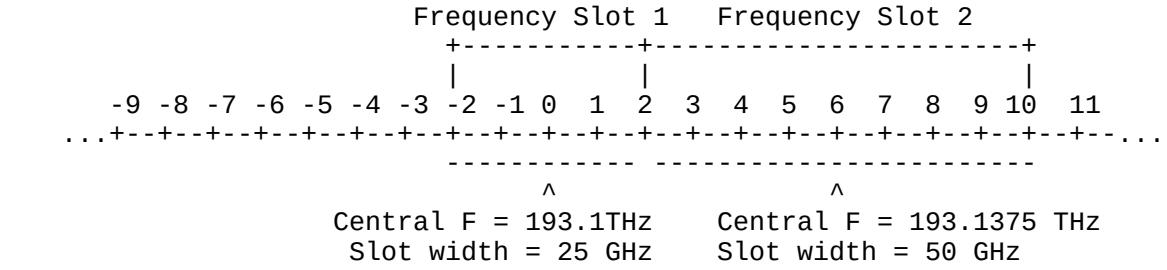


Figure 1 - Two Frequency Slots on a Link

Figure 1 shows two adjacent frequency slots on a link. The highest frequency of frequency slot 1 denoted by n=2 is the lowest frequency of slot 2. In this example, it means that the frequency range from n=-2 to n=10 is occupied and is unavailable to other flexi-LSPs.

Hence, in order to clearly show which LSPs can be supported and what frequency slots are unavailable, the available frequency ranges should be advertised by the routing protocol for the flexi-grid DWDM links. A set of non-overlapping available frequency ranges should be disseminated in order to allow efficient resource management of flexi-grid DWDM links and RSA procedures which are described in section 5.8 of [FLEX-FWK].

3.2. Application Compliance Considerations

As described in [G.694.1], devices or applications that make use of the flexi-grid may not be capable of supporting every possible slot width or position (i.e., central frequency). In other words, applications or implementations may be defined where only a subset of the possible slot widths and positions are required to be supported.

For example, an application could be defined where the nominal central frequency granularity is 12.5 GHz (by only requiring values of n that are even) and that only requires slot widths as a multiple of 25 GHz (by only requiring values of m that are even).

Hence, in order to support all possible applications and implementations the following information should be advertised for a flexi-grid DWDM link:

o Central frequency granularity: a multiplier of 6.25 GHz.

o Slot width granularity: a multiplier of 12.5 GHz.

o Slot width range: two multipliers of 12.5GHz, each indicate the minimal and maximal slot width supported by a port respectively.

The combination of slot width range and slot width granularity can be used to determine the slot widths set supported by a port.

3.3. Comparison with Fixed-grid DWDM Links

In the case of fixed-grid DWDM links, each wavelength has a pre-defined central frequency and each wavelength has the same frequency range (i.e., there is a uniform channel spacing). Hence all the wavelengths on a DWDM link can be identified uniquely simply by giving it an identifier (such as the central wavelength [RFC6205]), and the status of the wavelengths (available or not) can be advertised through a routing protocol.

Figure 2 shows a link that supports a fixed-grid with 50 GHz channel spacing. The central frequencies of the wavelengths are pre-defined by values of 'n' and each wavelength occupies a fixed 50 GHz frequency range as described in [G.694.1].

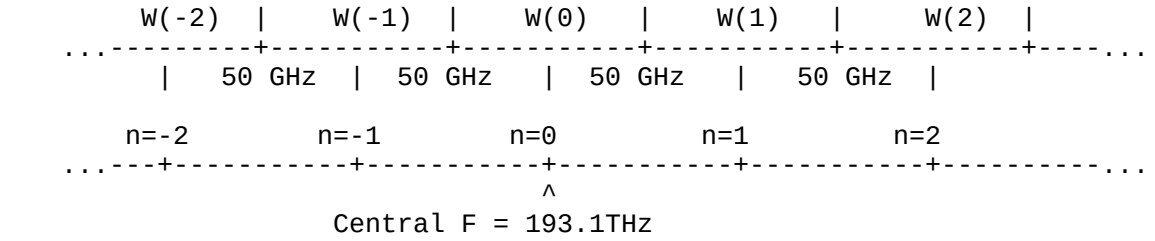


Figure 2 - A Link Supports Fixed Wavelengths with 50 GHz Channel Spacing

Unlike the fixed-grid DWDM links, on a flexi-grid DWDM link the slot width of the frequency slot are flexible as described in section 3.1. That is, the value of m in the formula is uncertain before a frequency slot is actually allocated. For this reason, the available frequency slot/ranges need to be advertised for a flexi-grid DWDM link instead of the specific "wavelengths" that are sufficient for a fixed-grid link.

4. Extensions

As described in [FLEX-FWK], the network connectivity topology constructed by the links/nodes and node capabilities are the same as

for WSON, and can be advertised by the GMPLS routing protocols (refer to section 6.2 of [RFC6163]). In the flexi-grid case, the available frequency ranges instead of the specific "wavelengths" are advertised for the link. This section defines the GMPLS OSPF-TE extensions in support of advertising the available frequency ranges for flexi-grid DWDM links.

4.1. ISCD for Flexi-grid

Value	Type
152 (TBA by IANA)	Flexi-Grid-LSC capable (DWDM-LSC)

Switching Capability and Encoding values MUST be used as follows:

Switching Capability = Flexi-Grid-LSC

Encoding Type = lambda [as defined in RFC3471]

When Switching Capability and Encoding fields are set to values as stated above, the Interface Switching Capability Descriptor MUST be interpreted as in RFC4203 with the optional inclusion of one or more Switching Capability Specific Information sub-TLVs.

4.2. Available Labels Set Sub-TLV

As described in section 3.1, the available frequency ranges other than the available frequency slots should be advertised for the flexi-grid DWDM links. The label encoding defined in [FLEX-LBL] is used to encode the label field in Available Labels Set sub-TLV [GEN-Encode].

4.2.1. Inclusive/Exclusive Label Range

The inclusive/exclusive label ranges format of the Available Labels Set sub-TLV defined in [GEN-ENCODE] can be used for specifying the frequency ranges of the flexi-grid DWDM links.

Note that multiple Available Labels Set sub-TLVs may be needed if there are multiple discontinuous frequency ranges on a link.

4.2.2. Inclusive/Exclusive Label Lists

The inclusive/exclusive label lists format of Available Labels Set sub-TLV defined in [GEN-ENCODE] can be used for specifying the available central frequencies of flexi-grid DWDM links.

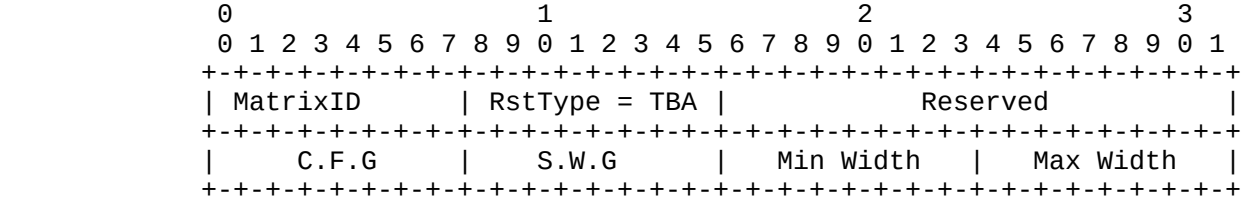
4.2.3. Bitmap

The bitmap format of Available Labels Set sub-TLV defined in [GEN-ENCODE] can be used for specifying the available central frequencies of the flexi-grid DWDM links.

Each bit in the bit map represents a particular central frequency with a value of 1/0 indicating whether the central frequency is in the set or not. Bit position zero represents the lowest central frequency and corresponds to the base label, while each succeeding bit position represents the next central frequency logically above the previous.

4.3. Extensions to Port Label Restriction sub-TLV

As described in Section 3.2, a port that supports flexi-grid may support only a restricted subset of the full flexible grid. The Port Label Restriction sub-TLV is defined in [GEN-ENCODE] and [GEN-OSPF]. It can be used to describe the label restrictions on a port. A new restriction type, the flexi-grid Restriction Type, is defined here to specify the restrictions on a port to support flexi-grid.



MatrixID (8 bits): As defined in [GEN-ENCODE].

RstType (Restriction Type, 8 bits): Takes the value (TBD) to indicate the restrictions on a port to support flexi-grid.

C.F.G (Central Frequency Granularity, 8 bits): A positive integer. Its value indicates the multiple of 6.25 GHz in terms of central frequency granularity.

S.W.G (Slot Width Granularity, 8 bits): A positive integer. Its value indicates the multiple of 12.5 GHz in terms of slot width granularity.

Min Width (8 bits): A positive integer. Its value indicates the multiple of 12.5 GHz in terms of the supported minimal slot width.

Max Width (8 bits): A positive integer. Its value indicates the multiple of 12.5 GHz in terms of the supported maximal slot width.

4.4. Examples for Available Label Set Sub-TLV

Figure 3 shows an example of available frequency range of a flexi-grid DWDM link.

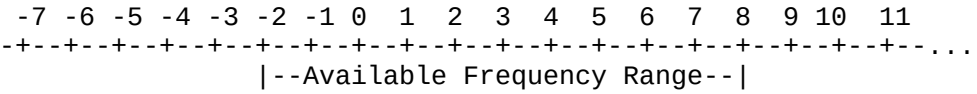


Figure 3 - Flexi-grid DWDM Link

The symbol '+' represents the allowed nominal central frequency. The symbol "--" represents a 6.25 GHz frequency unit. The number on the top of the line represents the 'n' in the frequency calculation formula (193.1 + n * 0.00625). The nominal central frequency is 193.1 THz when n equals zero.

Assume that the central frequency granularity is 6.25GHz, the label set can be encoded as follows:

Inclusive Label Range:

- Start Slot = -2;
- End Slot = 8.

The available central frequencies (-1, 0, 1, 2, 3, 4, 5, 6, 7) can be deduced by the Inclusive Label Range, because the Central Frequency Granularity is 6.25 GHz.

Inclusive Label Lists:

- List Entry 1 = slot -1;
- List Entry 2 = slot 0;
- List Entry 3 = slot 1;
- List Entry 4 = slot 2;
- List Entry 5 = slot 3;
- List Entry 6 = slot 4;
- List Entry 7 = slot 5;
- List Entry 8 = slot 6;
- List Entry 9 = slot 7.

Bitmap:

- Base Slot = -1;
- Bitmap = 111111111(padded out to a full multiple of 32 bits)

5. IANA Considerations

[GEN-OSPF] defines the Port label Restriction sub-TLV of OSPF TE Link TLV. It also creates a registry of values of the Restriction Type field of that sub-TLV

IANA is requested to assign a new value from that registry as follows:

Value	Meaning	Reference
TBD	Flexi-grid restriction	[This.I-D]

6. Implementation Status

[RFC Editor Note: Please remove this entire seciton prior to publication as an RFC.]

This section records the status of known implementations of the protocol defined by this specification at the time of posting of this Internet-Draft, and is based on a proposal described in RFC 6982[RFC6982]. The description of implementations in this section is intended to assist the IETF in its decision processes in progressing drafts to RFCs. Please note that the listing of any individual implementation here does not imply endorsement by the IETF. Furthermore, no effort has been spent to verify the information presented here that was supplied by IETF contributors. This is not intended as, and must not be construed to be, a catalog of available implementations or their features. Readers are advised to note that other implementations may exist.

According to RFC 6982, "this will allow reviewers and working groups to assign due consideration to documents that have the benefit of running code, which may serve as evidence of valuable experimentation and feedback that have made the implemented protocols more mature. It is up to the individual working groups to use this information as they see fit.

6.1. Centre Tecnologic de Telecomunicacions de Catalunya (CTTC)

Organization Responsible for the Implementation: CTTC - Centre Tecnologic de Telecomunicacions de Catalunya (CTTC), Optical Networks and Systems Department, http://wikiona.cttc.es.

Implementation Name and Details: ADRENALINE testbed, http://networks.cttc.es/experimental-testbeds/

Brief Description: Experimental testbed implementation of GMPLS/PCE control plane.

Level of Maturity: Implemented as extensions to a mature GMLPS/PCE control plane. It is limited to research / prototyping stages but it has been used successfully for more than the last five years.

Coverage: Support for the 64 bit label [FLEX-LBL] for flexi-grid as described in this document, with available label set encoded as bitmap. It is expected that this implementation will evolve to follow the evolution of this document.

Licensing: Proprietary

Implementation Experience: Implementation of this document reports no issues. General implementation experience has been reported in a number of journal papers. Contact Ramon Casellas for more information or see http://networks.cttc.es/publications/?search=GMPLS&research_area=optical-networks-systems

Contact Information: Ramon Casellas: ramon.casellas@cttc.es

Interoperability: No report.

7. Acknowledgments

This work was supported in part by the FP-7 IDEALIST project under grant agreement number 317999.

8. Security Considerations

This document does not introduce any further security issues other than those discussed in [RFC3630], [RFC4203].

9. References

9.1. Normative References

[RFC2119] S. Bradner, "Key words for use in RFCs to indicate requirements levels", RFC 2119, March 1997.

[G.694.1] ITU-T Recommendation G.694.1 (revision 2), "Spectral grids for WDM applications: DWDM frequency grid", February 2012.

[GEN-ENCODE] Bernstein, G., Lee, Y., Li, D., and W. Imajuku, "General Network Element Constraint Encoding for GMPLS Controlled Networks", draft-ietf-ccamp-general-constraint-encode, work in progress.

[GEN-OSPF] Fatai Zhang, Y. Lee, Jianrui Han, G. Bernstein and Yunbin Xu, " OSPF-TE Extensions for General Network Element Constraints ", draft-ietf-ccamp-gmpls-general-constraints-ospf-te, work in progress.

[RFC6205] T. Otani and D. Li, "Generalized Labels for Lambda-Switch-Capable (LSC) Label Switching Routers", RFC 6205, March 2011.

[FLEX-LBL] King, D., Farrel, A. and Y. Li, "Generalized Labels for the Flexi-Grid in Lambda Switch Capable (LSC) Label Switching Routers", draft-farrkingel-ccamp-flexigrid-lambda-label, work in progress.

9.2. Informative References

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

[FLEX-SIG] F.Zhang et al, "RSVP-TE Signaling Extensions in support of Flexible-grid", draft-zhang-ccamp-flexible-grid-rsvp-te-ext, work in progress.

[FLEX-FWK] Gonzalez de Dios, O,, Casellas R., Zhang, F., Fu, X., Ceccarelli, D., and I. Hussain, ''Framework and Requirements for GMPLS based control of Flexi-grid DWDM networks', draft-ogrcetal-cammp-flexi-grid-fwk, work in progress.

[WSON-OSPF] Y. Lee and G. Bernstein, "GMPLS OSPF Enhancement for Signal and Network Element Compatibility for Wavelength Switched Optical Networks ", draft-ietf-ccamp-wson-signal-compatibility-ospf, work in progress.

10. Authors' Addresses

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

Haomian Zheng
Huawei Technologies
Email: zhenghaomian@huawei.com

Ramon Casellas, Ph.D.
CTTC
Spain
Phone: +34 936452916
Email: ramon.casellas@cttc.es

Oscar Gonzalez de Dios
Telefonica Investigacion y Desarrollo
Emilio Vargas 6
Madrid, 28045
Spain
Phone: +34 913374013
Email: ogondio@tid.es

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

11. Contributors' Addresses

Adrian Farrel
Old Dog Consulting
Email: adrian@olddog.co.uk

Fatai Zhang
Huawei Technologies
Email: zhangfatai@huawei.com

Lei Wang,
ZTE
Email: wang.lei31@zte.com.cn

Network Working Group
Internet-Draft
Intended status: Standards Track

Fatai Zhang
Xian Zhang
Huawei
Adrian Farrel
Old Dog Consulting
Oscar Gonzalez de Dios
Telefonica
D. Ceccarelli
Ericsson

Expires: August 14, 2014

February 14, 2014

RSVP-TE Signaling Extensions in support of Flexible Grid

draft-zhang-ccamp-flexible-grid-rsvp-te-ext-04.txt

Abstract

This memo describes the extensions to RSVP-TE signaling to support Label Switched Paths in a GMPLS-controlled network that includes devices using the flexible optical grid.

Status of this Memo

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

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

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

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt.

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

This Internet-Draft will expire on August 12, 2014.

Copyright Notice

Table of Contents

1. Introduction

[G.694.1] defines the Dense Wavelength Division Multiplexing (DWDM) frequency grids for Wavelength Division Multiplexing (WDM) applications. A frequency grid is a reference set of frequencies used to denote allowed nominal central frequencies that may be used

for defining applications that utilize WDM transmission. The channel spacing is the frequency spacing between two allowed nominal central frequencies. All of the wavelengths on a fiber use different central frequencies and occupy a designated range of frequency.

Fixed grid channel spacing is selected from 12.5 GHz, 25 GHz, 50 GHz, 100 GHz and integer multiples of 100 GHz. But [G.694.1] also defines ''flexible grids'', known as ''flexi-grid''. The terms ''frequency slot'' (i.e., the frequency range allocated to a specific channel and unavailable to other channels within a flexible grid) and ''slot width'' (i.e., the full width of a frequency slot in a flexible grid) are introduced in [G.694.1] to define a flexible grid.

[FLEX-FWK] defines a framework and the associated control plane requirements for the GMPLS based control of flexi-grid DWDM networks.

[RFC6163] provides a framework for GMPLS and Path Computation Element (PCE) control of Wavelength Switched Optical Networks (WSONs), and [WSON-SIG] describes the requirements and protocol extensions for signaling to set up Label Switched Paths (LSPs) in WSONs.

This document describes the additional requirements and protocol extensions for RSVP-TE signaling to set up LSPs in networks that support the flexi-grid.

2. Terminology

For terminology related to flexi-grid, please refer to [FLEX-FWK] and [G.694.1].

2.1. Conventions used in this document

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

3. Requirements for Flexible Grid Signaling

The architecture for establishing LSPs in a flexi-grid network is described in [FLEX-FWK].

An optical spectrum LSP occupies a specific frequency slot, i.e., a range of frequencies. The process of computing a route and the allocation of a frequency slot is referred to as RSA (Routing and Spectrum Assignment). [FLEX-FWK] describes three architectural approaches to RSA: combined RSA, separated RSA, and distributed SA.

The first two approaches are referred to as ''centralized SA'' because both routing and spectrum (frequency slot) assignment are performed by a centralized entity before the signaling procedure.

In the case of centralized SA the assigned frequency slot is specified in the RSVP-TE Path message during LSP setup. In the case of distributed SA, the slot width of the flexi-grid LSP is specified in the Path message, allowing the network elements to select the frequency slot to be used when they process the RSVP-TE messages.

If the capability to switch or convert the whole optical spectrum allocated to an optical spectrum LSP is not available at some nodes along the path of the LSP, the LSP is subject to the Optical ''Spectrum Continuity Constraint'' as described in [FLEX-FWK].

The remainder of this section states the additional requirements for signaling in a flexi-grid network.

3.1. Slot Width

The slot width is an end-to-end parameter representing how much frequency resource is requested for a flexi-grid LSP. It is the equivalent of optical bandwidth, although the amount of bandwidth associated with a slot width will depend on the signal encoding.

Different LSPs may request different amounts of frequency resource in flexible grid networks, so the slot width needs to be carried in the signaling message during LSP establishment. This enables the nodes along the LSP to know how much frequency resource has been requested (in a Path message) and has been allocated (by a Resv message) for the LSP.

3.2. Frequency Slot

The frequency slot information identifies which part of the frequency spectrum is allocated on each link for an LSP in a flexi-grid network.

This information is required in a Resv message to indicate, hop-by-hop, the central frequency of the allocated resource. In combination with the slot width indicated in a Resv message (see Section 3.1) the central frequency carried in a Resv message identifies the resources reserved for the LSP (known as the frequency slot).

The frequency slot can be represented by the two parameters as follows:

```
Frequency slot = [(central frequency) - (slot width)/2] ~
                 [(central frequency) + (slot width)/2]
```

As is common with other resource identifiers (i.e., labels) in GMPLS signaling, it must be possible for the head-end LSP when sending a Path message to suggest or require the central frequency to be used for the LSP. Furthermore, for bidirectional LSPs, the Path message must be able to specify the central frequency to be used for reverse direction traffic.

As described in [G.694.1], the allowed frequency slots for the flexible DWDM grid have a nominal central frequency (in THz) defined by:

193.1 + n * 0.00625

where n is zero or a positive or negative integer.

The slot width (in GHz) is defined as:

12.5 * m

where m is a positive integer.

It is possible that implementing a subset of the possible slot widths and central frequencies are supported. For example, an implementation could built where the nominal central frequency granularity is 12.5 GHz (by only requiring values of n that are even) and that only supports slot widths as a multiple of 25 GHz (by only allowing values of m that are even).

Further details can be found in [FLEX-FWK].

4. Protocol Extensions

This section defines the extensions to RSVP-TE signaling for GMPLS [RFC3473] to support flexible grid networks.

4.1. Traffic Parameters

In RSVP-TE, the SENDER_TSPEC object in the Path message indicates the requested resource reservation. The FLOWSPEC object in the Resv message indicates the actual resource reservation.

As described in Section 3.1, the slot width represents how much frequency resource is requested for a flexi-grid LSP. That is, it

describes the end-to-end traffic profile of the LSP. Therefore, the traffic parameters for a flexi-grid LSP encode the slot width.

This document defines new C-Types for the SENDER_TSPEC and FLOWSPEC objects to carry Spectrum Switched Optical Network (SSON) traffic parameters:

SSON SENDER_TSPEC: Class = 12, C-Type = TBD1.

SSON FLOWSPEC: Class = 9, C-Type = TBD2.

The SSON traffic parameters carried in both objects have the same format as shown in Figure 1.

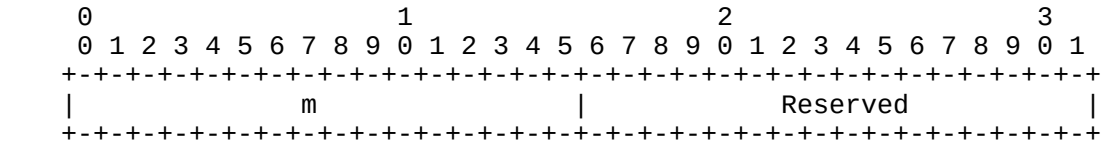


Figure 1: The SSON Traffic Parameters

m (16 bits): the slot width is specified by m*12.5 GHz.

The Reserved bits MUST be set to zero and ignored upon receipt.

4.1.1. Applicability to Fixed Grid Networks

Note that the slot width (i.e., traffic parameters) of a fixed grid defined in [G.694.1] can also be specified by using the SSON traffic parameters. The fixed grid channel spacings (12.5 GHz, 25 GHz, 50 GHz, 100 GHz and integer multiples of 100 GHz) are also the multiples of 12.5 GHz, so the m parameter can be used to represent these slot widths.

Therefore, it is possible to consider using the new traffic parameter object types in common signaling messages for flexi-grid and legacy DWDM networks.

4.2. Generalized Label

In the case of a flexible grid network, the labels that have been requested or allocated as signaled in the RSVP-TE objects are encoded as described in [FLEX-LBL]. This new label encoding can appear in any RSVP-TE object or sub-object that can carry a label.

As noted in Section 4.2 of [FLEX-LBL], the m parameter forms part of the label as well as part of the traffic parameters.

4.3. Signaling Procedures

There are no differences between the signaling procedure described for LSP control in [FLEX-FWK] and those required for use in a fixed-grid network [WSON-SIG]. Obviously, the TSpec, FlowSpec, and label formats described in Section 3 are used. The signaling procedures for distributed SA and centralized SA can be applied.

5. IANA Considerations

5.1. RSVP Objects Class Types

This document introduces two new Class Types for existing RSVP objects. IANA is requested to make allocations from the "Resource ReSerVation Protocol (RSVP) Parameters" registry using the "Class Names, Class Numbers, and Class Types" sub-registry.

Class Number	Class Name	Reference
9	FLOWSPEC	[RFC2205]
	Class Type (C-Type):	
	(TBD2) SSON FLOWSPEC	[This.I-D]

Class Number	Class Name	Reference
12	SENDER_TSPEC	[RFC2205]
	Class Type (C-Type):	
	(TBD1) SSON SENDER_TSPEC	[This.I-D]

IANA is requested to assign the same value for TBD1 and TBD2, and a value of 8 is suggested.

6. Manageability Considerations

This document makes minor modifications to GMPLS signaling, but does not change the manageability considerations for such networks. Clearly, protocol analysis tools and other diagnostic aids (including logging systems and MIB modules) will need to be enhanced to support the new traffic parameters and label formats.

7. Implementation Status

[RFC Editor Note: Please remove this entire seciton prior to publication as an RFC.]

This section records the status of known implementations of the protocol defined by this specification at the time of posting of this Internet-Draft, and is based on a proposal described in RFC 6982 [RFC6982]. The description of implementations in this section is intended to assist the IETF in its decision processes in progressing drafts to RFCs. Please note that the listing of any individual implementation here does not imply endorsement by the IETF. Furthermore, no effort has been spent to verify the information presented here that was supplied by IETF contributors. This is not intended as, and must not be construed to be, a catalog of available implementations or their features. Readers are advised to note that other implementations may exist.

According to RFC 6982, "this will allow reviewers and working groups to assign due consideration to documents that have the benefit of running code, which may serve as evidence of valuable experimentation and feedback that have made the implemented protocols more mature. It is up to the individual working groups to use this information as they see fit."

7.1. Centre Tecnologic de Telecomunicacions de Catalunya (CTTC)

Organization Responsible for the Implementation:
Centre Tecnologic de Telecomunicacions de Catalunya (CTTC)
Optical Networks and Systems Department

Implementation Name and Details:
ADRENALINE testbed
http://networks.cttc.es/experimental-testbeds/

Brief Description:
Experimental testbed implementation of GMPLS/PCE control plane.

Level of Maturity:
Implemented as extensions to a mature GMLPS/PCE control plane. It is limited to research / prototyping stages, but it has been used successfully for more than the last five years.

Coverage:
Support for the Tspec, FlowSpec, and label formats as described version 03 of this document. Label format support extends to the following RSVP-TE objects and sub-objects:

- Generalized Label Object
- Suggested Label Object
- Upstream Label Object
- ERO Label Subobjects

It is expected that this implementation will evolve to follow the evolution of this document.

Licensing:
Proprietary

Implementation Experience:
Implementation of this document reports no issues.
General implementation experience has been reported in a number of journal papers. Contact Ramon Casellas for more information or see

http://networks.cttc.es/publications/?search=GMPLS&research_area=optical-networks-systems

Contact Information:
Ramon Casellas: ramon.casellas@cttc.es

Interoperability:
No report.

8. Acknowledgments

This work was supported in part by the FP-7 IDEALIST project under grant agreement number 317999.

9. Security Considerations

This document introduces no new security considerations to [RFC3473].

See also [RFC5920] for a discussion of security considerations for GMPLS signaling.

10. References

10.1. Normative References

[RFC2119] S. Bradner, "Key words for use in RFCs to indicate requirements levels", RFC 2119, March 1997.

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

[G.694.1] ITU-T Recommendation G.694.1 (revision 2), ''Spectral grids for WDM applications: DWDM frequency grid'', February 2012.

[FLEX-LBL]King, D., Farrel, A. and Y. Li, ''Generalized Labels for the Flexi-Grid in Lambda Switched Capable (LSC) Label Switching Routers'', draft-farrkingel-ccamp-flexigrid-lambda-label, work in progress.

10.2. Informative References

[RFC2205] Braden, R., Zhang L., Berson, S., Herzog, S. and S. Jamin, ''Resource ReServation Protocol (RSVP) -- Version 1, Functional Specification', RFC2205, September 1997.

[RFC5920] L. Fang et al., "Security Framework for MPLS and GMPLS Networks", RFC 5920, July 2010.

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

[RFC6982] Sheffer, Y. and A. Farrel, "Improving Awareness of Running Code: The Implementation Status Section", RFC 6982, July 2013.

[RFC Editor Note: This reference can be removed when Section 7 is removed]

[FLEX-FWK] Gonzalez de Dios, O,, Casellas R., Zhang, F., Fu, X., Ceccarelli, D., and I. Hussain, ''Framework and Requirements for GMPLS based control of Flexi-grid DWDM networks', draft-ogrcetal-cammp-flexi-grid-fwk, work in progress.

[WSON-SIG] G. Bernstein, Sugang Xu, Y. Lee, G. Martinelli and Hiroaki Harai, "Signaling Extensions for Wavelength Switched Optical Networks", draft-ietf-ccamp-wson-signaling, work in progress.

11. Contributors' Address

Ramon Casellas
CTTC
Av. Carl Friedrich Gauss n7
Castelldefels, Barcelona 08860
Spain

Email: ramon.casellas@cttc.es

Felipe Jimenez Arribas
Telefonica Investigacion y Desarrollo
Emilio Vargas 6
Madrid, 28045
Spain
Email: felipej@tid.es

Yi Lin
Huawei Technologies Co., Ltd.
F3-5-B R&D Center, Huawei Base,
Bantian, Longgang District
Shenzhen 518129 P.R.China

Phone: +86-755-28972914
Email: yi.lin@huawei.com

Qilei Wang

ZTE

wang.qilei@zte.com.cn

Haomian Zheng

Huawei Technologies
zhenghaomian@huawei.com

12. Authors' Addresses

Fatai Zhang
Huawei Technologies
Email: zhangfatai@huawei.com

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

Adrian Farrel
Old Dog Consulting
Email: adrian@olddog.co.uk

Oscar Gonzalez de Dios
Telefonica Investigacion y Desarrollo
Emilio Vargas 6
Madrid, 28045
Spain

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

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

Network Working Group
Internet Draft
Category: Informational

Xian Zhang
Haomian Zheng
Huawei

Expires: August 14, 2014

February 14, 2014

Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Signaling Procedure for Resource Sharing-based LSP Setup/Teardown

draft-zhang-ccamp-gmpls-resource-sharing-proc-00.txt

Status of this Memo

Copyright Notice

Abstract

Generalized Multiprotocol Label Switching (GMPLS) defines a set of protocols for the creation of Label Switched Paths (LSPs) in various switching technologies. It can be used for different types of switching technologies.

This document compliments existing standards by explaining the missing pieces of information during the Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) signaling procedure in support of resource sharing-based LSP setup/teardown in GMPLS-controlled circuit networks.

Table of Contents

1. Introduction

Generalized Multiprotocol Label Switching (GMPLS) [RFC3945] defines a set of protocols, including Open Shortest Path Fist - 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 and Fiber Switching.

In MPLS networks, in order to avoid double booking of resource during the process of LSP restoration or LSP re-optimization, the Make-Before-Break (MBB) exploiting the Shared-Explicit (SE) reservation style can be employed, as specified in [RFC3209]. This method is also used in GMPLS-controlled networks [RFC4872] [RFC4873] for end-to-end and segment recoveries 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. For example, in MPLS networks, label has no meaning/match in the data plane but this is not the case in GMPLS-controlled circuit networks, such as Optical Transport Network (OTN) and Wavelength-Switched Optical Networks (WSON), where the label matches the resource used in the data plane. So, 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 has upon the traffic delivery. Some other issues are also discussed in Section 2.

The purpose of this draft is to describe the signaling process for resource sharing-based LSP setup/teardown for GMPLS-based circuit networks. This includes the node behavior description, besides clarifying some un-discussed points for this process. Two typical examples mentioned in this draft are LSP restoration and LSP re-optimization, where it is desirable to share resources. This draft does not define any RSVP-TE extensions. If necessary, discussions may be provided to identify potential extensions to the existing RSVP-TE protocol. It is expected that the extensions, if there is any, will be addressed in separate drafts.

2. Problem Statement

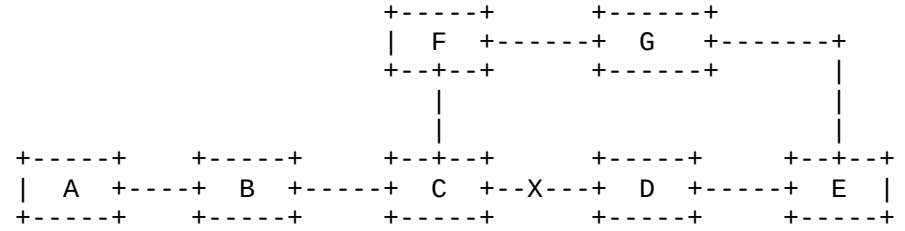


Figure 1: A Simple OTN Network

Using the network shown in Figure 1 as an example, 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 to which alternative path it will establish to reroute the traffic. In this case, A-B-C-F-G-E is chosen as the alternative path and the resource on the path segment A-B-C is re-used by this to-be-established path. Since this is an OTN network, different from packet-switching network, the label has a mapping into the data plane resource used and also the nodes along the path needs to send triggering commands to data plane nodes for setting up cross-connection accordingly during the RSVP-TE signaling process. So, the following issues are left un-described in the existing standards for resource sharing based LSP setup/teardown in GMPLS-controlled circuit networks:

o The purpose of using SE can still be fulfilled?

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, the first point may not be able to be fulfilled under certain scenarios. Thus, the name ''make before break'' may no longer holds true and worth discussion.

o Is the current defined MBB method sufficient in support of resource shared-based LSP setup/teardown?

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 LSP with different tunnel IDs, such as the one mentioned in [PCEP-RSO] and LSPs with LSP-stitching across multi-domains. Thus, how the signaling process can make intermediate nodes be aware of this resource sharing constraint and behavior accordingly is an issue that needs to be described and discussed.

O Other issues such as what is the reservation style assigned to the original LSP, and what is the node behavior during the traffic reversion, in the GMPLS-controlled circuit networks, are missing and should be explained.

3. RSVP-TE Signaling Procedure for Resource Sharing-based LSP Setup/Teardown

This section describes the signaling flow for resource sharing-based LSP setup/teardown in GMPLS-controlled circuit networks.

For LSP restoration upon failure, as explained in Section 11 of [RFC4872], the purpose of using MBB is to re-use existing resource. Thus, the behavior of the intermediate nodes during rerouting process will not impact on 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 optimization. Another dimension that needs separate attention is how to correlate the two LSPs sharing resource. For the ones sharing same tunnel ID, the majority description is provided in existing standards [RFC3209] [RFC4872]. For the ones with different Tunnel IDs, additional extensions are needed and discussed in this section.

3.1. LSPs with the Identical Tunnel ID

For this type of LSP resource sharing, SE flag and ASSOCIATION object are used together. The former is to enable sharing and the object is to identify the two correlated 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 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 phrase or the RESV message forwarding phrase.

3.1.1. LSP Restoration Setup and Reversion

For LSP restoration, the complete signaling flow processes for both LSP restoration upon failure and LSP reversion upon link failure recovery are described.

For LSP rerouting upon working LSP failure, using the network shown in Figure 1 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 nodes or a Path Computation Element (PCE) [RFC4655]. Assume 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 categorized into the following three categories:

Table 1: Node Behavior during LSP Restoration

Category	Node Behavior during LSP Reversion
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 resource 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.
   ---------+-------------------------------------------------------
```

As shown in Figure 2, 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 ASSOCATION object defined in [RFC4872] are followed.

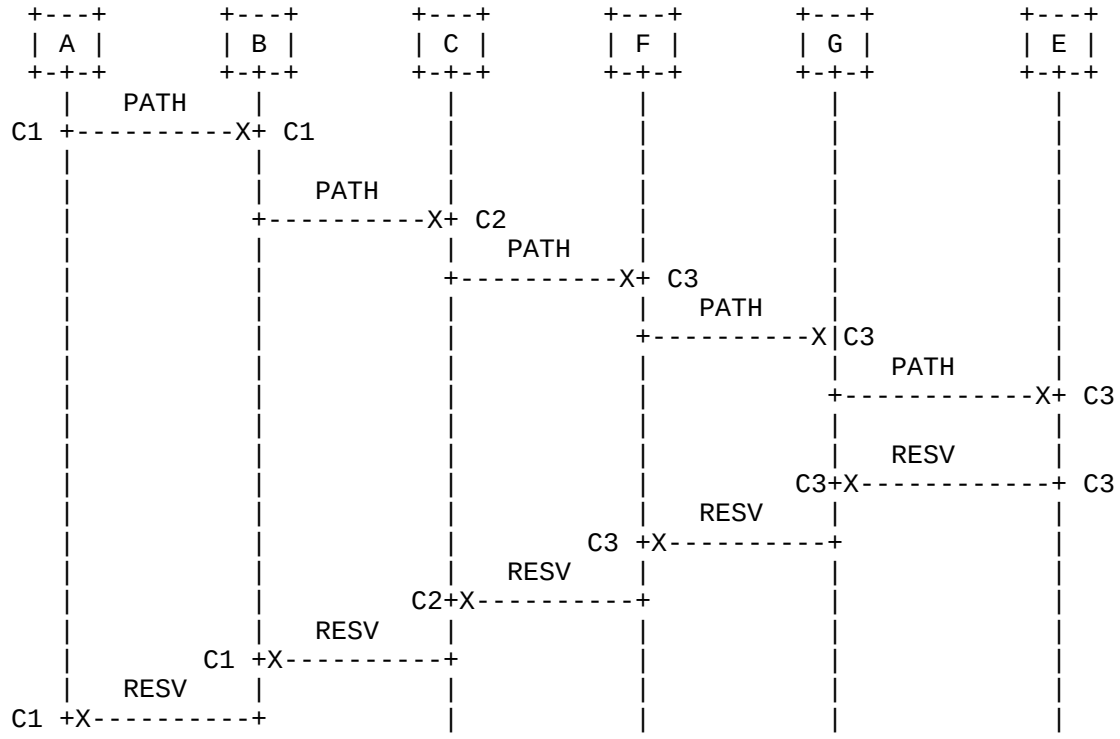


Figure 2: Restoration LSP Setup Signaling Procedure for LSP Restoration

If the LSP rerouting is revertive, which is a common requirement in transport networks [LSP-restoration], the traffic will be reverted to the working LSP if its failure is recovered. The three types of nodes classified above also have different behaviors during the process for tearing down the alternative LSP, as explained in Table 2.

Table 2: Node Behavior during LSP Reversion

Category	Node Behavior during LSP Reversion
D1	+ Resource reused on both interfaces. + When receiving PATH-TEAR, it only deletes the alternative + LSP state info in the control plane without changing the + cross-connection.
D2	+ Resource reused on only one interface. + When receiving PATH-TEAR, it deletes the alternative path + state information in the control plane as well as release + the resource on the interface that is not re-used between + the working and Restoration LSP.
D3	+ No resource are reused. + When receiving PATH-TEAR, it deletes the state information + related to the alternative LSP as well as tears down the + cross-connection to release the resource.

Note that before the working LSP failure recovers, the LSP in the control plane is still running and also it views the data plane resource still belongs to the working LSP. However, the re-used resource also belongs to the alternative LSP and these resources are actually used by the alternative LSP. So when the working LSP recovers, it needs to fresh the signaling messages to re-establish the working LSP cross-connection. The process would be similar to that shown in Figure 2, but running along the nodes on the working LSP path (i.e., A-B-C-D-E). Note this will interrupt the traffic delivery on the alternative LSP (i.e., Making the working LSP While Breaking the alternative LSP). This point is different from that of the MPLS networks. If no traffic interruption is mandated, mechanisms to ensure that the traffic can still be delivered should be employed and is outside the scope of this document.

Figure 3 shows the signaling process of the alternative LSP teardown during the LSP reversion. Similar to that of the alternative LSP setup process, the nodes may not need to reconfigure the cross-connection and the rationale is similar to that described above. For alarm-free LSP deletion in optical networks, the mechanisms described in Section 6 of [RFC4208] should be followed.

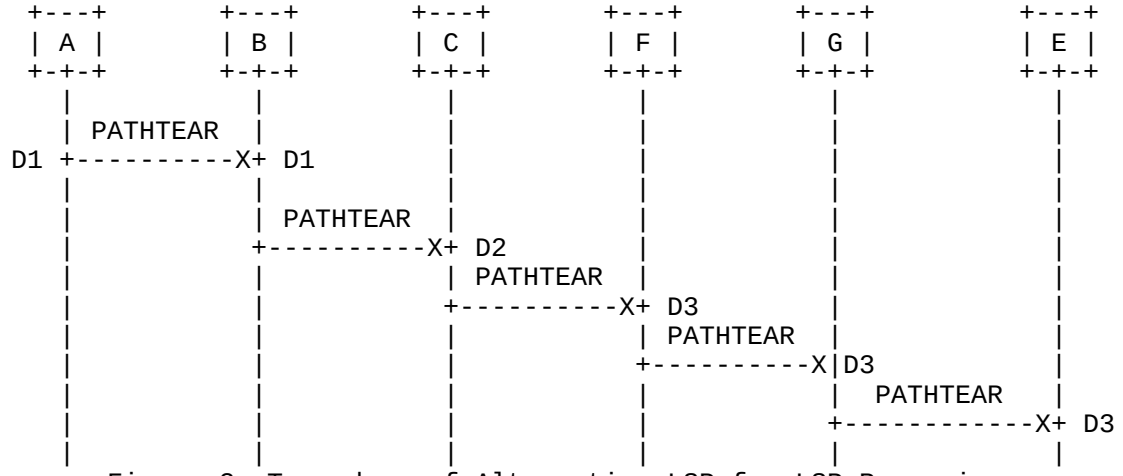


Figure 3: Tear-down of Alternative LSP for LSP Reversion

3.1.2. LSP Re-optimization 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 that are shown in Figure 2 and 3.

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 does not share the resource other than the source and destination nodes or using other mechanisms. This is out the scope of this draft.

3.2. LSPs with Different Tunnel IDs

For two LSPs with different Tunnel IDs, the ASSOCIATION object is used to both specify they are sharing resource (by setting ASSOCIATION type as 2) as well as 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 draft does not impose any constraint with regard to path computation.

In [RFC4873], it only considers resource sharing for LSP segment recovery. The ASSOCIATION object configuration is limited. [RFC6780] extends the usage of ASSOCIATION objects 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 obeyed. 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 so are outside of the scope of this document.

Other than this, the signaling flow for this type of resource sharing is similar to description provided in Section 3.1.1. 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 needed and are outside of the scope of this draft.

4. Security Considerations

This draft does not incur any new security issues other than those already covered in [RFC3209] [RFC4872] [RFC4873] and [RFC6780].

5. IANA Considerations

This informational document does not make any requests for IANA action.

6. References

6.1. Normative References

[RFC3209] D. Awduche et al, ''RSVP-TE: Extensions to RSVP for LSP Tunnels'', RFC3209, 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'', RFC4872, May 2007.

[RFC4873] L. Berger et al, ''GMPLS Segment Recovery'', RFC4873, May 2007.

[RFC6780] L. Berger et al, ''RSVP ASSOCIATION Object Extensions'', RFC6780, October 2012.

6.2. Informative References

[LSP-restoration] R. Gandhi, et al, ''RSVP-TE Signaling for GMPLS Restoration LSP'', work in progress, January 2014.

[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.

[RFC4655] A. Farrel et al, ''A Path Computation Element (PCE)-Based Architecture'', RFC4655, 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'', RFC4208, October 2005.

7. 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
Huawei Technologies
F3-1-B R&D Center, Huawei Base
Bantian, Longgang District
Shenzhen 518129 P.R.China

Email: zhenghaomian@huawei.com

CCAMP Working Group
Internet Draft
Category: Standards track

Xian Zhang
Fatai Zhang
Huawei
O. Gonzalez de Dios
Telefonica I+D
Igor Bryskin
ADVA Optical Networking
Dhruv Dhody
Huawei

Expires: August 13, 2014 February 14, 2014

Extensions to Resource ReSerVation Protocol-Traffic Engineering (RSVP-TE) to Support Route Exclusion Using Path Key Subobject

draft-zhang-ccamp-route-exclusion-pathkey-01.txt

Status of this Memo

Requirements Language

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

Abstract

This document extends the Resource ReSerVation Protocol-Traffic Engineering (RSVP-TE) eXclude Route Object (XRO) and Explicit eXclusion Route Subobject (EXRS) within Explicit Route Object (ERO) to support specifying route exclusion requirement using Path Key Subobject (PKS).

Table of Contents

1. Introduction

[RFC5520] defines the concept of a Path Key for confidentiality in a multi-domain environment. This can be used by a Path Computation Element (PCE) in place of a segment of a path that it wishes to keep confidential. The Path Key can be signaled in Resource ReSerVation Protocol-Traffic Engineering (RSVP-TE) protocol by placing it in an Explicit Route Object (ERO) as described in [RFC5553].

When establishing a set of LSPs to provide protection services [RFC4427], it is often desirable that the LSPs should take different paths through the network. This can be achieved by path computation entities that have full end-to-end visibility, but it is more complicated in multi-domain environments when segments of the path may be hidden so that they are not visible outside the domain they traverse.

This document describes how the Path Key object can be used in the RSVP-TE eXclude Route Object (XRO), and the Explicit eXclusion Route subobject (EXRS) of the ERO in order to facilitate path hiding, but allow diverse end-to-end paths to be established in multi-domain environments.

1.1. Example Use

Figure 1 shows a simple network with two domains. 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 will be signaled by the source with ERO {A, B, loose Dst} and will be set up with the path {Src, A, B, U, V, W, Dst}. But 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} which is not path-disjoint.

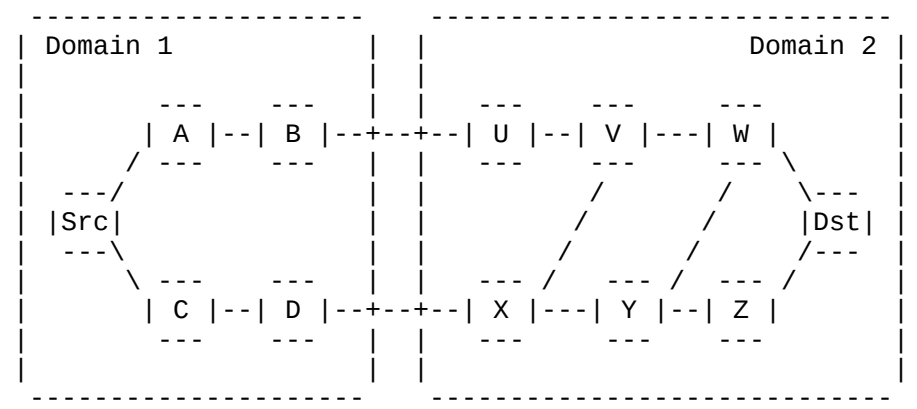


Figure 1: A Simple Multi-Domain Network

In order to improve the outcome, node U can replace the path segment {U, V, W} in the RRO with a Path Key Suboject. The Path Key Subobject assigns an identifier to the key and also indicates that it was node U that made the replacement.

With this additional information, the source is able to signal the second LSP with 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 so know 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.

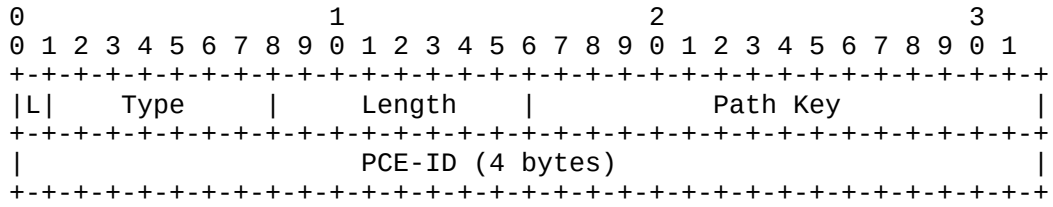
This example uses a PCE deployed in each border router, having at least the capability to expand PKS. Other deployment scenarios (Domain PCE, PCE being part of the Management system) may be used.

2. RSVP-TE Extensions

This section defines the Path Key Subobject that can be either in the XRO object or Explicit eXclusion Route subobject (EXRS) as defined in [RFC4874].

2.1. Path Key Subobject (PKS)

The IPv4 PKS has the same format as defined in [RFC5553] and is detailed as below:



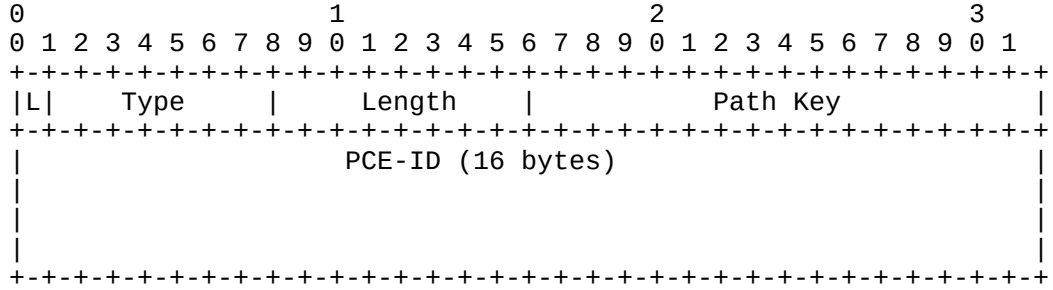
The meaning of the field Length and Path Key is defined in [RFC5553].

L: 0 indicates that the path or path segment hidden with the Path Key specified MUST be excluded. 1 indicates that the path or path segment hidden with the Path Key specified SHOULD be avoided.

Type: sub-object type for XRO Path Key; TBD.

PCE-ID: The IPv4 address of a node that assigned the Path Key identifier and that can return an expansion of the Path Key or use the Path Key as an exclusion in a path computation. Note this draft does not confine whether it is the network element or a dedicated server for path key generation and decoding.

Similarly, the format of IPv6 PKS is as follows:



2.2. PKS Processing Rules

The exclude route list is encoded as a series of subobjects contained in an EXCLUDE_ROUTE object or an EXRS of the ERO. Multiple Path-Keys may be included in XRO or EXRO of ERO if more than segment of a path are kept hidden during diverse path establishment. The procedure defined in [RFC4874] for processing XRO and EXRS is not changed by this document.

Irrespective of the L flag, if the node, receiving the PKS, cannot recognize the subobject, it will react according to [RFC4874] and SHOULD ignore the constraint.

Otherwise, if it decodes the PKS but cannot find a route/route segment meeting the constraint:

-if L flag is set to 0, it will react according to [RFC4874] and SHOULD send a PathErr message with the error code/value combination "Routing Problem" / "Route Blocked by Exclude Route".

-if L flag is set to 1, which means the node SHOULD try to be as much diversified as possible with the specified resource. If it cannot fully support the constraint, it SHOULD send a PathErr message with the error code/value combination "Notify Error" / "Fail to find diversified path" (TBD).

If it cannot decode the PKS, the error handling procedure defined in Section 3.1 of [RFC5553] is not changed by this document.

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

3. Other considerations

3.1. Path-Key Retention and Reuse

The use of the path key relies on the availability of a PCE function supporting [RFC5520] functionality.

Following [RFC4655] a simple deployment option is when the PCE function is collocated with each border domain node generating the RRO. This collocated PCE functionality can be restricted to only serve the PKS resolution. This PCE function is only required to be accessible to the nodes excluding this PKS, so this can be restricted to one domain. This option can very easily tie the lifetime of the PKS to the lifetime of the LSP.

Alternatively, if a dedicated server, such as a PCE, is in charge of this, it may need to be explicitly informed of the LSP tear-down in order to recycle the path key allocated already. This can be easily supported by a stateful PCE [Stateful-PCE]. Note this draft does not confine the methods for path key generation and decoding.

Last, options including allowing a LSR can use the PKS information to index a Path Segment previously supplied to it by the entity that

originated the PKS, for example the LSR that inserted the PKS in the RRO or a management system, can also be used.

3.2. Path-Key Uniqueness

In the CCAMP mailing list, there is concern about whether 16-bit Path key is still enough and future proof. This can be easily solved by confining the scope of a path key. If an ingress node is responsible for managing the Path Key, it should not be an issue since the LSP across domains do not expected to be larger than 65535. On the other hand, if a dedicated entity, such as a PCE server, is used to allocate and recycle the Path Key, it is advised to allocate the Path Key per ingress node basis to avoid the limitation of Path Key numbers facing a domain-based allocation space. These are only illustrative examples and other methods that can guarantee the uniqueness of Path-Key are not precluded.

3.3. PKS Update

When the information of a path is changed, the LSPs using that path and corresponding PKS should be aware of the changes. The procedures defined in Section 4.4.3 of RFC 3209 [RFC3209] MUST be used to refresh the PKS information if the PKS change is to be communicated to other nodes according to the local node's policy. If local policy is that the PKS change should be suppressed or would result in no change to the PKS expansion, the node does not need to send an update. This procedure allows for ingress node to react on path change.

4. Manageability Considerations

4.1. Control of Function through Configuration and Policy

In addition to the set of policies described in [RFC5553] the following policies (are local and domain-wide) SHOULD be available for configuration in an implementation:

- Handling a XRO or EXRS containing a PKS. As described in Section 2.2, an LSR that receives a Path message containing a PKS exclusion can be configured to reject the Path message according to policy.

- Hiding of reason codes. The policy described in [RFC5553] section 5.1 is also applicable to policies for PKS in XRO or EXRS.

This document makes no other new management consideration to RSVP and PCE, the existing consideration applies.

5. Security Considerations

The use of path keys proposed in this draft allows nodes to hide parts of the path as it is signaled. This can be used to improve the confidentially of the LSP setup. Moreover, it may serve to improve security of the control plane for the LSP as well as data plane traffic carried on this LSP. However, the benefits of using path key are lost unless there is an appropriate access control of any tool that allows expansion of the path key.

6. IANA Considerations

6.1. New Subobject Type

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 Type	Subobject Description
64(TBD by IANA)	IPv4 Path Key Subobject
65(TBD By IANA)	IPv6 Path Key Subobject

Note: [RFC5520] defines the PKS for use in PCEP. The above number suggestions for use in RSVP-TE follow that assigned for the PKS in PCEP [RFC5520].

6.2. New Error Code

IANA registry: RSVP PARAMETERS

Subsection: Error Codes and Globally-Defined Error Value Sub-Codes

New Error Values sub-codes have been registered for the Error Code ’Notify Error’ (25).

TBD = "Fail to find diversified path"

7. Acknowledgments

The authors would like to thank John Drake, Daniele Ceccarelli and Zafar Ali for their comments and dicussions.

8. References

8.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to indicate requirements levels", RFC 2119, March 1997.

[RFC3209] D. Awduche et al, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC3209, December 2001.

[RFC4874] CY. Lee, A. Farrel, S. De Cnodder, "Exclude Routes - Extension to Resource ReserVation Protocol-Traffic Engineering (RSVP-TE), RFC4874, April 2007.

[RFC5553] A. Farrel, Ed., "Resource Reservation Protocol (RSVP) Extensions for Path Key Support", RFC5553, May 2009.

8.2. Informative References

[RFC5520] R. Bradford, Ed., "Preserving Topology Confidentiality in Inter-Domain Path Computation Using a Path-Key-Based Mechanism", RFC5520, April 2009.

[RFC4427] E. Mannie, Ed., "Recovery (Protection and Restoration) Terminology for Generalized Multi-Protocol Label Switching (GMPLS)", RFC4427, March 2006.

[Stateful-PCE] Crabbe, E., Medved, J., Minei, I., and R. Varga, "PCEP Extensions for Stateful PCE", draft-ietf-pce-stateful-pce-06 (work in progress), August 2013.

9. Contributors

Cyril
cyril.margaria@gmail.com

10. Authors' Addresses

Xian Zhang
Huawei Technologies

Email: zhang.xian@huawei.com

Fatai Zhang
Huawei Technologies

Email: zhangfatai@huawei.com

Oscar Gonzalez de Dios
Telefonica I+D
Don Ramon de la Cruz
Madrid, 28006
Spain

Phone: +34 913328832
Email: ogondio@tid.es

Igor Bryskin
ADVA Optical Networking

Email: ibryskin@advaoptical.com

Dhruv Dhody
Huawei Technologies
Leela Palace
Bangalore, Karnataka 560008
INDIA

EMail: dhruv.ietf@gmail.com