CCAMP Working Group
Internet Draft
Intended status: Standards Track
Expires: April 21, 2012
October 21, 2011

V.Beeram
I.Bryskin
W.Doonan
ADVA Optical Networking
J.Drake
G.Grammel
Juniper Networks
Manuel Paul
Ruediger Kunze
Deutsche Telekom

GMPLS-UNI BCP
draft-beeram-ccamp-gmpls-uni-bcp-00.txt

Status of this Memo

Copyright Notice

Abstract

This document pools together the best current practices that are being used to apply the GMPLS Overlay model at the User-Network Interface (UNI) reference point (as defined in [G.8080])

Table of Contents

1. Introduction

Generalized Multiprotocol Label Switching (GMPLS) provides tools to create end-to-end services in various transport technologies. These tools can be used to support service management in different types of deployment models. RFC 4208 discusses how GMPLS can be applied to the overlay model. There are a good number of implementations that have built on the basic concepts discussed in RFC 4208 and have successfully demonstrated interoperability. This document is an attempt to pool together the best current practices that are being used to apply the GMPLS Overlay model at the User-Network Interface (UNI) reference point (as defined in [G.8080]).

RFC 4208 recommends the use of hierarchical service activation when GMPLS is used for the core network - this document takes this concept further, discusses the notion of "augmenting the client-layer TE topology" and explains how this augmentation enables client-layer networking in an overlay model. The concepts discussed in this document are based primarily on experiences drawn from interoperating GMPLS-enabled IP routers with Optical Transport elements.

2. Multi-Layered Approach

When an end-to-end service crosses a boundary between two regions of dissimilar transport technology, it is necessary to execute distinct forms of service activation within each region.

Fig 1: Sample Hybrid Topology
(See PDF version)

For example, in the hybrid network illustrated in Fig 1, provisioning a transport service between two GMPLS-enabled IP routers on either side of the optical WDM transport topology requires operations in two distinct layer networks; the client-layer network interconnecting the routers themselves, and the server-layer network interconnecting the optical transport elements in between the routers.

Activation of the end-to-end service begins with a path determination process, followed by the initiation of a signaling process from the ingress along the determined path, per the set of figures shown in Fig2.

Fig 2: Hierarchical Service Action
(See PDF version)

3. Traffic Engineering

The previous section outlines the basic method for activating end-to-end services across a multi-layer network. As a necessary part of that process an initial path selection process was performed, whereby an appropriate path between the desired endpoints was determined through some means. Further, per expectations set through current practices with regard to service provisioning in homogeneous networks, operators expect that the underlying control plane system will provide automated mechanisms for computing the desired path or paths between network endpoints.

In particular, operators do not expect under normal circumstances to be required to explicitly specify the end-to-end path; rather, operators expect to be able to specify just the endpoints of the path and rely on an automated computational process to identify and qualify all the elements and links on the path between them. Hence when operating a hybrid network such as that described in Fig 1, it is necessary to extend existing traffic engineering and path computation mechanisms to operate in a similar manner.

Path computation and qualification operations occur at the path computation element (PCE) selected by ingress element of an end-to-end service. In order to be able to compute and qualify paths, the PCE must be provided with information regarding the traffic engineering capabilities of the layer network to which it is associated with, in particular what the topology of the layer network is and what layer-specific transport capabilities exist at the various nodes and links in that topology.

It is important to note that topology information is layer-specific; e.g. path computation and qualification operations occur within a given layer, and hence information about topology and resource availability are required for the specific layer to which the connection belongs to. The topology and resource availability information required by elements in the client-layer is quite distinct from that required by the elements in the server-layer network. Hence, the server-layer traffic engineering links are of no importance for the client-layer network, and it is actually desirable to block their advertisements into the client TE domain by the server-layer border nodes.

In the sample hybrid network (Fig 1) there are multiple optical transport elements supporting the connection between the GMPLS-enabled IP routers, and hence the physical topology between them includes several nodes and links. However, the optical elements between the IP routers are not able to switch traffic within the client-layer network of routers (e.g. IP/MPLS), as the optical elements are lambda switches, not IP/MPLS switches. Hence while the intervening optical elements may physically exist along the path, they are not a part of the topology available to the IP/MPLS routers for the purposes of traffic engineering in the client-layer network.

An example of what the client-layer Traffic Engineering topology would look like for the sample hybrid network is shown in the top half of Fig 3.

Fig 3: Traffic Engineering - ERO with "loose hop"
(See PDF version)

In this example, the TE topology associated with the client-layer network is indicated by nodes [A, B, C, D, E, F, I, J] and links [A-F, E-B, J-C, I-D], whereas the TE topology associated with the server-layer network is indicated by nodes [E, F, G, H, I, J] and links [E-G, F-G, F-H, G-H, G-J, H-I, H-J, I-J]. The border nodes [E, F, I, J] at the core are visible in both the topologies.

In this example, if the "B" router attempts to determine a path to the "D" router it will be unable to do so, as the client-layer topology to which the B and D routers is connected does not include a full client-layer path between them. The only way to setup an end-to-end path in this case is to use an ERO with a "loose hop" across the server-layer domain as illustrated in Fig 3. This would cause the server-layer to create the necessary segment of the client-layer topology on the fly. However, this approach has a few drawbacks - [a] the necessity for the operator to specify the ERO with the "loose" hop; [b] potential sub-optimal usage of server-layer network resources; and [c] unpredictability with regard to the fate-sharing of the new segment (that is created on the fly) with other links of the client-layer topology.

In order to be able to compute a full path between the two client-layer endpoints, the client-layer topology must be sufficiently augmented to indicate where there are paths through the server-layer topology which can provide transport to services in the client-layer

topology. In other words, in order for a client to compute path(s) across the server-layer domain to other clients, the segment of the client-layer topology over the server-layer domain should be pre-planned and made available (in terms of TE links and nodes that exist in the client-layer) to all the clients. This is discussed in detail in the next section.

3.1. Augmenting the Client-Layer Topology

In the example hybrid network, consider a scenario where each GMPLS-enabled IP router is connected to the optical WDM transport network via a transponder. Further consider the situation where the transponder at node F can be connected to the transponder in node J via the optical pathway F-G-H-J. A WDM connection can be provisioned in the server-layer along this path, and then advertised as a TE link in the client-layer. With the availability of this link, the path computation function at node A is able to compute an end-to-end path from A to C.

Fig 4: Traffic Engineering - End to End Path Computation
(See PDF version)

In this case, in order for the TE link to be made available in the client-layer topology, the network resources corresponding to the underlying server-layer connection must be fully provisioned beforehand.

As another example, consider a network configuration where the transponders at nodes E, F, J and I are connected to each other via directionless ROADM components. It is physically possible to connect any transponder to any other transponder in the network. As there are transport capabilities available in the server-layer network between every element containing an adaptation function to the client-layer network, the operator in this case would not wish to reserve any network resources in the server-layer until the setup of the client-layer connection is initiated. The next section proposes a method to cater to this particular operational requirement.

3.1.1. Virtual TE Links

A "Virtual TE Link" is defined as a TE link that is advertised into the client-layer, with available but unreserved resources in the server-layer necessary to bring up the connection that supports that TE link. In other words, "Virtual TE Links" represent specific transport capabilities available in the server-layer network which can support services in the client-layer network. The two fundamental properties of a Virtual TE Link are: [a] It is advertised just like a real TE link and thus contributes to the buildup of the client-layer topology (and thus client-layer elements see no difference between virtual and real links); and [b] It does not require allocation of resources at the server-layer until used, thus allowing the sharing of server-layer resources with other virtual TE links.

Fig 5: Traffic Engineering - End to End Path Computation (w/ "Virtual TE Links"
(See PDF version)

In the example shown in Fig 5, the availability of a lambda channel along the path F-G-H-J results in there being a virtual traffic engineering link between F and J within the client-layer topology. With the availability of this link, the path computation function at node A is able to compute an end-to-end path from A to C.

When a virtual TE link gets selected and signaled in the ERO of the client-layer connection, it ceases to be "Virtual" and transforms into a regular TE-link. When this transformation takes place, the clients will notice the change in the advertised available bandwidth of this TE-link. Also, all other Virtual TE links that share resources with the TE-link in question start advertising "zero" available bandwidth. Likewise, the TE network image reverts back to the original form as soon as the last client-layer connection, going through the TE link in question, is released.

3.2. Macro SRLGs

The TE links that are added to the client-layer topology cannot be assumed to be totally independent. It is quite possible for a given TE link to share the same fate with one or more other TE link(s). This is because the underlying server-layer connections (real or potential) can traverse the same server-layer link and/or node, and

failure of any such shared link/node would make all such connections inoperable (along with the client-layer links they serve). If diverse end-to-end client-layer connections are to be computed, the fate-sharing information of the TE links needs to be accounted for. The standard way of addressing this problem is to use SRLGs as a part of TE link advertisements.

A traditional SRLG represents a shared physical network resource upon which normal function of a link depends. Such SRLGs can also be referred to as physical SRLGs. Zero, one or more physical SRLGs could be identified and advertised for every TE link in a given layer network. However, there is a scalability issue with physical SRLGs in multi-layer environments. For example, if a WDM layer connection serves an IP layer link, every WDM link and node traversed by the connection must be considered as a separate SRLG. The number of SRLGs to be advertised to client (e.g. IP) layer per link would be directly proportional to the number of hops traversed by the underlying server-layer connection.

The notion of Macro SRLGs addresses this scaling problem. Macro SRLGs have the same protocol format as that of their physical counterpart and can be assigned automatically for each TE link that is advertised into the client-layer as a result of the existence of an underlying server-layer connection (instantiated or otherwise). A Macro SRLG represents a set of shared path segments that are traversed by two or more of the underlying server-layer connections. Each shared path segment can be viewed as a sequence of shared resources where each individual resource has a physical SRLG associated with it (example depicted in Fig 6). The actual procedure for deriving these Macro SRLGs is beyond the scope of this document.

Fig 6: Macro SRLGs
(See PDF version)

3.3. Switching Constraints

Certain types of optical network configurations necessitate the specification of connectivity constraints in the TE advertisements. If the switching constraints associated with the binding between the TE link served by the server domain and its associated access TE link do not get advertised, there is a risk of an invalid path being picked (Fig 7). This document recommends the use of the extensions specified in [GEN_CNSTR] to address this.

Fig 7: Switching Constraints
(See PDF version)

4. Connection Setup

Experience with control plane operations in multi-layer networks indicates there are benefits to coordinating certain signaling operations, in the following manner. Consider the scenario where the core is a WDM network comprising of ROADMs. The set-up time for a service at the optical layer can be fairly long, as it can involve time-consuming power-equalization procedures, amongst other layer-specific operations. This means that at very least, the setup timers for the client-layer service would need to be somehow coordinated with that of the server-layer service. To avoid this operationally awkward issue, a phased connection setup process as depicted in Fig 8 is proposed.

Fig 8: Phased Connection Setup
(See PDF version)

As long as the server-layer connection is not completely "UP" (for example: Fully Power Equalized), the nodes at the edge of the core would signal the client-layer PATH/RESV messages with the T (Testing) bit set in the ADMIN_STATUS. The T bit would be cleared in these messages only after the underlying server-layer connection is deemed fully operable.

5. Security Considerations

TBD

6. Normative References

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

[RFC4208] G. Swallow, J.Drake, H. Ishimatsu, and Y. Rekhter, "GMPLS UNI: RSVP-TE Support for the Overlay Model", RFC 4208, October 2005.

7. Acknowledgments

TBD

Authors' Addresses

Vishnu Pavan Beeram
ADVA Optical Networking

Email: vbeeram@advaoptical.com

Igor Bryskin
ADVA Optical Networking

Email: ibryskin@advaoptical.com

Wes Doonan
ADVA Optical Networking

Email: wdoonan@advaoptical.com

John Drake
Juniper Networks

Email: jdrake@juniper.net

Gert Grammel
Juniper Networks

Email: ggrammel@juniper.net

Manuel Paul
Deutsche Telekom

Email: Manuel.Paul@telekom.de

Ruediger Kunze
Deutsche Telekom

Email: Ruediger.Kunze@telekom.de

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: May 3, 2012

M. Chen
L. Zheng
Huawei Technologies Co., Ltd

October 31, 2011

Multi-Protocol Label Switching Transport Profile (MPLS-TP) Operator Identifier Object
draft-chen-ccamp-mpls-tp-oio-01.txt

Abstract

Two formats of operator identifier are specified in Multi-Protocol Label Switching Transport Profile (MPLS-TP) networks, to uniquely identify an operator. One is Global_ID, the other is ICC_Operator_ID. In MPLS-TP networks, operator identifier as part of the global identifier of an MPLS-TP LSP may be required to be carried in the Path/Resv message when setting up the LSP.

This document defines a new RSVP-TE Object: Operator Identifier object (OIO). It has two types, the Global_ID object and the ICC_Operator_ID object. Similar as Session and Sender Template object, OIO could be carried in the Path or Resv message to communicate the Operator ID that the two ends of an LSP in use. At the same time, it makes sure all the nodes that the LSP traverses to use the same format of Operator ID.

Requirements Language

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

Status of this Memo

This Internet-Draft will expire on May 3, 2012.

Table of Contents

1. Introduction

RFC6370 [RFC6370] specifies an initial set of identifiers to be used in the Multiprotocol Label Switching Transport Profile (MPLS-TP). The Global_ID is defined in RFC6370 [RFC6370] to uniquely identify an operator. [I.D.draft-ietf-mpls-tp-itu-t-identifiers] [I-D.ietf-mpls-tp-itu-t-identifiers] specifies the ICC_Operator_ID, an alternative way to uniquely identify an operator based on ITU-T conventions. We call both Global_ID and ICC_Operator_ID as Operator ID in this document. The Operator ID is used to provide a globally unique context for other MPLS-TP identifiers.

Resource ReserVation Protocol Traffic Engineering (RSVP-TE) signaling [RFC3209][RFC3471] does not exchange Operator ID when setup an LSP. It is because in the traditional MPLS network the use of Operator ID is not necessary. When the IP addresses is globally unique, a Traffic Engineering (TE) LSP could be uniquely identified by the source IP address, destination IP address, Tunnel ID and Label Switching Path (LSP) ID of the LSP, which are carried in the Session and Sender Template/Filter Spec object. But in MPLS-TP networks, the Node Identifier (Node_ID) is unique within the scope of the Operator ID. In situations where a Node_ID needs to be globally unique, this is accomplished by prefixing the identifier with the Operator ID. This means the Operator ID as part of the global identifier of an MPLS-TP LSP may be required to be carried in the Path/Resv message when setting up the LSP.

In addition, since two formats of Operator ID are defined, i.e.Global_ID and ICC_Operator_ID, two things need to be determined when setting up a LSP in terms of Operator ID. One is that which format should be used, the other is that how to make sure both ends of the LSP using the same format. The former could be achieved by the configuration of the operators. The later may need some signaling exchange.

This document defines a new RSVP-TE object: Operator Identifier object (OIO). It has two types, the Global_ID object and ICC_Operator_ID object. Similar as the Session and Sender Template/ Filter Spec object, OIO could be carried in the Path/Resv message to communicate the Operator ID that the two ends of an LSP in use. At the same time, it makes sure all the nodes that the LSP traverses to use the same format of Operator ID.

2. Operator Identifier Object

Two types of Operator Identifier object (OIO) are defined in this document for two formats of Operator ID respectively: Global_ID object and ICC_Operator_ID object. Both Global_ID and ICC_Operator_ID object are OPTIONAL. When setting up an MPLS-TP LSP, one and only one type of the OIO MUST be included in the Path/Resv message if the Operator ID is required.

2.1. Global_ID Object

The encoding of the Global_ID object including the common object header is as follows:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|            Length             | Class-Num (TBA)|   C-Type (1) |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                           Global_ID                           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 1. Global_ID Object

Class-Num: To be allocated by IANA.

C-Type: To be allocated by IANA (0x01 is recommended).

Global_ID: Global_ID of the sender node. As defined in RFC6370 [RFC6370], a Global_ID is an unsigned 32-bit value and MUST be derived from a 4-octet AS number assigned to the operator. Note that 2-octet AS numbers have been incorporated in the 4-octet by placing the 2-octet AS number in the low-order octets and setting the two high-order octets to zero.

2.2. ICC_Operator_ID Object

The encoding of the ICC_Operator_ID object including the common object header is as follows:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|            Length             | Class-Num (TBA)|   C-Type (2) |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                       ICC_Operator_ID                         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                       ICC_Operator_ID(contd.)                 |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 2. ICC_Operator_ID Object

Class-Num: To be allocated by IANA.

C-Type: To be allocated by IANA (0x02 is recommended).

ICC_Operator_ID: ICC_Operator_ID of the sender node. As defined in [I-D.ietf-mpls-tp-itu-t-identifiers], the ICC_Operator_ID is formed by Country Code (CC) and ICC(ITU-T Carrier Code) as CC::ICC. The ICC itself is a string of one to six characters, global uniqueness is assured by concatenating the ICC with a CC. The Country Code (alpha-2) is a string of two alphabetic characters represented with upper case letters (i.e., A-Z). When the length of a ICC_Operator_ID string is less than 8 octets, the higher-order unused octets of the ICC_Operator_ID field MUST be set to zero.

2.3. Procedures

When signaling an MPLS-TP LSP, if the Operator ID is needed, e.g. LSRs on the LSP belong to different operators, the Operator Identifier object MUST be carried in the Path message, either the Global_ID or ICC_Operator_ID C-Type is used based on its local configuration. The Global_ID or ICC_Operator_ID field is filled by the sender node’s Global_ID or ICC_Operator_ID. When receiving such a Path message with OIO object, the node MUST check it’s C-Type to see whether it agree to use that format of Operator ID. If the node agrees, it MUST use the same format of Operator ID. The same C-Type of OIO MUST be carried in the Resv message, and the Global_ID or ICC_Operator_ID field is filled by its own Global_ID or ICC_Operator_ID.

If the receiving node does not recognize the Operator Identifier object, it sends a PathErr message with the error code "Unknown object class" towards the sender. If the receiving node recognizes the Operator Identifier object but does not recognize or support the C-Type, it sends a PathErr message with the error code "Unknown object C-Type" towards the sender. If the receiving node, for any reason, is not able to use the same C-Type as sender, e.g. due to fault configuration or requested C-Type not enabled, it MUST send a PathErr message with the error code "Wrong Operator Identifier C-Type" to notify the sender (Error Code to be allocated by IANA), and the LSP is not set up.

3. IANA Considerations

This document request IANA to assign new Class-Num, C-Types as follows:

Class-Num	Class Name	Class Types or C-Types:	
TBD	Operator Identifier	0x01(TBD)	Global_ID
		0x02(TBD)	ICC_Operator_ID

This document also request IANA to assign new Error-Code as follows:

Error Code	Meaning
TBD	Wrong Operator Identifier C-Type

4. Security Considerations

When Operator Identifier Object is in use, a form of source-validation checking may be enabled to ensure that the Global_ID or ICC_Operator_ID originated from a legitimate source, especially in the inter-provider case.

5. Acknowledgements

6. References

6.1. Normative References

[I-D.ietf-mpls-tp-itu-t-identifiers]
Winter, R., Gray, E., Helvoort, H., and M. Betts, "MPLS-TP Identifiers Following ITU-T Conventions", draft-ietf-mpls-tp-itu-t-identifiers-01 (work in progress), October 2011.

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

[RFC6370] Bocci, M., Swallow, G., and E. Gray, "MPLS Transport Profile (MPLS-TP) Identifiers", RFC 6370, September 2011.

6.2. Informative 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.

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

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

Authors' Addresses

Mach(Guoyi) Chen
Huawei Technologies Co., Ltd
Q14 Huawei Campus, No. 156 Beiqing Road, Hai-dian District
Beijing 100095
China

Email: mach@huawei.com

Lianshu Zheng
Huawei Technologies Co., Ltd
Q14 Huawei Campus, No. 156 Beiqing Road, Hai-dian District
Beijing 100095
China

Email: vero.zheng@huawei.com

Network Working Group Abinder Dhillon
Iftekhar Hussain
Rajan Rao
Infinera
Internet Draft
Intended status: Standard Track October 31, 2011
Expires: April 2012

OSPFTE extension to support GMPLS for Flex Grid
draft-dhillon-ccamp-super-channel-ospfte-ext-01.txt

Abstract

This document specifies the extension to TELINK LSA of OSPF routing protocol [RFC4203] [3] in support of GMPLS [1] for flex-grid networks [2].

Status of this Memo

Copyright Notice

Table of Contents

1. Introduction

To enable scaling of existing transport systems to ultra high data rates of 1 Tbps and beyond, next generation systems providing super-channel switching capability are currently being developed. To allow efficient allocation of optical spectral bandwidth for such high bit rate systems, International Telecommunication Union Telecommunication Standardization Sector (ITU-T) is extending the G.694.1 grid standard (termed ''Fixed-Grid'') to include flexible grid (termed ''Flex-Grid'') support.

This document defines OSPF-TE extensions in support of flex-grid networks.

Figure-1 shows a network consisting of Network Elements(NEs) with super channel switching capability. User can create super channel connections using GMPLS through these NEs.

To support the routing function in GMPLS for flex-grid network, NE models each flex-grid link (C-band or C-band-extended) with new switching capability and provides optical bandwidth in terms of 12.5 GHz spectral slices. This information is flooded in OSPFTE. During path calculation time, NE selects only that path where all the telinks support super channel switching and have required set of 12.5 GHz slices available. NE then signals along that path to establish super channel connection. Once the connection is established then spectral slice availability is updated in each telink and flooded back in OSFPTE.

```
+-------+              +-------+              +-------+
|  SC   |              |  SC   |              |  SC   |
|Switch |.---Link ---> |Switch |<- Link----- ->|Switch |
|   A   |              |   B   |              |   C   |
+-------+              +-------+              +-------+

        |<-- TE-Link -->|       |<-- TE-Link -->|
```

Figure 1: TE-Links

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].

3. Interface Switching Capability Descriptor

The Interface Switching Capability Descriptor describes switching capability of a
n
interface [RFC 4203]. This document defines a new Switching Capability value for
 Flex
Grid [G.694.1] as follows:

```
Value                        Type
-----                        ----
102 (TBA by IANA)    Super-Channel-Switch-Capable(SCSC)
```

Switching Capability and Encoding values MUST be used as follows:

```
Switching Capability = SCSC
Encoding Type = Lambda [as defined in RFC3471]
```

The Interface Switching Capability Descriptor is a sub-TLV (of type 15) of the Link TLV. The length is the length of value field in Octets. The format of the value field is as shown below:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Switching Cap |   Encoding    |           Reserved            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                  Max LSP Bandwidth at priority 0              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                  Max LSP Bandwidth at priority 1              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                  Max LSP Bandwidth at priority 2              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                  Max LSP Bandwidth at priority 3              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                  Max LSP Bandwidth at priority 4              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                  Max LSP Bandwidth at priority 5              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                  Max LSP Bandwidth at priority 6              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                  Max LSP Bandwidth at priority 7              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|        Switching Capability-specific information              |
|                  (variable)                                   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Maximum LSP BW is really not used for super channel connection. All the required information is in terms of spectral slices which are distributed within the ISCD specific portion of the ISCD for flex grid telink.

3.1. Switch Capability Specific Information

The technology specific part of the ISCD can include a variable number of sub-TLV s. We propose to include following sub-TLVs under SCSI field:

```
 0                   1                   2                   3
```

```
     0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |               Type              |             Length          |
    +---------------------------------------------------------------+
    |Grid |  S.S  |  Res  |   Priority    |          Reserved       |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                                                               |
    |               BW sub-TLV(s) one per priority                  |
    |                                                               |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 3: SCSI Format for ISCD=SCSD

```
     0                   1                   2                   3
     0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |               Type              |             Length          |
    +---------------------------------------------------------------+
    |               N-Start           |        Slice Count          |
    +---------------------------------------------------------------+
    |      Min S.W per LSP            |    Max S.W per LSP          |
    +---------------------------------------------------------------+
    |                                                               |
    |           Bit-Map showing Available 12.5GHz Slots             |
    |                       (upto 48 bytes per priority)            |
    |                                                               |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 4: BW sub-tlv

Various attributes in ISCD specific information TLV and Bandwidth sub-tlv are a s following:

- . Grid = FlexGrid;
- . Slice Spacing (S.S) = 12.5GHz;
- . Priority bit map to show priorities supported
  - o Up to 8 priorities can be supported
- . N-Start -   - integer, specifies start of the grid;

```
          o Use center freq formula to determine start of spectrum
     . Slice Count
          o Total number of slices advertised for the link  (available +  consume
d)
     . Min Slot Width per LSP
          o This is an integer value.  nxS.S;  n > 0 ;
     . Max Slot Width per LSP
          o This is an integer value;   nxS.S; n < some integer value up to Slice
            Count
     . Available BW encoded as bit-map
          o Each bit represents availability of one slice of width identified by
            S.S field
          o Zero -                     - Available ; One - occupied
```

3.2. BW advertisement procedure

This section describes bandwidth advertisement of telink when ISCD is of type SCSD.

Key points are:

- o An Optical node capable of Super Channel Switching advertises slices of certain width available based on the frequency spectrum supported by the node(e.g. C band, extended C-band). For example extended C-band will advertize 384 slices.
- o The BW advertisement involves a bit-map where each bit corresponds to a single slice of width as identified by S.S field.
- o The slice position/numbering in the bit-map is identified based on N-start field. The N-start field is derived based on ITU center frequency formula.
- o The advertising node MUST also set Slice-Count field.
- o Minimum & Maximum slot width fields are included to allow for any restrictions on the link for carrying super channel LSPs.
- o The BW advertisement is priority based and up to 8 priority levels are allowed.
- o The node capable of supporting one or more priorities MUST set the priority field and include BW-sub TLV for each of the priority supported.

3.2.1. Example - BW advertisement W/O any service present

Figure 5 shows an example of BW sub-tlv for a telink which has no service established over it yet. Attributes of BW sub-tlv in the telink are:

- o N-start=-142 for extended C-band

- o Total number of slices available on the link = 384 (based on Slice spacing = 12.5GHz)
- o Min SW field shows min consumption of 4 Slices per LSP (=50GHz)
- o Max SW field shows up to 400GHz BW allowed per LSP (32x12.5GHz)
- o 48 bytes showing that all 384 slices are available.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|               Type=1          |            Length             |
+---------------------------------------------------------------+
|           N-Start=-142        |        Slice Count=384        |
+---------------------------------------------------------------+
|      Min S.W per LSP=4        |    Max S.W per LSP=32         |
+---------------------------------------------------------------+
|                                                               |
|           Bit-Map with a bit for each 12.5GHz Slot.           |
|            (384 bits -                              - each bit-set to value
zero to show it  |
|            is unused                                          |
|                                                               |
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 5: TELINK BW sub-tlv w/o any service present

3.2.2. Example - How to use advertized Bandwidth

Assume user wants to setup Super Channel LSP over a single FlexGrid link with BW requirement = 250GHz and transponder fully tunable.

- o The path computing node performs the following:
  - o Determine the number of slices required for the LSP (250/S.S = 20)

- o Look for contiguous spectrum availability on each link from BW adv (both dir)

  - o Look for 20 contiguous bits in the BW advertisement TLV

  - o If available select the link for LSP creation.

. Signal for LSP creation. Once LSP is created , update BW available via new advertisement using the same Bandwidth sub-TLV.

4. Security Considerations

<Add any security considerations>

5. IANA Considerations

IANA needs to assign a new Grid field value to represent ITU-T Flex-Grid.

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.

6.2. Informative References

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

[2] Iftekhar H, Abinder, Zhong, Marco, Bert, Steve, Andrew, ’’Generalized Label for Super-Channel Assignment on Flexible Grid’’, draft-hussain-ccamp-super-channel-label-02.txt, October 2011.

[3] K. Kompella, Y., " OSPF Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4203, Oct 2005

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

[5] M. Jinno et. al., ''Spectrum-Efficient and Scalable Elastic Optical Path Network: Architecture, Benefits and Enabling Technologies'', IEEE Comm. Mag., Nov. 2009, pp. 66-73.

[6] S. Chandrasekhar and X. Liu, ''Terabit Super-Channels for High Spectral Efficiency Transmission '',in Proc. ECOC 2010, paper Tu.3.C.5, Torino (Italy), September 2010.

[7] ITU-T Recommendation G.694.1, "Spectral grids for WDM applications: DWDM frequency grid", June 2002

[8] A. Farrel, D King, ''Generalized Labels for the Flexi-Grid inLambda-Switch-Capable (LSC) Label Switching Routers'', Work in progress:draft-farrkingel-ccamp-flexigrid-lambda-label-00.txt - - October 2011.

[9] G. Bernstein, Y. Lee, D. Li, W. Imajuku, " General Network Element Constraint Encoding for GMPLS Controlled Networks", work in progress: draft-ietf-ccamp-general-constraint-encode-05, May 2011

7. Acknowledgments

<Add any acknowledgements>

Authors' Addresses

Abinder Dhillon
Infinera
140 Caspian Ct., Sunnyvale, CA 94089

Email: adhillon@infinera.com

Iftekhar Hussain
Infinera
140 Caspian Ct., Sunnyvale, CA 94089

Email: ihussain@infinera.com

Rajan Rao
Infinera
140 Caspian Ct., Sunnyvale, CA 94089

Email: rrao@infinera.com

Contributor's Addresses

Marco Sosa
Infinera
140 Caspian Ct., Sunnyvale, CA 94089

Email: msosa@infinera.com

Biao Lu
Infinera
140 Caspian Ct., Sunnyvale, CA 94089

Email: blu@infinera.com

Subhendu Chattopadhyay
Infinera
140 Caspian Ct., Sunnyvale, CA 94089

Email: schattopadhyay@infinera.com

Harpreet Uppal
Infinera
140 Caspian Ct., Sunnyvale, CA 94089

Email: harpreet.uppal@infinera.com

Network Working Group D. King (Editor)
Internet Draft A. Farrel
Updates: 3471, 6205 (if approved) Old Dog Consulting
Category: Standards Track Y. Li (Editor)
Expires: 17 April 2012 F. Zhang
ZTE
R. Casellas
CTTC
17 October 2011

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

draft-farrkingel-ccamp-flexigrid-lambda-label-01.txt

Abstract

A new flexible wavelength grid ("flexi-grid") is being developed within the ITU-T Study Group 15 to allow selection and switching of individual lambdas chosen flexibly from a detailed, fine granularity grid of available wavelengths. This document updates the definition of GMPLS lambda labels to support the flexi-grid.

This document updates RFC 3471 and updates RFC 6205.

Status of this Memo

Copyright Notice

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].

[RFC3471] states that wavelength labels "only have significance between two neighbors" (Section 3.2.1.1); global wavelength semantics are not considered. [RFC6205] defines a standard lambda label format 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].

A flexible grid network selects its data channels as arbitrarily assigned spectral slices. 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" is being developed within the ITU-T Study Group 15 to allow selection and switching of individual lambdas chosen flexibly from a detailed, fine granularity grid of wavelengths with arbitrary spectral bandwidth [G.FLEXIGRID]. This document updates the definition of GMPLS lambda labels provided in [RFC6205] to support the flexi-grid.

This document will not be put forward for publication as an RFC before the ITU-T have completed technical development of [G.FLEXIGRID], and the encoding specified in this document will also be communicated to the ITU-T for comment before publication as an RFC.

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.FLEXIGRID] extends DWDM fixed grids as defined in [G.694.1] 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.

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 either 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.

```
Frequency (THz) = 193.1 THz + n * channel spacing (THz)
```

4. Flexi-Label Format and Values

4.1 Flexi-Label Encoding

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

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

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

[Editors’ note: We considered the possibility of having a 40 bit label with no reserved bits, or a 48 bit label with 8 reserved bits. This would be somewhat more efficient for objects that carry multiple labels encoded as a sequence. However, since most uses of the label are in objects where the label is padded to a 32 bit boundary, there seemed little benefit.]

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|Grid | C.S.  |    Identifier   |              n                |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|       m       |                    Reserved                   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 1 : The Flexi-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. While, for flexible grid situation, this field should be used to represent the channel spacing granularity/or 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.FLEXIGRID] as follows:

Slot Width (GHz) = 12.5 GHz * m

The practical range of values for m is from 1 to 8 for slot widths ranging from 12.5 GHz to 100 GHz. Wider slot widths may be considered with larger values of 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 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.

5. Manageability Considerations

This document introduces no new elements for management. That is, labels will continue to be used in the same way by the GMPLS protocols and lambda labels have the same fields as previously defined so any management tools defined to handle [RFC6205] labels will be able to handle labels as specified in this document. However, it is obvious that the management tools will have to interpret the new values and meanings of label fields as defined in this document, and management tools that may have been (unwisely) coded to expect all lambda labels to be 32 bits, will need to be updated to handle these 64 bit labels.

6. Security Considerations

[RFC6205] notes that the definition of a new label encoding does not introduces 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].

7. IANA Considerations

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

7.1. 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]

7.2. 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]

8. Acknowledgments

Very many thanks to Lou Berger for discussions of labels of more than 32 bits.

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

9. References

9.1. Normative References

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

[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, March 2011.

[G.FLEXIGRID] Preliminary Draft revised G.694.1. Unpublished ITU-T Study Group 15, Question 6.

9.2. Informative References

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

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

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

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

Appendix A. Flexi-Grid Example

Considering the network displayed in Figure 2 (reproduced from [RFC6205]) it is possible to show an example of LSP setup using the lambda labels. The figure shows Reconfigurable Optical Add/Drop Multiplexers (ROADMs) and Wavelength Cross-Connects (WXCs) that operate at the wavelength switching level as well as PXCs

```
                                  |
      Domain A (or Vendor A)      |      Domain B (or Vendor B)
                                  |
     Node-1            Node-2     |         Node-6            Node-7
   +--------+        +--------+   |      +-------+ +-+     +-+ +-------+
   | ROADM  |        | ROADM  +---|------+  PXC  +-+D|     |D+-+  PXC  |
   | or WXC +========+ or WXC +---|------+       +-+W+=====+W+-+       |
   | (LSC)  |        | (LSC)  +---|------+ (LSC) +-+D|     |D+-+ (LSC) |
   +--------+        +--------+   |      |       +-|M|     |M+-+       |
       ||                ||       |      +++++++++ +-+     +-+ +++++++++
       ||     Node-3     ||       |       |||||||              |||||||
       ||   +--------+   ||       |      +++++++++           +++++++++
       ||===|  WXC   +===||       |      | DWDM  |           | DWDM  |
            | (LSC)  |            |      +--++---+           +--++---+
       ||===+        +===||       |         ||                  ||
       ||   +--------+   ||       |      +--++---+           +--++---+
       ||                ||       |      | DWDM  |           | DWDM  |
   +--------+        +--------+   |      +++++++++           +++++++++
   | ROADM  |        | ROADM  |   |       |||||||              |||||||
   | or WXC +========+ or WXC +=+ |  +-+ +++++++++ +-+     +-+ +++++++++
   | (LSC)  |        | (LSC)  | | |  |D|-|  PXC  +-+D|     |D+-+  PXC  |
   +--------+        +--------+ +=|==+W|-|       +-+W+=====+W+-+       |
     Node-4            Node-5     |  |D|-| (LSC) +-+D|     |D+-+ (LSC) |
                                  |  |M|-|       +-+M|     |M+-+       |
                                  |  +-+ +-------+ +-+     +-+ +-------+
                                  |        Node-8              Node-9
```

Figure 2 : Example Network

Node 1 receives the request for establishing an LSP from itself to Node 9. The ITU-T grid to be used is the Flexi-Grid, the channel spacing is 6.25 GHz, and the wavelength to be used is 193.05 THz. The cslot width to be used is 50 GHz. Node 1 signals the LSP using a Path message including a wavelength label structured as follows:

```
    0                   1                   2                   3
    0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |Grid |  C.S. |   Identifier    |              n                |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |       m       |                  Reserved                     |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Where:

Grid = 3 : ITU-T Flexi-Grid

C.S. = 5 : 6.25 GHz channel spacing

Identifier = local value indicating the laser in use

n = 24 :

Frequency (THz) = 193.1 THz + n * channel spacing (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 (Editor)
Old Dog Consulting
EMail: daniel@olddog.co.uk

Yao Li (Editor)
ZTE
EMail: li.yao3@zte.com.cn

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

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

Internet Draft
Updates: 2205, 3209, 3473
Category: Standards Track
Expiration Date: April 28, 2012

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

October 28, 2011

RSVP Association Object Extensions

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

Abstract

The RSVP ASSOCIATION object was defined in the context of GMPLS (Generalized Multi-Protocol Label Switching) controlled label switched paths (LSPs). In this context, the object is used to associate recovery LSPs with the LSP they are protecting. This object also has broader applicability as a mechanism to associate RSVP state, and this document defines how the ASSOCIATION object can be more generally applied. This document also defines extended ASSOCIATION objects which, in particular, can be used in the context of Transport Profile of Multiprotocol Label Switching (MPLS-TP). This document updates RFC 2205, RFC 3209, and RFC 3473.

Status of this Memo

Table of Contents

1. Introduction

End-to-end and segment recovery are defined for GMPLS (Generalized Multi-Protocol Label Switching) controlled label switched paths (LSPs) in [RFC4872] and [RFC4873] respectively. Both definitions use the ASSOCIATION object to associate recovery LSPs with the LSP they are protecting. Additional narrative on how such associations are to be identified is also provided in [ASSOC-INFO].

This document expands the possible usage of the ASSOCIATION object to non-GMPLS recovery contexts. This document reviews how association should be made in the case where the object is carried in a Path message and defines usage with Resv messages. This section also discusses usage of the ASSOCIATION object outside the context of GMPLS LSPs.

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

- o Voice Call-Waiting:
  A bidirectional voice call between two endpoints A and B is signaled using two separate unidirectional RSVP reservations for the flows A->B and B->A. If endpoint A wishes to put the A-B call on hold and join a separate A-C call, it is desirable that network resources on common links be shared between the A-B and A-C calls. The B->A and C->A subflows of the call can share resources using existing RSVP sharing mechanisms, but only if they use the same destination IP addresses and ports. However, there is no way in RSVP today to share the resources between the A->B and A->C subflows of the call since by definition the RSVP reservations for these subflows must have different IP addresses in the SESSION objects.

- o Voice Shared Line:
  A single number that rings multiple endpoints (which may be geographically diverse), such as phone lines on a manager's desk and their assistant. A VoIP system that models these calls as multiple P2P unicast pre-ring reservations would result in significantly over-counting bandwidth on shared links, since today unicast reservations to different endpoints cannot share bandwidth.

- o Symmetric NAT:
  RSVP permits sharing of resources between multiple flows addressed to the same destination D, even from different senders S1 and S2. However, if D is behind a NAT operating in symmetric mode [RFC5389], it is possible that the destination port of the flows S1->D and S2->D may be different outside the NAT. In this case, these flows cannot share resources using RSVP today, since the SESSION objects for these two flows outside the NAT would have different ports.

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

1.1. Conventions Used In This Document

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

2. Non-GMPLS Recovery Usage

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

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

2.1. Upstream Initiated Association

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

2.1.1. Path Message Format

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

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

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

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

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

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

2.1.2. Path Message Processing

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

A node that wishes to allow downstream nodes to associate Path state across RSVP sessions MUST include an ASSOCIATION object in the outgoing Path messages corresponding to the RSVP sessions to be associated. In the absence of Association Type-specific rules for identifying association, the included ASSOCIATION objects MUST be identical. When there is an Association Type-specific definition of association rules, the definition SHOULD allow for association based on identical ASSOCIATION objects. This document does not define any Association Type-specific rules. (See Section 3 for a discussion of an example of Association Type-specific rules which are derived from [RFC4872].)

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

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

Note that as more than one association may exist, the described matching MUST continue after a match is identified, and MUST be performed against all local Path state.

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

2.2. Downstream Initiated Association

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

2.2.1. Resv Message Format

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

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

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

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

2.2.2. Resv Message Processing

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

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

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

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

Note that as more than one association may exist, the described matching MUST continue after a match is identified, and MUST be performed against all local Resv state.

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

2.3. Association Types

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

2.3.1. Resource Sharing Association Type

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

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

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

3. IPv4 and IPv6 Extended ASSOCIATION Objects

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

An additional case that cannot be supported using the existing ASSOCIATION objects is presented by MPLS-TP LSPs. Per [RFC6370], MPLS-TP LSPs can be identified based on an operator unique global identifier. The [RFC6370] defined "global identifier", or Global_ID, is based on [RFC5003] and includes the operator's Autonomous System Number (ASN).

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

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

3.1. IPv4 and IPv6 Extended ASSOCIATION Object Format

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

```
     0                   1                   2                   3
     0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |             Length            | Class-Num(199)|  C-Type (TBA) |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |       Association Type        |       Association ID          |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                    IPv4 Association Source                    |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                   Global Association Source                   |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    :                               .                               :
    :                    Extended Association ID                    :
    :                               .                               :
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

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

```
     0                   1                   2                   3
     0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |             Length            | Class-Num(199)|  C-Type (TBA) |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |       Association Type        |       Association ID          |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                                                               |
    |                    IPv6 Association Source                    |
    |                                                               |
    |                                                               |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                   Global Association Source                   |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    :                               .                               :
    :                    Extended Association ID                    :
    :                               .                               :
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Association Type: 16 bits

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

Association ID: 16 bits

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

Association Source: 4 or 16 bytes

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

Global Association Source: 4 bytes

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

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

Please note that, as stated in [RFC6370], the use of the provider's ASN as a global identifier DOES NOT have anything at all to do with the use of the ASN in protocols such as BGP.

This field is based on the definition of Global_ID defined in [RFC5003] and used by [RFC6370].

Extended Association ID: variable, 4-byte aligned

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

3.2. Processing

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

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

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

- o The originator MUST set the Association Type, Association ID and Association Source fields as described in Section 4.

- o When ASN-based global identification of the Association Source is desired, the originator MUST set the Global Association Source field. When ASN-based global identification is not desired, the originator MUST set the Global Association Source field to zero (0).

- o The Extended ASSOCIATION object originator MAY include the Extended Association ID field. The field is included based on local policy. The field MUST be included when the Association ID field is insufficient to uniquely identify association within the scope of the source of the association. When included, this field MUST be set to a value that, when taken together with the other fields in the object, uniquely identifies the sessions to be associated.

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

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

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

4. Security Considerations

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

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

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

5. IANA Considerations

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

5.1. IPv4 and IPv6 Extended ASSOCIATION Objects

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

```
199  ASSOCIATION                          [RFC4872]

     Class Types or C-Types

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

5.2. Resource Sharing Association Type

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

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

There are no other IANA considerations introduced by this document.

6. Acknowledgments

Valuable comments and input was received from Dimitri Papadimitriou. We thank Subha Dhesikan for her contribution to the early work on sharing of resources across RSVP reservations.

7. References

7.1. Normative References

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

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

[RFC4872] Lang, J., Rekhter, Y., and Papadimitriou, D., "RSVP-TE Extensions in Support of End-to-End Generalized Multi-Protocol Label Switching (GMPLS) Recovery", RFC 4872, May 2007.

[RFC4873] Berger, L., Bryskin, I., Papadimitriou, D., Farrel, A., "GMPLS Segment Recovery", RFC 4873, May 2007.

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

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

[RFC5511] Farrel, A., "Routing Backus-Naur Form (RBNF): A Syntax Used to Form Encoding Rules in Various Routing Protocol Specifications", RFC 5511, April 2009

7.2. Informative References

[ASSOC-INFO] Berger, L., Faucheur, F., Narayanan, A., "Usage of The RSVP Association Object", work in progress, draft-ietf-ccamp-assoc-info.

[RFC2207] Berger., L., O'Malley., T., "RSVP Extensions for IPSEC RSVP Extensions for IPSEC Data Flows", RFC 2207, September 1997.

[RFC3175] Baker, F., Iturralde, C., Le, F., Davie, B., "Aggregation of RSVP for IPv4 and IPv6 Reservations", RFC 3175, September 2001.

[RFC4860] Le, F., Davie, B., Bose, P., Christou, C., Davenport, M., "Generic Aggregate Resource ReSerVation Protocol (RSVP) Reservations", RFC 4860, May 2007.

[RFC5003] Metz, C., Martini, L., Balus, F., Sugimoto, J., "Attachment Individual Identifier (AII) Types for Aggregation", RFC 5003, September 2007.

[RFC5389] Rosenberg, J., Mahy, R., Matthews, P., Wing, D., "Session Traversal Utilities for NAT (STUN)", RFC 5389, October 2008.

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

[RFC6370] Bocci, M., Swallow, G., Gray, E., "MPLS-TP Identifiers", RFC 6370, June 2011.

8. Authors’ Addresses

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

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

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

Generated on: Fri, Oct 28, 2011 8:00:39 AM

Fatai Zhang, Ed.
Dan Li
Huawei
Han Li
CMCC
S.Belotti
Alcatel-Lucent
D. Ceccarelli
Ericsson

Expires: March 9, 2012
September 9, 2011

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

draft-ietf-ccamp-gmpls-g709-framework-05.txt

Status of this Memo

Abstract

This document provides a framework to allow the development of protocol extensions to support Generalized Multi-Protocol Label Switching (GMPLS) and Path Computation Element (PCE) control of

Optical Transport Networks (OTN) as specified in ITU-T Recommendation G.709 as consented in October 2009.

Table of Contents

1. Introduction

OTN has become a mainstream layer 1 technology for the transport network. Operators want to introduce control plane capabilities based on Generalized Multi-Protocol Label Switching (GMPLS) to OTN networks, to realize the benefits associated with a high-function control plane (e.g., improved network resiliency, resource usage efficiency, etc.).

GMPLS extends MPLS to encompass time division multiplexing (TDM) networks (e.g., SONET/SDH, PDH, and G.709 sub-lambda), lambda switching optical networks, and spatial switching (e.g., incoming port or fiber to outgoing port or fiber). The GMPLS architecture is provided in [RFC3945], signaling function and Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) extensions are described in

[RFC3471] and [RFC3473], routing and OSPF extensions are described in [RFC4202] and [RFC4203], and the Link Management Protocol (LMP) is described in [RFC4204].

The GMPLS protocol suite including provision [RFC4328] provides the mechanisms for basic GMPLS control of OTN networks based on the 2001 revision of the G.709 specification [G709-V1]. Later revisions of the G.709 specification, including [G709-V3], have included some new features; for example, various multiplexing structures, two types of TSs (i.e., 1.25Gbps and 2.5Gbps), and extension of the Optical Data Unit (ODU) ODUj definition to include the ODUflex function.

This document reviews relevant aspects of OTN technology evolution that affect the GMPLS control plane protocols and examines why and how to update the mechanisms described in [RFC4328]. This document additionally provides a framework for the GMPLS control of OTN networks and includes a discussion of the implication for the use of the Path Computation Element (PCE) [RFC4655].

For the purposes of the control plane the OTN can be considered as being comprised of ODU and wavelength (OCh) layers. This document focuses on the control of the ODU layer, with control of the wavelength layer considered out of the scope. Please refer to [RFC6163] for further information about the wavelength layer.

2. Terminology

OTN: Optical Transport Network

ODU: Optical Channel Data Unit

OTU: Optical channel transport unit

OMS: Optical multiplex section

MSI: Multiplex Structure Identifier

TPN: Tributary Port Number

LO ODU: Lower Order ODU. The LO ODUj (j can be 0, 1, 2, 2e, 3, 4, flex.) represents the container transporting a client of the OTN that is either directly mapped into an OTUk (k = j) or multiplexed into a server HO ODUk (k > j) container.

HO ODU: Higher Order ODU. The HO ODUk (k can be 1, 2, 2e, 3, 4.) represents the entity transporting a multiplex of LO ODUj tributary signals in its OPUk area.

ODUflex: Flexible ODU. A flexible ODUk can have any bit rate and a bit rate tolerance up to +/-100 ppm.

3. G.709 Optical Transport Network (OTN)

This section provides an informative overview of those aspects of the OTN impacting control plane protocols. This overview is based on the ITU-T Recommendations that contain the normative definition of the OTN. Technical details regarding OTN architecture and interfaces are provided in the relevant ITU-T Recommendations.

Specifically, [G872-2001] and [G872Am2] describe the functional architecture of optical transport networks providing optical signal transmission, multiplexing, routing, supervision, performance assessment, and network survivability. [G709-V1] defines the interfaces of the optical transport network to be used within and between subnetworks of the optical network. With the evolution and deployment of OTN technology many new features have been specified in ITU-T recommendations, including for example, new ODU0, ODU2e, ODU4 and ODUflex containers as described in [G709-V3].

3.1. OTN Layer Network

The simplified signal hierarchy of OTN is shown in Figure 1, which illustrates the layers that are of interest to the control plane. Other layers below OCh (e.g. Optical Transmission Section - OTS) are not included in this Figure. The full signal hierarchy is provided in [G709-V3].

```
 Client signal
       |
      ODUj
       |
    OTU/OCh
      OMS
```

Figure 1 - Basic OTN signal hierarchy

Client signals are mapped into ODUj containers. These ODUj containers are multiplexed onto the OTU/OCh. The individual OTU/OCh signals are combined in the Optical Multiplex Section (OMS) using WDM multiplexing, and this aggregated signal provides the link between the nodes.

3.1.1. Client signal mapping

The client signals are mapped into a Low Order (LO) ODUj. Appendix A gives more information about LO ODU.

The current values of j defined in [G709-V3] are: 0, 1, 2, 2e, 3, 4, Flex. The approximate bit rates of these signals are defined in [G709-V3] and are reproduced in Tables 1 and 2.

ODU Type	ODU nominal bit rate
ODU0	1 244 160 kbits/s
ODU1	239/238 x 2 488 320 kbit/s
ODU2	239/237 x 9 953 280 kbit/s
ODU3	239/236 x 39 813 120 kbit/s
ODU4	239/227 x 99 532 800 kbit/s
ODU2e	239/237 x 10 312 500 kbit/s
ODUflex for CBR Client signals	239/238 x client signal bit rate
ODUflex for GFP-F Mapped client signal	Configured bit rate

Table 1 - ODU types and bit rates

NOTE - The nominal ODUk rates are approximately: 2 498 775.126 kbit/s (ODU1), 10 037 273.924 kbit/s (ODU2), 40 319 218.983 kbit/s (ODU3), 104 794 445.815 kbit/s (ODU4) and 10 399 525.316 kbit/s (ODU2e).

ODU Type	ODU bit-rate tolerance
ODU0	+- 20 ppm
ODU1	+- 20 ppm
ODU2	+- 20 ppm
ODU3	+- 20 ppm
ODU4	+- 20 ppm
ODU2e	+- 100 ppm
ODUflex for CBR Client signals	+- 100 ppm
ODUflex for GFP-F Mapped client signal	+- 100 ppm

Table 2 - ODU types and tolerance

One of two options is for mapping client signals into ODUflex depending on the client signal type:

- Circuit clients are proportionally wrapped. Thus the bit rate and tolerance are defined by the client signal.

- Packet clients are mapped using the Generic Framing Procedure (GFP). [G709-V3] recommends that the bit rate should be set to an integer multiplier of the High Order (HO) Optical Channel Physical Unit (OPU) OPUk TS rate, the tolerance should be +/-100ppm, and the bit rate should be determined by the node that performs the mapping.

[Editors’ Note: As outcome of ITU SG15/q11 expert meeting held in Vimercate in September 2010 it was decided that a resizable ODUflex(GFP) occupies the same number of TS on every link of the path (independently of the High Order (HO) OPUk TS rate). Please see WD07 and the meeting report of this meeting for more information.

The authors will update the above text related to Packet client mapping as soon as new version of G.709 will be updated accordingly with expert meeting decision reported here.]

3.1.2. Multiplexing ODUj onto Links

The links between the switching nodes are provided by one or more wavelengths. Each wavelength carries one OCh, which carries one OTU, which carries one ODU. Since all of these signals have a 1:1:1 relationship, we only refer to the OTU for clarity. The ODUjs are mapped into the TS of the OPUk. Note that in the case where j=k the ODUj is mapped into the OTU/OCh without multiplexing.

The initial versions of G.709 [G709-V1] only provided a single TS granularity, nominally 2.5Gb/s. [G709-V3], approved in 2009, added an additional TS granularity, nominally 1.25Gb/s. The number and type of TSs provided by each of the currently identified OTUk is provided below:

	2.5Gb/s	1.25Gb/s	Nominal Bit rate
OTU1	1	2	2.5Gb/s
OTU2	4	8	10Gb/s
OTU3	16	32	40Gb/s
OTU4	--	80	100Gb/s

To maintain backwards compatibility while providing the ability to interconnect nodes that support 1.25Gb/s TS at one end of a link and 2.5Gb/s TS at the other, the 'new' equipment will fall back to the use of a 2.5Gb/s TS if connected to legacy equipment. This information is carried in band by the payload type.

The actual bit rate of the TS in an OTUk depends on the value of k. Thus the number of TS occupied by an ODUj may vary depending on the values of j and k. For example an ODU2e uses 9 TS in an OTU3 but only 8 in an OTU4. Examples of the number of TS used for various cases are provided below:

- ODU0 into ODU1, ODU2, ODU3 or ODU4 multiplexing with 1,25Gbps TS granularity
  - o ODU0 occupies 1 of the 2, 8, 32 or 80 TS for ODU1, ODU2, ODU3 or ODU4
- ODU1 into ODU2, ODU3 or ODU4 multiplexing with 1,25Gbps TS granularity
  - o ODU1 occupies 2 of the 8, 32 or 80 TS for ODU2, ODU3 or ODU4

- ODU1 into ODU2, ODU3 multiplexing with 2.5Gbps TS granularity
  - o ODU1 occupies 1 of the 4 or 16 TS for ODU2 or ODU3

- ODU2 into ODU3 or ODU4 multiplexing with 1.25Gbps TS granularity
  - o ODU2 occupies 8 of the 32 or 80 TS for ODU3 or ODU4

- ODU2 into ODU3 multiplexing with 2.5Gbps TS granularity
  - o ODU2 occupies 4 of the 16 TS for ODU3

- ODU3 into ODU4 multiplexing with 1.25Gbps TS granularity
  - o ODU3 occupies 31 of the 80 TS for ODU4

- ODUflex into ODU2, ODU3 or ODU4 multiplexing with 1.25Gbps TS granularity
  - o ODUflex occupies n of the 8, 32 or 80 TS for ODU2, ODU3 or ODU4 (n <= Total TS numbers of ODUk)

- ODU2e into ODU3 or ODU4 multiplexing with 1.25Gbps TS granularity
  - o ODU2e occupies 9 of the 32 TS for ODU3 or 8 of the 80 TS for ODU4

In general the mapping of an ODUj (including ODUflex) into the OTUk TSs is determined locally, and it can also be explicitly controlled by a specific entity (e.g., head end, NMS) through Explicit Label Control [RFC3473].

3.1.2.1. Structure of MSI information

When multiplexing an ODUj into a HO ODUk (k>j), G.709 specifies the information that has to be transported in-band in order to allow for correct demultiplexing. This information, known as Multiplex Structure Information (MSI), is transported in the OPUk overhead and is local to each link. In case of bidirectional paths the association between TPN and TS MUST be the same in both directions.

The MSI information is organized as a set of entries, with one entry for each HO ODUj TS. The information carried by each entry is:

Payload Type: the type of the transported payload.

Tributary Port Number (TPN): the port number of the ODUj transported by the HO ODUk. The TPN is the same for all the TSs assigned to the transport of the same ODUj instance.

For example, an ODU2 carried by a HO ODU3 is described by 4 entries in the OPU3 overhead when the TS size is 2.5 Gbit/s, and by 8 entries when the TS size is 1.25 Gbit/s.

On each node and on every link, two MSI values have to be provisioned:

The TxMSI information inserted in OPU (e.g., OPU3) overhead by the source of the HO ODUk trail.
The expectedMSI information that is used to check the acceptedMSI information. The acceptedMSI information is the MSI valued received in-band, after a 3 frames integration.

The sink of the HO ODU trail checks the complete content of the acceptedMSI information (against the expectedMSI.
If the acceptedMSI is different from the expectedMSI, then the traffic is dropped and a payload mismatch alarm is generated.

Provisioning of TPN can be performed either by network management system or control plane. In the last case, control plane is also responsible for negotiating the provisioned values on a link by link base.

4. Connection management in OTN

OTN-based connection management is concerned with controlling the connectivity of ODU paths and optical channels (OCh). This document focuses on the connection management of ODU paths. The management of OCh paths is described in [RFC6163].

While [G872-2001] considered the ODU as a set of layers in the same way as SDH has been modeled, recent ITU-T OTN architecture progress [G872-Am2] includes an agreement to model the ODU as a single layer network with the bit rate as a parameter of links and connections. This allows the links and nodes to be viewed in a single topology as a common set of resources that are available to provide ODUj connections independent of the value of j. Note that when the bit rate of ODUj is less than the server bit rate, ODUj connections are supported by HO-ODU (which has a one-to-one relationship with the OTU).

From an ITU-T perspective, the ODU connection topology is represented by that of the OTU link layer, which has the same topology as that of the OCh layer (independent of whether the OTU supports HO-ODU, where multiplexing is utilized, or LO-ODU in the case of direct mapping). Thus, the OTU and OCh layers should be visible in a single topological representation of the network, and from a logical perspective, the OTU and OCh may be considered as the same logical, switchable entity.

Note that the OTU link layer topology may be provided via various infrastructure alternatives, including point-to-point optical connections, flexible optical connections fully in the optical domain, flexible optical connections involving hybrid sub-lambda/lambda nodes involving 3R, etc.

The document will be updated to maintain consistency with G.872 progress when it is consented for publication.

4.1. Connection management of the ODU

LO ODUj can be either mapped into the OTUk signal (j = k), or multiplexed with other LO ODUjs into an OTUk (j < k), and the OTUk is mapped into an OCh. See Appendix A for more information.

From the perspective of control plane, there are two kinds of network topology to be considered.

(1) ODU layer

In this case, the ODU links are presented between adjacent OTN nodes, which is illustrated in Figure 2. In this layer there are ODU links with a variety of TSs available, and nodes that are ODXCs. Lo ODU connections can be setup based on the network topology.

```
                  Link #5       +--+---+--+        Link #4
     +--------------------------|         |---------------------------+
     |                          |  ODXC   |                           |
     |                          +---------+                           |
     |                             Node E                             |
     |                                                                |
   +-++---+--+        +--+---+--+        +--+---+--+        +--+---+-++
   |         |Link #1 |         |Link #2 |         |Link #3 |         |
   |         |--------|         |--------|         |--------|         |
   |  ODXC   |        |  ODXC   |        |  ODXC   |        |  ODXC   |
   +---------+        +---------+        +---------+        +---------+
      Node A             Node B              Node C             Node D
```

Figure 2 - Example Topology for LO ODU connection management

If an ODUj connection is requested between Node C and Node E routing/path computation must select a path that has the required number of TS available and that offers the lowest cost. Signaling is then invoked to set up the path and to provide the information (e.g., selected TS) required by each transit node to allow the configuration of the ODUj to OTUk mapping (j = k) or multiplexing (j < k), and demapping (j = k) or demultiplexing (j < k).

(2) ODU layer with OCh switching capability

In this case, the OTN nodes interconnect with wavelength switched node (e.g., ROADM,OXC) that are capable of OCh switching, which is illustrated in Figure 3 and Figure 4. There are ODU layer and OCh layer, so it is simply a MLN. OCh connections may be created on demand, which is described in section 5.1.

In this case, an operator may choose to allow the underlined OCh layer to be visible to the ODU routing/path computation process in which case the topology would be as shown in Figure 4. In Figure 3 below, instead, a cloud representing OCH capable switching nodes is represented. In Figure 3, the operator choice is to hide the real RWA network topology.

```
                                Node E
         Link #5             +---------+       Link #4
     +-----------------------|         |-----------------------+
     |                         ------                          |
     |                       //      \\                        |
     |                      ||        ||                       |
     |                      | RWA domain |                     |
   +-+-----+        +----+- ||        || ------+       +-----+-+
   |       |        |        \\      //        |       |       |
   |       |Link #1 |          --------        |Link #3 |      |
   |       +--------+          |       |       +--------+      +
   | ODXC  |        |  ODXC   +---------+ ODXC  |       | ODXC  |
   +-------+        +---------+Link #2 +--------+       +-------+
     Node A            Node B            Node C           Node D
```

Figure 3 - RWA Hidden Topology for LO ODU connection management

```
           Link #5            +---------+            Link #4
     +------------------------|         |-----------------------+
     |                   +----| ODXC    |----+                  |
     |                 +-++   +---------+   ++-+                |
     |         Node f  |  |     Node E      |  |  Node g        |
     |                 +-++                 ++-+                |
     |                   |       +--+        |                  |
   +-+-----+        +----+----+--|  |--+-----+---+        +-----+-+
   |       |Link #1 |         |  +--+  |         |Link #3 |       |
   |       +--------+         | Node h |         +--------+       +
   | ODXC  |        | ODXC    +--------+ ODXC    |        | ODXC  |
   +-------+        +---------+ Link #2+---------+        +-------+
     Node A            Node B            Node C             Node D
```

Figure 4 - RWA Visible Topology for LO ODUj connection management

In Figure 4, the cloud of previous figure is substitute by the real topology. The nodes f, g, h are nodes with OCH switching capability.

In the examples (i.e., Figure 3 and Figure 4), we have considered the case in which LO-ODUj connections are supported by OCh connection, and the case in which the supporting underlying connection can be also made by a combination of HO-ODU/OCh connections.

In this case, the ODU routing/path selection process will request an HO-ODU/OCh connection between node C and node E from the RWA domain. The connection will appear at ODU level as a Forwarding Adjacency, which will be used to create the ODU connection.

5. GMPLS/PCE Implications

The purpose of this section is to provide a set of requirements to be evaluated for extensions of the current GMPLS protocol suite and the PCE applications and protocols to encompass OTN enhancements and connection management.

5.1. Implications for LSP Hierarchy with GMPLS TE

The path computation for ODU connection request is based on the topology of ODU layer, including OCh layer visibility.

The OTN path computation can be divided into two layers. One layer is OCh/OTUk, the other is ODUj. [RFC4206] and [RFC6107] define the mechanisms to accomplish creating the hierarchy of LSPs. The LSP management of multiple layers in OTN can follow the procedures defined in [RFC4206], [RFC6107] and related MLN drafts.

As discussed in section 4, the route path computation for OCh is in the scope of WSON [RFC6163]. Therefore, this document only considers ODU layer for ODU connection request.

LSP hierarchy can also be applied within the ODU layers. One of the typical scenarios for ODU layer hierarchy is to maintain compatibility with introducing new [G709-V3] services (e.g., ODU0, ODUflex) into a legacy network configuration (containing [G709-V1] or [G709-V2] OTN equipment). In this scenario, it may be needed to consider introducing hierarchical multiplexing capability in specific network transition scenarios. One method for enabling multiplexing hierarchy is by introducing dedicated boards in a few specific places in the network and tunneling these new services through [G709-V1] or [G709-V2] containers (ODU1, ODU2, ODU3), thus postponing the need to upgrade every network element to [G709-V3] capabilities.

In such case, one ODUj connection can be nested into another ODUk (j<k) connection, which forms the LSP hierarchy in ODU layer. The creation of the outer ODUk connection can be triggered via network planning, or by the signaling of the inner ODUj connection. For the former case, the outer ODUk connection can be created in advance based on network planning. For the latter case, the multi-layer network signaling described in [RFC4206], [RFC6107] and [RFC6001] (including related modifications, if needed) are relevant to create the ODU connections with multiplexing hierarchy. In both cases, the outer ODUk connection is advertised as a Forwarding Adjacency (FA).

5.2. Implications for GMPLS Signaling

The signaling function and Resource reSerVation Protocol-Traffic Engineering (RSVP-TE) extensions are described in [RFC3471] and [RFC 3473]. For OTN-specific control, [RFC4328] defines signaling extensions to support G.709 Optical Transport Networks Control as defined in [G709-V1].

As described in Section 3, [G709-V3] introduced some new features that include the ODU0, ODU2e, ODU4 and ODUflex containers. The mechanisms defined in [RFC4328] do not support such new OTN features, and protocol extensions will be necessary to allow them to be controlled by a GMPLS control plane.

[RFC4328] defines the LSP Encoding Type, the Switching Type and the Generalized Protocol Identifier (Generalized-PID) constituting the common part of the Generalized Label Request. The G.709 Traffic Parameters are also defined in [RFC4328]. The following signaling aspects should be considered additionally since [RFC4328] was published:

- Support for backward compatibility with [RFC4328]

  A new Switching Capability type needs to be defined for control of [G709-V3] in the routing, so the Switching Type used when signalling of LSPs for [G709-V3] should be consistent with the Switching Type in the routing information.

  Assume [RFC4328] has been deployed to control the OTN networks supporting [G709-V1], control plane backward compatibility needs to be taken into consideration when interworking with legacy nodes only supporting [RFC4328] and [G709-V1].

- Support for specifying the new signal types and the related traffic information

  The traffic parameters should be extended in signaling message to support the new optical Channel Data Unit (ODUj) including:

  - ODU0
  - ODU2e
  - ODU4
  - ODUflex

  For ODUflex, since it has a variable bandwidth/bit rate BR and a bit rate tolerance T, the (node local) mapping process must be aware of the bit rate and tolerance of the ODUj being multiplexed in order to select the correct number of TS and the fixed/variable stuffing bytes. Therefore, bit rate and bit rate tolerance should also be carried in the Traffic Parameter in the signaling of connection setup request.

  For other ODU signal types, the bit rates and tolerances of them are fixed and can be deduced from the signal types.

- Support for LSP setup using different Tributary Slot granularity

  The signaling protocol should be able to identify the type of TS (i.e., the 2.5 Gbps TS granularity and the new 1.25 Gbps TS granularity) to be used for establishing an H-LSP which will be used to carry service LSP(s) requiring specific TS type.

- Support for LSP setup of new ODUk/ODUflex containers with related mapping and multiplexing capabilities

  New label should be defined to carry the exact TS allocation information related to the extended mapping and multiplexing hierarchy (For example, ODU0 into ODU2 multiplexing (with 1,25Gbps TS granularity)), in order to setting up the ODU connection.

- Support for Tributary Port Number allocation and negotiation

  Tributary Port Number needs to be configured as part of the MSI information (See more information in Section 3.1.2.1). A new extension object has to be defined to carry TPN information if control plane is used to configure MSI information.

- Support for ODU Virtual Concatenation (VCAT) and Link Capacity Adjustment Scheme (LCAS)

  GMPLS signaling should support the creation of Virtual Concatenation of ODUk signal with k=1, 2, 3. The signaling should also support the control of dynamic capacity changing of a VCAT container using LCAS ([G.7042]). [RFC6344] has a clear description of VCAT and LCAS control in SONET/SDH and OTN networks.

- Support for constraint signaling

  How an ODUk connection service is transported within an operator network is governed by operator policy. For example, the ODUk connection service might be transported over an ODUk path over an OTUk section, with the path and section being at the same rate as that of the connection service. In this case, an entire lambda of capacity is consumed in transporting the ODUk connection service. On the other hand, the operator might leverage sub-lambda multiplexing capabilities in the network to improve infrastructure efficiencies within any given networking domain. In this case, ODUk multiplexing may be performed prior to transport over various rate ODU servers over associated OTU sections.

  The identification of constraints and associated encoding in the signaling for differentiating full lambda LSP or sub lambda LSP is for further study.

- Support for Control of Hitless Adjustment of ODUflex (GFP)

  [G.7044] has been created in ITU-T to specify hitless adjustment of ODUflex (GFP) (HAO) that is used to increase or decrease the bandwidth of an ODUflex (GFP) that is transported in an OTN network.

The procedure of ODUflex (GFP) adjustment requires the participation of every node along the path. Therefore, it is recommended to use the control plane signaling to initiate the adjustment procedure in order to avoid the manual configuration at each node along the path.

Since the [G.7044] is being developed currently, the control of HAO is for further study.

All the extensions above should consider the extensibility to match future evolvement of OTN.

5.3. Implications for GMPLS Routing

The path computation process should select a suitable route for an ODUj connection request. In order to perform the path computation, it must evaluate the available bandwidth on each candidate link. The routing protocol should be extended to convey some information to represent ODU TE topology.

GMPLS Routing [RFC4202] defines Interface Switching Capability Descriptor of TDM which can be used for ODU. However, some issues discussed below, should also be considered.

Interface Switching Capability Descriptors present a new constraint for LSP path computation. [RFC4203] defines the switching capability and related Maximum LSP Bandwidth and the Switching Capability specific information. When the Switching Capability field is TDM the Switching Capability Specific Information field includes Minimum LSP Bandwidth, an indication whether the interface supports Standard or Arbitrary SONET/SDH, and padding. Hence a new Switching Capability value needs to be defined for [G709-V3] ODU switching in order to allow the definition of a new Switching Capability Specific Information field definition. The following requirements should be considered:

- Support for carrying the link multiplexing capability

  As discussed in section 3.1.2, many different types of ODUj can be multiplexed into the same OTUk. For example, both ODU0 and ODU1 may be multiplexed into ODU2. An OTU link may support one or more types of ODUj signals. The routing protocol should be capable of carrying this multiplexing capability.

- Support any ODU and ODUflex

  The bit rate (i.e., bandwidth) of TS is dependent on the TS granularity and the signal type of the link. For example, the bandwidth of a 1.25G TS in an OTU2 is about 1.249409620 Gbps, while the bandwidth of a 1.25G TS in an OTU3 is about 1.254703729 Gbps.

  One LO ODU may need different number of TSs when multiplexed into different HO ODUs. For example, for ODU2e, 9 TSs are needed when multiplexed into an ODU3, while only 8 TSs are needed when multiplexed into an ODU4. For ODUflex, the total number of TSs to be reserved in a HO ODU equals the maximum of [bandwidth of ODUflex / bandwidth of TS of the HO ODU].

  Therefore, the routing protocol must be capable of carrying the necessary and sufficient link bandwidth information for performing accurate route computation for any of the fixed rate ODUs as well as ODUflex.

- Support for differentiating between terminating and switching capability

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

  Therefore, to allow the required flexibility, the routing protocol should clearly distinguish the terminating and switching capability.

- Support for Tributary Slot Granularity advertisement

  [G709-V3] defines two types of TS but each link can only support a single type at a given time. In order to perform a correct path computation (i.e. the LSP end points have matching Tributary Slot Granularity values) the Tributary Slot Granularity needs to be advertised.

- Support different priorities for resource reservation

How many priorities levels should be supported depends on the operator’s policy. Therefore, the routing protocol should be capable of supporting either no priorities or up to 8 priority levels as defined in [RFC4202].

- Support link bundling

  Link bundling can improve routing scalability by reducing the amount of TE links that has to be handled by routing protocol. The routing protocol must be capable of supporting bundling multiple OTU links, at the same line rate and muxing hierarchy, between a pair of nodes as a TE link. Note that link bundling is optional and is implementation dependent.

- Support for Control of Hitless Adjustment of ODUflex (GFP)

  As described in Section 5.2, the routing requirements for supporting hitless adjustment of ODUflex (GFP) (G.7044) are for further study.

As mentioned in Section 5.1, one method of enabling multiplexing hierarchy is via usage of dedicated boards to allow tunneling of new services through legacy ODU1, ODU2, ODU3 containers. Such dedicated boards may have some constraints with respect to switching matrix access; detection and representation of such constraints is for further study.

5.4. Implications for Link Management Protocol (LMP)

As discussed in section 5.3, Path computation needs to know the interface switching capability of links. The switching capability of two ends of the link may be different, so the link capability of two ends should be correlated.

The Link Management Protocol (LMP) [RFC4204] provides a control plane protocol for exchanging and correlating link capabilities.

It is not necessary to use LMP to correlate link-end capabilities if the information is available from another source such as management configuration or automatic discovery/negotiation within the data plane.

Note that LO ODU type information can be, in principle, discovered by routing. Since in certain case, routing is not present (e.g. UNI case) we need to extend link management protocol capabilities to cover this aspect. In case of routing presence, the discovering procedure by LMP could also be optional.

- Correlating the granularity of the TS

  As discussed in section 3.1.2, the two ends of a link may support different TS granularity. In order to allow interconnection the node with 1.25Gb/s granularity must fall back to 2.5Gb/s granularity.

  Therefore, it is necessary for the two ends of a link to correlate the granularity of the TS. This ensures the correct use and of the TE link.

- Correlating the supported LO ODU signal types and multiplexing hierarchy capability

  Many new ODU signal types have been introduced in [G709-V3], such as ODU0, ODU4, ODU2e and ODUflex. It is possible that equipment does not support all the LO ODU signal types introduced by those new standards or drafts. Furthermore, since multiplexing hierarchy is not allowed before [G709-V3], it is possible that only one end of an ODU link can support multiplexing hierarchy capability, or the two ends of the link support different multiplexing hierarchy capabilities (e.g., one end of the link supports ODU0 into ODU1 into ODU3 multiplexing while the other end supports ODU0 into ODU2 into ODU3 multiplexing).

  For the control and management consideration, it is necessary for the two ends of an HO ODU link to correlate which types of LO ODU can be supported and what multiplexing hierarchy capabilities can be provided by the other end.

5.5. Implications for Path Computation Elements

[PCE-APS] describes the requirements for GMPLS applications of PCE in order to establish GMPLS LSP. PCE needs to consider the GMPLS TE attributes appropriately once a PCC or another PCE requests a path computation. The TE attributes which can be contained in the path calculation request message from the PCC or the PCE defined in [RFC5440] includes switching capability, encoding type, signal type, etc.

As described in section 5.2.1, new signal types and new signals with variable bandwidth information need to be carried in the extended signaling message of path setup. For the same consideration, PCECP also has a desire to be extended to carry the new signal type and related variable bandwidth information when a PCC requests a path computation.

6. Data Plane Backward Compatibility Considerations

The node supporting 1.25Gbps TS can interwork with the other nodes that supporting 2.5Gbps TS by combining Specific TSs together in data plane. The control plane MUST support this TS combination.

Take Figure 5 as an example. Assume that there is an ODU2 link between node A and B, where node A only supports the 2.5Gbps TS while node B supports the 1.25Gbps TS. In this case, the TS#i and TS#i+4 (where i<=4) of node B are combined together. When creating an ODU1 service in this ODU2 link, node B reserves the TS#i and TS#i+4 with the granularity of 1.25Gbps. But in the label sent from B to A, it is indicated that the TS#i with the granularity of 2.5Gbps is reserved.

```
                                Path
            +----------+   ------------>    +----------+
            |     TS1==|===========\--------+--TS1     |
            |     TS2==|=========\--\-------+--TS2     |
            |     TS3==|=======\--\--\------+--TS3     |
            |     TS4==|=====\--\--\--\-----+--TS4     |
            |          |      \  \  \  \----+--TS5     |
            |          |       \  \  \------+--TS6     |
            |          |        \  \--------+--TS7     |
            |          |         \----------+--TS8     |
            +----------+   <------------    +----------+
               node A           Resv           node B
```

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

In the contrary direction, when receiving a label from node A indicating that the TS#i with the granularity of 2.5Gbps is reserved, node B will reserved the TS#i and TS#i+4 with the granularity of 1.25Gbps in its control plane.

7. Security Considerations

The use of control plane protocols for signaling, routing, and path computation opens an OTN to security threats through attacks on those protocols. The data plane technology for an OTN does not introduce any specific vulnerabilities, and so the control plane may be secured using the mechanisms defined for the protocols discussed.

For further details of the specific security measures refer to the documents that define the protocols ([RFC3473], [RFC4203], [RFC4205], [RFC4204], and [RFC5440]). [GMPLS-SEC] provides an overview of security vulnerabilities and protection mechanisms for the GMPLS control plane.

8. IANA Considerations

This document makes not requests for IANA action.

9. Acknowledgments

We would like to thank Maarten Vissers for his review and useful comments.

10. References

10.1. Normative References

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

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

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

[RFC4201] K. Kompella, Y. Rekhter, Ed., "Link Bundling in MPLS Traffic Engineering (TE)", RFC 4201, October 2005.

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

[RFC4203] K. Kompella, Y. Rekhter, Ed., "OSPF Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4203, October 2005.

[RFC4205] K. Kompella, Y. Rekhter, Ed., "Intermediate System to Intermediate System (IS-IS) Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4205, October 2005.

[RFC4204] Lang, J., Ed., "Link Management Protocol (LMP)", RFC 4204, October 2005.

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

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

[RFC6001] Dimitri Papadimitriou et al, "Generalized Multi-Protocol Label Switching (GMPLS) Protocol Extensions for Multi-Layer and Multi-Region Networks (MLN/MRN)", RFC6001, February 21, 2010.

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

[RFC6344] G. Bernstein et al, "Operating Virtual Concatenation (VCAT) and the Link Capacity Adjustment Scheme (LCAS) with Generalized Multi-Protocol Label Switching (GMPLS)", RFC6344, August, 2011.

[G709-V3] ITU-T, "Interfaces for the Optical Transport Network (OTN)", G.709 Recommendation, December 2009.

10.2. Informative References

[G709-V1] ITU-T, "Interface for the Optical Transport Network (OTN)," G.709 Recommendation and Amendment1, November 2001.

[G709-V2] ITU-T, "Interface for the Optical Transport Network (OTN)," G.709 Recommendation, March 2003.

[G7042] ITU-T G.7042/Y.1305, "Link capacity adjustment scheme (LCAS) for virtual concatenated signals", March 2006.

[G872-2001] ITU-T, "Architecture of optical transport networks", November 2001 (11 2001).

[G872-Am2] Draft Amendment 2, ITU-T, "Architecture of optical transport networks".

[G.7044] TD 382 (WP3/15), 31 May - 11 June 2010, Q15 Plenary Meeting in Geneva, Initial draft G.7044 "Hitless Adjustment of ODUflex (HAO)".

[HZang00] H. Zang, J. Jue and B. Mukherjeee, "A review of routing and wavelength assignment approaches for wavelength-routed optical WDM networks", Optical Networks Magazine, January 2000.

[RFC6163] Y. Lee, G. Bernstein, W. Imajuku, "Framework for GMPLS and PCE Control of Wavelength Switched Optical Networks (WSON)", RFC6163, April 2011.

[PCE-APS] Tomohiro Otani, Kenichi Ogaki, Diego Caviglia, and Fatai Zhang, "Requirements for GMPLS applications of PCE", draft-ietf-pce-gmpls-aps-req-04.txt, May 30,2011.

[GMPLS-SEC] Fang, L., Ed., "Security Framework for MPLS and GMPLS Networks", Work in Progress, October 2009.

11. Authors' Addresses

Fatai Zhang (editor)
Huawei Technologies
F3-5-B R&D Center, Huawei Base
Bantian, Longgang District
Shenzhen 518129 P.R.China

Phone: +86-755-28972912
Email: zhangfatai@huawei.com

Dan Li
Huawei Technologies Co., Ltd.
F3-5-B R&D Center, Huawei Base
Bantian, Longgang District
Shenzhen 518129 P.R.China

Phone: +86-755-28973237
Email: huawei.danli@huawei.com

Han Li
China Mobile Communications Corporation
53 A Xibianmennei Ave. Xuanwu District

Beijing 100053 P.R. China

Phone: +86-10-66006688
Email: lihan@chinamobile.com

Sergio Belotti
Alcatel-Lucent
Optics CTO
Via Trento 30 20059 Vimercate (Milano) Italy
+39 039 6863033

Email: sergio.belotti@alcatel-lucent.it

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

12. Contributors

Jianrui Han
Huawei Technologies Co., Ltd.
F3-5-B R&D Center, Huawei Base
Bantian, Longgang District
Shenzhen 518129 P.R.China

Phone: +86-755-28972913
Email: hanjianrui@huawei.com

Malcolm Betts
Huawei Technologies Co., Ltd.

Email: malcolm.betts@huawei.com

Pietro Grandi
Alcatel-Lucent
Optics CTO
Via Trento 30 20059 Vimercate (Milano) Italy
+39 039 6864930

Email: pietro_vittorio.grandi@alcatel-lucent.it

Eve Varma
Alcatel-Lucent
1A-261, 600-700 Mountain Av
PO Box 636
Murray Hill, NJ 07974-0636
USA
Email: eve.varma@alcatel-lucent.com

APPENDIX A: ODU connection examples

This appendix provides a description of ODU terminology and connection examples. This section is not normative, and is just intended to facilitate understanding.

In order to transmit a client signal, an ODU connection must first be created. From the perspective of [G709-V3] and [G872-Am2], some types of ODUs (i.e., ODU1, ODU2, ODU3, ODU4) may assume either a client or server role within the context of a particular networking domain:

(1) An ODUj client that is mapped into an OTUk server. For example, if a STM-16 signal is encapsulated into ODU1, and then the ODU1 is mapped into OTU1, the ODU1 is a LO ODU (from a multiplexing perspective).

(2) An ODUj client that is mapped into an ODUk (j < k) server occupying several TSs. For example, if ODU1 is multiplexed into ODU2, and ODU2 is mapped into OTU2, the ODU1 is a LO ODU and the ODU2 is a HO ODU (from a multiplexing perspective).

Thus, a LO ODUj represents the container transporting a client of the OTN that is either directly mapped into an OTUk (k = j) or multiplexed into a server HO ODUk (k > j) container. Consequently, the HO ODUk represents the entity transporting a multiplex of LO ODUj tributary signals in its OPUk area.

In the case of LO ODUj mapped into an OTUk (k = j) directly, Figure 6 give an example of this kind of LO ODU connection.

In Figure 6, The LO ODUj is switched at the intermediate ODXC node. OCh and OTUk are associated with each other. From the viewpoint of connection management, the management of OTUk is similar with OCh. LO ODUj and OCh/OTUk have client/server relationships.

For example, one LO ODU1 connection can be setup between Node A and Node C. This LO ODU1 connection is to be supported by OCh/OTU1 connections, which are to be set up between Node A and Node B and between Node B and Node C. LO ODU1 can be mapped into OTU1 at Node A, demapped from it in Node B, switched at Node B, and then mapped into the next OTU1 and demapped from this OTU1 at Node C.

```
      |                            LO ODUj                         |
      +------------------------------(b)---------------------------+
      |      |      OCh/OTUk      |     |    OCh/OTUk       |     |
      |      +---------(a)---------+     +---------(a)---------+     |
      |      |                    |     |                   |     |
     +------++-+              +--+---+--+              +-++------+
     |      |EO|              |OE|   |EO|              |OE|      |
     |      +--+----------------+--+   +--+----------------+--+      |
     |  ODXC   |              |  ODXC   |              |  ODXC   |
     +---------+              +---------+              +---------+
      Node A                   Node B                   Node C
```

Figure 6 - Connection of LO ODUj (1)

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

In Figure 7, OCh, OTUk, HO ODUk are associated with each other. The LO ODUj is multiplexed/de-multiplexed into/from the HO ODU at each ODXC node and switched at each ODXC node (i.e. trib port to line port, line card to line port, line port to trib port). From the viewpoint of connection management, the management of these HO ODUk and OTUk are similar to OCh. LO ODUj and OCh/OTUk/HO ODUk have client/server relationships. When a LO ODU connection is setup, it will be using the existing HO ODUk (/OTUk/OCh) connections which have been set up. Those HO ODUk connections provide LO ODU links, of which the LO ODU connection manager requests a link connection to support the LO ODU connection.

For example, one HO ODU2 (/OTU2/OCh) connection can be setup between Node A and Node B, another HO ODU3 (/OTU3/OCh) connection can be setup between Node B and Node C. LO ODU1 can be generated at Node A, switched to one of the 10G line ports and multiplexed into a HO ODU2 at Node A, demultiplexed from the HO ODU2 at Node B, switched at Node B to one of the 40G line ports and multiplexed into HO ODU3 at Node B, demultiplexed from HO ODU3 at Node C and switched to its LO ODU1 terminating port at Node C.

```
     |                          LO ODUj                            |
     +--------------------------(b)----------------------------+
     |      |  OCh/OTUk/HO ODUk  |     | OCh/OTUk/HO ODUk   |     |
     |      +---------(c)---------+     +----------(c)--------+     |
     |      |                    |     |                    |     |
    +------++-+                +--+---+--+              +-++------+
    |      |EO|                |OE|   |EO|              |OE|      |
    |      +--+----------------+--+   +--+----------------+--+      |
    |  ODXC   |                |  ODXC   |              |  ODXC   |
    +---------+                +---------+              +---------+
      Node A                     Node B                   Node C
```

Figure 7 - Connection of LO ODUj (2)

Intellectual Property

Network work group
Internet Draft
Intended status: Standards Track

Fatai Zhang
Young Lee
Jianrui Han
Huawei
G. Bernstein
Grotto Networking
Yunbin Xu
CATR

Expires: March 22, 2012
September 22, 2011

OSPF-TE Extensions for General Network Element Constraints

draft-ietf-ccamp-gmpls-general-constraints-ospf-te-02.txt

Status of this Memo

Abstract

Generalized Multiprotocol Label Switching can be used to control a wide variety of technologies including packet switching (e.g., MPLS), time-division (e.g., SONET/SDH, OTN), wavelength (lambdas), and

spatial switching (e.g., incoming port or fiber to outgoing port or fiber). In some of these technologies network elements and links may impose additional routing constraints such as asymmetric switch connectivity, non-local label assignment, and label range limitations on links. This document describes OSPF routing protocol extensions to support these kinds of constraints under the control of Generalized MPLS (GMPLS).

Conventions used in this document

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

Table of Contents

1. Introduction

Some data plane technologies that wish to make use of a GMPLS control plane contain additional constraints on switching capability and label assignment. In addition, some of these technologies should be capable of performing non-local label assignment based on the nature of the technology, e.g., wavelength continuity constraint in WSON

[RFC6163]. Such constraints can lead to the requirement for link by link label availability in path computation and label assignment.

[GEN-Encode] provides efficient encodings of information needed by the routing and label assignment process in technologies such as WSON and are potentially applicable to a wider range of technologies.

This document defines extensions to the OSPF routing protocol based on [GEN-Encode] to enhance the Traffic Engineering (TE) properties of GMPLS TE which are defined in [RFC3630], [RFC4202], and [RFC4203]. The enhancements to the Traffic Engineering (TE) properties of GMPLS TE links can be announced in OSPF TE LSAs. The TE LSA, which is an opaque LSA with area flooding scope [RFC3630], has only one top-level Type/Length/Value (TLV) triplet and has one or more nested sub-TLVs for extensibility. The top-level TLV can take one of three values (1) Router Address [RFC3630], (2) Link [RFC3630], (3) Generic Node Attribute defined in Section 2. In this document, we enhance the sub-TLVs for the Link TLV and define a new top-level TLV (Generic Node Attribute TLV) in support of the general network element constraints under the control of GMPLS.

The detailed encoding of OSPF extensions are not defined in this document. [GEN-Encode] provides encoding detail.

2. Node Information

According to [GEN-Encode], the additional node information representing node switching asymmetry constraints includes Node ID, connectivity matrix. Except for the Node ID which should comply with Routing Address described in [RFC3630], the other pieces of information are defined in this document.

This document defines a new top TLV named the Generic Node Attribute TLV which carries attributes related to a general network element. This Generic Node Attribute TLV contains one or more sub-TLVs

Per [GEN-Encode], we have identified the following new Sub-TLVs to the Generic Node Attribute TLV. Detail description for each newly defined Sub-TLV is provided in subsequent sections:

Sub-TLV Type	Length	Name
TBD	variable	Connectivity Matrix

In some specific technologies, e.g., WSON networks, Connectivity Matrix sub-TLV may be optional, which depends on the control plane implementations. Usually, for example, in WSON networks, Connectivity

Matrix sub-TLV may appear in the LSAs because WSON switches are asymmetric at present. It is assumed that the switches are symmetric switching, if there is no Connectivity Matrix sub-TLV in the LSAs.

2.1. Connectivity Matrix

It is necessary to identify which ingress ports and labels can be switched to some specific labels on a specific egress port, if the switching devices in some technology are highly asymmetric.

The Connectivity Matrix is used to identify these restrictions, which can represent either the potential connectivity matrix for asymmetric switches (e.g. ROADMs and such) or fixed connectivity for an asymmetric device such as a multiplexer as defined in [WSON-Info].

The Connectivity Matrix is a sub-TLV (the type is TBD by IANA) of the Generic Node Attribute TLV. The length is the length of value field in octets. The meaning and format of this sub-TLV are defined in Section 5.3 of [GEN-Encode]. One sub-TLV contains one matrix. The Connectivity Matrix sub-TLV may occur more than once to contain multi-matrices within the Generic Node Attribute TLV. In addition a large connectivity matrix can be decomposed into smaller separate matrices for transmission in multiple LSAs as described in Section 5.

3. Link Information

The most common link sub-TLVs nested to link top-level TLV are already defined in [RFC3630], [RFC4203]. For example, Link ID, Administrative Group, Interface Switching Capability Descriptor (ISCD), Link Protection Type, Shared Risk Link Group Information (SRLG), and Traffic Engineering Metric are among the typical link sub-TLVs.

Per [GEN-Encode], we add the following additional link sub-TLVs to the link-TLV in this document.

Sub-TLV Type	Length	Name
TBD	variable	Port Label Restrictions
TBD	variable	Available Labels
TBD	variable	Shared Backup Labels

Generally all the sub-TLVs above are optional, which depends on the control plane implementations. If it is default no restrictions on labels, Port Label Restrictions sub-TLV may not appear in the LSAs.

In order to be able to compute label assignment, Available Labels sub-TLV may appear in the LSAs. For example, in WSON networks, without available wavelength information, path computation need guess what lambdas may be available (high blocking probability or distributed wavelength assignment may be used).

3.1. Port Label Restrictions

Port label restrictions describe the label restrictions that the network element (node) and link may impose on a port. These restrictions represent what labels may or may not be used on a link and are intended to be relatively static. More dynamic information is contained in the information on available labels. Port label restrictions are specified relative to the port in general or to a specific connectivity matrix for increased modeling flexibility.

For example, Port Label Restrictions describes the wavelength restrictions that the link and various optical devices such as OXCs, ROADMs, and waveband multiplexers may impose on a port in WSON. These restrictions represent what wavelength may or may not be used on a link and are relatively static. The detailed information about Port label restrictions is described in [WSON-Info].

The Port Label Restrictions is a sub-TLV (the type is TBD by IANA) of the Link TLV. The length is the length of value field in octets. The meaning and format of this sub-TLV are defined in Section 5.4 of [GEN-Encode]. The Port Label Restrictions sub-TLV may occur more than once to specify a complex port constraint within the link TLV.

3.2. Available Labels

Available Labels indicates the labels available for use on a link as described in [GEN-Encode]. The Available Labels is a sub-TLV (the type is TBD by IANA) of the Link TLV. The length is the length of value field in octets. The meaning and format of this sub-TLV are defined in Section 5.1 of [GEN-Encode]. The Available Labels sub-TLV may occur at most once within the link TLV.

Note that there are five approaches for Label Set which is used to represent the Available Labels described in [GEN-Encode]. Usually, it depends on the implementation to one of the approaches. In WSON networks, considering that the continuity of the available or unavailable wavelength set can be scattered for the dynamic wavelength availability, so it may burden the routing to reorganize the wavelength set information when the Inclusive (/Exclusive) List (/Range) approaches are used to represent Available Wavelengths information. Therefore, it is RECOMMENDED that only the Bitmap Set be used for representation Available Wavelengths information.

The "Base Label" and "Last Label" in label set defined in [GEN-Encode] corresponds to base wavelength label and last wavelength label in WSON, the format of which is described as follows:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|Grid |  C.S. |      Reserved     |    n                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

The detailed information related to wavelength label can be referred to [RFC6205].

3.3. Shared Backup Labels

Shared Backup Labels indicates the labels available for shared backup use on a link as described in [GEN-Encode].

The Shared Backup Labels is a sub-TLV (the type is TBD by IANA) of the Link TLV. The length is the length of value field in octets. The meaning and format of this sub-TLV are defined in Section 5.2 of [GEN-Encode]. The Shared Backup Labels sub-TLV may occur at most once within the link TLV.

4. Routing Procedures

All the sub-TLVs are nested to top-level TLV(s) and contained in Opaque LSAs. The flooding of Opaque LSAs must follow the rules specified in [RFC2328], [RFC5250], [RFC3630], [RFC4203].

Considering the routing scalability issues in some cases, the routing protocol should be capable of supporting the separation of dynamic information from relatively static information to avoid unnecessary updates of static information when dynamic information is changed. A standard-compliant approach is to separate the dynamic information sub-TLVs from the static information sub-TLVs, each nested to top-level TLV ([RFC3630 and RFC5876]), and advertise them in the separate OSPF TE LSAs.

For node information, since the Connectivity Matrix information is static, the LSA containing the Generic Node Attribute TLV can be updated with a lower frequency to avoid unnecessary updates.

For link information, a mechanism MAY be applied such that static information and dynamic information of one TE link are contained in separate Opaque LSAs. For example, the Port Label Restrictions information sub-TLV and Available Labels information sub-TLV can be nested to the top level link TLVs and advertised in the separate LSAs.

Note that as with other TE information, an implementation SHOULD take measures to avoid rapid and frequent updates of routing information that could cause the routing network to become swamped. A threshold mechanism MAY be applied such that updates are only flooded when a number of changes have been made to the label availability information (e.g., wavelength availability) within a specific time. Such mechanisms MUST be configurable if they are implemented.

5. Scalability and Timeliness

This document has defined four sub-TLVs for describing generic routing contraints. The examples given in [Gen-Encode] show that very large systems, in terms of label count or ports can be very efficiently encoded. However there has been concern expressed that some possible systems may produce LSAs that exceed the IP Maximum Transmission Unit (MTU) and that methods be given to allow for the splitting of general constraint LSAs into smaller LSA that are under the MTU limit. This section presents a set of techniques that can be used for this purpose.

5.1. Different Sub-TLVs into Multiple LSAs

Four sub-TLVs are defined in this document:

1. Connectivity Matrix (Generic Node Attribute TLV)
2. Port Label Restrictions (Link TLV)
3. Available Labels (Link TLV)
4. Shared Backup Labels (Link TLV)

Except for the Connectivity Matrix all these are carried in an Link TLV of which there can be at most one in an LSA [RFC3630]. Of these sub-TLVs the Port Label Restrictions are relatively static, i.e., only would change with hardware changes or significant system reconfiguration. While the Available Labels and Shared Backup Labels are dynamic, meaning that they may change with LSP setup or teardown through the system. The most important technique for scalability and OSPF bandwidth reduction is to separate the dynamic information sub-TLVs from the static information sub-TLVs and advertise them in separate OSPF TE LSAs[RFC3630 and RFC5250].

5.2. Decomposing a Connectivity Matrix into Multiple Matrices

In the highly unlikely event that a Connectivity matrix sub-TLV by itself would result in an LSA exceeding the MTU, a single large matrix can be decomposed into sub-matrices. Per [GEN-Encode] a connectivity matrix just consists of pairs of input and output ports that can reach each other and hence such this decomposition would be straightforward. Each of these sub-matrices would get a unique matrix identifier per [GEN-Encode].

From the point of view of a path computation process, prior to receiving an LSA with a Connectivity Matrix sub-TLV, no connectivity restrictions are assumed, i.e., the standard GMPLS assumption of any port to any port reachability holds. Once a Connectivity Matrix sub-TLV is received then path computation would know that connectivity is restricted and use the information from all Connectivity Matrix sub-TLVs received to understand the complete connectivity potential of the system. Prior to receiving any Connectivity Matrix sub-TLVs path computation may compute a path through the system when in fact no path exists. In between the reception of an additional Connectivity Matrix sub-TLV path computation may not be able to find a path through the system when one actually exists. Both cases are currently encountered and handled with existing GMPLS mechanisms. Due to the reliability mechanisms in OSPF the phenomena of late or missing Connectivity Matrix sub-TLVs would be relatively rare.

6. Security Considerations

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

7. IANA Considerations

[RFC3630] says that the top level Types in a TE LSA and Types for sub-TLVs for each top level Types must be assigned by Expert Review, and must be registered with IANA.

IANA is requested to allocate new Types for the TLV or sub-TLVs as defined in Sections 2 and 3 as follows:

7.1. Node Information

This document introduces a new Top Level Node TLV (Generic Node Attribute TLV) under the OSPF TE LSA defined in [RFC3630].

```
   Value    TLV Type
```

TBA   Generic Node Attribute

This document also introduces the following sub-TLVs of Generic Node Attribute TLV:

Type	sub-TLV
TBD	Connectivity Matrix

7.2. Link Information

This document introduces the following sub-TLVs of TE Link TLV (Value 2):

Type	sub-TLV
TBD	Port Label Restrictions
TBD	Available Labels
TBD	Shared Backup Labels

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.

[RFC2328] Moy, J., "OSPF Version 2", STD 54, RFC 2328, April 1998.

[RFC5250] L. Berger, I. Bryskin, A. Zinin, R. Coltun "The OSPF Opaque LSA Option", RFC 5250, July 2008.

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

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

[RFC4203] Kompella, K., Ed., and Y. Rekhter, Ed., "OSPF Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4203, October 2005.

[GEN-Encode] G. Bernstein, Y. Lee, D. Li, W. Imajuku, " General Network Element Constraint Encoding for GMPLS Controlled Networks", work in progress: draft-ietf-ccamp-general-constraint-encode-05.txt, May 2011.

[RFC6205] T. Otani, H. Guo, K. Miyazaki, D. Caviglia, " Generalized Labels for Lambda-Switching Capable Label Switching Routers", work in progress: draft-ietf-ccamp-gmpls-g-694-lambda-labels-11.txt, January 2011.

8.2. Informative References

[RFC6163] Y. Lee, G. Bernstein, W. Imajuku, "Framework for GMPLS and PCE Control of Wavelength Switched Optical Networks (WSON)", work in progress: draft-ietf-ccamp-rwa-WSON-Framework-12.txt, February 2011.

[WSON-Info] Y. Lee, G. Bernstein, D. Li, W. Imajuku, "Routing and Wavelength Assignment Information Model for Wavelength Switched Optical Networks", work in progress: draft-ietf-ccamp-rwa-info-12.txt, September 2011.

9. Authors' Addresses

Fatai Zhang
Huawei Technologies
F3-5-B R&D Center, Huawei Base
Bantian, Longgang District
Shenzhen 518129 P.R.China

Phone: +86-755-28972912
Email: zhangfatai@huawei.com

Young Lee
Huawei Technologies
1700 Alma Drive, Suite 100
Plano, TX 75075
USA

Phone: (972) 509-5599 (x2240)

Email: ylee@huawei.com

Jianrui Han
Huawei Technologies Co., Ltd.
F3-5-B R&D Center, Huawei Base
Bantian, Longgang District
Shenzhen 518129 P.R.China

Phone: +86-755-28977943
Email: hanjianrui@huawei.com

Greg Bernstein
Grotto Networking
Fremont CA, USA

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

Yunbin Xu
China Academy of Telecommunication Research of MII
11 Yue Tan Nan Jie Beijing, P.R.China
Phone: +86-10-68094134
Email: xuyunbin@mail.ritt.com.cn

Guoying Zhang
China Academy of Telecommunication Research of MII
11 Yue Tan Nan Jie Beijing, P.R.China
Phone: +86-10-68094272
Email: zhangguoying@mail.ritt.com.cn

Dan Li
Huawei Technologies Co., Ltd.
F3-5-B R&D Center, Huawei Base
Bantian, Longgang District
Shenzhen 518129 P.R.China

Phone: +86-755-28973237
Email: danli@huawei.com

Ming Chen
European Research Center
Huawei Technologies
Riesstr. 25, 80992 Munchen, Germany

Phone: 0049-89158834072
Email: minc@huawei.com

Yabin Ye
European Research Center
Huawei Technologies
Riesstr. 25, 80992 Munchen, Germany

Phone: 0049-89158834074
Email: yabin.ye@huawei.com

Acknowledgment

We thank Ming Chen and Yabin Ye from DICONNET Project who provided valuable information for this document.

Intellectual Property

Disclaimer of Validity

Full Copyright Statement

CCAMP Working Group   D. Ceccarelli, Ed.
Internet-Draft   D. Caviglia
Intended status: Standards Track   Ericsson
Expires: April 15, 2012   F. Zhang
                                                                    D. Li
                                                      Huawei Technologies
                                                               S. Belotti
                                                                P. Grandi
                                                           Alcatel-Lucent
                                                                   R. Rao
                                                              K. Pithewan
                                                     Infinera Corporation
                                                                 J. Drake
                                                                  Juniper
                                                         October 13, 2011

Traffic Engineering Extensions to OSPF for Generalized MPLS (GMPLS) Control of Evolving G.709 OTN Networks

draft-ietf-ccamp-gmpls-ospf-g709v3-00

Abstract

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

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

Status of this Memo

Copyright Notice

Table of Contents

1. Introduction

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

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

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

1.1. Terminology

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

2. OSPF-TE Extensions

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

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

As discussed in [OTN-FWK] and [OTN-INFO], the OSPF-TE must be extended so to be able to advertise the termination and switching capabilites related to each different ODUj and ODUk/OTUk and the advertisement of related multiplexing capabilities. This leads to the need to define a new Switching Capability value and associated new Switching Capability for the ISCD.

In the following we will use ODUj to indicate a service type that is multiplexed into an higher order ODU, ODUk an higher order ODU including an ODUj and ODUk/OTUk to indicate the layer mapped into the OTUk. Moreover ODUj(S) and ODUk(S) are used to indicate ODUj and ODUk supporting switching capability only, and the ODUj->ODUk format is used to indicate the ODUj into ODUk multiplexing capability.

This notation can be iterated as needed depending on the number of multiplexing levels. In the following the term "multiplexing tree" is used to identify a multiplexing hierarchy where the root is always a server ODUk/OTUk and any other supported multiplexed container is represented with increasing granularity until reaching the leaf of the tree. The tree can be structured with more than one branch if the server ODUk/OTUk supports more than one hierarchy.

If for example a multiplexing hierarchy like the following one is considered:

```
ODU2 ODU0    ODUflex ODU0
   \ /           \ /
    |             |
  ODU3          ODU2
     \           /
      \         /
       \       /
        \     /
         ODU4
```

The ODU4 is the root of the muxing tree, ODU3 and ODU2 are containers directly multiplexed into the server and then ODU2, ODU0 are the leaves of the ODU3 branch, while ODUflex and ODU0 are the leaves of the ODU2 one. This means that on this traffic card it is possible to have the following multiplexing capabilities:

```
ODU2->ODU3->ODU4
ODU0->ODU3->ODU4
ODUflex->ODU2->ODU4
ODU0->ODU2->ODU4
```

3. TE-Link Representation

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

OTUk physical Link(s) can be modeled as a TE-Link(s). The TE-Link is termed as OTUk-TE-Link. The OTUk-TE-Link advertises ODUj switching capacity. The advertised capacity could include ODUk switching capacity. Figure-1 below provides an illustration of one hop ODUk TE-links.

```
+-------+               +-------+               +-------+
|  OTN  |               |  OTN  |               |  OTN  |
|Switch |<- OTUk Link ->|Switch |<- OTUk Link ->|Switch |
|   A   |               |   B   |               |   C   |
+-------+               +-------+               +-------+

        |<-- TE-Link -->|       |<-- TE-Link -->|
```

Figure 1: ODUk TE-Links

It is possible to create TE-Links that span more than one hop by creating FA between non-adjacent nodes. Such TE-Links are also termed ODUk-TE-Links. As in the one hop case, these types of ODUk-TE-Links also advertise ODUj switching capacity. The advertised capacity could include ODUk switching capacity.

```
+-------+               +-------+               +-------+
|  OTN  |               |  OTN  |               |  OTN  |
|Switch |<- OTUk Link ->|Switch |<- OTUk Link ->|Switch |
|   A   |               |   B   |               |   C   |
+-------+               +-------+               +-------+
                      ODUk Switched

        |<------------- ODUk Link ------------->|
        |<-------------- TE-Link--------------->|
```

Figure 2: Multiple hop TE-Link

4. ISCD format extensions

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

Value	Type
101 (TBA by IANA)	OTN-TDM capable (OTN-TDM)

Switching Capability and Encoding values MUST be used as follows:

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

Both fixed and flexible ODUs use the same switching type and encoding values. When Switching Capability and Encoding fields are set to values as stated above, the Interface Switching Capability Descriptor should be interpreted as follows:

```
 0                   1                   2                   3
  0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 | Switching Cap |   Encoding    |           Reserved            |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                  Max LSP Bandwidth at priority 0              |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                  Max LSP Bandwidth at priority 1              |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                  Max LSP Bandwidth at priority 2              |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                  Max LSP Bandwidth at priority 3              |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                  Max LSP Bandwidth at priority 4              |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                  Max LSP Bandwidth at priority 5              |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                  Max LSP Bandwidth at priority 6              |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                  Max LSP Bandwidth at priority 7              |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                Switch Capability Specific Information         |
 |                        (variable length)                      |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Maximum LSP Bandwidth

The MAX LSP bandwidth field is used according to [RFC4203]: i.e. 0 <= Max LSP Bandwidth <= ODUk/OTUk and intermediate values are those on the branch of OTN switching hierarchy supported by the interface. E.g. in the OTU4 link it could be possible to have ODU4 as MAX LSP Bandwidth for some priorities, ODU3 for others, ODU2 for some others etc. The bandwidth unit is in bytes per second and the encoding is in IEEE floating point format. The discrete values for various ODUs is shown in the table below.

ODU Type	ODU nominal bit rate	Value in Byte/Sec
ODU0	1 244 160 kbits/s	0x4D1450C0
ODU1	239/238 x 2 488 320 kbit/s	0x4D94F048
ODU2	239/237 x 9 953 280 kbit/s	0x4E959129
ODU3	239/236 x 39 813 120 kbit/s	0X4F963367
ODU4	239/227 x 99 532 800 kbit/s	0x504331E3
ODU2e	239/237 x 10 312 500 kbit/s	0x4E9AF70A
ODUflex for CBR Client signals	239/238 x client signal bit rate	MAX LSP BANDWIDTH
ODUflex for GFP-F Mapped client signal	Configured bit rate	MAX LSP BANDWIDTH
ODU flex resizable	Configured bit rate	MAX LSP BANDWIDTH

A single ISCD MAY be used for the advertisement of unbundled or bundled links supporting homogeneous multiplexing hierarchies and the same Tributary Slot Granularity (TSG). A different ISCD MUST be used for each different muxing hierarchy (muxing tree in the following examples) and different TSG supported within the TE Link, if it includes component links with differing characteristics.

4.1. Switch Capability Specific Information

The technology specific part of the OTN ISCD can include a variable number of sub-TLVs called Bandwidth sub-TLVs. The muxing hierarchy tree is encoded as an order independent list of them. Three types of Bandwidth TLV are defined (TBA by IANA):

- Type 1 - Unreserved Bandwidth for fixed containers

- Type 2 - Unreserved Bandwidth for flexible containers

- Type 3 - Max LSP Bandwidth for flexible containers

The format of the SCSI is depicted in the following figure:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|        Type = 1 (Unres-fix)   |           Length              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
|                          Fixed Container                      |
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                               ...                             ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|        Type = 2 (Unres-var)   |           Length              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
|                       Variable Container                      |
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                                                               ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|        Type = 3 (Max-var)     |           Length              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
|                       Variable Container                      |
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 3: SCSI format

The format of the three different types of Bandwidth TLV are depicted in the following figures:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|        Type = 1 (Unres-fix)   |           Length              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Signal type  | Num of stages |TSG|T|S|  Res  |   Priority    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    Stage#1    |      ...      |   Stage#N     |    Padding    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     Unres ODUj at Prio 0      |     Unres ODUj at Prio 1      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     Unres ODUj at Prio 2      |     Unres ODUj at Prio 3      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     Unres ODUj at Prio 4      |     Unres ODUj at Prio 5      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     Unres ODUj at Prio 6      |     Unres ODUj at Prio 7      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 4: Bandwidth TLV - Type 1 -

The values of the fields shown in figure 4 are explained after figure 6.

```
    0                   1                   2                   3
    0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |        Type = 2 (Unres-var)   |           Length              |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |  Signal type  | Num of stages |TSG|T|S|  Res  |   Priority    |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |    Stage#1    |      ...      |   Stage#N     |    Padding    |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                Unreserved Bandwidth at priority 0             |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                Unreserved Bandwidth at priority 1             |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                Unreserved Bandwidth at priority 2             |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                Unreserved Bandwidth at priority 3             |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                Unreserved Bandwidth at priority 4             |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                Unreserved Bandwidth at priority 5             |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                Unreserved Bandwidth at priority 6             |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                Unreserved Bandwidth at priority 7             |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 5: Bandwidth TLV - Type 2 -

While Bandwidth TLV Type 3 is as follows:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|        Type = 3 (Max-var)     |           Length              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Signal type  | Num of stages |TSG|T|S|  Res  |   Priority    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    Stage#1    |      ...      |  Stage#N      |    Padding    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                 MAX LSP Bandwidth at priority 0               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                 MAX LSP Bandwidth at priority 1               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                 MAX LSP Bandwidth at priority 2               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                 MAX LSP Bandwidth at priority 3               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                 MAX LSP Bandwidth at priority 4               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                 MAX LSP Bandwidth at priority 5               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                 MAX LSP Bandwidth at priority 6               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                 MAX LSP Bandwidth at priority 7               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 6: Bandwidth TLV - Type 3 -

- Signal Type: Indicates the ODU type being advertised

Value	Type
1	ODU1
2	ODU2
3	ODU3
4	ODU4
10	ODU0
11	ODU2e
20	ODUflex CBR
21	ODUflex GFP-F resizable
22	ODUflex GFP-F non resizable
60000-65535	Experimental

- Number of stages: Indicates the number of multiplexing stages level. It is equal to 0 when a server layer is being advertised, 1 in case of single stage muxing, 2 in case of dual stage muxing, etc.

- TSG: Tributary Slot Granularity (2bit): Used for the advertisement of the supported Tributary Slot granularity

   - 00 - Reserved

   - 01 - 1.25 Gbps

   - 10 - 2.5 Gbps

   - 11 - Reserved

- Flags:

   - T Flag (bit 17): Indicates whether the advertised bandwidth can be terminated. When T=1, the signal type can be terminated, when T=0, the signal type cannot be terminated.

   - S Flag (bit 18): Indicates whether the advertised bandwidth can be switched. When S=1, the signal type can be switched, when S=0, the signal type cannot be switched.

   The value 00 in both T and S bits is not permitted.

- Priority :8 bits field with 1 flag for each priority. Bit set indicates priority supported, bit cleared means priority not supported. The priority 0 is related to the most significant bit. When no priority is supported, priority 0 MUST be advertised.

- Stage#1 ... Stage#N : These fields are 8 bits long. Their number is variable and a field is present for each stage of the muxing hierarchy. The last one is always indicating the server ODU container (ODUk/OTUk). The values of the Stage fields are the same ones defined for the Signal Type field. If the number of stages is 0, then no Stage fields are included.

- Padding: Given that the number of Stages is variable, padding to 32 bits field is used as needed.

- Unreserved Bandwidth/Max LSP BW : In case of fixed containers (Type=1) the Unreserved Bandwidth field MUST be 16 bits long and indicates the Unreserved Bandwidth in number of available containers. Only Unreserved/MAX LSP BW fields for supported priorities are included, in order of increasing prioritiy (0 to

7). In case the number of supported priorities is odd, a 16 bits all zeros padding field is added. On the other hand, in case of variable containers (Type 2-3) the Unreserved/MAX LSP Bandwidth fields MUST be 32 bits long and expressed in IEEE floating point format. Only Unreserved/MAX LSP bandwidth for supported priorities MUST be advertised.

5. Examples

The examples in the following pages are not normative and are not intended to infer or mandate any specific implementation.

5.1. MAX LSP Bandwidth fields in the ISCD

This example shows how the MAX LSP Bandwidth fields of the ISCD are filled accordingly to the evolving of the TE-link bandwidth occupancy. In the example an OTU4 link is considered, with supported priorities 0,2,4,7 and muxing hierarchy ODU1->ODU2->ODU3->ODU4.

At time T0, with the link completely free, the advertisement would be:

```
 0                   1                   2                   3
  0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 | Switching Cap |   Encoding    |           Reserved            |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |             Max LSP Bandwidth at priority 0 = 100Gbps         |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |             Max LSP Bandwidth at priority 1 = 0               |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |             Max LSP Bandwidth at priority 2 = 100Gbps         |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |             Max LSP Bandwidth at priority 3 = 0               |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |             Max LSP Bandwidth at priority 4 = 100Gbps         |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |             Max LSP Bandwidth at priority 5 = 0               |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |             Max LSP Bandwidth at priority 6 = 0               |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |             Max LSP Bandwidth at priority 7 = 100Gbps         |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                Switch Capability Specific Information         |
 |                        (variable length)                      |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 7: Example 1 - MAX LSP Bandwidth fields in the ISCD @T0

At time T1 an ODU3 at priority 2 is set-up, so for priority 0 the MAX LSP Bandwidth is still equal to the ODU4 bandwidth, while for priorities from 2 to 7 (excluding the non supported ones) the MAX LSP Bandwidth is equal to ODU3, as no more ODU4s are available and the next supported ODUj in the hierarchy is ODU3.The advertisement is updated as follows:

```
 0                   1                   2                   3
  0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 | Switching Cap |   Encoding    |           Reserved            |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |             Max LSP Bandwidth at priority 0 = 100Gbps         |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |             Max LSP Bandwidth at priority 1 = 0               |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |             Max LSP Bandwidth at priority 2 = 40Gbps          |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |             Max LSP Bandwidth at priority 3 = 0               |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |             Max LSP Bandwidth at priority 4 = 40Gbps          |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |             Max LSP Bandwidth at priority 5 = 0               |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |             Max LSP Bandwidth at priority 6 = 0               |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |             Max LSP Bandwidth at priority 7 = 40Gbps          |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                Switch Capability Specific Information         |
 |                        (variable length)                      |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 8: Example 1 - MAX LSP Bandwidth fields in the ISCD @T1

At time T2 an ODU2 at priority 4 is set-up. The first ODU3 is no longer available since T1 as it was kept by the ODU3 LSP, while the second is no more available and just 3 ODU2 are left in it. ODU2 is now the MAX LSP bandwidth for priorities higher than 4. The advertisement is updated as follows:

```
         0                   1                   2                   3
          0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
         +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
         | Switching Cap |   Encoding    |           Reserved            |
         +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
         |             Max LSP Bandwidth at priority 0 = 100Gbps         |
         +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
         |             Max LSP Bandwidth at priority 1 = 0               |
         +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
         |             Max LSP Bandwidth at priority 2 = 40Gbps          |
         +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
         |             Max LSP Bandwidth at priority 3 = 0               |
         +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
         |             Max LSP Bandwidth at priority 4 = 10Gbps          |
         +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
         |             Max LSP Bandwidth at priority 5 = 0               |
         +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
         |             Max LSP Bandwidth at priority 6 = 0               |
         +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
         |             Max LSP Bandwidth at priority 7 = 10Gbps          |
         +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
         |                Switch Capability Specific Information         |
         |                        (variable length)                      |
         +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 9: Example 1 - MAX LSP Bandwidth fields in the ISCD @T2

5.2. Example of TSG, T and S utilization

In this example an interface with Tributary Slot Type 1.25 Gbps is considered. It supports the simple ODU1->ODU2->ODU3 hierarchy and priorities 0 and 3. Suppose that in this interface the ODU3 signal type can be both switched or terminated, the ODU2 can only be terminated and the ODU1 switched only. For the advertisement of the capabilities of such interface a single ISCD is used and its format is as follows:

```
    0                   1                   2                   3
    0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |        Type = 1 (Unres-fix)   |           Length              |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |Sig type=ODU1  |  #stages= 2   |0 1|T=0|S=1|Res|1|0|0|1|0|0|0|0|
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   | Stage#1=ODU2  | Stage#2=ODU3  |             Padding           |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |     Unres ODU1 at Prio 0      |     Unres ODU1 at Prio 3      |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |        Type = 1 (Unres-fix)   |           Length              |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |Sig type=ODU2  |  #stages= 1   |0 1|T=1|S=0|Res|1|0|0|1|0|0|0|0|
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   | Stage#1=ODU3  |                  Padding                      |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |     Unres ODU2 at Prio 0      |     Unres ODU2 at Prio 3      |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |        Type = 1 (Unres-fix)   |           Length              |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |Sig type=ODU3  |  #stages= 0   |0 1|T=1|S=1|Res|1|0|0|1|0|0|0|0|
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |     Unres ODU3 at Prio 0      |     Unres ODU3 at Prio 3      |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 10: Example 2 - TSG, T and S utilization

5.3. Example of ODUflex advertisement

In this example the advertisement of an ODUflex->ODU3 hierarchy is shown. In case of ODUflex advertisement the MAX LSP bandwidth needs to be advertised but in some cases also information about the Unreserved bandwidth could be useful. The amount of Unreserved bandwidth does not give a clear indication of how many ODUflex LSP can be set up either at the MAX LSP Bandwidth or at different rates, as it gives no information about the spatial allocation of the free TSs.

An indication of the amount of Unreserved bandwidth could be useful during the path computation process, as shown in the following example. Supposing there are two TE-links (A and B) with MAX LSP Bandwidth equal to 10 Gbps each. In case 50Gbps of Unreserved Bandwidth are available on Link A, 10Gbps on Link B and 3 ODUflex LSPs of 10 GBps each, have to be restored, for sure only one can be restored along Link B and it is probable (but not sure) that two of them can be restored along Link A.

In the case of ODUflex advertisement both the Type 2 and Type 3 Bandwidth TLVs are used.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|        Type = 2 (Unres-var)   |           Length              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|S. type=ODUflex|  #stages= 1   |TSG|T|S|  Res  |   Priority    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Stage#1=ODU3 |                   Padding                     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|               Unreserved Bandwidth at priority 0              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|               Unreserved Bandwidth at priority 1              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|               Unreserved Bandwidth at priority 2              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|               Unreserved Bandwidth at priority 3              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|               Unreserved Bandwidth at priority 4              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|               Unreserved Bandwidth at priority 5              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|               Unreserved Bandwidth at priority 6              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|               Unreserved Bandwidth at priority 7              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|        Type = 3 (Max-var)     |           Length              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|S. type=ODUflex|  #stages= 1   |TSG|T|S|  Res  |   Priority    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Stage#1=ODU3 |                   Padding                     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                 MAX LSP Bandwidth at priority 0               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                 MAX LSP  Bandwidth at priority 1              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                 MAX LSP  Bandwidth at priority 2              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                 MAX LSP  Bandwidth at priority 3              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                 MAX LSP  Bandwidth at priority 4              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                 MAX LSP  Bandwidth at priority 5              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                 MAX LSP  Bandwidth at priority 6              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                 MAX LSP  Bandwidth at priority 7              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 11: Example 3 - ODUflex advertisement

5.4. Example of single stage muxing

Supposing there is 1 OTU4 component link supporting single stage muxing of ODU1, ODU2, ODU3 and ODUflex, the supported hierarchy can be summarized in a tree as in the following figure. For sake of simplicity we assume that also in this case only priorities 0 and 3 are supported.

```
        ODU1 ODU2  ODU3 ODUflex
          \    \    /    /
           \    \  /    /
                ODU4
```

and the related SCSIs as follows:

```
0                   1                   2                   3
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|        Type = 1 (Unres-fix)   |           Length              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|Sig type=ODU4  |  #stages= 0   |TSG|T|S|  Res  |1|0|0|1|0|0|0|0|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Unres ODU4 at Prio 0 =1     |    Unres ODU4 at Prio 3 =1    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|        Type = 1 (Unres-fix)   |           Length              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|Sig type=ODU1  |  #stages= 1   |TSG|T|S|  Res  |1|0|0|1|0|0|0|0|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Stage#1=ODU4  |                  Padding                      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Unres ODU1 at Prio 0 =40    |    Unres ODU1 at Prio 3 =40   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|        Type = 1 (Unres-fix)   |           Length              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|Sig type=ODU2  |  #stages= 1   |TSG|T|S|  Res  |1|0|0|1|0|0|0|0|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Stage#1=ODU4  |                  Padding                      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Unres ODU2 at Prio 0 =10    |    Unres ODU2 at Prio 3 =10   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|        Type = 1 (Unres-fix)   |           Length              |
```

```
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|Sig type=ODU3  |  #stages= 1   |TSG|T|S|  Res  |1|0|0|1|0|0|0|0|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Stage#1=ODU4  |                  Padding                      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Unres ODU3 at Prio 0 =2     |    Unres ODU3 at Prio 3 =2    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|        Type = 2 (Unres-var)   |           Length              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|S. type=ODUflex|  #stages= 1   |TSG|T|S|  Res  |1|0|0|1|0|0|0|0|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Stage#1=ODU4 |                   Padding                     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|            Unreserved Bandwidth at priority 0 =100Gbps        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|            Unreserved Bandwidth at priority 3 =100Gbps        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|        Type = 3 (Max-var)     |           Length              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|S. type=ODUflex|  #stages= 1   |TSG|T|S|  Res  |1|0|0|1|0|0|0|0|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Stage#1=ODU4 |                   Padding                     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|              MAX LSP Bandwidth at priority 0 =100Gbps         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|              MAX LSP Bandwidth at priority 3 =100Gbps         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 12: Example 4 - Single stage muxing

5.5. Example of multi stage muxing - Unbundled link

Supposing there is 1 OTU4 component link with muxing capabilities as shown in the following figure:

```
ODU2 ODU0    ODUflex ODU0
   \ /            \ /
    |              |
  ODU3           ODU2
     \            /
      \          /
       \        /
        \      /
          ODU4
```

and supported pririties 0 and 3, the advertisement is composed by the following Bandwidth TLVs:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|        Type = 1 (Unres-fix)   |           Length              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|Sig type=ODU4  |  #stages= 0   |TSG|T|S|  Res  |1|0|0|1|0|0|0|0|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Unres ODU4 at Prio 0 =1     |    Unres ODU4 at Prio 3 =1    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|        Type = 1 (Unres-fix)   |           Length              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|Sig type=ODU3  |  #stages= 1   |TSG|T|S|  Res  |1|0|0|1|0|0|0|0|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Stage#1=ODU4 |                   Padding                     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Unres ODU3 at Prio 0 =2     |    Unres ODU3 at Prio 3 =2    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|        Type = 1 (Unres-fix)   |           Length              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|Sig type=ODU2  |  #stages= 1   |TSG|T|S|  Res  |1|0|0|1|0|0|0|0|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Stage#1=ODU4 |                   Padding                     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Unres ODU2 at Prio 0 =10    |    Unres ODU2 at Prio 3 =10   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|        Type = 1 (Unres-fix)   |           Length              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|Sig type=ODU2  |  #stages= 2   |TSG|T|S|  Res  |1|0|0|1|0|0|0|0|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Stage#1=ODU3  | Stage#2=ODU4  |             Padding           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Unres ODU2 at Prio 0 =8     |    Unres ODU2 at Prio 3 =8    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|        Type = 1 (Unres-fix)   |           Length              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|Sig type=ODU0  |  #stages= 2   |TSG|T|S|  Res  |1|0|0|1|0|0|0|0|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Stage#1=ODU3  | Stage#2=ODU4  |             Padding           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Unres ODU0 at Prio 0 =64    |    Unres ODU0 at Prio 3 =64   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|        Type = 1 (Unres-fix)   |           Length              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|Sig type=ODU0  |  #stages= 2   |TSG|T|S|  Res  |1|0|0|1|0|0|0|0|
```

```
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Stage#1=ODU2  | Stage#2=ODU4  |             Padding            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Unres ODU0 at Prio 0 =80    |    Unres ODU0 at Prio 3 =80   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|        Type = 2 (Unres-var)   |           Length               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|S.type=ODUflex |  #stages= 2   |TSG|T|S|  Res  |1|0|0|1|0|0|0|0|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Stage#1=ODU2  | Stage#2=ODU4  |             Padding            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|            Unreserved Bandwidth at priority 0 =100Gbps         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|            Unreserved Bandwidth at priority 3 =100Gbps         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|        Type = 2 (Max-var)     |           Length               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|S.type=ODUflex |  #stages= 2   |TSG|T|S|  Res  |1|0|0|1|0|0|0|0|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Stage#1=ODU2  | Stage#2=ODU4  |             Padding            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|            MAX LSP Bandwidth at priority 0 =10Gbps             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|            MAX LSP Bandwidth at priority 3 =10Gbps             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 13: Example 5 - Multi stage muxing - Unbundled link

5.6. Example of multi stage muxing - Bundled links

In this example 2 OTU4 component links with the same supported TSG and homogeneous muxing hierarchies are considered. The following muxing capabilities trees are supported:

```
Component Link#1      Component Link#2
   ODU2 ODU0             ODU2 ODU0
      \ /                   \ /
       |                     |
      ODU3                  ODU3
       |                     |
      ODU4                  ODU4
```

Considering only supported priorities 0 and 3, the advertisement is as follows:

```
    0                   1                   2                   3
    0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |         Type = 1 (Unres-fix)  |           Length              |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |Sig type=ODU4  |  #stages= 0   |TSG|T|S|  Res  |1|0|0|1|0|0|0|0|
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |   Unres ODU4 at Prio 0 =2     |    Unres ODU4 at Prio 3 =2    |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |         Type = 1 (Unres-fix)  |           Length              |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |Sig type=ODU3  |  #stages= 1   |TSG|T|S|  Res  |1|0|0|1|0|0|0|0|
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |  Stage#1=ODU4 |                   Padding                     |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |   Unres ODU3 at Prio 0 =4     |    Unres ODU3 at Prio 3 =4    |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |         Type = 1 (Unres-fix)  |           Length              |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |Sig type=ODU2  |  #stages= 2   |TSG|T|S|  Res  |1|0|0|1|0|0|0|0|
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   | Stage#1=ODU3  | Stage#2=ODU4  |             Padding           |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |   Unres ODU2 at Prio 0 =16    |    Unres ODU2 at Prio 3 =16   |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |         Type = 1 (Unres-fix)  |           Length              |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |Sig type=ODU0  |  #stages= 2   |TSG|T|S|  Res  |1|0|0|1|0|0|0|0|
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   | Stage#1=ODU3  | Stage#2=ODU4  |             Padding           |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |   Unres ODU0 at Prio 0 =128   |    Unres ODU0 at Prio 3 =128  |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 14: Example 6 - Multi stage muxing - Bundled lilnks

6. Compatibility

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

7. Security Considerations

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

8. IANA Considerations

TBD

9. Contributors

Xiaobing Zi, Huawei Technologies

Email: zixiaobing@huawei.com

Francesco Fondelli, Ericsson

Email: francesco.fondelli@ericsson.com

Marco Corsi, Altran Italia

EMail: marco.corsi@altran.it

Eve Varma, Alcatel-Lucent

EMail: eve.varma@alcatel-lucent.com

Jonathan Sadler, Tellabs

EMail: jonathan.sadler@tellabs.com

Lyndon Ong, Ciena

EMail: lyong@ciena.com

Ashok Kunjidhapatham

akunjidhapatham@infinera.com

Snigdho Bardalai

sbardalai@infinera.com

Steve Balls

Steve.Balls@metaswitch.com

Jonathan Hardwick

Jonathan.Hardwick@metaswitch.com

Xihua Fu

fu.xihua@zte.com.cn

Cyril Margaria

cyril.margaria@nsn.com

10. Acknowledgements

11. References

11.1. Normative References

[OTN-FWK] F.Zhang, D.Li, H.Li, S.Belotti, D.Ceccarelli, "Framework for GMPLS and PCE Control of G.709 Optical Transport networks, work in progress draft-ietf-ccamp-gmpls-g709-framework-04", March 2011.

[OTN-INFO] S.Belotti, P.Grandi, D.Ceccarelli, D.Caviglia, F.Zhang, D.Li, "Information model for G.709 Optical Transport Networks (OTN), work in progress draft-ietf-ccamp-otn-g709-info-model-00", April 2011.

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

[RFC2154] Murphy, S., Badger, M., and B. Wellington, "OSPF with Digital Signatures", RFC 2154, June 1997.

[RFC2328] Moy, J., "OSPF Version 2", STD 54, RFC 2328, April 1998.

[RFC2370] Coltun, R., "The OSPF Opaque LSA Option", RFC 2370, July 1998.

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

[RFC4201] Kompella, K., Rekhter, Y., and L. Berger, "Link Bundling in MPLS Traffic Engineering (TE)", RFC 4201, October 2005.

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

[RFC4203] Kompella, K. and Y. Rekhter, "OSPF Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4203, October 2005.

[RFC5250] Berger, L., Bryskin, I., Zinin, A., and R. Coltun, "The OSPF Opaque LSA Option", RFC 5250, July 2008.

[RFC5340] Coltun, R., Ferguson, D., Moy, J., and A. Lindem, "OSPF for IPv6", RFC 5340, 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.

11.2. Informative References

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

[G.709-v3] ITU-T, "Draft revised G.709, version 3", consented by ITU-T on Oct 2009.

[Gsup43] ITU-T, "Proposed revision of G.sup43 (for agreement)", December 2008.

Authors' Addresses

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

Email: daniele.ceccarelli@ericsson.com

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

Email: diego.caviglia@ericsson.com

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

Email: zhangfatai@huawei.com

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

Email: danli@huawei.com

Sergio Belotti
Alcatel-Lucent
Via Trento, 30
Vimercate
Italy

Email: sergio.belotti@alcatel-lucent.com

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

Email: pietro_vittorio.grandi@alcatel-lucent.com

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

Email: rrao@infinera.com

Khuzema Pithewan
Infinera Corporation
169, Java Drive
Sunnyvale, CA-94089
USA

Email: kpithewan@infinera.com

John E Drake
Juniper

Email: jdrake@juniper.net

Network Working Group   Fatai Zhang, Ed.
Internet Draft   Huawei
Category: Standards Track   Guoying Zhang
                                                                    CATR
                                                          Sergio Belotti
                                                          Alcatel-Lucent
                                                           D. Ceccarelli
                                                                Ericsson
                                                        Khuzema Pithewan
                                                                Infinera
Expires: April 26, 2012   October 26, 2011

Generalized Multi-Protocol Label Switching (GMPLS) Signaling Extensions for the evolving G.709 Optical Transport Networks Control

draft-ietf-ccamp-gmpls-signaling-g709v3-01.txt

Status of this Memo

Abstract

Recent progress in ITU-T Recommendation G.709 standardization has introduced new ODU containers (ODU0, ODU4, ODU2e and ODUflex) and

enhanced Optical Transport Networking (OTN) flexibility. Several recent documents have proposed ways to modify GMPLS signaling protocols to support these new OTN features.

It is important that a single solution is developed for use in GMPLS signaling and routing protocols. This solution must support ODUk multiplexing capabilities, address all of the new features, be acceptable to all equipment vendors, and be extensible considering continued OTN evolution.

This document describes the extensions to the Generalized Multi-Protocol Label Switching (GMPLS) signaling to control the evolving Optical Transport Networks (OTN) addressing ODUk multiplexing and new features including ODU0, ODU4, ODU2e and ODUflex.

Conventions used in this document

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

Table of Contents

1. Introduction

Generalized Multi-Protocol Label Switching (GMPLS) [RFC3945] extends MPLS to include Layer-2 Switching (L2SC), Time-Division Multiplex (e.g., SONET/SDH, PDH, and ODU), Wavelength (OCh, Lambdas) Switching, and Spatial Switching (e.g., incoming port or fiber to outgoing port or fiber). [RFC3471] presents a functional description of the extensions to Multi-Protocol Label Switching (MPLS) signaling required to support Generalized MPLS. RSVP-TE-specific formats and mechanisms and technology specific details are defined in [RFC3473].

With the evolution and deployment of G.709 technology, it is necessary that appropriate enhanced control technology support be provided for G.709. [RFC4328] describes the control technology details that are specific to foundation G.709 Optical Transport Networks (OTN), as specified in the ITU-T Recommendation G.709 [G709-V1], for ODUk deployments without multiplexing.

In addition to increasing need to support ODUk multiplexing, the evolution of OTN has introduced additional containers and new flexibility. For example, ODU0, ODU2e, ODU4 containers and ODUflex are developed in [G709-V3].

In addition, the following issues require consideration:

- Support for hitless adjustment of ODUflex, which is to be specified in ITU-T G.hao.

- Support for Tributary Port Number. The Tributary Port Number has to be negotiated on each link for flexible assignment of tributary ports to tributary slots in case of LO-ODU over HO-ODU (e.g., ODU2 into ODU3).

Therefore, it is clear that [RFC4328] has to be updated or superceded in order to support ODUk multiplexing, as well as other ODU enhancements introduced by evolution of OTN standards.

This document updates [RFC4328] extending the G.709 ODUk traffic parameters and also presents a new OTN label format which is very flexible and scalable.

2. Terminology

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

3. GMPLS Extensions for the Evolving G.709 - Overview

New features for the evolving OTN, for example, new ODU0, ODU2e, ODU4 and ODUflex containers are specified in [G709-V3]. The corresponding new signal types are summarized below:

- Optical Channel Transport Unit (OTUk):
  - . OTU4

- Optical Channel Data Unit (ODUk):
  - . ODU0
  - . ODU2e
  - . ODU4
  - . ODUflex

A new Tributary Slot (TS) granularity (i.e., 1.25 Gbps) is also described in [G709-V3]. Thus, there are now two TS granularities for the foundation OTN ODU1, ODU2 and ODU3 containers. The TS granularity at 2.5 Gbps is used on legacy interfaces while the new 1.25 Gbps is used on the new interfaces.

In addition to the support of ODUk mapping into OTUk (k = 1, 2, 3, 4), the evolving OTN [G.709-V3] encompasses the multiplexing of ODUj (j = 0, 1, 2, 2e, 3, flex) into an ODUk (k > j), as described in Section 3.1.2 of [OTN-FWK].

Virtual Concatenation (VCAT) of OPUk (OPUk-Xv, k = 1/2/3, X = 1...256) is also supported by [OTN-V3]. Note that VCAT of OPU0 / OPU2e / OPU4 / OPUflex is not supported per [OTN-V3].

[RFC4328] describes GMPLS signaling extensions to support the control for G.709 Optical Transport Networks (OTN) [G709-V1]. However, [RFC4328] needs to be updated because it does not provide the means to signal all the new signal types and related mapping and multiplexing functionalities. Moreover, it supports only the deprecated auto-MSI mode which assumes that the Tributary Port Number is automatically assigned in the transmit direction and not checked in the receive direction.

This document extends the G.709 traffic parameters described in [RFC4328] and presents a new flexible and scalable OTN label format.

Additionally, procedures about Tributary Port Number assignment through control plane are also provided in this document.

4. Generalized Label Request

The Generalized Label Request, as described in [RFC3471], carries the LSP Encoding Type, the Switching Type and the Generalized Protocol Identifier (G-PID).

[RFC4328] extends the Generalized Label Request, introducing two new code-points for the LSP Encoding Type (i.e., G.709 ODUk (Digital Path) and G.709 Optical Channel) and adding a list of G-PID values in order to accommodate [G709-v1].

This document follows these extensions and a new Switching Type is introduced to indicate the ODUk switching capability [G709-V3] in order to support backward compatibility with [RFC4328], as described in [OTN-FWK]. The new Switching Type (101, TBA by IANA) is defined in [OTN-OSPF].

5. Extensions for Traffic Parameters for the Evolving G.709

The traffic parameters for G.709 are defined as follows:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Signal Type  |   Reserved    |           NMC/ Tolerance      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|              NVC              |        Multiplier (MT)        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                            Bit_Rate                           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

The Signal Type needs to be extended in order to cover the new Signal Type introduced by the evolving OTN. The new Signal Type values are extended as follows:

Value	Type
0	Not significant
1	ODU1 (i.e., 2.5 Gbps)
2	ODU2 (i.e., 10 Gbps)
3	ODU3 (i.e., 40 Gbps)

4	ODU4 (i.e., 100 Gbps)
5	Reserved (for future use)
6	OCh at 2.5 Gbps
7	OCh at 10 Gbps
8	OCh at 40 Gbps
9	OCh at 100 Gbps
10	ODU0 (i.e., 1.25 Gbps)
11	ODU2e (i.e., 10Gbps for FC1200 and GE LAN)
12~19	Reserved (for future use)
20	ODUflex(CBR) (i.e., 1.25*N Gbps)
21	ODUflex(GFP-F), resizable (i.e., 1.25*N Gbps)
22	ODUflex(GFP-F), non resizable (i.e., 1.25*N Gbps)
23~255	Reserved (for future use)

NMC/Tolerance:

This field is redefined from the original definition in [RFC4328]. NMC field defined in [RFC4328] cannot be fixed value for an end-to-end circuit involving dissimilar OTN link types. For example, ODU2e requires 9 TS on ODU3 and 8 TS on ODU4. Usage of NMC field is deprecated and should be used only with [RFC4328] generalized label format for backwards compatibility reasons. For the new generalized label format as defined in this document this field is interpreted as Tolerance.

In case of ODUflex(CBR), the Bit_Rate and Tolerance fields MUST be used together to represent the actual bandwidth of ODUflex, where:

- The Bit_Rate field indicates the nominal bit rate of ODUflex(CBR) expressed in bytes per second, encoded as a 32-bit IEEE single-precision floating-point number (referring to [RFC4506] and [IEEE]). The value contained in the Bit Rate field has to keep into account both 239/238 factor and the Transcoding factor.

- The Tolerance field indicates the bit rate tolerance (part per million, ppm) of the ODUflex(CBR) encoded as an unsigned integer, which is bounded in 0~100ppm.

For example, for an ODUflex(CBR) service with Bit_Rate = 2.5Gbps and Tolerance = 100ppm, the actual bandwidth of the ODUflex is:

2.5Gbps * (1 +/- 100ppm)

In case of ODUflex(GFP), the Bit_Rate field is used to indicate the nominal bit rate of the ODUflex(GFP), which implies the number of tributary slots requested for the ODUflex(GFP). Since the tolerance of ODUflex(GFP) makes no sense on tributary slot resource reservation, the Tolerance field for ODUflex(GFP) is not necessary and MUST be filled with 0.

In case of other ODUk signal types, the Bit_Rate and Tolerance fields are not necessary and MUST be set to 0.

The usage of the NVC and Multiplier (MT) fields are the same as [RFC4328].

5.1. Usage of ODUflex(CBR) Traffic Parameter

In case of ODUflex(CBR), the information of Bit_Rate and Tolerance in the ODUflex traffic parameter MUST be used to determine the total number of tributary slots N in the HO ODUk link to be reserved. Here:

```
      N = Ceiling of

ODUflex(CBR) nominal bit rate * (1 + ODUflex(CBR) bit rate tolerance)
---------------------------------------------------------------------
    ODTUk.ts nominal bit rate * (1 - HO OPUk bit rate tolerance)
```

In this formula, the ODUflex(CBR) nominal bit rate is the bit rate of the ODUflex(CBR) on the line side, i.e., the client signal bit rate after applying the 239/238 factor (according to clause 7.3 table 7.2 of [G709-V3]) and the transcoding factor T (if needed) on the CBR client. According to clauses 17.7.3, 17.7.4 and 17.7.5 of [G709-V3]:

ODUflex(CBR) nominal bit rate = CBR client bit rate * (239/238) / T

The ODTUk.ts nominal bit rate is the nominal bit rate of the tributary slot of ODUk, as shown in Table 1 (referring to [G709-V3]).

Table 1 - Actual TS bit rate of ODUk (in Gbps)

ODUk.ts	Minimum	Nominal	Maximum
ODU2.ts	1.249 384 632	1.249 409 620	1.249 434 608
ODU3.ts	1.254 678 635	1.254 703 729	1.254 728 823
ODU4.ts	1.301 683 217	1.301 709 251	1.301 735 285

Note that:

Minimum bit rate of ODUTk.ts =
ODTUk.ts nominal bit rate * (1 - HO OPUk bit rate tolerance)

Maximum bit rate of ODTUk.ts =
ODTUk.ts nominal bit rate * (1 + HO OPUk bit rate tolerance)

Where: HO OPUk bit rate tolerance = 20ppm

Therefore, a node receiving a PATH message containing ODUflex(CBR) nominal bit rate and tolerance can allocate precise number of tributary slots and set up the cross-connection for the ODUflex service.

Note that for different ODUk, the bit rates of the tributary slots are different, and so the total number of tributary slots to be reserved for the ODUflex(CBR) may not be the same on different HO ODUk links.

An example is given below to illustrate the usage of ODUflex(CBR) traffic parameter.

As shown in Figure 1, assume there is an ODUflex(CBR) service requesting a bandwidth of (2.5Gbps, +/-100ppm) from node A to node C. In other words, the ODUflex traffic parameters indicate that Signal Type is 20 (ODUflex(CBR)), Bit_Rate is 2.5Gbps and Tolerance is 100ppm.

```
+-----+             +---------+            +-----+
|     +-------------+ +-----+ +-------------+     |
|     +=============+\| ODU |/+=============+     |
|     +=============+/| flex+-+=============+     |
|     +-------------+ |     |\+=============+     |
|     +-------------+ +-----+ +-------------+     |
|     |             |         |             |     |
|     |   .......   |         |   .......   |     |
|  A  +-------------+    B    +-------------+  C  |
+-----+   HO ODU4   +---------+   HO ODU2   +-----+

  =========: TS occupied by ODUflex
  ---------: free TS
```

Figure 1 - Example of ODUflex(CBR) Traffic Parameters

- On the HO ODU4 link between node A and B:

  The maximum bit rate of the ODUflex(CBR) equals 2.5Gbps * (1 + 100ppm), and the minimum bit rate of the tributary slot of ODU4 equals 1.301 683 217Gbps, so the total number of tributary slots N1 to be reserved on this link is:

N1 = ceiling (2.5Gbps * (1 + 100ppm) / 1.301 683 217Gbps) = 2

- On the HO ODU2 link between node B and C:

  The maximum bit rate of the ODUflex equals 2.5Gbps * (1 + 100ppm), and the minimum bit rate of the tributary slot of ODU2 equals 1.249 384 632Gbps, so the total number of tributary slots N2 to be reserved on this link is:

  N2 = ceiling (2.5Gbps * (1 + 100ppm) / 1.249 384 632Gbps) = 3

5.2. Usage of ODUflex(GFP) Traffic Parameters

[G709-V3-A2] recommends that the ODUflex(GFP) will fill an integral number of tributary slots of the smallest HO ODUk path over which the ODUflex(GFP) may be carried, as shown in Table 2.

Table 2 - Recommended ODUflex(GFP) bit rates and tolerance

ODU type	Nominal bit-rate	Tolerance
ODUflex(GFP) of n TS, 1<=n<=8	n * ODU2.ts	+/-100 ppm
ODUflex(GFP) of n TS, 9<=n<=32	n * ODU3.ts	+/-100 ppm
ODUflex(GFP) of n TS, 33<=n<=80	n * ODU4.ts	+/-100 ppm

Accoding to this table, the Bit_Rate field for ODUflex(GFP) MUST equal to one of the 80 values listed below:

1 * ODU2.ts; 2 * ODU2.ts; ...; 8 * ODU2.ts;
9 * ODU3.ts; 10 * ODU3.ts, ...; 32 * ODU3.ts;
33 * ODU4.ts; 34 * ODU4.ts; ...; 80 * ODU4.ts.

In this way, the number of required tributary slots for the ODUflex(GFP) (i.e., the value of "n" in Table 2) can be deduced from the Bit_Rate field.

6. Generalized Label

[RFC3471] has defined the Generalized Label which extends the traditional label by allowing the representation of not only labels which are sent in-band with associated data packets, but also labels which identify time-slots, wavelengths, or space division multiplexed positions. The format of the corresponding RSVP-TE Generalized Label object is defined in the Section 2.3 of [RFC3473].

However, for different technologies, we usually need use specific label rather than the Generalized Label. For example, the label format described in [RFC4606] could be used for SDH/SONET, the label format in [RFC4328] for G.709.

6.1. New definition of ODU Generalized Label

In order to be compatible with new types of ODU signal and new types of tributary slot, the following new ODU label format MUST be used:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|         TPN           |   Reserved    |        Length         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~             Bit Map         .........                         ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

The ODU Generalized Label is used to indicate how the LO ODUj signal is multiplexed into the HO ODUk link. Note that the LO OUDj signal type is indicated by traffic parameters, while the type of HO ODUk link can be figured out locally according to the identifier of the selected interface carried in the IF_ID RSVP_HOP Object.

TPN (12 bits): indicates the Tributary Port Number (TPN) for the assigned Tributary Slot(s).

- In case of LO ODUj multiplexed into HO ODU1/ODU2/ODU3, only the lower 6 bits of TPN field are significant and the other bits of TPN MUST be set to 0.

- In case of LO ODUj multiplexed into HO ODU4, only the lower 7 bits of TPN field are significant and the other bits of TPN MUST be set to 0.

- In case of ODUj mapped into OTUk (j=k), the TPN is not needed and this field MUST be set to 0.

As per [G709-V3], The TPN is used to allow for correct demultiplexing in the data plane. When an LO ODUj is multiplexed into HO ODUk occupying one or more TSs, a new TPN value is configured at the two ends of the HO ODUk link and is put into the related MSI byte(s) in the OPUk overhead at the (traffic) ingress end of the link, so that the other end of the link can learn which TS(s) is/are used by the LO ODUj in the data plane.

According to [G709-V3], the TPN field MUST be set as according to the following tables:

Table 3 - TPN Assignment Rules (2.5Gbps TS granularity)

HO ODUk	LO ODUj	TPN	TPN Assignment Rules
ODU2	ODU1	1~4	Fixed, = TS# occupied by ODU1
ODU3	ODU1	1~16	Fixed, = TS# occupied by ODU1
	ODU2	1~4	Flexible, != other existing LO ODU2s' TPNs

Table 4 - TPN Assignment Rules (1.25Gbps TS granularity)

HO ODUk	LO ODUj	TPN	TPN Assignment Rules
ODU1	ODU0	1~2	Fixed, = TS# occupied by ODU0
ODU2	ODU1	1~4	Flexible, != other existing LO ODU1s' TPNs
	ODU0 & ODUflex	1~8	Flexible, != other existing LO ODU0s and ODUflexes' TPNs
ODU3	ODU1	1~16	Flexible, != other existing LO ODU1s' TPNs
	ODU2	1~4	Flexible, != other existing LO ODU2s' TPNs
	ODU0 & ODU2e & ODUflex	1~32	Flexible, != other existing LO ODU0s and ODU2es and ODUflexes' TPNs
ODU4	Any ODU	1~80	Flexible, != ANY other existing LO ODUs' TPNs

Note that in the case of "Flexible", the value of TPN is not corresponding to the TS number as per [G709-V3].

Length (12 bits): indicates the number of bit of the Bit Map field, i.e., the total number of TS in the HO ODUk link.

In case of an ODUk mapped into OTUk, there is no need to indicate which tributary slots will be used, so the length field MUST be set to 0.

Bit Map (variable): indicates which tributary slots in HO ODUk that the LO ODUj will be multiplexed into. The sequence of the Bit Map is consistent with the sequence of the tributary slots in HO ODUk. Each bit in the bit map represents the corresponding tributary slot in HO ODUk with a value of 1 or 0 indicating whether the tributary slot will be used by LO ODUj or not.

Padded bits are added behind the Bit Map to make the whole label a multiple of four bytes if necessary. Padded bit MUST be set to 0 and MUST be ignored.

Note that the Length field in the label format can also be used to indicate the TS type of the HO ODUk (i.e., TS granularity at 1.25Gbps or 2.5Gbps) since the HO ODUk type can be known from IF_ID RSVP_HOP Object. In some cases when there is no LMP or routing to make the two end points of the link to know the TSG, the TSG information used by another end can be deduced from the label format. For example, for HO ODU2 link, the value of the length filed will be 4 or 8, which indicates the TS granularity is 2.5Gbps or 1.25Gbps, respectively.

6.2. Examples

The following examples are given in order to illustrate the label format described in the previous sections of this document.

(1) ODUk into OTUk mapping:

In such conditions, the downstream node along an LSP returns a label indicating that the ODUk (k=1, 2, 3, 4) is directly mapped into the corresponding OTUk. The following example label indicates an ODU1 mapped into OTU1.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|       TPN = 0         |   Reserved    |     Length = 0        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

(2) ODUj into ODUk multiplexing:

In such conditions, this label indicates that an ODUj is multiplexed into several tributary slots of OPUk and then mapped into OTUk. Some instances are shown as follow:

- ODU0 into ODU2 Multiplexing:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|       TPN = 2         |   Reserved    |     Length = 8        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|0 1 0 0 0 0 0 0|             Padded Bits (0)                   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

This above label indicates an ODU0 multiplexed into the second tributary slot of ODU2, wherein there are 8 TS in ODU2 (i.e., the type of the tributary slot is 1.25Gbps), and the TPN value is 2.

- ODU1 into ODU2 Multiplexing with 1.25Gbps TS granularity:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|       TPN = 1         |   Reserved    |     Length = 8        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|0 1 0 1 0 0 0 0|             Padded Bits (0)                   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

This above label indicates an ODU1 multiplexed into the 2nd and the 4th tributary slot of ODU2, wherein there are 8 TS in ODU2 (i.e., the type of the tributary slot is 1.25Gbps), and the TPN value is 1.

- ODU2 into ODU3 Multiplexing with 2.5Gbps TS granularity:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|       TPN = 1         |   Reserved    |     Length = 16       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|0 1 1 0 1 0 1 0 0 0 0 0 0 0 0 0|       Padded Bits (0)         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

This above label indicates an ODU2 multiplexed into the 2nd, 3rd, 5th and 7th tributary slot of ODU3, wherein there are 16 TS in ODU3 (i.e., the type of the tributary slot is 2.5Gbps), and the TPN value is 1.

6.3. Label Distribution Procedure

This document does not change the existing label distribution procedures [RFC4328] for GMPLS except that the new ODUk label MUST be processed as follows.

When a node receives a generalized label request for setting up an ODUj LSP from its upstream neighbor node, the node MUST generate an ODU label according to the signal type of the requested LSP and the free resources (i.e., free tributary slots of ODUk) that will be reserved for the LSP, and send the label to its upstream neighbor node.

In case of ODUj to ODUk multiplexing, the node MUST firstly determine the size of the Bit Map field according to the signal type and the tributary slot type of ODUk, and then set the bits to 1 in the Bit Map field corresponding to the reserved tributary slots. The node MUST also assign a valid TPN, which does not collide with other TPN value used by existing LO ODU connections in the selected HO ODU link, and configure the expected multiplex structure identifier (ExMSI) using this TPN. Then, the assigned TPN is filled into the label.

In case of ODUk to OTUk mapping, the node only needs to fill the ODUj and the ODUk fields with corresponding values in the label. Other bits are reserved and MUST be set to 0.

In order to process a received ODU label, the node MUST firstly learn which ODU signal type is multiplexed or mapped into which ODU signal type accordingly to the traffic parameters and the IF_ID RSVP_HOP Object in the received message.

In case of ODUj to ODUk multiplexing, the node MUST retrieve the reserved tributary slots in the ODUk by its downstream neighbor node according to the position of the bits that are set to 1 in the Bit Map field. The node determines the TS type (according to the total TS number of the ODUk, or pre-configured TS type), so that the node, based on the TS type, can multiplex the ODUj into the ODUk. The node MUST also retrieve the TPN value assigned by its downstream neighbor node from the label, and fill the TPN into the related MSI byte(s) in the OPUk overhead in the data plane, so that the downstream neighbor node can check whether the TPN received from the data plane is consistent with the ExMSI and determine whether there is any mismatch defect.

In case of ODUk to OTUk mapping, the size of Bit Map field MUST be 0 and no additional procedure is needed.

Note that the procedures of other label related objects (e.g., Upstream Label, Label Set) are similar to the one described above.

Note also that the TPN in the label_ERO MAY not be assigned (i.e., TPN field = 0) if the TPN is requested to be assigned locally.

6.3.1. Notification on Label Error

When receiving an ODUk label from the neighbor node, the node SHOULD check the integrity of the label. An error message containing an "Unacceptable label value" indication ([RFC3209]) SHOULD be sent if one of the following cases occurs:

- Invalid value in the length field.

- The selected link only supports 2.5Gbps TS granularity while the Length field in the label along with ODUk signal type indicates the 1.25Gbps TS granularity;

- The label includes an invalid TPN value that breaks the TPN assignment rules;

- The reserved resources (i.e., the number of "1" in the Bit Map field) do not match with the Traffic Parameters.

6.4. Supporting Virtual Concatenation and Multiplication

As per [RFC6344], the VCGs can be created using Co-Signaled style or Multiple LSPs style.

In case of Co-Signaled style, the explicit ordered list of all labels reflects the order of VCG members, which is similar to [RFC4328]. In case of multiplexed virtually concatenated signals (NVC > 1), the first label indicates the components of the first virtually concatenated signal; the second label indicates the components of the second virtually concatenated signal; and so on. In case of multiplication of multiplexed virtually concatenated signals (MT > 1), the first label indicates the components of the first multiplexed virtually concatenated signal; the second label indicates components of the second multiplexed virtually concatenated signal; and so on.

In case of Multiple LSPs style, multiple control plane LSPs are created with a single VCG and the VCAT Call can be used to associate the control plane LSPs. The procedures are similar to section 6 of [RFC6344].

6.5. Control Plane Backward Compatibility Considerations

Since the [RFC4328] has been deployed in the network for the nodes that support [G709-V1], we call nodes supporting [RFC4328] "legacy nodes". Backward compatibility SHOULD be taken into consideration when the new nodes (i.e., nodes that support RSVP-TE extensions defined in this document) and the legacy nodes are interworking.

For backward compatibility consideration, the new node SHOULD have the ability to generate and parse legacy labels.

- o A legacy node always generates and sends legacy label to its upstream node, no matter the upstream node is new or legacy, as described in [RFC4328].

- o A new node SHOULD generate and send legacy labels if its upstream node is a legacy one, and generate and send new label if its upstream node is a new one.

One backward compatibility example is shown in Figure 2:

```
        Path          Path           Path           Path
+-----+ ----> +-----+ ----> +------+ ----> +------+ ----> +-----+
|     |       |     |       |      |       |      |       |     |
|  A  +-------+  B  +-------+   C  +-------+   D  +-------+  E  |
| new |       | new |       |legacy|       |legacy|       | new |
+-----+ <---- +-----+ <---- +------+ <---- +------+ <---- +-----+
         Resv          Resv           Resv           Resv
     (new label)  (legacy label) (legacy label)  (legacy label)
```

Figure 2 - Backwards compatibility example

As described above, for backward compatibility considerations, it is necessary for a new node to know whether the neighbor node is new or legacy.

One optional method is manual configuration, but it is recommended to use LMP to discover the capability of the neighbor node automatically, as described in [OTN-LMP].

When performing the HO ODU link capability negotiation:

- o If the neighbor node only support the 2.5Gbps TS and only support ODU1/ODU2/ODU3, the neighbor node SHOULD be treated as a legacy node.

o If the neighbor node can support the 1.25Gbps TS, or can support other LO ODU types defined in [G709-V3]), the neighbor node SHOULD be treated as new node.

o If the neighbor node returns a LinkSummaryNack message including an ERROR_CODE indicating nonsupport of HO ODU link capability negotiation, the neighbor node SHOULD be treated as a legacy node.

7. Supporting Multiplexing Hierarchy

As described in [OTN-FWK], one ODUj connection can be nested into another ODUk (j<k) connection, which forms the multiplexing hierarchy in the ODU layer. This is useful if there are some intermediate nodes in the network which only support ODUk but not ODUj switching.

For example, in Figure 3, assume that N3 is a legacy node which only supports [G709-V1] and does not support ODU0 switching. If an ODU0 connection between N1 and N5 is required, then we can create an ODU2 connection between N2 and N4 (or ODU1 / ODU3 connection, depending on policies and the capabilities of the two ends of the connection), and nest the ODU0 into the ODU2 connection. In this way, N3 only needs to perform ODU2 switching and does not need to be aware of the ODU0 connection.

```
   |                                                         |
   |<------------------- ODU0 Connection ------------------->|
   |              |                             |            |
   |              |<---- ODU2 Connection ----->|            |
   |              |                             |            |
+----+         +----+         +----+         +----+         +----+
| N1 +---------+ N2 +=========+ N3 +=========+ N4 +---------+ N5 |
+----+         +----+         +----+         +----+         +----+
      ODU3 link      ODU3 link      ODU3 link      ODU3 link
```

Figure 3 - Example of multiplexing hierarchy

The control plane signaling should support the provisioning of hierarchical multiplexing. Two methods are provided below (taking Figure 3 as example):

- Using the multi-layer network signaling described in [RFC4206], [RFC6107] and [RFC6001] (including related modifications, if needed). That is, when the signaling message for ODUO connection arrives at N2, a new RSVP session between N2 and N4 is triggered to create the ODU2 connection. This ODU2 connection is treated as a Forwarding Adjacency (FA) after it is created. And then the signaling procedure for the ODU0 connection can be continued using the resource of the ODU2 FA.

- The ODU2 FA-LSP is created in advance based on network planning, which is treated as an FA. Then the ODU0 connection can be created using the resource of the ODU2 FA. In this case, the ODU2 FA-LSP and inner ODU0 connections are created separately.

For both methods, when creating an FA-LSP(e.g., ODU2 FA-LSP), the penultimate hop needs to choose a correct outgoing interface for the ODU2 connection, so that the destination node can support multiplexing and de-multiplexing LO ODU signal(e.g., ODU0). In order to choose a correct outgoing interface for the penultimate hop of the FA-LSP, multiplexing capability (i.e., what client signal type that can be adapted directly to this FA-LSP) should be carried in the signaling to setup this FA-LSP. In addition, when Auto_Negotiation in the data plane is not enabled, TS granularity may also be needed.

7.1. ODU FA-LSP Creation

The required hierarchies and TS type for both ends of an FA-LSP is for further study.

8. Security Considerations

This document introduces no new security considerations to the existing GMPLS signaling protocols. Referring to [RFC3473], further details of the specific security measures are provided. Additionally, [GMPLS-SEC] provides an overview of security vulnerabilities and protection mechanisms for the GMPLS control plane.

9. IANA Considerations

- G.709 SENDER_TSPEC and FLOWSPEC objects:

  The traffic parameters, which are carried in the G.709 SENDER_TSPEC and FLOWSPEC objects, do not require any new object class and type based on [RFC4328]:

  o G.709 SENDER_TSPEC Object: Class = 12, C-Type = 5 [RFC4328]

  o G.709 FLOWSPEC Object: Class = 9, C-Type = 5 [RFC4328]

- Generalized Label Object:

  The new defined ODU label (Section 6) is a kind of generalized label. Therefore, the Class-Num and C-Type of the ODU label is the same as that of generalized label described in [RFC3473], i.e., Class-Num = 16, C-Type = 2.

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.

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

[RFC3209] D. Awduche et al, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC3209, December 2001.

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

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

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

[RFC6344] G. Bernstein et al, "Operating Virtual Concatenation (VCAT) and the Link Capacity Adjustment Scheme (LCAS) with Generalized Multi-Protocol Label Switching (GMPLS)", RFC6344, August 2011.

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

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

[RFC6001] Dimitri Papadimitriou et al, "Generalized Multi-Protocol Label Switching (GMPLS) Protocol Extensions for Multi-Layer and Multi-Region Networks (MLN/MRN)", RFC6001, February 21, 2010.

[OTN-FWK] Fatai Zhang et al, "Framework for GMPLS and PCE Control of G.709 Optical Transport Networks", draft-ietf-ccamp-gmpls-g709-framework-05.txt, September 9, 2011.

[OTN-INFO] S. Belotti et al, "Information model for G.709 Optical Transport Networks (OTN)", draft-ietf-ccamp-otn-g709-info-model-01.txt, September 21, 2011.

[OTN-OSPF] D. Ceccarelli et al, "Traffic Engineering Extensions to OSPF for Generalized MPLS (GMPLS) Control of Evolving G.709 OTN Networks", draft-ietf-ccamp-gmpls-ospf-g709v3-00.txt, October 13, 2011

[OTN-LMP] Fatai Zhang, Ed., "Link Management Protocol (LMP) extensions for G.709 Optical Transport Networks", draft-zhang-ccamp-gmpls-g.709-lmp-discovery-04.txt, April 6, 2011.

[G709-V3] ITU-T, "Interfaces for the Optical Transport Network (OTN) ", G.709/Y.1331, December 2009.

[G709-V3-A2] ITU-T, "Interfaces for the Optical Transport Network (OTN) Amendment 2", G.709/y.1331 Amendment 2, April 2011.

10.2. Informative References

[G709-V1] ITU-T, "Interface for the Optical Transport Network (OTN)," G.709 Recommendation (and Amendment 1), February 2001 (November 2001).

[G709-V2] ITU-T, "Interface for the Optical Transport Network (OTN)," G.709 Recommendation, March 2003.

[G798-V2] ITU-T, "Characteristics of optical transport network hierarchy equipment functional blocks", G.798, December 2006.

[G798-V3] ITU-T, "Characteristics of optical transport network hierarchy equipment functional blocks", G.798v3, consented June 2010.

[RFC4506] M. Eisler, Ed., "XDR: External Data Representation Standard", RFC 4506, May 2006.

[IEEE] "IEEE Standard for Binary Floating-Point Arithmetic", ANSI/IEEE Standard 754-1985, Institute of Electrical and Electronics Engineers, August 1985.

[GMPLS-SEC] Fang, L., Ed., "Security Framework for MPLS and GMPLS Networks", Work in Progress, October 2009.

11. Contributors

Jonathan Sadler, Tellabs
Email: jonathan.sadler@tellabs.com

Kam LAM, Alcatel-Lucent
Email: kam.lam@alcatel-lucent.com

Xiaobing Zi, Huawei Technologies
Email: zixiaobing@huawei.com

Francesco Fondelli, Ericsson
Email: francesco.fondelli@ericsson.com

Lyndon Ong, Ciena
Email: lyong@ciena.com

Biao Lu, infinera
Email: blu@infinera.com

12. Authors' Addresses

Fatai Zhang (editor)
Huawei Technologies
F3-5-B R&D Center, Huawei Base
Bantian, Longgang District
Shenzhen 518129 P.R.China
Phone: +86-755-28972912
Email: zhangfatai@huawei.com

Guoying Zhang
China Academy of Telecommunication Research of MII
11 Yue Tan Nan Jie Beijing, P.R.China
Phone: +86-10-68094272
Email: zhangguoying@mail.ritt.com.cn

Sergio Belotti
Alcatel-Lucent
Optics CTO
Via Trento 30 20059 Vimercate (Milano) Italy
+39 039 6863033
Email: sergio.belotti@alcatel-lucent.it

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

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

Yi Lin
Huawei Technologies
F3-5-B R&D Center, Huawei Base
Bantian, Longgang District
Shenzhen 518129 P.R.China
Phone: +86-755-28972914
Email: yi.lin@huawei.com

Yunbin Xu
China Academy of Telecommunication Research of MII
11 Yue Tan Nan Jie Beijing, P.R.China
Phone: +86-10-68094134
Email: xuyunbin@mail.ritt.com.cn

Pietro Grandi
Alcatel-Lucent
Optics CTO
Via Trento 30 20059 Vimercate (Milano) Italy
+39 039 6864930
Email: pietro_vittorio.grandi@alcatel-lucent.it

Diego Caviglia
Ericsson
Via A. Negrone 1/A
Genova - Sestri Ponente
Italy
Email: diego.caviglia@ericsson.com

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

John E Drake
Juniper
Email: jdrake@juniper.net

Igor Bryskin
Adva Optical
EMail: IBryskin@advaoptical.com

13. Acknowledgment

The authors would like to thank Lou Berger and Deborah Brungard for their useful comments to the document.

Intellectual Property

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: January 12, 2012

A. Takacs
Ericsson
D. Fedyk
Alcatel-Lucent
J. He
Huawei
July 11, 2011

GMPLS RSVP-TE extensions for OAM Configuration
draft-ietf-ccamp-oam-configuration-fwk-06

Abstract

OAM is an integral part of transport connections, hence it is required that OAM functions are activated/deactivated in sync with connection commissioning/decommissioning; avoiding spurious alarms and ensuring consistent operation. In certain technologies OAM entities are inherently established once the connection is set up, while other technologies require extra configuration to establish and configure OAM entities. This document specifies extensions to RSVP-TE to support the establishment and configuration of OAM entities along with LSP signaling.

Requirements Language

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

Status of this Memo

This Internet-Draft will expire on January 12, 2012.

Copyright Notice

Table of Contents

1. Introduction

GMPLS is designed as an out-of-band control plane supporting dynamic connection provisioning for any suitable data plane technology; including spatial switching (e.g., incoming port or fiber to outgoing port or fiber), wavelength-division multiplexing (e.g., DWDM), time-division multiplexing (e.g., SONET/SDH, G.709), and Ethernet Provider Backbone Bridging -- Traffic Engineering (PBB-TE) and MPLS. In most of these technologies there are Operations, Administration and Maintenance (OAM) functions employed to monitor the health and performance of the connections and to trigger data plane (DP) recovery mechanisms. Similarly to connections, OAM functions follow general principles but also have some technology specific characteristics.

OAM is an integral part of transport connections, hence it is required that OAM functions are activated/deactivated in sync with connection commissioning/decommissioning; avoiding spurious alarms and ensuring consistent operation. In certain technologies OAM entities are inherently established once the connection is set up, while other technologies require extra configuration to establish and configure OAM entities. In some situations the use of OAM functions, like those of Fault- (FM) and Performance Management (PM), may be optional confirming to actual network management policies. Hence the network operator must be able to choose which kind of OAM functions to apply to specific connections and with what parameters the selected OAM functions should be configured and operated. To achieve this objective OAM entities and specific functions must be selectively configurable.

In general, it is required that the management plane and control plane connection establishment mechanisms are synchronized with OAM establishment and activation. In particular, if the GMPLS control plane is employed it is desirable to bind OAM setup and configuration to connection establishment signaling to avoid two separate management/configuration steps (connection setup followed by OAM configuration) which increases delay, processing and more importantly may be prune to misconfiguration errors. Once OAM entities are setup and configured, pro-active as well as on-demand OAM functions can be activated via the management plane. On the other hand, it should be possible to activate/deactivate pro-active OAM functions via the GMPLS control plane as well.

This document describes requirements on OAM configuration and control via RSVP-TE, and specifies extensions to the RSVP-TE protocol providing a framework to configure and control OAM entities along with the capability to carry technology specific information. Extensions can be grouped into generic elements that are applicable to any OAM solution and technology specific elements that provide additional configuration parameters, which are only needed for a specific OAM technology. This document specifies the technology agnostic elements, which alone can be used to establish and control OAM entities in the case no technology specific information is needed, and specifies the way additional technology specific OAM parameters are provided.

This document addresses end-to-end OAM configuration, that is, the setup of OAM entities bound to an end-to-end LSP, and configuration and control of OAM functions running end-to-end in the LSP. Configuration of OAM entities for LSP segments and tandem connections are out of the scope of this document.

The mechanisms described in this document provide an additional option for bootstrapping OAM that is not intended to replace or deprecate the use of other technology specific OAM bootstrapping techniques; e.g., LSP Ping [RFC4379] for MPLS networks. The procedures specified in this document are intended only for use in environments where RSVP-TE signaling is already in use to set up the LSPs that are to be monitored using OAM.

2. Requirements

MPLS OAM requirements are described in [RFC4377], which provides requirements to create consistent OAM functionality for MPLS networks.

The following list is an excerpt of MPLS OAM requirements documented in [RFC4377]. Only a few requirements are discussed that bear a direct relevance to the discussion set forth in this document.

- It is desired to support the automation of LSP defect detection. It is especially important in cases where large numbers of LSPs might be tested.

- In particular some LSPs may require automated ingress-LSR to egress-LSR testing functionality, while others may not.

- Mechanisms are required to coordinate network responses to defects. Such mechanisms may include alarm suppression, translating defect signals at technology boundaries, and synchronizing defect detection times by setting appropriately bounded detection timeframes.

MPLS-TP defines a profile of MPLS targeted at transport applications [RFC5921]. This profile specifies the specific MPLS characteristics and extensions required to meet transport requirements, including providing additional OAM, survivability and other maintenance functions not currently supported by MPLS. Specific OAM requirements for MPLS-TP are specified in [RFC5654] [RFC5860]. MPLS-TP poses requirements on the control plane to configure and control OAM entities:

- From [RFC5860]: OAM functions MUST operate and be configurable even in the absence of a control plane. Conversely, it SHOULD be possible to configure as well as enable/disable the capability to operate OAM functions as part of connectivity management, and it SHOULD also be possible to configure as well as enable/disable the capability to operate OAM functions after connectivity has been established.

- From [RFC5654]: The MPLS-TP control plane MUST support the configuration and modification of OAM maintenance points as well as the activation/ deactivation of OAM when the transport path or transport service is established or modified.

Ethernet Connectivity Fault Management (CFM) defines an adjunct connectivity monitoring OAM flow to check the liveliness of Ethernet networks [IEEE-CFM]. With PBB-TE [IEEE-PBBTE] Ethernet networks support explicitly-routed Ethernet connections. CFM can be used to track the liveliness of PBB-TE connections and detect data plane failures. In IETF the GMPLS controlled Ethernet Label Switching (GELS) (see [RFC5828] and [RFC6060]) work extended the GMPLS control plane to support the establishment of PBB-TE data plane connections. Without control plane support separate management commands would be needed to configure and start CFM.

GMPLS based OAM configuration and control should be general to be applicable to a wide range of data plane technologies and OAM solutions. There are three typical data plane technologies used for transport application, which are wavelength based such as WSON, TDM based such as SDH/SONET, packet based such as MPLS-TP [RFC5921] and Ethernet PBB-TE [IEEE-PBBTE]. In all these data planes, the operator MUST be able to configure and control the following OAM functions.

- o It MUST be possible to explicitly request the setup of OAM entities for the signaled LSP and provide specific information for the setup if this is required by the technology.

- o Control of alarms is important to avoid false alarm indications and reporting to the management system. It MUST be possible to enable/disable alarms generated by OAM functions. In some cases selective alarm control may be desirable when, for instance, the operator is only concerned about critical alarms thus the non-service affecting alarms should be inhibited.

- o When periodic messages are used for liveliness check (continuity check) of LSPs it MUST be possible to set the frequency of messages allowing proper configuration for fulfilling the requirements of the service and/or meeting the detection time boundaries posed by possible congruent connectivity check operations of higher layer applications. For a network operator to be able to balance the trade-off in fast failure detection and overhead it is beneficial to configure the frequency of continuity check messages on a per LSP basis.

- o Pro-active Performance Monitoring (PM) functions are continuously collecting information about specific characteristics of the connection. For consistent measurement of Service Level Agreements (SLAs) measurement points must use common probing rate to avoid measurement errors.

- o The extensions MUST allow the operator to use only a minimal set of OAM configuration and control features if the data plane technology, the OAM solution or network management policy allows. The extensions must be reusable as much as reasonably possible. That is generic OAM parameters and data plane or OAM technology specific parameters must be separated.

3. RSVP-TE based OAM Configuration

In general, two types of Maintenance Points (MPs) can be distinguished: Maintenance End Points (MEPs) and Maintenance Intermediate Points (MIPs). MEPs reside at the ends of an LSP and are capable of initiating and terminating OAM messages for Fault Management (FM) and Performance Monitoring (PM). MIPs on the other hand are located at transit nodes of an LSP and are capable of reacting to some OAM messages but otherwise do not initiate messages. Maintenance Entity (ME) refers to an association of MEPs and MIPs that are provisioned to monitor an LSP. The ME association is achieved by configuring MPs to belong to the same ME.

When an LSP is signaled, forwarding association is established between endpoints and transit nodes via label bindings. This association creates a context for the OAM entities monitoring the LSP. On top of this association OAM entities may be configured to unambigously identify MPs and MEs.

In addition to MP and ME identification parameters pro-active OAM functions (e.g., Continuity Check (CC), Performance Monitoring) may have specific parameters requiring configuration as well. In particular, the frequency of periodic CC packets and the measurement interval for loss and delay measurements may need to be configured.

In some cases all the above parameters may be either derived from some exiting information or pre-configured default values can be used. In the simplest case the control plane needs to provide information whether or not OAM entities need to be setup for the signaled LSP. If OAM entities are created signaling must provide means to activate/deactivate OAM message flows and associated alarms.

OAM identifiers as well as the configuration of OAM functions are technology specific, i.e., vary depending on the data plane technology and the chosen OAM solution. In addition, for any given data plane technology a set of OAM solutions may be applicable. The OAM configuration framework allows selecting a specific OAM solution to be used for the signaled LSP and provides technology specific TLVs to carry further detailed configuration information.

3.1. Establishment of OAM Entities and Functions

In order to avoid spurious alarms OAM functions should be setup and enabled in the appropriate order. When using the GMPLS control plane, establishment and enabling of OAM functions MUST be bound to RSVP-TE message exchanges.

An LSP can be signaled and established without OAM configuration first, and OAM entities can be added later with a subsequent re-signaling of the LSP. Alternatively, the LSP can be setup with OAM entities right with the first signaling of the LSP. The below procedures apply to both cases.

Before the initiator first sends a Path messages with OAM Configuration information, it MUST establish and configure the corresponding OAM entities locally, however OAM source functions MUST NOT start sending any OAM messages. In the case of bidirectional connections, the initiator node MUST setup the OAM sink function to be prepared to receive OAM messages but MUST suppress any OAM alarms (e.g., due to missing or unidentified OAM messages). The Path message MUST be sent with the "OAM Alarms Enabled" ADMIN_STATUS flag cleared, i.e, data plane OAM alarms are suppressed.

When the Path message arrives at the receiver, the remote end MUST establish and configure OAM entities according to the OAM information provided in the Path message. If this is not possible a PathErr SHOULD be sent and neither the OAM entities nor the LSP SHOULD be established. If OAM entities are established successfully, the OAM sink function MUST be prepared to receive OAM messages but MUST not generate any OAM alarms (e.g., due to missing or unidentified OAM messages). In the case of bidirectional connections, an OAM source function MUST be setup and, according to the requested configuration, the OAM source function MUST start sending OAM messages. Then a Resv message is sent back, including the OAM Configuration TLV that corresponds to the actually established and configured OAM entities and functions. Depending on the OAM technology, some elements of the OAM Configuration TLV MAY be updated/changed; i.e., if the remote end is not supporting a certain OAM configuration it may suggest an alternative setting, which may or may not be accepted by the initiator of the Path message. If it is accepted, the initiator will reconfigure its OAM functions according to the information received in the Resv message. If the alternate setting is not acceptable a ResvErr may be sent tearing down the LSP. Details of this operation are technology specific and should be described in accompanying technology specific documents.

When the initiating side receives the Resv message it completes any pending OAM configuration and enables the OAM source function to send OAM messages.

After this round, OAM entities are established and configured for the LSP and OAM messages are already exchanged. OAM alarms can now be enabled. The initiator, while still keeping OAM alarms disabled sends a Path message with "OAM Alarms Enabled" ADMIN_STATUS flag set. The receiving node enables the OAM alarms after processing the Path message. The initiator enables OAM alarms after it receives the Resv message. Data plane OAM is now fully functional.

3.2. Adjustment of OAM Parameters

There may be a need to change the parameters of an already established and configured OAM function during the lifetime of the LSP. To do so the LSP needs to be re-signaled with the updated parameters. OAM parameters influence the content and timing of OAM messages and identify the way OAM defects and alarms are derived and generated. Hence, to avoid spurious alarms, it is important that both sides, OAM sink and source, are updated in a synchronized way. First, the alarms of the OAM sink function should be suppressed and only then should expected OAM parameters be adjusted. Subsequently, the parameters of the OAM source function can be updated. Finally, the alarms of the OAM sink side can be enabled again.

In accordance with the above operation, the LSP MUST first be re-signaled with "OAM Alarms Enabled" ADMIN_STATUS flag cleared and including the updated OAM Configuration TLV corresponding to the new parameter settings. The initiator MUST keep its OAM sink and source functions running unmodified, but it MUST suppress OAM alarms after the updated Path message is sent. The receiver MUST first disable all OAM alarms, then update the OAM paramaters according to the information in the Path message and reply with a Resv message acknowledging the changes by including the OAM Configuration TLV. Note that the receiving side has the possibility to adjust the requested OAM configuration parameters and reply with and updated OAM Configuration TLV in the Resv message, reflecting the actually configured values. However, in order to avoid an extensive negotiation phase, in the case of adjusting already configured OAM functions, the receiving side SHOULD NOT update the parameters requested in the Path message to an extent that would provide lower performance than what has been configured previously.

The initiator MUST only update its OAM sink and source functions after it received the Resv message. After this Path/Resv message exchange (in both unidirectional and bidirectional LSP cases) the OAM parameters are updated and OAM is running according the new parameter settings. However OAM alarms are still disabled. A subsequent Path/Resv message exchange with "OAM Alarms Enabled" ADMIN_STATUS flag set is needed to enable OAM alarms again.

3.3. Deleting OAM Entities

In some cases it may be useful to remove some or all OAM entities and functions from an LSP without actually tearing down the connection.

To avoid any spurious alarm, first the LSP SHOULD be re-signaled with

"OAM Alarms Enabled" ADMIN_STATUS flag cleared but unchanged OAM configuration. Subsequently, the LSP is re-signaled with "OAM MEP Entities desired" and "OAM MIP Entities desired" LSP ATTRIBUTES flags cleared, and without the OAM Configuration TLV, this MUST result in the deletion of all OAM entities associated with the LSP. All control and data plane resources in use by the OAM entities and functions SHOULD be freed up. Alternatively, if only some OAM functions need to be removed, the LSP is re-signalled with the updated OAM Configuration TLV. Changes between the contents of the previously signalled OAM Configuration TLV and the currently received TLV represent which functions SHOULD be removed/added.

First, OAM source functions SHOULD be deleted and only after that SHOULD the associated OAM sink functions be removed, this will ensure that OAM messages do not leak outside the LSP. To this end the initiator, before sending the Path message, SHOULD remove the OAM source, hence terminating the OAM message flow associated to the downstream direction. In the case of a bidirectional connection, it SHOULD leave in place the OAM sink functions associated to the upstream direction. The remote end, after receiving the Path message, SHOULD remove all associated OAM entities and functions and reply with a Resv message without an OAM Configuration TLV. The initiator completely removes OAM entities and functions after the Resv message arrived.

4. RSVP-TE Extensions

4.1. LSP Attributes Flags

In RSVP-TE the Flags field of the SESSION_ATTRIBUTE object is used to indicate options and attributes of the LSP. The Flags field has 8 bits and hence is limited to differentiate only 8 options. [RFC5420] defines new objects for RSVP-TE messages to allow the signaling of arbitrary attribute parameters making RSVP-TE easily extensible to support new applications. Furthermore, [RFC5420] allows options and attributes that do not need to be acted on by all Label Switched Routers (LSRs) along the path of the LSP. In particular, these options and attributes may apply only to key LSRs on the path such as the ingress LSR and egress LSR. Options and attributes can be signaled transparently, and only examined at those points that need to act on them. The LSP_ATTRIBUTES and the LSP_REQUIRED_ATTRIBUTES objects are defined in [RFC5420] to provide means to signal LSP attributes and options in the form of TLVs. Options and attributes signaled in the LSP_ATTRIBUTES object can be passed transparently through LSRs not supporting a particular option or attribute, while the contents of the LSP_REQUIRED_ATTRIBUTES object must be examined and processed by each LSR. One TLV is defined in [RFC5420]: the Attributes Flags TLV.

One bit (IANA to assign): "OAM MEP entities desired" is allocated in the LSP Attributes Flags TLV to be used in the LSP_ATTRIBUTES object. If the "OAM MEP entities desired" bit is set it is indicating that the establishment of OAM MEP entities are required at the endpoints of the signaled LSP. If the establishment of MEPs is not supported an error must be generated: "OAM Problem/MEP establishment not supported".

If the "OAM MEP entities desired" bit is set and additional parameters need to be configured, an OAM Configuration TLV MAY be included in the LSP_ATTRIBUTES Object.

One bit (IANA to assign): "OAM MIP entities desired" is allocated in the LSP Attributes Flags TLV to be used in the LSP_ATTRIBUTES or LSP_REQUIRED_ATTRIBUES objects. This bit can only be set if the "OAM MEP entities desired" bit is set in. If the "OAM MIP entities desired" bit is set in the LSP_ATTRIBUTES Flags TLV in the LSP_REQUIRED_ATTRIBUTES Object, it is indicating that the establishment of OAM MIP entities is required at every transit node of the signalled LSP. If the establishment of a MIP is not supported an error MUST be generated: "OAM Problem/MIP establishment not supported".

4.2. OAM Configuration TLV

This TLV provides information about which OAM technology/method should be used and carries sub-TLVs for any additional OAM configuration information. The OAM Configuration TLV MAY be carried in the LSP_ATTRIBUTES or LSP_REQUIRED_ATTRIBUTES object in Path and Resv messages. When carried in the LSP_REQUIRED_ATTRIBUTES object it is indicating that intermediate nodes MUST recognize and eventually react on the OAM configuration onformation.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           Type (IANA)         |           Length              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    OAM Type   |                 Reserved                      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
~                           sub-TLVs                            ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

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

OAM Type: specifies the technology specific OAM method. When carried in the LSP_REQUIRED_ATTRIBUTES Object, if the requested OAM method is not supported at a given node an error MUST be generated: "OAM Problem/Unsupported OAM Type". When carried in the LSP_ATTRIBUTES Object, intermediate nodes not supporting the OAM Type pass the object forward unchanged as specified in [RFC5420] only Label Edge Nodes MUST generate the error if the OAM Type is not supported.

OAM Type	Description
0-255	Reserved

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

The receiving node based on the OAM Type will check if a corresponding technology specific OAM configuration sub-TLV is included. If the included technology specific OAM configuration sub-TLV is different than what is specified in the OAM Type an error MUST be generated: "OAM Problem/OAM Type Mismatch". IANA is requested to maintain the sub-TLV space in the new "RSVP-TE OAM Configuration

Registry".

Note that there is a hierarchical dependency in between the OAM configuration elements. First, the "OAM MEP (and MIP) entities desired" flag needs to be set. Only when that is set MAY an "OAM Configuration TLV" be included in the LSP_ATTRIBUTES or LSP_REQUIRED_ATTRIBUTES Object. When this TLV is present, based on the "OAM Type" field, it MAY carry a technology specific OAM configuration sub-TLV. If this hierarchy is broken (e.g., "OAM MEP entities desired" flag is not set but an OAM Configuration TLV is present) an error MUST be generated: "OAM Problem/Configuration Error".

4.2.1. OAM Function Flags Sub-TLV

As the first sub-TLV the "OAM Function Flags sub-TLV" MUST be always included in the "OAM Configuration TLV". "OAM Function Flags" specifies which pro-active OAM functions (e.g., connectivity monitoring, loss and delay measurement) and which fault management signals MUST be established and configured. If the selected OAM Function(s) is(are) not supported, an error MUST be generated: "OAM Problem/Unsupported OAM Function".

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           Type (1) (IANA)     |           Length              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
~                      OAM Function Flags                       ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

OAM Function Flags is bitmap with extensible length based on the Lenght field of the TLV. Bits are numbered from left to right. IANA is requested to maintain the OAM Function Flags in the new "RSVP-TE OAM Configuration Registry". This document defines the following flags.

OAM Function Flag bit#	Description
0	Continuity Check (CC)
1	Connectivity Verification (CV)
2	Fault Monitoring Signal (FMS)
3	Performance Monitoring/Loss (PM/Loss)
4	Performance Monitoring/Delay (PM/Delay)
5	Performance Monitoring/Throughput Measurement (PM/Throughput)

4.2.2. Technology Specific sub-TLVs

One technology specific sub-TLV MAY be defined for each "OAM Type". This sub-TLV MUST contain any further OAM configuration information for that specific "OAM Type". The technology specific sub-TLV, when used, MUST be carried within the OAM Configuration TLV. IANA is requested to maintain the sub-TLV space in the new "RSVP-TE OAM Configuration Registry".

4.3. Administrative Status Information

Administrative Status Information is carried in the ADMIN_STATUS Object. The Administrative Status Information is described in [RFC3471], the ADMIN_STATUS Object is specified for RSVP-TE in [RFC3473].

Two bits are allocated for the administrative control of OAM monitoring. Two bits (IANA to assign) are allocated by this draft: the "OAM Flows Enabled" (M) and "OAM Alarms Enabled" (O) bits. When the "OAM Flows Enabled" bit is set, OAM packets are sent if it is cleared no OAM packets are emitted. When the "OAM Alarms Enabled" bit is set OAM triggered alarms are enabled and associated consequent actions are executed including the notification of the management system. When this bit is cleared, alarms are suppressed and no action is executed and the management system is not notified.

4.4. Handling OAM Configuration Errors

To handle OAM configuration errors a new Error Code (IANA to assign) "OAM Problem" is introduced. To refer to specific problems a set of Error Values is defined.

If a node does not support the establishment of OAM MEP or MIP entities it must use the error value (IANA to assign): "MEP establishment not supported" or "MIP establishment not supported" respectively in the PathErr message.

If a node does not support a specific OAM technology/solution it must use the error value (IANA to assign): "Unsupported OAM Type" in the PathErr message.

If a different technology specific OAM configuration TLV is included than what was specified in the OAM Type an error must be generated with error value: "OAM Type Mismatch" in the PathErr message.

There is a hierarchy in between the OAM configuration elements. If this hierarchy is broken the error value: "Configuration Error" must be used in the PathErr message.

If a node does not support a specific OAM Function it must use the error value: "Unsupported OAM Function" in the PathErr message.

4.5. Considerations on Point-to-Multipoint OAM Configuration

RSVP-TE extensions for the establishment of point-to-multipoint (P2MP) LSPs are specified in [RFC4875]. A P2MP LSP is comprised of multiple source-to-leaf (S2L) sub-LSPs. These S2L sub-LSPs are set up between the ingress and egress LSRs and are appropriately combined by the branch LSRs using RSVP semantics to result in a P2MP TE LSP. One Path message may signal one or multiple S2L sub-LSPs for a single P2MP LSP. Hence the S2L sub-LSPs belonging to a P2MP LSP can be signaled using one Path message or split across multiple Path messages.

P2MP OAM mechanisms are very specific to the data plane technology, hence in this document we only highlight basic operations for P2MP OAM configuration. We consider only the configuration of the root to leaves OAM flows of P2MP LSPs and as such aspects of any return path are outside the scope of our discussions. We also limit our consideration to cases where all leaves must successfully establish OAM entities in order a P2MP OAM is successfully established. In any case, the discussion set forth below provides only guidelines for P2MP OAM configuration, details SHOULD be specified in technology specific documents.

The root node may select if it uses a single Path message or multiple Path messages to setup the whole P2MP tree. In the case when multiple Path messages are used the root node is responsible also to keep the OAM Configuration information consistent in each of the sent Path messages, i.e., the same information MUST be included in all Path messages used to construct the multicast tree. Each branching node will propagate the Path message downstream on each of the branches, when constructing a Path message the OAM Configuration information MUST be copied unchanged from the received Path message, including the related ADMIN_STATUS bits, LSP Attribute Flags and the OAM Configuration TLV. The latter two also imply that the

LSP_ATTRIBUTES and LSP_REQUIRED_ATTRIBUTES Object MUST be copied for the upstream Path message to the subsequent downstream Path messages.

Leaves MUST create and configure OAM sink functions according to the parameters received in the Path message, for P2MP OAM configuration there is no possibility for parameter negotiation on a per leaf basis. This is due to the fact that the only OAM source function, residing in the root of the tree, can only operate with a single configuration which must be obeyed by all leaves. If a leaf cannot accept the OAM parameters it MUST use the RRO Attributes sub-object [RFC5420] to notify the root of the problem. In particular, if the OAM configuration was successful the leaf would set the "OAM MEP entities desired" flag in the RRO Attributes sub-object in the Resv message, while, if due to any reason, OAM entities could not be established the Resv message should be sent with the "OAM MEP entities desired" bit cleared in the RRO Attributes sub-object. Branching nodes should collect and merge the received RROs according to the procedures described in [RFC4875]. This way, the root when receiving the Resv message (or messages if multiple Path messages were used to setup the tree) will have a clear information on which of the leaves could the OAM sink functions be established. If all leaves established OAM entities successfully, the root can enable the OAM message flow. On the other hand, if at some leaves the establishment was unsuccessful additional actions will be needed before the OAM message flow can be enabled. Such action could be to setup two independent P2MP LSPs. One with OAM Configuration information towards leaves which successfully setup OAM. This can be done by prunning the leaves which failed to setup OAM of the previously signalled P2MP LSP. The other P2MP LSP could be constructed for leaves without OAM entities. What exact procedures are needed are technology specific and SHOULD be described in technology specific documents.

5. IANA Considerations

Two bits ("OAM Alarms Enabled" (O) and "OAM Flows Enabled" (M)) needs to be allocated in the ADMIN_STATUS Object.

Two bits ("OAM MEP entities desired" and "OAM MIP entities desired") needs to be allocated in the LSP Attributes Flags Registry.

This document specifies one new TLV to be carried in the LSP_ATTRIBUTES and LSP_REQUIRED_ATTRIBUTES objects in Path and Resv messages: OAM Configuration TLV.

One new Error Code: "OAM Problem" and a set of new values: "MEP establishment not supported", "MIP establishment not supported", "Unsupported OAM Type", "Configuration Error" and "Unsupported OAM Function" needs to be assigned.

IANA is requested to open a new registry: "RSVP-TE OAM Configuration Registry" that maintains the "OAM Type" code points, an associated sub-TLV space, and the allocations of "OAM Function Flags" within the OAM Configuration TLV.

6. Security Considerations

The signaling of OAM related parameters and the automatic establishment of OAM entities based on RSVP-TE messages adds a new aspect to the security considerations discussed in [RFC3473]. In particular, a network element could be overloaded, if a remote attacker could request liveliness monitoring, with frequent periodic messages, for a high number of LSPs, targeting a single network element. Such an attack can efficiently be prevented when mechanisms for message integrity and node authentication are deployed. Since the OAM configuratiuon extensions rely on the hop-by-hop exchange of exiting RSVP-TE messages, procedures specified for RSVP message security in [RFC2747] can be used to mitigate possible attacks.

For a more comprehensive discussion on GMPLS security please see the Security Framework for MPLS and GMPLS Networks [RFC5920]. Cryptography can be used to protect against many attacks described in [RFC5920].

7. Acknowledgements

The authors would like to thank Francesco Fondelli, Adrian Farrel, Loa Andersson, Eric Gray and Dimitri Papadimitriou for their useful comments.

8. References

8.1. Normative References

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

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

[RFC5420] "Encoding of Attributes for Multiprotocol Label Switching (MPLS) Label Switched Path (LSP) Establishment Using Resource ReserVation Protocol-Traffic Engineering (RSVP-TE)", RFC 5420, February 2009.

8.2. Informative References

[IEEE-CFM] "IEEE 802.1ag, Draft Standard for Connectivity Fault Management", work in progress.

[IEEE-PBBTE] "IEEE 802.1Qay Draft Standard for Provider Backbone Bridging Traffic Engineering", work in progress.

[RFC2747] "RSVP Cryptographic Authentication", RFC 2747, January 2000.

[RFC3469] "Framework for Multi-Protocol Label Switching (MPLS)-based Recovery", RFC 3469, February 2003.

[RFC4377] "Operations and Management (OAM) Requirements for Multi-Protocol Label Switched (MPLS) Networks", RFC 4377, February 2006.

[RFC4379] "Detecting Multi-Protocol Label Switched (MPLS) Data Plane Failures", RFC 4379, February 2006.

[RFC4875] "Extensions to Resource Reservation Protocol - Traffic Engineering (RSVP-TE) for Point-to-Multipoint TE Label Switched Paths (LSPs)", RFC 4875, May 2007.

[RFC5654] "Requirements of an MPLS Transport Profile", RFC 5654, September 2009.

[RFC5828] "GMPLS Ethernet Label Switching Architecture and Framework", RFC 5828, March 2010.

[RFC5860] "Requirements for OAM in MPLS Transport Networks", RFC 5860, May 2010.

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

[RFC5921] "A Framework for MPLS in Transport Networks", RFC 5921, July 2010.

[RFC6060] "Generalized Multiprotocol Label Switching (GMPLS) Control of Ethernet Provider Backbone Traffic Engineering (PBB-TE)", RFC 6060.

Authors' Addresses

Attila Takacs
Ericsson
Konyves Kalman krt. 11.
Budapest,   1097
Hungary

Email: attila.takacs@ericsson.com

Don Fedyk
Alcatel-Lucent
Groton, MA 01450
USA

Email: donald.fedyk@alcatel-lucent.com

Jia He
Huawei

Email: hejia@huawei.com

CCAMP Working Group S. Belotti, Ed.
Internet-Draft P. Grandi
Intended status: Informational Alcatel-Lucent
Expires: April 30, 2012 D. Ceccarelli, Ed.
D. Caviglia
Ericsson
F. Zhang
D. Li
Huawei Technologies
October 28, 2011

Information model for G.709 Optical Transport Networks (OTN)
draft-ietf-ccamp-otn-g709-info-model-02

Abstract

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

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

Status of this Memo

Copyright Notice

Table of Contents

1. Introduction

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

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

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

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

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

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

OSPF-TE and RSVP-tE requirements are defined in [OTN-FWK], while protocol extensions are defined in [OTN-OSPF] and [OTN-RSVP].

1.1. Terminology

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

2. OSPF-TE requirements overview

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

- Support for link multiplexing capability advertisement: The routing protocol has to be able to carry information regarding the capability of an OTU link to support different type of ODUs

- Support of any ODUk and ODUflex: The routing protocol must be capable of carrying the required link bandwidth information for performing accurate route computation for any of the fixed rate ODUs as well as ODUflex.

- Support for differentiation between switching and terminating capacity

- Support for the client server mappings as required by [G.7715.1]. The list of different mappings methods is reported in [G.709-v3]. Since different methods exist for how the same client layer is mapped into a server layer, this needs to be captured in order to avoid the set-up of connections that fail due to incompatible mappings.

- Support different priorities for resource reservation. How many priorities levels should be supported depends on operator policies. Therefore, the routing protocol should be capable of supporting either no priorities or up to 8 priority levels as defined in [RFC4202].

- Support link bundling of component links at the same line rate and with same muxing hierarchy.

- Support for Tributary Slot Granularity (TSG) advertisement.

3. RSVP-TE requirements overview

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

- Support for LSP setup of new ODUk/ODUflex containers with related mapping and multiplexing capabilities

- Support for LSP setup using different Tributary Slot granularity

- Support for Tributary Port Number allocation and negoziation

- Support for constraint signaling

- Support for TSG signaling

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

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

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

Figure 1: OTN mapping capability

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

Figure 2: OTN multiplexing capability

How an ODUk connection service is transported within an operator network is governed by operator policy. For example, the ODUk connection service might be transported over an ODUk path over an OTUk section, with the path and section being at the same rate as that of the connection service (see Table 1). In this case, an entire lambda of capacity is consumed in transporting the ODUk connection service. On the other hand, the operator might exploit different multiplexing capabilities in the network to improve infrastructure efficiencies within any given networking domain. In this case, ODUk multiplexing may be performed prior to transport over various rate ODU servers (as per Table 2) over associated OTU sections.

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

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

- o Case A) one or more wavelength sub-networks connected by optical links or

- o Case B) one or more ODU links (having sub-lambda and/or lambda bandwidth granularity)

- o Case C) a mix of ODU links and wavelength sub-networks.

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

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

4.1. Tributary Slot Granularity

ITU-T recommendation defines two type of TS granularity. This TS granularity is defined per layer, meaning that both ends of a link can select proper TS granularity differently for each supported layer, based on the rules below:

- If both ends of a link are new cards supporting both 1.25Gbps TS and 2.5Gbps TS, then the link will work with 1.25Gbps TS.

- If one end is a new card supporting both the 1.25Gbps and 2,5Gbps TS, and the other end is an old card supporting just the 2.5Gbps TS, the link will work with 2.5Gbps TS.

When setting up an ODUj over an ODUk, it is possible to identify two types of TSG, the server and the client one. The server TSG is used to map an end to end ODUj onto a server ODUk LSP or links. This parameter can not be influenced in any way from the ODUj LSP: ODUj LSP will be mapped on tributary slots available on the different links/ODUk LSPs. When setting up an ODUj at a given rate, the fact that it is carried over a path composed by links/FAs structured with 1.25Gbps or 2.5Gbps TS size is completely transparent to the end to end ODUj.

On the other side the client TSG is the tributary slot size that is exported towards the client layer. The client TSG information is one of the parameters needed to correctly select the adaptation towards the client layers at the end nodes and this is the only thing that the ODUj has to guarantee. When setting up an HO-ODUk/OTUk LSP or an H-LSP/FA, in the case where the egress interface cannot be identified from the ERO, it is necessary for the penultimate node to select an interface on the egress node that supports the TSG and ODU client hierarchy specified in signaling. It must then select an interface on itself that can be paired with the interface it selected.

In figure 4 an example of client and server TSG utilization in a scenario with mixed G.709 v2 and G.709 v3 interfaces is shown.

```
                            ODU1-LSP
           .........................................
      TSG-C|                                       |TSG-C
       1.25|                 ODU2-H-LSP            |1.25
           +------------X--------------------------+
           |       TSG-S|                          |TSG-S
           |         2.5|                          |2.5
           |            |       ODU3-H-LSP         |
           |            |------------X-------------|
           |            |                          |
        +--+--+      +--+--+                    +---+-+
        |     |      |     |     +-+   +-+     |     |
        |  A  +------+  B  +-----+ +***+ +-----+  Z  |
        | V.3 | OTU2 | V.2 |OTU3 +-+   +-+ OTU3| V.3 |
        +-----+      +-----+                    +-----+

         ... Service LSP
         --- H-LSP
```

Figure 3: Client-Server TSG example

In this scenario, an ODU3 LSP is setup from node B to Z. Node B has an old interface able to support 2.5 TSG granularity, hence only client TSG equal to 2.5Gbps can be exported to ODU3 H-LSP possible clients. An ODU2 LSP is setup from node A to node Z with client TSG 1.25 signaled and exported towards clients. The ODU2 LSP is carried by ODU3 H-LSP from B to Z. Due to the limitations of old node B interface, the ODU2 LSP is mapped with 2.5Gbps TSG over the ODU3 H-LSP. Then an ODU1 LSP is setup from A to Z, carried by the ODU2 H-LSP and mapped over it using a 1.25Gbps TSG.

What is shown in the example is that the TSG processing is a per layer issue: even if the ODU3 H-LSP is created with TSG client at 2.5Gbps, the ODU2 H-LSP must guarantee a 1.25Gbps TSG client. ODU3 H-LSP is elegible from ODU2 LSP perspective since from the routing it is known that this ODU3 interface at node Z, supports an ODU2 termination exporting a TSG 1.25/2.5.

Moreover, with respect to the penultimate hop implications let’s consider a further example in which the setup of an ODU3 path that is going to carry an ODU0 is considered. In this case it is needed the support of 1,25 GBps TS. The information related to the TSG is carried in the signaling and node C, having two different interfaces toward D with different TSGs, can choose the right one as depicted in the following figure. In case the full ERO is provided in the signaling with explicit interface declaration, there is no need for C to choose the right interface as it has been already decided by the ingress node or the PCE.

```
                         ODU0
         ________________________________________
         |                                      |
+--------+      +--------+      +--------+      +--------+
|        |      |        |      |        | 1.25 |        |
|  Node  |      |  Node  |      |  Node  +------+  Node  |
|   A    +------+   B    +------+   C    | ODU3 |   D    |
|        | ODU3 |        | ODU3 |        +------+        |
+--------+ 1.25 +--------+ 2.5  +--------+ 2.5  +--------+
```

Figure 4: TSG in signaling

The TSG information is needed also in the routing protocol as the ingress node (A in the previous example) needs to know if the interfaces between C and D can support the required TSG. In case they cannot, A will compute an alternate path from itself to D.

In a multi-stage multiplexing environment any layer can have a different TSG structure, e.g. in a multiplexing hierarchy like ODU0->ODU2->ODU3, the ODU3 can be structured at TSG=2.5 in order to support an ODU2 connection, but this ODU2 connection can be a tunnel for ODU0, and hence structured with 1.25 TSG. Therefore any multiplexing level has to advertise his TSG capabilities in order to allow a correct path computation by the end nodes (both of the ODUk trail and of the H-LSP/FA).

The following table shows the different mapping possibilities depending on the TSG types. The client types are shown in the left column, while the different OPUk server and related TSGs are listed in the top row. The table also shows the relationship between the TSG and the payload type.

	2.5G TS		1.25G TS			
	OPU2	OPU3	OPU1	OPU2	OPU3	OPU4
ODU0	-	-	AMP PT=20	GMP PT=21	GMP PT=21	GMP PT=21
ODU1	AMP PT=20	AMP PT=20	-	AMP PT=21	AMP PT=21	GMP PT=21
ODU2	-	AMP PT=20	-	-	AMP PT=21	GMP PT=21
ODU2e	-	-	-	-	GMP PT=21	GMP PT=21
ODU3	-	-	-	-	-	GMP PT=21
ODUfl	-	-	-	GMP PT=21	GMP PT=21	GMP PT=21

Figure 5: ODUj into OPUk mapping types

The signaled TSGs information is not enough to have a complete choice since the penultimate hop node has to distinguish between interfaces with the same TSG (e.g. 1.25Gbps) whether the interface is able to support the right hierarchy, i.e. it is possible to have two interfaces both at 1.25 TSG but only one is supporting ODU0.

A dedicated optional object could be defined in order to carry the multiplexing hierarchy and adaptation information (i.e. TSG/PT, AMP/GMP) so to have a more precise choice capability. In this way, when the penultimate node receives such object, together with the Traffic Parameters Object, is allowed to choose the correct interface towards the egress node.

In conclusion both routing and signaling will need to be extended to appropriately represent the TSG/PT information. Routing will need to represent a link’s TSG and PT capabilities as well as the supported multiplexing hierarchy. Signaling will need to represent the TSG/PT and multiplexing hierarchy encoding.

4.1.1. Fall-back procedure

SG15 ITU-T G.798 recommendation describes the so called PT=21-to-PT=20 interworking process that explains how two equipments with interfaces with different PayloadType, and hence different TS granularity (1.25Gbps vs. 2.5Gbps), can be coordinated so to permit the equipment with 1.25 TS granularity to adapt his TS allocation accordingly to the different TS granularity (2.5Gbps) of a neighbour.

Therefore, in order to let the NE change TS granularity accordingly to the nieghbour requirements, the AUTOpayloadtype needs to be set. When both the neighbors (link or trail) have been configured as structured, the payload type received in the overhead is compared to the transmitted PT. If they are different and the transmitted PT=21, the node must fallback to PT=20. In this case the fall-back process makes the system self consistent and the only reason for signaling the TS granularity is to provide the correct label (i.e. label for PT=21 has twice the TS number of PT=20). On the other side, if the AUTOpayloadtype is not configured, the RSVP-TE consequent actions in case of TS mismatch need to be defined.

4.2. Tributary Port Number

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

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

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

4.3. Signal type

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

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

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

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

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

MAX LSP Bw: ODU1

MIN LSP Bw: ODU1

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

- Unreserved Bandwidth (of the bundle) set to ODU2

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

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

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

On the other hand, from a singaling point of view, [RFC4328] describes GMPLS signaling extensions to support the control for G.709 OTNs [G709-V1]. However,[RFC4328] needs to be updated because it does not provide the means to signal all the new signal types and related mapping and multiplexing functionalities.

4.4. Bit rate and tolerance

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

4.5. Unreserved Resources

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

```
+------+ component link 1 +------+
|      +------------------+      |
|      | component link 2 |      |
|  N1  +------------------+  N2  |
|      | component link 3 |      |
|      +------------------+      |
+------+                  +---+--+
```

Figure 6: Concurrent path computation

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

4.6. Maximum LSP Bandwidth

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

4.7. Distinction between terminating and switching capability

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

```
             MATRIX                   LINE INTERFACE
       +-----------------+          +-----------------+
       |    +-------+    |   ODU2   |                 |
      ----->| ODU-2 |----|----------|--------\        |
       |    +-------+    |          |      +----+     |
       |                 |          |       \__/      |
       |                 |          |        \/       |
       |    +-------+    |   ODU3   |         | ODU3  |
      ----->| ODU-3 |----|----------|------\  |       |
       |    +-------+    |          |       \ |       |
       |                 |          |        \|       |
       |                 |          |      +----+     |
       |                 |          |       \__/      |
       |                 |          |        \/       |
       |                 |          |         ---------> OTU-3
       +-----------------+          +-----------------+
```

Figure 7: Switching and Terminating capabilities

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

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

- Map an ODU3 coming from the switching matrix into an OTU3

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

This piece of information needs to be advertised together with the related unreserved bandwidth and signal type. As a consequence signaling must have the possibility to setup an LSP allowing the local selection of resources consistent with the limitations considered during the path computation.

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

```
               ODU0 transparently transported
       +++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
       |           ODU2 LSP Carrying ODU0 service                  |
       |       |''''''''''''''''''''''''''''''''''''''''''''|       |
       |       |                                            |       |
       |  +-----++  OTU2   +------+   OTU2  +------+  OTU2   ++-----+  |
     ODU0 |      |  Link   |      |   Link  |      |  Link   |      | ODU0
     ---->|  A   |_________|  B   |_________|  C   |_________|  D   |---->
          |      |         |      |         |      |         |      |
          +------+         +--+---+         +------+         ++--+-+
                              |                               |  |
                          OTU3|                               |  |
                          Link|    +------+___________________|  |
                              |    |      |    OTU3 Link         |
                              |____|  E   |                      |
                                   |      |______________________|
                                   +------+    OTU2 Link
```

Figure 8: Switching and Terminating capabilities - Example 1

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

```
               ODU0 transparently transported
       ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
       |           ODU2 LSP Carrying ODU0 service                 |
       |       |''''''''''''''''''''''''''''''''''''''''''|       |
       |       |                                          |       |
       |  +----++  OTU2   +-----+   OTU2 +-----+   OTU2  ++----+  |
     ODU0 |     |  Link   |     |   Link |     |   Link  |     | ODU0
     ---->|  A  |_________|  B  |________|  C  |_________|  D  |---->
          |     |         |     |        |     |         |     |
          +-----+         ++-+-++        +-----+         +--+--+
                           | | |                            |
                       OTU2| | |                            |
             +-----+   Link| | |   OTU3   +-----+           |
             |     |       | | |   Link   |     |           |
             |  F  |_______| | |__________|  E  |___________|
             |     |         |____________|     | OTU2 Link
             +-----+            OTU2 Link +-----+
```

Figure 9: Switching and Terminating capabilities - Example 2

4.8. Priority Support

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

4.9. Multi-stage multiplexing

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

With respect to the routing, please note that in case of multi stage muxing hierarchy (e.g. ODU1->ODU2->ODU3), not only the ODUk/OTUk bandwidth (ODU3) and service layer bandwidth (ODU1) are needed, but also the intermediate one (ODU2). This is a typical case of spatial allocation problem.

Suppose in this scenario to have the following advertisement:

Hierarchy: ODU1->ODU2->ODU3

Number of ODU1==5

The number of ODU1 suggests that it is possible to have an ODU2 FA, but it depends on the spatial allocation of such ODU1s.

It is possible that 2 links are bundled together and 3 ODU1->ODU2->ODU3 are available on a component link and 2 on the other one, in such a case no ODU2 FA could be set up. The advertisement of the ODU2 is needed because in case of ODU1 spatial allocation (3+2), the ODU2 available bandwidth would be 0 (no ODU2 FA can be created), while in case of ODU1 spatial allocation (4+1) the ODU2 available bandwidth would be 1 (1 ODU2 FA can be created).

4.10. Generalized Label

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

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

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

5. Security Considerations

TBD

6. IANA Considerations

TBD

7. Contributors

Jonathan Sadler, Tellabs

EMail: jonathan.sadler@tellabs.com

John Drake, Juniper

EMail: jdrake@juniper.net

8. Acknowledgements

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

9. References

9.1. Normative References

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

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

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

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

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

[RFC4203] Kompella, K. and Y. Rekhter, "OSPF Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4203, October 2005.

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

[RFC5250] Berger, L., Bryskin, I., Zinin, A., and R. Coltun, "The OSPF Opaque LSA Option", RFC 5250, July 2008.

[RFC5339] Le Roux, JL. and D. Papadimitriou, "Evaluation of Existing GMPLS Protocols against Multi-Layer and Multi-Region Networks (MLN/MRN)", RFC 5339, September 2008.

9.2. Informative References

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

[G.709-v2] ITU-T, "Interface for the Optical Transport Network (OTN)", G.709 Recommendation (and Amendment 1), March 2003.

[G.709-v3] ITU-T, "Rec G.709, version 3", approved by ITU-T on December 2009.

[G.872-am2] ITU-T, "Amendment 2 of G.872 Architecture of optical transport networks for consent", consented by ITU-T on June 2010.

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

Authors' Addresses

Sergio Belotti (editor)
Alcatel-Lucent
Via Trento, 30
Vimercate
Italy

Email: sergio.belotti@alcatel-lucent.com

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

Email: pietro_vittorio.grandi@alcatel-lucent.com

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

Email: daniele.ceccarelli@ericsson.com

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

Email: diego.caviglia@ericsson.com

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

Email: zhangfatai@huawei.com

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

Email: danli@huawei.com

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: January 7, 2012
A. Takacs
B. Gero
Ericsson
H. Long
Huawei
July 11, 2011

GMPLS RSVP-TE Extensions for Ethernet OAM Configuration
draft-ietf-ccamp-rsvp-te-eth-oam-ext-06

Abstract

The GMPLS controlled Ethernet Label Switching (GELS) work extended GMPLS RSVP-TE to support the establishment of Ethernet LSPs. IEEE Ethernet Connectivity Fault Management (CFM) specifies an adjunct OAM flow to check connectivity in Ethernet networks. CFM can be also used with Ethernet LSPs for fault detection and triggering recovery mechanisms. The ITU-T Y.1731 specification builds on CFM and specifies additional OAM mechanisms, including Performance Monitoring, for Ethernet networks. This document specifies extensions of GMPLS RSVP-TE to support the setup of the associated Ethernet OAM (CFM and Y.1731) entities defining Ethernet technology specific TLV based on [OAM-CONF-FWK].

Requirements Language

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

Status of this Memo

This Internet-Draft will expire on January 7, 2012.

Copyright Notice

Table of Contents

1. Background

Provider Backbone Bridging - Traffic Engineering (PBB-TE) [IEEE-PBBTE] decouples the Ethernet data and control planes by explicitly supporting external control/management mechanisms to configure static filtering entries in bridges and create explicitly routed Ethernet connections. In addition PBB-TE defines mechanisms for protection switching of bidirectional Ethernet connections. Ethernet Connectivity Fault Management (CFM) defines an adjunct connectivity monitoring OAM flow to check the liveliness of Ethernet networks [IEEE-CFM], including the monitoring of explicitly-routed Ethernet connections.

In IETF the GMPLS controlled Ethernet Label Switching (GELS) work extended the GMPLS control plane to support the establishment of explicitly routed Ethernet connections [RFC5828][RFC6060]. We refer to GMPLS established Ethernet connections as Ethernet LSPs. GELS enables the application of MPLS-TE and GMPLS provisioning and recovery features in Ethernet networks.

2. Overview of Ethernet OAM operation

For the purposes of this document, we only discuss Ethernet OAM [IEEE-CFM] aspects that are relevant for the connectivity monitoring of Ethernet LSPs.

PBB-TE [IEEE-PBBTE] defines point-to-point Ethernet Switched Paths (ESPs) as a provisioned traffic engineered unidirectional connectivity, identified by the 3-tuple [ESP-MAC DA, ESP-MAC SA, ESP-VID] where the ESP-MAC DA is the destination address of the ESP, the ESP-MAC SA is the source address of the ESP, and the ESP-VID is a VLAN identifier allocated for explicitly routed connections. To form a bidirectional PBB-TE connection two co-routed point-to-point ESPs are combined. The combined ESPs must have the same ESP-MAC addresses but may have different ESP-VIDs.

Note that although it would be possible to use GMPLS to setup a single unidirectional ESP, the Ethernet OAM mechanisms are only full functional when bidirectional connections are established with co-routed ESPs. Hence, we focus on bidirectional point-to-point PBB-TE connections only.

At both ends of the bidirectional point-to-point PBB-TE connection one Maintenance Endpoint (MEP) is configured. The MEPs monitoring a PBB-TE connection must be configured with the same Maintenance Domain Level (MD Level) and Maintenance Association Identifier (MAID). Each MEP has a unique identifier, the MEP ID. Besides these identifiers a MEP monitoring a PBB-TE connection must be provisioned with the 3-tuples [ESP-MAC DA, ESP-MAC SA, ESP-VID] of the two ESPs.

In the case of point-to-point VLAN connections, the connection is identified with a single VLAN forwarding traffic in both directions or with two VLANs each forwarding traffic in a single direction. Hence instead of the 3-tuples of the PBB-TE case MEPs must be provisioned with the proper VLAN information, otherwise the same MD Level, MAID, MEP ID configuration is required in this case as well.

MEPs exchange Connectivity Check Messages (CCMs) periodically with fixed intervals. Eight distinct intervals are defined in [IEEE-CFM]:

#	CCM Interval (CCI)	3 bit encoding
0	Reserved	000
1	3 1/3 ms	001
2	10 ms	010
3	100 ms	011
4	1 s	100
5	10 s	101
6	1 min	110
7	10 min	111

Table 1: CCM Interval encoding

If 3 consecutive CCM messages are not received by one of the MEPs it declares a connectivity failure and signals the failure in subsequent CCM messages, by setting the Remote Defect Indicator (RDI) bit, to the remote MEP. If a MEP receives a CCM message with RDI set it immediately declares failure. The detection of a failure may trigger protection switching mechanisms or may be signaled to a management system. However, what happens once a failure is detected is out of the scope of this document.

At each transit node Maintenance Intermediate Points (MIPs) can be established to help failure localization by supporting link trace and loop back functions. MIPs need to be provisioned with a subset of MEP identification parameters described above.

3. GMPLS RSVP-TE Extensions

3.1. Operation overview

To simplify the configuration of connectivity monitoring, when an Ethernet LSP is signaled the associated MEPs should be automatically established. To monitor an Ethernet LSP a set of parameters must be provided to setup a Maintenance Association and related MEPs. Optionally, MIPs may be created at the transit nodes of the Ethernet LSP. The LSP Attributes Flags: "OAM MEP entities desired" and "OAM MIP entities desired", described in [OAM-CONF-FWK] are used to signal that the respective OAM entities must be established. Subsequently, an OAM Configuration TLV is added to the LSP_ATTRIBUTES Object specifying that Ethernet OAM is to be setup for the LSP. The below detailed Ethernet OAM specific information is carried in the new Ethernet OAM Configuration sub-TLV.

- A unique MAID must be allocated for the PBB-TE connection and both MEPs must be configured with the same information. The MAID consists of an optional Maintenance Domain Name (MD Name) and a mandatory Short Maintenance Association Name (Short MA Name). Various formatting rules for these names have been defined by [IEEE-CFM]. Since these information is also carried in all CCM messages, the combined length of the Names is limited to 44 bytes. How these parameters are determined is out of scope of this document.

- Each MEP must be provisioned with a MEP ID. The MEP ID uniquely identifies a given MEP within a Maintenance Association. That is, the combination of MAID and MEP ID must uniquely identify a MEP. How the value of the MEP ID is determined is out of scope of this document.

- The Maintenance Domain Level (MD Level) allows hierarchical separation of monitoring entities. [IEEE-CFM] allows differentiation of 8 levels. How the value of the MD Level is determined is out of scope of this document. Note that most probably for all Ethernet LSPs a single (default) MD Level will be used within a network domain.

- The desired CCM Interval must be specified by the management system based on service requirements or operator policy. The same CCM Interval must be set in each of the MEPs monitoring a given Ethernet LSP. How the value of the CCM Interval is determined is out of scope of this document.

- The desired CCM priority to be set by MEPs for the CCM frames can be specified. The same CCM priority must be set in each of the

MEPs monitoring a given Ethernet LSP. How CCM priority is determined is out of scope of this document. Note that the highest priority is used as the default CCM priority.

- MEPs must be aware of their own and the reachability parameters of the remote MEP. In the case of bidirectional point-to-point PBB-TE connections this requires that the 3-tuples [ESP-MAC A, ESP-MAC B, ESP-VID1] and [ESP-MAC B, ESP-MAC A, ESP-VID2] are configured in each MEP, where the ESP-MAC A is the same as the local MEP's MAC address and ESP-MAC B is the same as remote MEP's MAC address. The GMPLS Ethernet Label for forwarding, as defined in [RFC6060], consists of the ESP-MAC DA and ESP-VID. Hence the necessary reachability parameters for the MEPs can be obtained from Ethernet Labels (i.e., carried in the "downstream" and upstream labels). In the case of point-to-point VLAN connections, MEPs need to be provisioned with the VLAN identifiers, which can be derived similarly from the Ethernet Label.

Assuming the procedures described in [RFC6060] for bidirectional PBB-TE Ethernet LSP establishment the MEP configuration should be as follows. When the RSVP-TE signaling is initiated for the bidirectional Ethernet LSP the local node generates a Path message and:

- Allocates an Upstream Label from its MAC address (ESP-MAC A) and locally selected VID (ESP-VID1), which will be used to receive traffic;

- Inserts the OAM Configuration TLV with OAM Type set to Ethernet OAM in the LSP_ATTRIBUTES object;

- Adds the OAM Function Flags sub-TLV in the OAM Configuration TLV and sets the OAM function flags as needed;

- Adds an Ethernet OAM Configuration sub-TLV in the OAM Configuration TLV that specifies the CCM Interval and MD Level;

- Adds an MD Name Sub-TLV (optional) and a Short MA Name Sub-TLV to the Ethernet OAM Configuration TLV, that will unambiguously identify a Maintenance Association for this specific PBB-TE connection. Note that values for these parameters may be derived from the GMPLS LSP identification parameters;

- Adds a MEP ID Sub-TLV to the Ethernet OAM Configuration TLV. It selects two distinct integer values to identify the local and remote MEPs within the Maintenance Association created for monitoring of the point-to-point PBB-TE connection.

Once the remote node receives the Path message it can use the UPSTREAM_LABEL to extract the reachability information of the initiator. Then it allocates a Label by selecting the MAC address (ESP-MAC B) and VID (ESP-VID2) it would like to use to receive traffic. These parameters determine the reachability information of the local MEP. That is, the 3-tuples [ESP-MAC A, ESP-MAC B, ESP-VID1] and [ESP-MAC B, ESP-MAC A, ESP-VID2] are derived from the Ethernet Labels. In addition the information received in the Ethernet OAM Configuration TLV is used to configure the local MEP.

Once the Resv message successfully arrives to the initiator it can extract the remote side's reachability information from the Label Object whereby this node has also obtained all the information needed to establish its local MEP.

3.2. OAM Configuration TLV

This TLV is specified in [OAM-CONF-FWK] and is used to select which OAM technology/method should be used for the LSP. In this document a new OAM Type: Ethernet OAM is defined.

OAM Type	Description
0	Reserved
1	Ethernet OAM
2-256	Reserved

The receiving node when the Ethernet OAM Type is requested should look for the corresponding technology specific Ethernet OAM Configuration TLV.

3.3. Ethernet OAM Configuration TLV

The Ethernet OAM Configuration TLV (depicted below) is defined for Ethernet OAM specific configuration parameters. The Ethernet OAM Configuration TLV is carried within the OAM Configuration TLV in the LSP_ATTRIBUTES or LSP_REQUIRED_ATTRUIBUTES object in Path messages. This new TLV accommodates generic Ethernet OAM information and carries sub-TLVs.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           Type  (IANA)        |           Length              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Version |MD L.|           Reserved (set to all 0s)            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
~                           sub TLVs                            ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Type: indicates a new type: the Ethernet OAM Configuration TLV. IANA is requested to assigne a value from the "OAM Type sub-TLV" space in the "RSVP-TE OAM Configuration Registry".

Length: indicates the total length including sub-TLVs.

Version: identifies the CFM protocol version according to [IEEE-CFM]. If a node does not support a specific CFM version an error must be generated: "OAM Problem/Unsupported OAM Version"

MD L. (MD Level): indicates the desired MD Level. The values are according to [IEEE-CFM]. If a node does not support a specific MD Level an error must be generated: "OAM Problem/Unsupported OAM Level".

3.3.1. MD Name Sub-TLV

The optional MD Name sub-TLV is depicted below.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           Type (1)            |           Length              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    Format     |  Name Length  |   Reserved (set to all 0s)    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
~                            MD Name                            ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Type: 1, MD Name Sub-TLV.

Length: indicates the total length of the TLV including padding.

Format: according to [IEEE-CFM].

Name Length: the length of the MD Name field in bytes. This is necessary to allow non 4 byte padded MD Name lengths.

MD Name: variable length field, formatted according to the format specified in the Format field.

If an undefined Format is specified an error must be generated: "OAM Problem/Unknown MD Name Format". Also the combined length of MD Name and Short MA Name must be less or equal to 44bytes, if this is violated an error must be generated: "OAM Problem/Name Length Problem". Note that it is allowed to have no MD Name, as such the MD Name sub-TLV is optional. In this case the MA Name must uniquely identify a Maintenance Association.

3.3.2. Short MA Name Sub-TLV

The Short MA Name sub-TLV is depicted below.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           Type (2)            |           Length              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    Format     |  Name Length  |   Reserved (set to all 0s)    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
~                        Short MA Name                          ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Type: 2, Short MA Name Sub-TLV.

Length: indicates the total length of the TLV including padding.

Format: according to [IEEE-CFM].

Name Length: the length of the MA Name field in bytes. This is necessary to allow non 4 byte padded MA Name lengths.

Short MA Name: variable length field formatted according to the format specified in the Format field.

If an undefined Format is specified an error must be generated: "OAM Problem/Unknown MA Name Format". Also the combined length of MD Name and Short MA Name must be less or equal to 44bytes, if this is violated an error must be generated: "OAM Problem/Name Length Problem". Note that it is allowed to have no MD Name, in this case the MA Name must uniquely identify a Maintenance Association.

3.3.3. MEP ID Sub-TLV

The MEP ID Sub-TLV is depicted below.

```
    0                   1                   2                   3
    0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |           Type (3)            |           Length (4)          |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |        Local MEP ID           |T|R|      Reserved             |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |        Remote MEP ID          |T|R|      Reserved             |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Type: 3, MEP ID Sub-TLV.

Length: indicates the total length of the TLV including padding.

Local MEP ID: a 16 bit integer value in the range 1-8191 of the MEP ID on the initiator side.

Remote MEP ID: a 16 bit integer value in the range 1-8191 of the MEP ID to be set for the MEP established at the receiving side. This value is determined by the initiator node. This is possible, since a new MAID is assigned to each PBB-TE connection, and MEP IDs must be only unique within the scope of the MAID.

Two flags are defined Transmit (T) and Receive (R). When T is set the corresponding MEP must send OAM packets. When R is set the corresponding MEP must expect to receive OAM packets. These flags are used to configure the role of MEPs.

3.3.4. Continuity Check (CC) Sub-TLV

The Continuity Check (CC) sub-TLV is depicted below.

```
    0                   1                   2                   3
    0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |           Type (4)            |            Length             |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   | Prio  | CCM I |           Reserved (set to all 0s)            |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Type: 4, Continuity Check (CC) sub-TLV.

Prio: Indicates the priority to be set for CCM frames. In Ethernet 3 bits carried in VLAN TAGs identify priority information.

CCM I (CCM Interval): CCM Interval, according to the 3 bit encoding [IEEE-CFM] shown in Table 1. If a node does not support the requested CCM Interval an error must be generated: "OAM Problem/ Unsupported CC Interval".

3.4. Pro-active Performance Monitoring

Ethernet OAM functions for Performance Monitoring (PM) allow measurements of different performance parameters including Frame Loss Ratio, Frame Delay and Frame Delay variation as defined in the ITU-T Y.1731 recommendation. Only a subset of PM functions are operated in a pro-active fashion to monitor the performance of the connection continuously. Pro-active PM supports Fault Management functions, by providing an indication of decreased service performance and as such may provide triggers to initiate recovery procedures.

While on demand PM functions are always initiated by management commands, for pro-active PM it may be desirable to utilize the control plane for configuration and activation together with Fault Management functions such as Continuity Check.

ITU-T Y.1731 defines dual-ended Loss Measurement as pro-active OAM for performance monitoring and as a PM function applicable to fault management. For dual-ended Loss Measurement each MEP piggy-backs transmitted and received frame counters on CC messages; to support and synchronize bidirectional Loss Measurements at the MEPs. Dual-ended Loss Measurement is invoked by setting the Performance Monitoring/Loss OAM Function Flag in the OAM Function Flags Sub-TLV [OAM-CONF-FWK]. Besides configuring the Continuity Check functionality, no additional configuration is required for this type of Loss Measurement.

3.5. Ethernet OAM configuration errors

In addition to error values specified in [OAM-CONF-FWK] this document defines the following values for the "OAM Problem" Error Code.

- o If a node does not support a specific CFM version an error must be generated: "OAM Problem/Unsupported OAM Version".

- o If a node does not support a specific MD Level an error must be generated: "OAM Problem/Unsupported OAM Level".

- o If an undefined MD name format is specified an error must be generated: "OAM Problem/Unknown MD Name Format".

- If an undefined MA name format is specified an error must be generated: "OAM Problem/Unknown MA Name Format".

- If the combined length of MD Name and Short MA Name must be less or equal to 44bytes, if this is violated an error must be generated: "OAM Problem/Name Length Problem".

- If a node does not support the requested CCM Interval an error must be generated: "OAM Problem/Unsupported CC Interval".

4. IANA Considerations

This document specifies the Ethernet OAM Configuration sub-TLV to be carried in the OAM Configuration TLV in LSP_ATTRIBUTES and LSP_REQUIRED_ATTRIBUTES objects in Path messages.

IANA is requested to allocate the value 1 for Ethernet OAM from the OAM Type space in the "RSVP-TE OAM Configuration Registry" and allocate type 1 for the Ethernet OAM Configuration sub-TLV from the OAM Type sub-TLV space in the "RSVP-TE OAM Configuration Registry".

The following values need to be assigned under the Error Code: "OAM Problem": "Unsupported OAM Version", "Unsupported OAM Level", "Unknown MD Name Format", "Unknown MA Name Format", "Name Length Problem", "Unsupported CC Interval".

5. Security Considerations

This document does not introduce any additional security issuse to those discussed in [OAM-CONF-FWK].

6. Acknowledgements

The authors would like to thank Francesco Fondelli, Adrian Farrel, Loa Andersson, Eric Gray and Dimitri Papadimitriou for their useful comments.

Contributors

Don Fedyk
Alcatel-Lucent
Groton, MA 01450
USA
Email: donald.fedyk@alcatel-lucent.com

Dinesh Mohan

7. References

7.1. Normative References

[OAM-CONF-FWK] "OAM Configuration Framework for GMPLS RSVP-TE", Internet Draft, work in progress.

[RFC5828] "GMPLS Ethernet Label Switching Architecture and Framework", RFC 5828, March 2010.

[RFC6060] "Generalized Multiprotocol Label Switching (GMPLS) Control of Ethernet Provider Backbone Traffic Engineering (PBB-TE)", RFC 6060.

7.2. Informative References

[IEEE-CFM] "IEEE 802.1ag, Draft Standard for Connectivity Fault Management", work in progress.

[IEEE-PBBTE] "IEEE 802.1Qay Draft Standard for Provider Backbone Bridging Traffic Engineering", work in progress.

Authors' Addresses

Attila Takacs
Ericsson
Laborc u. 1.
Budapest,   1037
Hungary

Email: attila.takacs@ericsson.com

Balazs Gero
Ericsson
Laborc u. 1.
Budapest,   1037
Hungary

Email: balazs.gero@ericsson.com

Hao Long
Huawei

Email: lonho@huawei.com

CCAMP Working Group E. Bellagamba, Ed.
Internet-Draft L. Andersson, Ed.
Intended status: Standards Track Ericsson
Expires: January 12, 2012 P. Skoldstrom, Ed.
Acreo AB
D. Ward
Juniper
A. Takacs
Ericsson
July 11, 2011

Configuration of Pro-Active Operations, Administration, and Maintenance (OAM) Functions for MPLS-based Transport Networks using RSVP-TE
draft-ietf-ccamp-rsvp-te-mpls-tp-oam-ext-06

Abstract

This specification describes the configuration of pro-active MPLS-TP Operations, Administration, and Maintenance (OAM) Functions for a given LSP using a set of TLVs that are carried by the RSVP-TE protocol.

This document is a product of a joint Internet Engineering Task Force (IETF) / International Telecommunication Union Telecommunication Standardization Sector (ITU-T) effort to include an MPLS Transport Profile within the IETF MPLS and PWE3 architectures to support the capabilities and functionalities of a packet transport network.

Status of this Memo

Copyright Notice

Table of Contents

1. Introduction

This document describes the configuration of pro-active MPLS-TP Operations, Administration, and Maintenance (OAM) Functions for a given LSP using TLVs carried by RSVP-TE [RFC3209]. In particular it specifies the mechanisms necessary to establish MPLS-TP OAM entities for monitoring and performing measurements on an LSP, as well as defining information elements and procedures to configure pro-active MPLS OAM functions. Initialization and control of on-demand MPLS OAM functions are expected to be carried out by directly accessing network nodes via a management interface; hence configuration and control of on-demand OAM functions are out-of-scope for this document.

The Transport Profile of MPLS must, by definition [RFC5654], be capable of operating without a control plane. Therefore there are three options for configuring MPLS-TP OAM, without a control plane by either using an NMS or LSP Ping, or with a control plane using GMPLS (specifically RSVP-TE) .

Pro-active MPLS OAM is performed by three different protocols, Bidirectional Forwarding Detection (BFD) [RFC5880] for Continuity Check/Connectivity Verification, the delay measurement protocol (DM) [MPLS-PM] for delay and delay variation (jitter) measurements, and the loss measurement protocol (LM) [MPLS-PM] for packet loss and throughput measurements. Additionally there is a number of Fault Management Signals that can be configured.

BFD is a protocol that provides low-overhead, fast detection of failures in the path between two forwarding engines, including the interfaces, data link(s), and to the extent possible the forwarding engines themselves. BFD can be used to track the liveliness and detect data plane failures of MPLS-TP point-to-point and might also be extended to support point-to-multipoint connections.

The delay and loss measurements protocols [MPLS-PM] use a simple query/response model for performing bidirectional measurements that allows the originating node to measure packet loss and delay in both directions. By timestamping and/or writing current packet counters to the measurement packets at four times (Tx and Rx in both directions) current delays and packet losses can be calculated. By performing successive delay measurements the delay variation (jitter) can be calculated. Current throughput can be calculated from the packet loss measurements by dividing the number of packets sent/ received with the time it took to perform the measurement, given by the timestamp in LM header. Combined with a packet generator the throughput measurement can be used to measure the maximum capacity of a particular LSP.

MPLS Transport Profile (MPLS-TP) describes a profile of MPLS that enables operational models typical in transport networks, while providing additional OAM, survivability and other maintenance functions not currently supported by MPLS. [RFC5860] defines the requirements for the OAM functionality of MPLS-TP.

This document is a product of a joint Internet Engineering Task Force (IETF) / International Telecommunication Union Telecommunication Standardization Sector (ITU-T) effort to include an MPLS Transport Profile within the IETF MPLS and PWE3 architectures to support the capabilities and functionalities of a packet transport network.

1.1. Contributing Authors

This document is the result of a large team of authors and contributors. The following is a list of the co-authors:

John Drake

Benoit Tremblay

1.2. Requirements Language

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

2. Overview of MPLS OAM for Transport Applications

[MPLS-TP-OAM-FWK] describes how MPLS OAM mechanisms are operated to meet transport requirements outlined in [RFC5860].

[BFD-CCCV] specifies two BFD operation modes: 1) "CC mode", which uses periodic BFD message exchanges with symmetric timer settings, supporting Continuity Check, 2) "CV/CC mode" which sends unique maintenance entity identifiers in the periodic BFD messages supporting Connectivity Verification as well as Continuity Check.

[MPLS-PM] specifies mechanisms for performance monitoring of LSPs, in particular it specifies loss and delay measurement OAM functions.

[MPLS-FMS] specifies fault management signals with which a server LSP can notify client LSPs about various fault conditions to suppress alarms or to be used as triggers for actions in the client LSPs. The following signals are defined: Alarm Indication Signal (AIS), Link Down Indication (LDI) and Locked Report (LKR). To indicate client faults associated with the attachment circuits Client Signal Failure

Indication (CSF) can be used. CSF is described in [MPLS-TP-OAM-FWK] and in the context of this document is for further study.

[MPLS-TP-OAM-FWK] describes the mapping of fault conditions to consequent actions. Some of these mappings may be configured by the operator, depending on the application of the LSP. The following defects are identified: Loss Of Continuity (LOC), Misconnectivity, MEP Misconfiguration and Period Misconfiguration. Out of these defect conditions, the following consequent actions may be configurable: 1) whether or not the LOC defect should result in blocking the outgoing data traffic; 2) whether or not the "Period Misconfiguration defect" should result in a signal fail condition.

3. Theory of Operations

3.1. MPLS OAM Configuration Operation Overview

RSVP-TE, or alternatively LSP Ping [LSP-PING CONF], can be used to simply enable the different OAM functions, by setting the corresponding flags in the "OAM Functions TLV". Additionally one may include sub-TLVs for the different OAM functions in order to specify different parameters in detail.

3.1.1. Configuration of BFD sessions

For this specification, BFD MUST be run in either one of the two modes:

- Asynchronous mode, where both sides should be in active mode.

- Unidirectional mode

In the simplest scenario LSP Ping, or alternatively RSVP-TE [RSVP-TE CONF], is used only to bootstrap a BFD session for an LSP, without any timer negotiation.

Timer negotiation can be performed either in subsequent BFD control messages (in this case the operation is similar to LSP Ping based bootstrapping described in [RFC5884]) or directly in the LSP ping configuration messages.

When BFD Control packets are transported in the G-ACh they are not protected by any end-to-end checksum, only lower-layers are providing error detection/correction. A single bit error, e.g. a flipped bit in the BFD State field could cause the receiving end to wrongly conclude that the link is down and in turn trigger protection switching. To prevent this from happening the "BFD Configuration sub-TLV" has an Integrity flag that when set enables BFD Authentication using Keyed SHA1 with an empty key (all 0s) [RFC5880]. This would make every BFD Control packet carry an SHA1 hash of itself that can be used to detect errors.

If BFD Authentication using a pre-shared key / password is desired (i.e. authentication and not only error detection) the "BFD Authentication sub-TLV" MUST be included in the "BFD Configuration sub-TLV". The "BFD Authentication sub-TLV" is used to specify which authentication method that should be used and which pre-shared key / password that should be used for this particular session. How the key exchange is performed is out of scope of this document.

3.1.2. Configuration of Performance Monitoring

It is possible to configure Performance Monitoring functionalities such as Loss, Delay and Throuhput as described in [MPLS-PM].

When configuring Performance monitoring functionalities it can be choosen either the default configuration (by only setting the respective flags in the "OAM functions TLV") or a customized configuration (by including the respective Loss and/or Delay sub-TLVs).

3.1.3. Configuration of Measurements and FMS

Additional OAM functions may be configured by setting the appropriate flags in the "OAM Functions TLV", these include Performance Measurements (packet loss, throughput, delay, and delay variation) and Fault Management Signal handling.

By setting the PM Loss flag in the "OAM Functions TLV" and including the "MPLS OAM PM Loss sub-TLV" one can configure the measurement interval and loss threshold values for triggering protection.

Delay measurements are configured by setting PM Delay flag in the "OAM Functions TLV" and including the "MPLS OAM PM Loss sub-TLV" one can configure the measurement interval and the delay threshold values for triggering protection.

To configure Fault Monitoring Signals and their refresh time the FMS flag in the "OAM Functions TLV" MUST be set and the "MPLS OAM FMS sub-TLV" included.

3.2. OAM Configuration TLV

The "OAM Configuration TLV" is depicted in the following figure. It specifies the OAM functions that are to be used for the LSP and it is defined in [OAM-CONF-FWK]. The "OAM Configuration TLV" is carried in the LSP_ATTRIBUTES object in Path and Resv messages.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           Type (2) (IANA)     |           Length              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    OAM Type   |                Reserved                       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
~                           sub-TLVs                            ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Type: indicates the "OAM Configuration TLV" (2) (IANA to assign).

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

This document defines a new OAM Type: "MPLS OAM" (suggested value 2, IANA to assign) from the "RSVP-TE OAM Configuration Registry". The "MPLS OAM" type is set to request the establishment of OAM functions for MPLS-TP LSPs. The specific OAM functions are specified in the "Function Flags" sub-TLV as depicted in [OAM-CONF-FWK].

The receiving edge LSR when the MPLS-TP OAM Type is requested should check which OAM Function Flags are set in the "Function Flags TLV" (also defined in [OAM-CONF-FWK]) and look for the corresponding technology specific configuration TLVs.

Additional corresponding sub-TLVs are as follows:

- "BFD Configuration sub-TLV", which MUST be included if the CC and/or the CV OAM Function flag is set. This sub-TLV MUST carry a "BFD Local Discriminator sub-TLV" and a "Timer Negotiation Parameters sub-TLV" if the N flag is cleared. If the I flag is set, the "BFD Authentication sub-TLV" may be included.

- "MPLS OAM PM Loss sub-TLV" within the "Performance Monitoring sub-TLV", which MAY be included if the PM/Loss OAM Function flag is set. If the "MPLS OAM PM Loss sub-TLV" is not included, default configuration values are used. Such sub-TLV MAY also be included in case the Throughput function flag is set and there is the need to specify measurement interval different from the default ones. In fact the throughput measurement make use of the same tool as the loss measurement, hence the same TLV is used.

- "MPLS OAM PM Delay sub-TLV" within the "Performance Monitoring sub-TLV", which MAY be included if the PM/Delay OAM Function flag is set. If the "MPLS OAM PM Delay sub-TLV" is not included, default configuration values are used.

- "MPLS OAM FMS sub-TLV", which MAY be included if the FMS OAM Function flag is set. If the "MPLS OAM FMS sub-TLV" is not included, default configuration values are used.

Moreover, if the CV or CC flag is set, the CC flag MUST be set at the same time. The format of an MPLS-TP CV/CC message is shown in [BFD-CCCV] and it requires, together with the BFD control packet information, the "Unique MEP-ID of source of BFD packet". [MPLS-TP-IDENTIF] defines the composition of such identifier as:

<"Unique MEP-ID of source of BFD packet"> ::=
<src_node_id><src_tunnel_num><lsp_num>

GMPLS signaling [RFC3473] uses a 5-tuple to uniquely identify an LSP within an operator’s network. This tuple is composed of a Tunnel Endpoint Address, Tunnel_ID, Extended Tunnel ID, and Tunnel Sender Address and (GMPLS) LSP_ID.

Hence, the following mapping is used without the need of redefining a new TLV for MPLS-TP proactive CV purpose.

- Tunnel ID = src_tunnel_num

- Tunnel Sender Address = src_node_id

- LSP ID = LSP_Num

"Tunnel ID" and "Tunnel Sender Address" are included in the "SESSION" object [RFC3209], which is mandatory in both Path and Resv messages.

"LSP ID" will be the same on both directions and it is included in the "SENDER_TEMPLATE" object [RFC3209] which is mandatory in Path messages.

[Author’s note: the same "Unique MEP-ID of source" will be likely required for Performance monitoring purposes. This need to be agreed with [MPLS-PM] authors.]

3.3. BFD Configuration sub-TLV

The "BFD Configuration sub-TLV" (depicted below) is defined for BFD OAM specific configuration parameters. The "BFD Configuration sub-TLV" is carried as a sub-TLV of the "OAM Configuration TLV".

This TLV accommodates generic BFD OAM information and carries sub-TLVs.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           Type (3) (IANA)     |           Length              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|Vers.| PHB |N|S|I|    Reserved (set to all 0s)                 |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
~                           sub-TLVs                            ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Type: indicates a new type, the "BFD Configuration sub-TLV" (IANA to define).

Length: indicates the total length including sub-TLVs.

Version: identifies the BFD protocol version. If a node does not support a specific BFD version an error must be generated: "OAM Problem/Unsupported OAM Version".

PHB: Identifies the Per-Hop Behavior (PHB) to be used for periodic continuity monitoring messages.

BFD Negotiation (N): If set timer negotiation/re-negotiation via BFD Control Messages is enabled, when cleared it is disabled.

Symmetric session (S): If set the BFD session MUST use symmetric timing values.

Integrity (I): If set BFD Authentication MUST be enabled. If the "BFD Configuration sub-TLV" does not include a "BFD Authentication sub-TLV" the authentication MUST use Keyed SHA1 with an empty pre-shared key (all 0s).

Encapsulation Capability (G): if set, it shows the capability of encapsulating BFD messages into G-Ach channel. If both the G bit and U bit are set, configuration gives precedence to the G bit.

Encapsulation Capability (U): if set, it shows the capability of encapsulating BFD messages into UDP packets. If both the G bit and U bit are set, configuration gives precedence to the G bit.

Bidirectional (B): if set, it configures BFD in the Bidirectional mode. If it is not set it configures BFD in unidirectional mode. In the second case, the source node does not expect any Discriminator values back from the destination node.

The "BFD Configuration sub-TLV" MUST include the following sub-TLVs in the Path message:

- "Local Discriminator sub-TLV";

- "Negotiation Timer Parameters sub-TLV" if the N flag is cleared.

The "BFD Configuration sub-TLV" MUST include the following sub-TLVs in the Resv message:

- "Local Discriminator sub-TLV;"

- "Negotiation Timer Parameters sub-TLV" if:

  - the N and S flags are cleared

  - the N flag is cleared and the S flag is set and a timing interval larger than the one received needs to be used

Reserved: Reserved for future specification and set to 0.

3.3.1. Local Discriminator sub-TLV

The "Local Discriminator sub-TLV" is carried as a sub-TLV of the "BFD Configuration sub-TLV" and is depicted below.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           Type (1) (IANA)     |         Length = 8            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                      Local Discriminator                      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Type: indicates a new type, the Local Discriminator sub-TLV (1) (IANA to define).

Length: indicates the TLV total length in octets.

Local Discriminator: A unique, nonzero discriminator value generated by the transmitting system and referring to itself, used to demultiplex multiple BFD sessions between the same pair of systems.

3.3.2. Negotiation Timer Parameters sub-TLV

The "Negotiation Timer Parameters sub-TLV" is carried as a sub-TLV of the "BFD Configuration sub-TLV" and is depicted below.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Timer Neg. Type (2) (IANA)   |          Length = 20          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|         Acceptable Min. Asynchronous TX interval              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|         Acceptable Min. Asynchronous RX interval              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|               Required Echo TX Interval                       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Type: indicates a new type, the "Negotiation Timer Parameters sub-TLV" (IANA to define).

Length: indicates the TLV total length in octets. (20)

Acceptable Min. Asynchronous TX interval: in case of S (symmetric) flag set in the "BFD Configuration sub-TLV", it expresses the desired time interval (in microseconds) at which the ingress LER intends to both transmit and receive BFD periodic control packets. If the receiving edge LSR can not support such value, it can reply with an interval greater than the one proposed.

In case of S (symmetric) flag cleared in the "BFD Configuration sub-TLV", this field expresses the desired time interval (in microseconds) at which a edge LSR intends to transmit BFD periodic control packets in its transmitting direction.

Acceptable Min. Asynchronous RX interval: in case of S (symmetric) flag set in the "BFD Configuration sub-TLV", this field MUST be equal to "Acceptable Min. Asynchronous TX interval" and has no additional meaning respect to the one described for "Acceptable Min. Asynchronous TX interval".

In case of S (symmetric) flag cleared in the "BFD Configuration sub-TLV", it expresses the minimum time interval (in microseconds) at which edge LSRs can receive BFD periodic control packets. In case this value is greater than the "Acceptable Min. Asynchronous TX interval" received from the other edge LSR, such edge LSR MUST adopt the interval expressed in this "Acceptable Min. Asynchronous RX interval".

Required Echo TX Interval: the minimum interval (in microseconds) between received BFD Echo packets that this system is capable of supporting, less any jitter applied by the sender as described in [RFC5880] sect. 6.8.9. This value is also an indication for the receiving system of the minimum interval between transmitted BFD Echo packets. If this value is zero, the transmitting system does not support the receipt of BFD Echo packets. If the receiving system can not support this value an error MUST be generated "Unsupported BFD TX rate interval".

3.3.3. BFD Authentication sub-TLV

The "BFD Authentication sub-TLV" is carried as a sub-TLV of the "BFD Configuration sub-TLV" and is depicted below.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    BFD Auth. Type (3) (IANA)  |         Length = 8            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Auth Type   |  Auth Key ID  |         Reserved (0s)         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Type: indicates a new type, the "BFD Authentication sub-TLV" (IANA to define).

Length: indicates the TLV total length in octets. (8)

Auth Type: indicates which type of authentication to use. The same values as are defined in section 4.1 of [RFC5880] are used.

Auth Key ID: indicates which authentication key or password (depending on Auth Type) should be used. How the key exchange is performed is out of scope of this document.

Reserved: Reserved for future specification and set to 0.

3.4. Performance Monitoring sub-TLV

If the "OAM functions TLV" has either the L (Loss), D (Delay) or T (Throughput) flag set, the "Performance Monitoring sub-TLV" MUST be present.

In case the vlues needs to be different than the default ones the "Performance Monitoring sub-TLV", "MPLS OAM PM Loss sub-TLV" MAY include the following sub-TLVs:

- "MPLS OAM PM Loss sub-TLV" if the L flag is set in the "OAM functions TLV";

- "MPLS OAM PM Delay sub-TLV" if the D flag is set in the "OAM functions TLV";

The "Performance Monitoring sub-TLV" depicted below is carried as a sub-TLV of the "OAM Functions TLV".

```
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    Perf Monitoring Type (IANA)|          Length               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|D|L|J|Y|K|C|            Reserved (set to all 0s)               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
~                           sub-TLVs                            ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Configuration Flags, for the specific function description please refer to [MPLS-PM]:

- D: Delay inferred/direct (0=INFERRED, 1=DIRECT)

- L: Loss inferred/direct (0=INFERRED, 1=DIRECT)

- J: Delay variation/jitter (1=ACTIVE, 0=NOT ACTIVE)

- Y: Dyadic (1=ACTIVE, 0=NOT ACTIVE)

- K: Loopback (1=ACTIVE, 0=NOT ACTIVE)

- C: Combined (1=ACTIVE, 0=NOT ACTIVE)

3.4.1. MPLS OAM PM Loss sub-TLV

The "MPLS OAM PM Loss sub-TLV" depicted below is carried as a sub-TLV of the "Performance Monitoring sub-TLV".

```
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  PM Loss Type (1) (IANA)      |          Length               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| OTF |T|B|                   RESERVED                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                    Measurement Interval                       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                       Test Interval                           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                      Loss Threshold                           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Type: indicates a new type, the "MPLS OAM PM Loss sub-TLV" (IANA to define, suggested value 1).

Length: indicates the length of the parameters in octets (12).

OTF: Origin Timestamp Format of the Origin Timestamp field described in [MPLS-PM]. By default it is set to IEEE 1588 version 1.

Configuration Flags, please refer to [MPLS-PM] for further details:

- T: Traffic-class-specific measurement indicator. Set to 1 when the measurement operation is scoped to packets of a particular traffic class (DSCP value), and 0 otherwise. When set to 1, the DS field of the message indicates the measured traffic class. By default it is set to 1.

- B: Octet (byte) count. When set to 1, indicates that the Counter 1-4 fields represent octet counts. When set to 0, indicates that the Counter 1-4 fields represent packet counts. By default it is set to 0.

Measurement Interval: the time interval (in microseconds) at which Loss Measurement query messages MUST be sent on both directions. If the edge LSR receiving the Path message can not support such value, it can reply back with a higher interval. By default it is set to (TBD).

Test Interval: test messages interval as described in [MPLS-PM]. By default it is set to (TBD).

Loss Threshold: the threshold value of lost packets over which protections MUST be triggered. By default it is set to (TBD).

3.4.2. MPLS OAM PM Delay sub-TLV

The "MPLS OAM PM Delay sub-TLV" depicted below is carried as a sub-TLV of the "OAM Functions TLV".

```
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  PM Delay Type (2) (IANA)     |          Length               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| OTF |T|B|                   RESERVED                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     Measurement Interval                      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                        Test Interval                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                       Delay Threshold                         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Type: indicates a new type, the "MPLS OAM PM Loss sub-TLV" (IANA to define, suggested value 1).

Length: indicates the length of the parameters in octets (12).

OTF: Origin Timestamp Format of the Origin Timestamp field described in [MPLS-PM]. By default it is set to IEEE 1588 version 1.

Configuration Flags, please refer to [MPLS-PM] for further details:

- T: Traffic-class-specific measurement indicator. Set to 1 when the measurement operation is scoped to packets of a particular traffic class (DSCP value), and 0 otherwise. When set to 1, the DS field of the message indicates the measured traffic class. By default it is set to 1.

- B: Octet (byte) count. When set to 1, indicates that the Counter 1-4 fields represent octet counts. When set to 0, indicates that the Counter 1-4 fields represent packet counts. By default it is set to 0.

Measurement Interval: the time interval (in microseconds) at which Delay Measurement query messages MUST be sent on both directions. If the edge LSR receiving the Path message can not support such value, it can reply back with a higher interval. By default it is set to (TBD).

Test Interval: test messages interval as described in [MPLS-PM]. By default it is set to (TBD).

Delay Threshold: the threshold value of measured delay (in microseconds) over which protections MUST be triggered. By default it is set to (TBD).

3.5. MPLS OAM FMS sub-TLV

The "MPLS OAM FMS sub-TLV" depicted below is carried as a sub-TLV of the "OAM Configuration sub-TLV".

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           Type (5)  (IANA)    |        Length = 12            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|A|D|L|C|            Reserved   (set all to 0s)        |E| PHB |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                       Refresh Timer                           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Type: indicates a new type, the "MPLS OAM FMS sub-TLV" (IANA to define).

Length: indicates the TLV total length in octets.

Signal Flags: are used to enable the following signals:

- A: Alarm Indication Signal (AIS) as described in [MPLS-FMS]

- D: Link Down Indication (LDI) as described in [MPLS-FMS]

- L: Locked Report (LKR) as described in [MPLS-FMS]

- C: Client Signal Failure (CSF) as described in [MPLS-CSF]

- Remaining bits: Reserved for future specification and set to 0.

Configuration Flags:

- E: used to enable/disable explicitly clearing faults

- PHB: identifies the per-hop behavior of packets with fault management information

Refresh Timer: indicates the refresh timer (in microseconds) of fault indication messages. If the edge LSR receiving the Path message can not support such value, it can reply back with a higher interval.

4. IANA Considerations

This document specifies the following new TLV types:

- "BFD Configuration" type: 2;

- "MPLS OAM PM Loss" type: 3;

- "MPLS OAM PM Delay" type: 4;

- "MPLS OAM FMS" type: 5.

sub-TLV types to be carried in the "BFD Configuration sub-TLV":

- "Local Discriminator" sub-TLV type: 1;

- "Negotiation Timer Parameters" sub-TLV type: 2.

- "BFD Authentication" sub-TLV type: 3.

5. BFD OAM configuration errors

In addition to error values specified in [OAM-CONF-FWK] and [ETH-OAM] this document defines the following values for the "OAM Problem" Error Code:

- "MPLS OAM Unsupported Functionality";

- "OAM Problem/Unsupported TX rate interval".

6. Acknowledgements

The authors would like to thank David Allan, Lou Berger, Annamaria Fulignoli, Eric Gray, Andras Kern, David Jocha and David Sinicrope for their useful comments.

7. Security Considerations

The signaling of OAM related parameters and the automatic establishment of OAM entities introduces additional security considerations to those discussed in [RFC3473]. In particular, a network element could be overloaded if an attacker were to request high frequency liveliness monitoring of a large number of LSPs, targeting a single network element.

Security aspects will be covered in more detailed in subsequent versions of this document.

8. References

8.1. Normative References

[MPLS-FMS]
Swallow, G., Fulignoli, A., Vigoureux, M., Boutros, S., and D. Ward, "MPLS Fault Management OAM", 2009, <draft-ietf-mpls-tp-fault>.

[MPLS-PM] Bryant, S. and D. Frost, "Packet Loss and Delay Measurement for the MPLS Transport Profile", 2010, <draft-ietf-mpls-loss-delay>.

[MPLS-PM-Profile]
Bryant, S. and D. Frost, "A Packet Loss and Delay Measurement Profile for MPLS-based Transport Networks", 2010, <draft-ietf-mpls-tp-loss-delay-profile>.

[MPLS-TP-IDENTIF]
Bocci, M., Swallow, G., and E. Gray, "MPLS-TP Identifiers", 2010, <draft-ietf-mpls-tp-identifiers>.

[OAM-CONF-FWK]
Takacs, A., Fedyk, D., and J. van He, "OAM Configuration Framework for GMPLS RSVP-TE", 2009, <draft-ietf-ccamp-oam-configuration-fwk>.

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

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

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

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

[RFC5586] Bocci, M., Vigoureux, M., and S. Bryant, "MPLS Generic Associated Channel", RFC 5586, June 2009.

[RFC5654] Niven-Jenkins, B., Brungard, D., Betts, M., Sprecher, N., and S. Ueno, "Requirements of an MPLS Transport Profile", RFC 5654, September 2009.

[RFC5860] Vigoureux, M., Ward, D., and M. Betts, "Requirements for Operations, Administration, and Maintenance (OAM) in MPLS Transport Networks", RFC 5860, May 2010.

[RFC5880] Katz, D. and D. Ward, "Bidirectional Forwarding Detection (BFD)", RFC 5880, June 2010.

[RFC5884] Aggarwal, R., Kompella, K., Nadeau, T., and G. Swallow, "Bidirectional Forwarding Detection (BFD) for MPLS Label Switched Paths (LSPs)", RFC 5884, June 2010.

8.2. Informative References

[BFD-CCCV] Allan, D., Swallow, G., and J. Drake, "Proactive Connectivity Verification, Continuity Check and Remote Defect indication for MPLS Transport Profile", 2010, <draft-ietf-mpls-tp-bfd-cc-cv-rdi>.

[BFD-Ping] Bahadur, N., Aggarwal, R., Ward, D., Nadeau, T., Sprecher, N., and Y. Weingarten, "LSP Ping and BFD encapsulation over ACH", 2010, <draft-ietf-mpls-tp-lsp-ping-bfd-procedures-02>.

[ETH-OAM] Takacs, A., Gero, B., Fedyk, D., Mohan, D., and D. Long, "GMPLS RSVP-TE Extensions for Ethernet OAM", 2009, <draft-ietf-ccamp-rsvp-te-eth-oam-ext>.

[LSP Ping] Kompella, K. and G. Swallow, "Detecting Multi-Protocol Label Switched (MPLS) Data Plane Failures", 2006, <RFC 3479>.

[LSP-PING CONF] Bellagamba, E., Andersson, L., Ward, D., and P. Skoldstrom, "Configuration of pro-active MPLS-TP Operations, Administration, and Maintenance (OAM) Functions Using LSP Ping", 2010, <draft-ietf-mpls-lsp-ping-mpls-tp-oam-conf>.

[MPLS-TP OAM Analysis]
Sprecher, N., Weingarten, Y., and E. Bellagamba, "MPLS-TP OAM Analysis", 2011, <draft-ietf-mpls-tp-oam-analysis>.

[MPLS-TP-OAM-FWK]
Bocci, M. and D. Allan, "Operations, Administration and Maintenance Framework for MPLS-based Transport Networks", 2010, <draft-ietf-mpls-tp-oam-framework>.

[RFC4447] Martini, L., Rosen, E., El-Aawar, N., Smith, T., and G. Heron, "Pseudowire Setup and Maintenance Using the Label Distribution Protocol (LDP)", RFC 4447, April 2006.

[RFC5921] Bocci, M., Bryant, S., Frost, D., Levrau, L., and L. Berger, "A Framework for MPLS in Transport Networks", RFC 5921, July 2010.

Authors' Addresses

Elisa Bellagamba (editor)
Ericsson
Torshamnsgatan 48
Kista, 164 40
Sweden

Email: elisa.bellagamba@ericsson.com

Loa Andersson (editor)
Ericsson
Torshamnsgatan 48
Kista, 164 40
Sweden

Phone:
Email: loa.andersson@ericsson.com

Pontus Skoldstrom (editor)
Acreo AB
Electrum 236
Kista, 164 40
Sweden

Phone: +46 8 6327731
Email: pontus.skoldstrom@acreo.se

Dave Ward
Juniper

Phone:
Email: dward@juniper.net

Attila Takacs
Ericsson
1. Laborc u.
Budapest,
HUNGARY

Phone:
Email: attila.takacs@ericsson.com

Network Working Group A. Kern
Internet-Draft A. Takacs
Intended status: Standards Track Ericsson
Expires: April 2012 October 26, 2011

GMPLS RSVP-TE Extensions for SONET/SDH and OTN OAM Configuration
draft-ietf-ccamp-rsvp-te-sdh-otn-oam-ext-03

Abstract

GMPLS has been extended to support connection establishment in both SONET/SDH [RFC4606] and OTN [RFC4328] networks. However support for the configuration of the supervision functions is not specified. Both SONET/SDH and OTN implement supervision functions to qualify the transported signals. This document defines extensions to RSVP-TE for SONET/SDH and OTN OAM configuration based on the OAM Configuration Framework defined in [GMPLS-OAM-FWK].

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].

Table of Contents

1. Introduction

Both SONET/SDH and OTN implement supervision functions to qualify the transported signals. Supervision functions include continuity, connectivity, signal quality, alignment and payload supervision. The ITU-T G.806 [G.806] recommendation defines the generic framework of the supervision functions, which are then further specified for SONET/SDH and OTN in technology specific documents.

GMPLS has been extended to support connection establishment in both SONET/SDH [RFC4606] and OTN [RFC4328] networks. These documents however do not support the configuration of the respective supervision functions.

[GMPLS-OAM-FWK] defines a technology-agnostic framework for GMPLS to support the establishment and configuration of the pro-active OAM functions of signalled connections. The properties of the OAM functions are exchanged during connection establishment and may be modified during the life of the connection. The technology specific parameters to be exchanged are to be described in accompanying documents. This document defines the extensions for SONET/SDH and OTN OAM configuration for end-to-end monitoring.

2. Overview of SONET/SDH and OTN OAM related functions

SONET/SDH [G.707] and OTN [G.709] provide a variety of supervision functions. Here we only consider continuity, connectivity and signal quality supervision functions, as these are the candidates for GMPLS based configuration.

2.1. Continuity supervision

Continuity supervision provides methods monitoring the health of a connection (trail).

2.2. Connectivity supervision

The connectivity supervision function provides a method to detect misconnections. The detection procedure is based on emitting a Trace Trail Identifier (TTI) known by both endpoints. The TTI is included by the source node as an overhead signal for each connection. The receiver node then compares the received TTI with the expected value and decides if a miss-connection occurred.

2.2.1. SONET/SDH

In case of SONET/SDH, connectivity supervision is implemented in the Regeneration Section (RS) and in the lower and higher order path layers (LOVC and HOVC). In all layers the TTI encodes only the Access Point Identifier (API) of the source node. In the various layers the lengths of these TTIs are different. In RS the TTI (encoded in J0 octet) is either 1 or 16 octets long. In higher order paths the TTI (encoded in J1), is either 16 or 64 octet long. In lower order paths the TTI is transmitted in the J2 byte and is 16 octet long.

2.2.2. OTN

In case of OTN, connectivity supervision is supported by the OTUk and ODUk digital hierarchy layers. In both layers, the length of the TTI is 64 octets, but only the first 32 octets are considered for connectivity supervision. This first part is further divided into a Source Access Point Identifier (SAPI) and a Destination Access Point Identifier (DAPI). Connectivity supervision may consider either the SAPI or DAPI only or both. The structure of the SAPI and DAPI is specified in [G.709].

2.3. Signal quality supervision

The quality of the transmitted signal is monitored as a ratio of bad frames. If the number of such frames reaches a threshold a defect state is declared. To detect the correctness of the frames an Error Detection Code (EDC), such as Bit Interleaved Parity (BIP), is used. The distribution of the errors is assumed to follow either Poisson or a bursty distribution. For Poisson distribution an EDC violation ratio is defined as the threshold; while for the bursty model the threshold is defined as a number of consecutive 1-second time intervals in which the EDC violation exceeds a predefined ratio. In case of Poisson error distribution two defect state levels are defined: the Excessive Error and Degraded Signal defect. In the case of the bursty model, only the Degraded Signal defect level is considered.

2.3.1. SONET/SDH

SONET/SDH supports both Excessive Error and Degraded Signal defect levels and supports both Poisson and bursty error distribution models. These signal quality parameters are configured for the Multiplexing Section (MS) and the LOVC and HOVC path layers.

2.3.2. OTN

For OTN, in the digital transport layers (OTUk and ODUk) only the bursty error distribution model errors with the Degraded Signal defect level is supported. Two parameters are defined: Ratio of the bad frames in a one second interval (0% to 100% or 0 to number of frames per 1-second interval) and Number of consecutive intervals (between 2 and 10). Signal quality supervision in the optical transport layers is not specified by [G.798], it is indicated to be for further study.

3. RSVP-TE signaling extensions

3.1. Operation overview

RFC 4606 and RFC 4328 define the RSVP-TE extensions necessary to manage SDH/SONET and OTN optical and digital hierarchy connections. The monitoring functions associated to these connections may be configured together with configuring the connection itself.

The LSP Attribute Flag "OAM MEP entities desired" [GMPLS-OAM-FWK] is used to signal that the monitoring functions at the endpoints must be established. The "OAM MIP entities desired" flag must be set to 0 and must be ignored.

To configure OAM parameters the OAM Configuration TLV can be included in the LSP_ATTRIBUTES object. The TLV identifies which OAM technology ("OAM Type" field) to be used as well as which OAM functions are to be enabled (OAM Function Flags sub-TLV). For SONET/SDH and OTN the "Continuity Check" and "Connectivity Verification" flags control the Continuity and Connectivity supervision functions, while the "Performance Monitoring/Loss" flag enables the Signal Quality supervision function. Since delay monitoring is not used for SONET/SDH or OTN the "Performance Monitoring/Delay" flag must be cleared.

For additional details the appropriate technology specific sub-TLV can be carried in the OAM Configuration TLV.

3.1.1. Continuity Check supervision

In case of both discussed technologies, setting up continuity supervision function for a connection does not need further configuration besides enabling it. Therefore, by setting the "Connectivity Monitoring" Flag of OAM Function implicitly enables the continuity supervision function as well.

3.1.2. Connectivity Monitoring supervision

3.1.2.1. SDH/SONET

[G.707] defines three bytes (signals) for connectivity supervision purposes: the J0 byte in RS layer, the J1 and J2 bytes in HOVC and LOVC layers. These bytes encode 1 octet, 16 octet or 64 octet long unstructured octet streams. The source node emits this stream and the destination node matches it with an expected one. When the destination detects mismatch, defect state will be declared.

Since these streams encode an identifier defined for the source node, different stream will to be emitted in the upstream and downstream directions for bidirectional connections. During the configuration the egress node has to be configured with the TTI value to be expected in the downstream direction and the TTI value to be emitted in the upstream direction. Therefore the SONET/SDH OAM Configuration TLV carries two Connectivity Supervision TLVs.

3.1.2.2. OTN

[G.709] defines a 64 octet long TTI, where the first 32 octets have a generic structure: a zero octet, a 15 octet long SAPI, a second zero octet and finally the 15 octet long DAPI.

For a unidirectional connection a single Connection Supervision TLV encodes elements of the TTI to be emitted. This TLV also specifies which parts of the TTI are compared to the expected values (only SAPI, only DAPI, both SAPI and DAPI).

In case of a bidirectional connection an endpoint can use a common API value for SAPI (for transmitted signal) and DAPI (for received signal). (See Figure 1.) The TTI values used in downstream and upstream directions are derived from the two API values: the downstream TTI will have the form of [0, API_a, 0, API_z] while the upstream TTI will use the form of [0, API_z, 0 API_a].

```
Ingress Node                                          Egress Node
 ( API_a )    TTI_upstream = [0, API_z, 0, API_a ]      ( API_z )
| Rx port | -- < -- < -- < -- < -- < -- < -- < -- < -- | Tx Port |
| Tx port | -- > -- > -- > -- > -- > -- > -- > -- > -- | Rx Port |
              TTI_downstream = [0, API_a, 0, API_z ]
```

Figure 1: TTI construction when a single API identifies the receiver and transmitter interfaces

Then, a single Connectivity Supervision TLV is defined. The SAPI field carries the API of the ingress node (API_a) that initiates the signaling, while the DAPI carries the API of the egress node (API_z).

On the other hand, it is possible that the endpoints use different values as SAPI and DAPI to identify the transmitter and receiver ports of a bidirectional connection (See Figure 2). In this case the TTIs to be used in the two directions are independent, thus, they must be explicitly configured. Therefore, two Connectivity Supervision TLVs are added to the OTN OAM Configuration TLV. Each TLV encodes whether it defines the downstream or the upstream TTI.

```
Ingress Node                                          Egress Node
 ( API_a1 )    TTI_upstream = [0, API_z1, 0, API_a1 ]   ( API_z1 )
| Rx port | -- < -- < -- < -- < -- < -- < -- < -- < -- | Tx Port |
| Tx port | -- > -- > -- > -- > -- > -- > -- > -- > -- | Rx Port |
 ( API_a2 )    TTI_downstream = [0, API_a2, 0, API_z2 ] ( API_z2 )
```

Figure 2: TTI construction when dedicated APIs identify the receiver and transmitter interfaces

3.1.3. Signal quality supervision

Signal quality supervision function is implemented in MS, HOVC, LOVC layers of SDH/SONET. All three layers support exceeded error level with Poisson error distribution model and degraded signal defect level with both, Poisson and bursty error distribution model. Dedicated Signal quality supervision TLVs encode each level, therefore when the "Performance Monitoring/Loss" flag is set; several such TLVs can be added to the SONET/SDH OAM Configuration TLV. If a configuration TLV for a particular level is missing the default parameters for that level is to be applied.

The OTN supports only Degraded Signal defect with bursty error model in OTUk and ODUk layers. Thus, the only parameters to be encoded are: the threshold for bad frames in a 1-second interval and the number of consecutive 1-second intervals with excessive bad frames. Furthermore, as only one level is allowed a single Signal quality supervision TLV is added to the OTN OAM Configuration TLV.

3.2. Signaling support of Virtual Concatenation Groups (VCG)

A key capability of both, SONET/SDH and OTN is the support of virtual concatenation. This inverse multiplexing method uses multiplicity of parallel basic signals. The supervision function parameters of these basic signals can be different.

[GMPLS-VCAT-LCAS] summarises GMPLS signaling capabilities to support virtual concatenation and proposes extensions to that. A Virtual Concatenated Group (VCG) is constructed from several individual data plane signals. The co-routed signals of a VCG could be provisioned together using a single RSVP-TE session (co-signaled). As different OAM configuration may be applied to each of these individual signals, the OAM configuration extension is applied as follows.

We assume that the same OAM type and the same set of OAM functions apply to every individual signal of the VCG. A single OAM Configuration TLV is carried in the LSP_ATTRIBUTES Object, while multiple instances of technology specific OAM configuration sub-TLVs are added: one instance per individual signal. The order of these

TLVs refers to the logical order of the basic signals (as they are listed in the Label Object).

[GMPLS-VCAT-LCAS] allows extension/pruning of a VCG. To achieve it the traffic descriptor, which encodes how the VCG is structured, in the RSVP-TE session is updated. If the VCG is updated the contents of the OAM Configuration TLV needs to be updated accordingly.

3.3. OAM types and functions

This document defines two new code points for the "OAM Type" field of the OAM Configuration TLV, defined in [GMPLS-OAM-FWK]: SONET/SDH OAM and OTN Digital Hierarchy OAM.

OAM Type	Description
3	SONET/SDH OAM
4	OTN Digital Hierarchy OAM

The "OAM Function Flags sub-TLV", defined in [GMPLS-OAM-FWK]. SONET/SDH and OTN supervision functions are defined in this document for the following flags: "Continuity Check", "Connectivity Verification" and "Performance Monitoring/Loss". As delay measurement is not supported, requesting that function SHOULD generate an error with code/value "OAM Problem/Unsupported OAM Function".

3.4. SONET/SDH OAM Configuration sub-TLV

SONET/SDH OAM Configuration sub-TLV is defined to encode the parameters of continuity, connectivity and signal quality supervision functions for SONET/SDH networks.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|         Type (4) (IANA)       |           Length              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
~                           sub TLVs                            ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Type: indicates a new type: the SONET/SDH OAM Configuration TLV (IANA to define).

Length: indicates the total length including sub-TLVs

3.5. OTN OAM Configuration sub-TLV

OTN OAM Configuration TLV is defined to encode the parameters of continuity, connectivity and signal quality supervision functions for OTN.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           Type (5) (IANA)     |           Length              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
~                            sub TLVs                           ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Type: indicates a new type: the OTN OAM Configuration TLV (IANA to define).

Length: indicates the total length including sub-TLVs

3.6. TTI Configuration Sub-TLV

3.6.1. SDH TTI Configuration Sub-TLV

This sub-TLV is carried in the SONET/SDH OAM Configuration sub-TLV, if the Connectivity Verification OAM Function Flag is set. In every supporting layers the TTI identifies the source interface (SAPI); however, the length of this identifier varies layer-by-layer (See Section 2.2.1). Therefore, a generic TLV is defined supporting various TTI lengths.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|         Type (1) (IANA)       |         Length                |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|A|U|           Reserved        |             TTI               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
~                             TTI cont                         ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Flag "A", when set enables the AIS insertion on detecting TTI mismatch.

Flat "U" encodes if the TTI refers to the downstream TTI (U=0) or the upstream one (U=1).

The TTI field carries the TTI to be transmitted by the source node and to be expected by the sink. The TLV is padded to 4-octets.

If the specified length and format of the TTI carried in this TLV is not supported by the referred SONET/SDH layer, error must be generated: "OAM Problem/TTI Length Mismatch".

3.6.2. OTN TTI Configuration Sub-TLV

This sub-TLV is carried in the OTN OAM Configuration sub-TLV, if the Connectivity Verification OAM Function Flag is set.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|         Type (1) (IANA)       |         Length = 32           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|A|S|D|APP|      Reserved       |             SAPI              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                            SAPI cont                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                            SAPI cont                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                            SAPI cont                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     SAPI      |                    DAPI                       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                            DAPI cont                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                            DAPI cont                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                            DAPI cont                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Three control flags are defined. Flag "A" indicates that AIS insertion on detecting TTI mismatch (failing the connectivity verification) is required (A=1) or not (A=0). The next two flags define which parts of the received TTI are compared to the expected one. If flag "S" is set the TTI octets 1 to 15 are matched to the expected SAPI value. If the flag "D" is set the TTI octets 17 to 31 are matched to the expected DAPI value. If both "S" and "D" are set both parts of TTI are compared to SAPI and DAPI values. Setting both "S" and "D" bits to 0 is invalid, and if encountered error must be generated: "OAM Problem/Invalid CC/CV configuration".

The next two bits "APP" encode the applicability of the TTI configuration and the following code points are defined:

0 - Single TTI configuration: the TTI configuration is done according only to this TLV and no further TTI configuration TLVs are expected. This code point is used for unidirectional connections and for bidirectional connections with common APIs (See Figure 1)

1 - Downstream TTI for double TTI configuration: the current TLV instruct the configuration of the TTI to be used in downstream direction (See Figure 2).

2 - Upstream TTI for double TTI configuration: the current TLV instruct the configuration of the TTI to be used in upstream direction (See Figure 2).

3 - Invalid.

If the APP is set to 1 and the next or the previous sub-TLV is not an OTN TTI Configuration TLV with APP code point 2, then an error must be generated "OAM Problem/Invalid OTN TTI Configuration/Missing Upstream TTI configuration".

If the APP is set to 2 and the next or the previous sub-TLV is not an OTN TTI Configuration TLV with APP code point 1, then an error must be generated "OAM Problem/Invalid OTN TTI Configuration/Missing Downstream TTI configuration".

If the APP is set to either 1 or 2 and the unidirectional LSP is signaled (no UPSTREAM_LABEL is added to the message) or the APP is set to 3, an error must be generated "OAM Problem/Invalid OTN TTI Configuration/Invalid applicability code"

3.7. Degraded signal thresholds Sub-TLV

The Degraded signal thresholds Sub-TLV instructs the configuration of the signal quality supervision function. This sub-TLV is applicable in both SONET/SDH and OTN cases. This sub-TLV can be carried in both the SONET/SDH OAM Configuration sub-TLV or OTN OAM Configuration sub-TLV, if the PerformanceMonitoring/Loss OAM Function Flag is set.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|         Type (2) (IANA)       |         Length = 4            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|B|L|         Reserved          |    DEG_THR    |    DEG_M      |
```

```
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Two flags are defined to encode the signal quality measurement. The bit "B" encodes if distribution of errors is either Poisson (B=0) or Bursty (B=1). In case of Poisson distribution of errors two levels of defects are defined and encoded with bit "L": excessive error (L=0) and degraded signal (L=1). Since in case of Bursty distribution of errors only degraded signal defect is to be detected, therefore, in this latter case (B=1) the "L" bit must be set. Otherwise error must be generated: "OAM Problem/Invalid Performance Monitoring/Loss configuration".

The field "DEG_THR" defines the threshold for the bad frames (BIP-8 violations) in both, Poisson and bursty distributions of errors. In the first case (B=0) this field encodes the quotient of the threshold 10e-X. The possible values for excessive error are 3,4 and 5, while for degraded signal defect are 6,7,8 and 9.

In the second case (B=1) it encodes ratio of the bad frames in a 1-second period and can be set between 0 and 100, interpreted as ratios in percentage.

The field "DEG_M" defines monitoring time-frame in 1 second periods assuming bursty distribution of errors. The valid values are 2 to 10 periods.

4. Error handling

In addition to error values specified in [GMPLS-OAM-FWK] this document defines the following values for the "OAM Problem" Error Code.

- If Performance Measurement/Delay flag is set in the OAM Functions Flag sub-TLV, an error must be generated "OAM Problem/Unsupported OAM Function".

- In case of SONET/SDH OAM when the length or format of the TTI to be configured is not supported by the referred SONET/SDH layer, an error must be generated: "OAM Problem/TTI Length Mismatch".

- If both "S" and "D" bits in OTN TTI Configuration TLV are set to 0, error must be generated: "OAM Problem/Invalid CC/CV configuration"

- If the APP is set to 1 and the next or the previous sub-TLV is not an OTN TTI Configuration TLV with APP code point 2, then an error must be generated "OAM Problem/Invalid OTN TTI Configuration/ Missing Upstream TTI configuration".

- If the APP is set to 2 and the next or the previous sub-TLV is not an OTN TTI Configuration TLV with APP code point 1, then an error must be generated "OAM Problem/Invalid OTN TTI Configuration/ Missing Downstream TTI configuration".

- If the APP is set to either 1 or 2 and the unidirectional LSP is signaled (no UPSTREAM_LABEL is added to the message) or the APP is set to 3, an error must be generated "OAM Problem/Invalid OTN TTI Configuration/Invalid applicability code"

- If flag "B" in Degraded signal thresholds Sub-TLV is set to 1 and flag "L" in the same sub-TLV is set to 0 error must be generated "OAM Problem/Invalid Performance Monitoring/Loss configuration".

5. IANA Considerations

This document specifies two new sub-TLVs to be carried in the OAM Configuration TLV in the LSP_ATTRIBUTES or LSP_REQUIRED_ATTRIBUTES Objects in Path and Resv messages. The document assigns values 3 and 4 from the "OAM Type" field of the OAM Configuration TLV.

The following error values need to be assigned under "OAM Problem" error code: "OAM Problem/Unsupported OAM Function", "OAM Problem/TTI Length Mismatch", "OAM Problem/Invalid CC/CV configuration", "OAM Problem/Invalid OTN TTI Configuration/Missing Upstream TTI configuration", "OAM Problem/Invalid OTN TTI Configuration/Missing Downstream TTI configuration", "OAM Problem/Invalid OTN TTI Configuration/Invalid applicability code", "OAM Problem/Invalid Performance Monitoring/Loss configuration".

6. Security Considerations

Security aspects are addressed in the OAM configuration framework document [GMPLS-OAM-FWK]

7. Acknowledgements

The authors would like to thank Francesco Fondelli for his useful comments.

8. References

8.1. Normative References

[GMPLS-OAM-FWK]
Takacs, A., Fedyk, D., and H. Jia, "OAM Configuration Framework and Requirements for GMPLS RSVP-TE", draft-ietf-ccamp-oam-configuration-fwk-01 (work in progress), March 2009.

8.2. Informative References

[G.707] International Telecommunications Union, "Network node interface for the synchronous digital hierarchy (SDH)", ITU-T Recommendation G.707, January 2007.

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

[G.798] International Telecommunications Union, "Characteristics of optical transport network hierarchy equipment functional blocks", ITU-T Recommendation G.798, December 2006.

[G.806] International Telecommunications Union, "Characteristics of transport equipment - Description methodology and generic functionality", ITU-T Recommendation G.806, January 2009.

[GMPLS-VCAT-LCAS]
Bernstein, G., Rabbat, R., and H. Helvoort, "Operating Virtual Concatenation (VCAT) and the Link Capacity Adjustment Scheme (LCAS) with Generalized Multi-Protocol Label Switching (GMPLS)", draft-ietf-ccamp-gmpls-vcat-lcas-07 (work in progress), December 2008.

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

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

[RFC4606] Mannie, E. and D. Papadimitriou, "Generalized Multi-Protocol Label Switching (GMPLS) Extensions for Synchronous Optical Network (SONET) and Synchronous

Digital Hierarchy (SDH) Control", RFC 4606, August 2006.

Authors' Addresses

Andras Kern
Ericsson
Laborc u. 1.
Budapest,   1037
Hungary

Email: andras.kern@ericsson.com

Attila Takacs
Ericsson
Laborc u. 1.
Budapest,   1037
Hungary

Email: attila.takacs@ericsson.com

Network Working Group G. Bernstein
Internet Draft Grotto Networking
Intended status: Standards Track Y. Lee
Expires: April 2012 D. Li
Huawei
W. Imajuku
NTT

October 31, 2011

Routing and Wavelength Assignment Information Encoding for Wavelength Switched Optical Networks

draft-ietf-ccamp-rwa-wson-encode-13.txt

Status of this Memo

This Internet-Draft will expire on February 31, 2012.

Copyright Notice

Abstract

A wavelength switched optical network (WSON) requires that certain key information elements are made available to facilitate path computation and the establishment of label switching paths (LSPs). The information model described in "Routing and Wavelength Assignment Information for Wavelength Switched Optical Networks" shows what information is required at specific points in the WSON. Part of the WSON information model contains aspects that may be of general applicability to other technologies, while other parts are fairly specific to WSONs.

This document provides efficient, protocol-agnostic encodings for the WSON specific information elements. It is intended that protocol-specific documents will reference this memo to describe how information is carried for specific uses. Such encodings can be used to extend GMPLS signaling and routing protocols. In addition these encodings could be used by other mechanisms to convey this same information to a path computation element (PCE).

Conventions used in this document

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

Table of Contents

1. Introduction

A Wavelength Switched Optical Network (WSON) is a Wavelength Division Multiplexing (WDM) optical network in which switching is performed selectively based on the center wavelength of an optical signal.

[RFC6163] describes a framework for Generalized Multiprotocol Label Switching (GMPLS) and Path Computation Element (PCE) control of a WSON. Based on this framework, [WSON-Info] describes an information model that specifies what information is needed at various points in a WSON in order to compute paths and establish Label Switched Paths (LSPs).

This document provides efficient encodings of information needed by the routing and wavelength assignment (RWA) process in a WSON. Such encodings can be used to extend GMPLS signaling and routing protocols. In addition these encodings could be used by other mechanisms to convey this same information to a path computation element (PCE). Note that since these encodings are relatively efficient they can provide more accurate analysis of the control plane communications/processing load for WSONs looking to utilize a GMPLS control plane.

Note that encodings of information needed by the routing and label assignment process applicable to general networks beyond WSON are addressed in a separate document [Gen-Encode].

1.1. Revision History

1.1.1. Changes from 00 draft

Edits to make consistent with update to [RFC6205], i.e., removal of sign bit.

Clarification of TBD on connection matrix type and possibly numbering.

New sections for wavelength converter pool encoding: Wavelength Converter Set Sub-TLV, Wavelength Converter Accessibility Sub-TLV, Wavelength Conversion Range Sub-TLV, WC Usage State Sub-TLV.

Added optional wavelength converter pool TLVs to the composite node TLV.

1.1.2. Changes from 01 draft

The encoding examples have been moved to an appendix. Classified and corrected information elements as either reusable fields or sub-TLVs. Updated Port Wavelength Restriction sub-TLV. Added available wavelength and shared backup wavelength sub-TLVs. Changed the title and scope of section 6 to recommendations since the higher level TLVs that this encoding will be used in is somewhat protocol specific.

1.1.3. Changes from 02 draft

Removed inconsistent text concerning link local identifiers and the link set field.

Added E bit to the Wavelength Converter Set Field.

Added bidirectional connectivity matrix example. Added simple link set example. Edited examples for consistency.

1.1.4. Changes from 03 draft

Removed encodings for general concepts to [Gen-Encode].

Added in WSON signal compatibility and processing capability information encoding.

1.1.5. Changes from 04 draft

Added encodings to deal with access to resource blocks via shared fiber.

1.1.6. Changes from 05 draft

Revised the encoding for the "shared access" indicators to only use one bit each for ingress and egress.

1.1.7. Changes from 06 draft

Removed section on "WSON Encoding Usage Recommendations"

1.1.8. Changes from 07 draft

Section 3: Enhanced text to clarify relationship between pools, blocks and resources. Section 3.1, 3.2: Change title to clarify Pool-Block relationship. Section 3.3: clarify block-resource state.

Section 4: Deleted reference to previously removed RBNF element. Fixed TLV figures and descriptions for consistent sub-sub-TLV nomenclature.

1.1.9. Changes from 08 draft

Fixed ordering of fields in second half of sub-TLV example in Appendix A.1.

Clarifying edits in section 3 on pools, blocks, and resources.

1.1.10. Changes from 09 draft

Fixed the "Block Shared Access Wavelength Availability sub-TLV" of section 3.4 to use an "RB set field" rather than a single RB ID. Removed all 1st person idioms.

1.1.11. Changes from 10 draft

Removed remaining 1st person idioms. Updated IANA section. Update references for newly issued RFCs.

1.1.12. Changes from 11 draft

Fixed length fields in section 4 to be 16 bits, correcting errors in TLV and field figures. Added a separate section on resources, blocks, sets and the resource pool. Moved definition of the resource block set field to this new section.

1.1.13. Changes from 12 draft

RB Identifier field in Section 3.1 to be 32 bits from 16 bits. Added Editorial changes and updated the contributor list.

2. Terminology

CWDM: Coarse Wavelength Division Multiplexing.

DWDM: Dense Wavelength Division Multiplexing.

FOADM: Fixed Optical Add/Drop Multiplexer.

ROADM: Reconfigurable Optical Add/Drop Multiplexer. A reduced port count wavelength selective switching element featuring ingress and egress line side ports as well as add/drop side ports.

RWA: Routing and Wavelength Assignment.

Wavelength Conversion. The process of converting an information bearing optical signal centered at a given wavelength to one with "equivalent" content centered at a different wavelength. Wavelength conversion can be implemented via an optical-electronic-optical (OEO) process or via a strictly optical process.

WDM: Wavelength Division Multiplexing.

Wavelength Switched Optical Network (WSON): A WDM based optical network in which switching is performed selectively based on the center wavelength of an optical signal.

3. Resources, Blocks, Sets, and the Resource Pool

The optical system to be encoded may contain a pool of resources of different types and properties for processing optical signals. For the purposes here a "resource" is an individual entity such as a wavelength converter or regenerator within the optical node that acts on an individual wavelength signal.

Since resources tend to be packaged together in blocks of similar devices, e.g., on line cards or other types of modules, the fundamental unit of identifiable resource in this document is the "resource block". A resource block may contain one or more resources. As resource blocks are the smallest identifiable unit of processing resource, one should group together resources into blocks if they have similar characteristics relevant to the optical system being modeled, e.g., processing properties, accessibility, etc.

This document defines the following sub-TLVs pertaining to resources within an optical node:

- Resource Pool Accessibility Sub-TLV

- Resource Block Wavelength Constraints Sub-TLV

- Resource Pool State Sub-TLV

- Block Shared Access Wavelength Availability Sub-TLV

- Resource Block Information Sub-TLV

Each of these sub-TLVs works with one or more sets of resources rather than just a single resource block. This motivates the following field definition.

3.1. Resource Block Set Field

In a WSON node that includes resource blocks (RB), denoting subsets of these blocks allows one to efficiently describe common properties the blocks and to describe the structure and characteristics, if non-trivial, of the resource pool. The RB Set field is defined in a similar manner to the label set concept of [RFC3471].

The information carried in a RB set field is defined by:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    Action     |C|   Reserved  |       Length                  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                         RB Identifier 1                       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
:                               :                               :
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                         RB Identifier n                       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Action: 8 bits

0 - Inclusive List

Indicates that the TLV contains zero or more RB elements that are included in the list.

1 - Reserved

2 - Inclusive Range

Indicates that the TLV contains a range of RBs. The object/TLV contains two WC elements. The first element indicates the start of the range. The second element indicates the end of the range. A value of zero indicates that there is no bound on the corresponding portion of the range.

3 - Reserved

C (Connectivity bit): Set to 0 to denote fixed (possibly multi-cast) connectivity; Set to 1 to denote potential (switched) connectivity. Used in resource pool accessibility sub-TLV. Ignored elsewhere.

Reserved: 7 bits

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

Length: 16 bits

The total length of this field in bytes.

RB Identifier:

The RB identifier represents the ID of the resource block which is a 32 bit integer.

Usage Note: the inclusive range "Action" can result in very compact encoding of resource sets and it can be advantages to number resource blocks in such a way so that status updates (dynamic information) can take advantage of this efficiency.

4. Resource Pool Accessibility/Availability

This section defines the sub-TLVs for dealing with accessibility and availability of resource blocks within a pool of resources. These include the ResourceBlockAccessibility, ResourceWaveConstraints, and RBPoolState sub-TLVs.

4.1. Resource Pool Accessibility Sub-TLV

This sub-TLV describes the structure of the resource pool in relation to the switching device. In particular it indicates the ability of an ingress port to reach sets of resources and of a sets of resources to reach a particular egress port. This is the PoolIngressMatrix and PoolEgressMatrix of [WSON-Info].

The resource pool accessibility sub-TLV is defined by:

```
     0                   1                   2                   3
     0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    | Connectivity  |                   Reserved                    |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                   Ingress Link Set Field A #1                 |
    :                                                               :
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                         RB Set Field A #1                     |
    :                                                               :
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |         Additional Link set and RB set pairs as needed to     |
    :                   specify PoolIngressMatrix                   :
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |               Egress Link Set Field B #1                      |
    :                                                               :
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |            RB Set B Field #1 (for egress connectivity)        |
    :                                                               :
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |         Additional Link Set and RB set pairs as needed to     |
    :                   specify PoolEgressMatrix                    :
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Where

Connectivity indicates how the ingress/egress ports connect to the resource blocks.

0 -- the device is fixed (e.g., a connected port must go through the resource block)

1 -- the device is switched (e.g., a port can be configured to go through a resource but isn't required)

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

Note that the direction parameter within the Link Set Field is used to indicate whether the link set is an ingress or egress link set, and the bidirectional value for this parameter is not permitted in this sub-TLV.

See Appendix A.1 for an illustration of this encoding.

4.2. Resource Block Wavelength Constraints Sub-TLV

Resources, such as wavelength converters, etc., may have a limited input or output wavelength ranges. Additionally, due to the structure of the optical system not all wavelengths can necessarily reach or leave all the resources. These properties are described by using one or more resource wavelength restrictions sub-TLVs as defined below:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     RB Set Field                              |
:                                                               :
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                Input Wavelength Set Field                     |
:                                                               :
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                Output Wavelength Set Field                    |
:                                                               :
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

RB Set Field:

A set of resource blocks (RBs) which have the same wavelength restrictions.

Input Wavelength Set Field:

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

Output Wavelength Set Field:

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

4.3. Resource Pool State Sub-TLV

The state of the pool is given by the number of resources available with particular characteristics. A resource block set is used to encode all or a subset of the resources of interest. The usage state of resources within a resource block set is encoded as either a list of 16 bit integer values indicating the number of available resources in the resource block, or a bit map indicating whether a particular resource is available or in use. The bit map encoding is appropriate when resource blocks consist of a single resource. This information can be relatively dynamic, i.e., can change when a connection (LSP is established or torn down.

```
 0                   1                   2                   3
  0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 | Action        |    Reserved                                   |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                          RB Set Field                         |
 :                                                               :
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                    RB Usage state                             |
 :                                                               :
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Where Action = 0 denotes a list of 16 bit integers and Action = 1 denotes a bit map. In both cases the elements of the RB Set field are in a one-to-one correspondence with the values in the usage RB usage state area.

```
  0                   1                   2                   3
  0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 | Action = 0    |    Reserved                                   |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                          RB Set Field                         |
 :                                                               :
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                    RB#1 state   |      RB#2 state             |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 :                                                               :
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                   RB#n-1 state  |   RB#n state or Padding     |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Whether the last 16 bits is a wavelength converter (RB) state or padding is determined by the number of elements in the RB set field.

```
  0                   1                   2                   3
  0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 | Action = 1    |    Reserved                                   |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                     RB Set Field                              |
 :                                                               :
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                  RB Usage state bitmap                        |
 :                                                               :
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                     ......    |         Padding bits          |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

RB Usage state: Variable Length but must be a multiple of 4 bytes.

Each bit indicates the usage status of one RB with 0 indicating the RB is available and 1 indicating the RB is in used. The sequence of the bit map is ordered according to the RB Set field with this sub-TLV.

Padding bits: Variable Length

4.4. Block Shared Access Wavelength Availability sub-TLV

Resources blocks may be accessed via a shared fiber. If this is the case, then wavelength availability on these shared fibers is needed to understand resource availability.

```
     0                   1                   2                   3
     0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |I|E|                          Reserved                         |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                      RB Set Field                             |
    :                                                               :
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |            Ingress Available Wavelength Set Field             |
    :                          (Optional)                           :
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |             Egress Available Wavelength Set Field             |
    :                          (Optional)                           :
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

I bit:

Indicates whether the ingress available wavelength set field is included (1) or not (0).

E bit:

Indicates whether the egress available wavelength set field is included (1) or not (0).

RB Set Field:

A Resource Block set in which all the members share the same ingress or egress fiber or both.

Ingress Available Wavelength Set Field:

Indicates the wavelengths currently available (not being used) on the ingress fiber to this resource block.

Egress Available Wavelength Set Field:

Indicates the wavelengths currently available (not being used) on the egress fiber from this resource block.

5. Resource Properties Encoding

Within a WSON network element (NE) there may be resources with signal compatibility constraints. These resources be regenerators, wavelength converters, etc... Such resources may also constitute the network element as a whole as in the case of an electro optical switch. This section primarily focuses on the signal compatibility and processing properties of such a resource block.

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

(a) Input constraints (shared ingress, modulation, FEC, bit rate, GPID)

(b) Processing capabilities (number of resources in a block, regeneration, performance monitoring, vendor specific)

(c) Output Constraints (shared egress, modulation, FEC)

5.1. Resource Block Information Sub-TLV

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

This sub-TLV has the following format:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     RB Set Field                              |
:                                                               :
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|I|E|                        Reserved                           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           Input Modulation Type List Sub-Sub-TLV  (opt)       |
:                                                               :
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           Input FEC Type List Sub-Sub-TLV    (opt)            |
:                                                               :
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|         Input Client Signal Type Sub-Sub-TLV      (opt)       |
:                                                               :
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|         Input Bit Rate Range List  Sub-Sub-TLV (opt)          |
:                                                               :
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|         Processing Capabilities List Sub-Sub-TLV (opt)        |
:                                                               :
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|         Output Modulation Type List Sub-Sub-TLV  (opt)        |
:                                                               :
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|         Output FEC Type List Sub-Sub-TLV  (opt)               |
:                                                               :
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

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

5.2. Input Modulation Format List Sub-Sub-TLV

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

Type := Input Modulation Format List

Value := A list of Modulation Format Fields

5.2.1. Modulation Format Field

Two different types of modulation format fields are defined: a standard modulation field and a vendor specific modulation field. Both start with the same 32 bit header shown below.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|S|I|      Modulation ID        |          Length               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

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

Where I bit set to 1 indicates it is an input modulation constraint and I bit set to 0 indicates it is an output modulation constraint.

Note that if an output modulation is not specified then it is implied that it is the same as the input modulation. In such case, no modulation conversion is performed.

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

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|1|1|       Modulation ID       |             Length            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Possible additional modulation parameters depending upon    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
:   the modulation ID                                           :
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Modulation ID (S bit = 1); Input modulation (I bit = 1)

Takes on the following currently defined values:

0   Reserved

1   optical tributary signal class NRZ 1.25G

2   optical tributary signal class NRZ 2.5G

3   optical tributary signal class NRZ 10G

4   optical tributary signal class NRZ 40G

5   optical tributary signal class RZ 40G

Note that future modulation types may require additional parameters in their characterization.

The format for vendor specific modulation field (for input constraint) is given by:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|0|1|  Vendor Modulation ID     |          Length               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                       Enterprise Number                       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
:   Any vendor specific additional modulation parameters        :
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Vendor Modulation ID

This is a vendor assigned identifier for the modulation type.

Enterprise Number

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

Vendor Specific Additional parameters

There can be potentially additional parameters characterizing the vendor specific modulation.

5.3. Input FEC Type List Sub-Sub-TLV

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

Type := Input FEC Type field List

Value := A list of FEC type Fields

5.3.1. FEC Type Field

The FEC type Field may consist of two different formats of fields: a standard FEC field or a vendor specific FEC field. Both start with the same 32 bit header shown below.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|S|I|      FEC ID               |          Length               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Possible additional FEC parameters depending upon           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
:   the FEC ID                                                  :
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

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

Where I bit set to 1 indicates it is an input FEC constraint and I bit set to 0 indicates it is an output FEC constraint.

Note that if an output FEC is not specified then it is implied that it is the same as the input FEC. In such case, no FEC conversion is performed.

The length is the length in bytes of the entire FEC type field.

The format for input standard FEC field is given by:

```
      0                   1                   2                   3
      0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
     |1|1|       FEC ID              |             Length            |
     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
     |   Possible additional FEC parameters depending upon           |
     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
     :   the FEC ID                                                  :
     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

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

0 Reserved

1 G.709 RS FEC

2 G.709V compliant Ultra FEC

3 G.975.1 Concatenated FEC (RS(255,239)/CSOC(n0/k0=7/6,J=8))

4 G.975.1 Concatenated FEC (BCH(3860,3824)/BCH(2040,1930))

5 G.975.1 Concatenated FEC (RS(1023,1007)/BCH(2407,1952))

6 G.975.1 Concatenated FEC (RS(1901,1855)/Extended Hamming Product Code (512,502)X(510,500))

7 G.975.1 LDPC Code

8 G.975.1 Concatenated FEC (Two orthogonally concatenated BCH codes)

9 G.975.1 RS(2720,2550)

10 G.975.1 Concatenated FEC (Two interleaved extended BCH (1020,988) codes)

Where RS stands for Reed-Solomon and BCH for Bose-Chaudhuri-Hocquengham.

The format for input vendor-specific FEC field is given by:

```
     0                   1                   2                   3
     0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |0|1|       Vendor FEC ID       |             Length            |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                       Enterprise Number                       |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    :   Any vendor specific additional FEC parameters               :
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Vendor FEC ID

This is a vendor assigned identifier for the FEC type.

Enterprise Number

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

Vendor Specific Additional FEC parameters

There can be potentially additional parameters characterizing the vendor specific FEC.

5.4. Input Bit Range List Sub-Sub-TLV

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

Type := Input Bit Range List

Value := A list of Bit Range Fields

5.4.1. Bit Range Field

The bit rate range list sub-TLV makes use of the following bit rate range field:

```
     0                   1                   2                   3
     0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                          Starting Bit Rate                    |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                            Ending Bit Rate                    |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

The starting and ending bit rates are given as 32 bit IEEE floating point numbers in bits per second. Note that the starting bit rate is less than or equal to the ending bit rate.

The bit rate range list sub-TLV is then given by:

```
     0                   1                   2                   3
     0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                                                               |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+ Bit Range Field #1  +-+-+-+-+-+-+-+-+
    |                                                               |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    :                               :                               :
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                                                               |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+ Bit Range Field #M  +-+-+-+-+-+-+-+-+
    |                                                               |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

5.5. Input Client Signal List Sub-Sub-TLV

This sub-sub-TLV contains a list of acceptable input client signal types.

Type := Input Client Signal List

Value := A list of GPIDs

The acceptable client signal list sub-TLV is a list of Generalized Protocol Identifiers (GPIDs). GPIDs are assigned by IANA and many are defined in [RFC3471] and [RFC4328].

```
      0                   1                   2                   3
      0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
     |       Number of GPIDs         |          GPID #1              |
     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
     :                               |                               :
     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
     |            GPID #N            |                               |
     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Where the number of GPIDs is an integer greater than or equal to one.

5.6. Processing Capability List Sub-Sub-TLV

This sub-sub-TLV contains a list of resource block processing capabilities.

Type := Processing Capabilities List

Value := A list of Processing Capabilities Fields

The processing capability list sub-TLV is a list of WSON network element (NE) that can perform signal processing functions including:

1. Number of Resources within the block

2. Regeneration capability

3. Fault and performance monitoring

4. Vendor Specific capability

Note that the code points for Fault and performance monitoring and vendor specific capability are subject to further study.

5.6.1. Processing Capabilities Field

The processing capability field is then given by:

```
    0                   1                   2                   3
    0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |          Processing Cap ID    |          Length               |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |   Possible additional capability parameters depending upon    |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   :   the processing ID                                           :
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

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

```
    0                   1                   2                   3
    0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |          Processing Cap ID    |          Length = 8           |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                  Number of resources per block                |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

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

```
    0                   1                   2                   3
    0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |  T  | C |                 Reserved                            |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Where T bit indicates the type of regenerator:

T=0: Reserved

T=1: 1R Regenerator

T=2: 2R Regenerator

T=3: 3R Regenerator

Where C bit indicates the capability of regenerator:

C=0: Reserved

C=1: Fixed Regeneration Point

C=2: Selective Regeneration Point

Note that when the capability of regenerator is indicated to be Selective Regeneration Pools, regeneration pool properties such as ingress and egress restrictions and availability need to be specified. This encoding is to be determined in the later revision.

5.7. Output Modulation Format List Sub-Sub-TLV

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

Type := Output Modulation Format List

Value := A list of Modulation Format Fields

5.8. Output FEC Type List Sub-Sub-TLV

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

Type := Output FEC Type field List

Value := A list of FEC type Fields

6. Security Considerations

This document defines protocol-independent encodings for WSON information and does not introduce any security issues.

However, other documents that make use of these encodings within protocol extensions need to consider the issues and risks associated with, inspection, interception, modification, or spoofing of any of this information. It is expected that any such documents will describe the necessary security measures to provide adequate protection.

7. IANA Considerations

This document provides general protocol independent information encodings. There is no IANA allocation request for the TLVs defined in this document. IANA allocation requests will be addressed in protocol specific documents based on the encodings defined here.

8. Acknowledgments

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

APPENDIX A: Encoding Examples

A.1. Wavelength Converter Accessibility Sub-TLV

Example:

Figure 1 shows a wavelength converter pool architecture know as "shared per fiber". In this case the ingress and egress pool matrices are simply:

```
              +-----+       +-----+
              | 1 1 |       | 1 0 |
          WI =|     |,  WE =|     |
              | 1 1 |       | 0 1 |
              +-----+       +-----+
```

```
                    +-----------+                      +------+
                    |           |--------------------->|      |
                    |           |--------------------->|  C   |
              /|    |           |--------------------->|  o   |
             /D+--->|           |--------------------->|  m   |
            + e+--->|           |                      |  b   |=======>
   ========>| M|    |  Optical  |    +-----------+     |  i   | Port E1
   Port I1  + u+--->|   Switch  |    |  WC Pool  |     |  n   |
             \x+--->|           |    |  +-----+  |     |  e   |
              \|    |           +----+->|WC #1|--+---->|  r   |
                    |           |    |  +-----+  |     +------+
                    |           |    |           |     +------+
              /|    |           |    |  +-----+  |     |      |
             /D+--->|           +----+->|WC #2|--+---->|  C   |
            + e+--->|           |    |  +-----+  |     |  o   |
   ========>| M|    |           |    +-----------+     |  m   |=======>
   Port I2  + u+--->|           |                      |  b   | Port E2
             \x+--->|           |--------------------->|  i   |
              \|    |           |--------------------->|  n   |
                    |           |--------------------->|  e   |
                    |           |--------------------->|  r   |
                    +-----------+                      +------+
```

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

This wavelength converter pool can be encoded as follows:

```
       0                   1                   2                   3
       0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      | Connectivity=1|                    Reserved                   |
                  Note: I1,I2 can connect to either WC1 or WC2
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |  Action=0     |0 1|0 0 0 0 0 0|            Length = 12        |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |                     Link Local Identifier = #1                |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |                     Link Local Identifier = #2                |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |  Action=0     |1|  Reserved   |            Length = 12        |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |                        RB ID = #1                             |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |                        RB ID = #2                             |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

                        Note: WC1 can only connect to E1
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |  Action=0     |1 0|0 0 0 0 0 0|            Length = 8         |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |                     Link Local Identifier = #1                |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |  Action=0     |0|  Reserved   |            Length = 8         |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |                        RB ID = #1                             |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

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

A.2. Wavelength Conversion Range Sub-TLV

Example:

This example, based on figure 1, shows how to represent the wavelength conversion range of wavelength converters. Suppose the wavelength range of input and output of WC1 and WC2 are {L1, L2, L3, L4}:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
                      Note: WC Set
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Action=0     |1| Reserved    |     Length = 12               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                           WC ID = #1                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                           WC ID = #2                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

                      Note: wavelength input range
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| 2   | Num Wavelengths = 4     |          Length = 8             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|Grid |  C.S. |     Reserved    |  n for lowest frequency = 1   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

                      Note: wavelength output range
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| 2   | Num Wavelengths = 4     |          Length = 8             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|Grid |  C.S. |     Reserved    |  n for lowest frequency = 1   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

A.3. An OEO Switch with DWDM Optics

Figure 2 shows an electronic switch fabric surrounded by DWDM optics. In this example the electronic fabric can handle either G.709 or SDH signals only (2.5 or 10 Gbps). To describe this node, the following information is needed:

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

In this case there is complete port to port connectivity so the <ConnectivityMatrix> is not required. In addition since there are sufficient ports to handle all wavelength signals the <RBPoolState> element is not needed.

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

<ResourcePool> ::=
<ResourceBlockInfo>[<ResourceBlockAccessibility>...][<ResourceWaveCo
nstraints>...]

```
           /|    +-----------+    +-------------+   +------+
          /D+--->|           +--->|Tunable Laser|-->|      |
         + e+--->|           |    +-------------+   |  C   |
========>| M|    |           |       ...            |  o   |=======>
Port I1  + u+--->|           |    +-------------+   |  m   | Port E1
          \x+--->|           |--->|Tunable Laser|-->|  b   |
           \|    |  Electric |    +-------------+   +------+
                 |   Switch  |
           /|    |           |    +-------------+   +------+
          /D+--->|           +--->|Tunable Laser|-->|      |
         + e+--->|           |    +-------------+   |  C   |
========>| M|    |           |       ...            |  o   |=======>
Port I2  + u+--->|           |    +-------------+   |  m   | Port E2
          \x+--->|           +--->|Tunable Laser|-->|  b   |
           \|    |           |    +-------------+   +------+
                 |           |
           /|    |           |    +-------------+   +------+
          /D+--->|           |--->|Tunable Laser|-->|      |
         + e+--->|           |    +-------------+   |  C   |
========>| M|    |           |       ...            |  o   |=======>
Port I3  + u+--->|           |    +-------------+   |  m   | Port E3
          \x+--->|           |--->|Tunable Laser|-->|  b   |
           \|    +-----------+    +-------------+   +------+
```

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

The resource block information will tell us about the processing constraints of the receivers, transmitters and the electronic switch. The resource availability information, although very simple, tells us that all signals must traverse the electronic fabric (fixed connectivity). The resource wavelength constraints are not needed since there are no special wavelength constraints for the resources that would not appear as port/wavelength constraints.

<ResourceBlockInfo>:

```
       0                   1                   2                   3
       0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |                     RB Set Field                              |
      :  (only one resource block in this example with shared         |
      |                     input/output case)                        |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |0|0|                             Reserved                      |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |            Input Modulation Type List Sub-Sub-TLV             |
      :         (The receivers can only process NRZ)                  :
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |            Input FEC Type List Sub-Sub-TLV                    |
      :            (Only Standard SDH and G.709 FECs   )              :
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |          Input Client Signal Type Sub-TLV                     |
      :               (GPIDs for SDH and G.709)                       :
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |          Input Bit Rate Range List  Sub-Sub-TLV               |
      :                           (2.5Gbps, 10Gbps)                   :
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |          Processing Capabilities List Sub-Sub-TLV             |
      :                     Fixed (non optional) 3R regeneration      :
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |          Output Modulation Type List Sub-Sub-TLV              |
      :                           NRZ                                 :
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |          Output FEC Type List Sub-Sub-TLV                     |
      :                  Standard SDH, G.709 FECs                     :
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Since there is fixed connectivity to resource blocks (the electronic switch) the <ResourceBlockAccessibility> is:

```
        0                   1                   2                   3
        0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       | Connectivity=1|Reserved       |                               |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |                     Ingress Link Set Field A #1               |
       :                     (All ingress links connect to resource)   :
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |                           RB Set Field A #1                   |
       :               (trivial set only one resource block)           :
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |                 Egress Link Set Field B #1                    |
       :                     (All egress links connect to resource)    :
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

9. References

9.1. Normative References

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

[RFC2578] McCloghrie, K., Perkins, D., and J. Schoenwaelder, "Structure of Management Information Version 2 (SMIv2)", STD 58, RFC 2578, April 1999.

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

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

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

9.2. Informative References

[G.694.1] ITU-T Recommendation G.694.1, Spectral grids for WDM applications: DWDM frequency grid, June 2002.

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

[Gen-Encode]   G. Bernstein, Y. Lee, D. Li, W. Imajuku, "General Network Element Constraint Encoding for GMPLS Controlled Networks", work in progress: draft-ietf-ccamp-general-constraint-encode.

[RFC6205]   T. Otani, H. Guo, K. Miyazaki, D. Caviglia, "Generalized Labels for G.694 Lambda-Switching Capable Label Switching Routers", RFC 6205, March 2011.

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

[WSON-Info] G. Bernstein, Y. Lee, D. Li, W. Imajuku, "Routing and Wavelength Assignment Information Model for Wavelength Switched Optical Networks", work in progress: draft-ietf-ccamp-rwa-info, March 2009.

10. Contributors

Diego Caviglia
Ericsson
Via A. Negrone 1/A 16153
Genoa Italy
Phone: +39 010 600 3736
Email: diego.caviglia@(marconi.com, ericsson.com)

Anders Gavler
Acreo AB
Electrum 236
SE - 164 40 Kista Sweden
Email: Anders.Gavler@acreo.se

Jonas Martensson
Acreo AB
Electrum 236
SE - 164 40 Kista, Sweden
Email: Jonas.Martensson@acreo.se

Itaru Nishioka
NEC Corp.
1753 Simonumabe, Nakahara-ku, Kawasaki, Kanagawa 211-8666
Japan
Phone: +81 44 396 3287
Email: i-nishioka@cb.jp.nec.com

Cyril Margaria
Nokia Siemens Networks
St Martin Strasse 76
Munich,   81541
Germany
Phone: +49 89 5159 16934
Email: cyril.margaria@nsn.com

Authors' Addresses

Greg M. Bernstein (ed.)
Grotto Networking
Fremont California, USA

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

Young Lee (ed.)
Huawei Technologies
1700 Alma Drive, Suite 100
Plano, TX 75075
USA

Phone: (972) 509-5599 (x2240)
Email: ylee@huawei.com

Dan Li
Huawei Technologies Co., Ltd.
F3-5-B R&D Center, Huawei Base,
Bantian, Longgang District
Shenzhen 518129 P.R.China

Phone: +86-755-28973237
Email: danli@huawei.com

Wataru Imajuku
NTT Network Innovation Labs
1-1 Hikari-no-oka, Yokosuka, Kanagawa
Japan

Phone: +81-(46) 859-4315
Email: imajuku.wataru@lab.ntt.co.jp

Jianrui Han
Huawei Technologies Co., Ltd.
F3-5-B R&D Center, Huawei Base,
Bantian, Longgang District
Shenzhen 518129 P.R.China

Phone: +86-755-28972916
Email: hanjianrui@huawei.com

Intellectual Property Statement

Acknowledgment

Funding for the RFC Editor function is currently provided by the Internet Society.

Network Working Group
Internet Draft
Intended status: Standards Track
Expires: April 2012

Y. Lee
Huawei
G. Bernstein
Grotto Networking

October 30, 2011

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

draft-ietf-ccamp-wson-signal-compatibility-ospf-07.txt

Status of this Memo

Copyright Notice

Abstract

This document provides GMPLS OSPF routing enhancements to support signal compatibility constraints associated with WSON network elements. These routing enhancements are required in common optical or hybrid electro-optical networks where not all of the optical signals in the network are compatible with all network elements participating in the network.

This compatibility constraint model is applicable to common optical or hybrid electro optical systems such as OEO switches, regenerators, and wavelength converters since such systems can be limited to processing only certain types of WSON signals.

Conventions used in this document

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

Table of Contents

1. Introduction

The documents [RFC6163, WSON-Info, WSON-Encode] explain how to extend the wavelength switched optical network (WSON) control plane to allow both multiple WSON signal types and common hybrid electro optical systems as well hybrid systems containing optical switching and electro-optical resources. In WSON, not all of the optical signals in the network are compatible with all network elements participating in the network. Therefore, signal compatibility is an important constraint in path computation in a WSON.

This document provides GMPLS OSPF routing enhancements to support signal compatibility constraints associated with general WSON network elements. These routing enhancements are required in common optical or hybrid electro-optical networks where not all of the optical signals in the network are compatible with all network elements participating in the network.

This compatibility constraint model is applicable to common optical or hybrid electro optical systems such as OEO switches, regenerators, and wavelength converters since such systems can be limited to processing only certain types of WSON signals.

1.1. Revision History

From 00 to 01: The details of the encodings for compatibility moved from this document to [WSON-Encode].

From 01 to 02: Editorial changes.

From 02 to 03: Add a new Top Level Node TLV, Optical Node Property TLV to carry WSON specific node information.

From 03 to 04: Add a new sub-TLV, Block Shared Access Wavelength Availability TLV to be consistent with [WSON-Encode] and editorial changes.

From 04 to 05: Add a new section that discusses OSPF scalability and timeliness and editorial changes.

From 05 to 06: Change the title of the draft to "GMPLS OSPF Enhancement" from "OSPF Enhancement" to make sure the changes apply to the GMPLS OSPF rather than the base OSPF. Add specific OSPF procedures on how sub-TLVs are packaged per [RFC3630] and editorial changes.

From 06 to 07: Add clarifying texts on how to sub-divide the Optical Node TLV in case it exceeds the IP MTU fragmentation limit. Delete Section 3.2. to avoid multiple rules so as to avoid confusion.

2. The Optical Node Property TLV

[RFC3630] defines OSPF TE LSA using an opaque LSA. This document adds a new top level TLV for use in the OSPF TE LSA: the Optical Node Property TLV. The Optical Node property TLV describes a single node. It is constructed of a set of sub-TLVs. There are no ordering requirements for the sub-TLVs. Only one Optical Node TLV shall be advertised in each LSA.

The Optical Node Property TLV contains all WSON-specific node properties and signal compatibility constraints. The detailed encodings of these properties are defined in [WSON-Encode].

The following sub-TLVs of the Optical Node Property TLV are defined:

Value	Length	Sub-TLV Type

TBA	variable	Resource Block Information
TBA	variable	Resource Pool Accessibility
TBA	variable	Resource Block Wavelength Constraints
TBA	variable	Resource Pool State
TBA	variable	Block Shared Access Wavelength Availability

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

Sub-TLV Type	Section [WSON-Encode]
Resource Block Information	4.1
Resource Pool Accessibility	3.1
Resource Block Wavelength Constraints	3.2
Resource Pool State	3.3
Block Shared Access Wavelength Availability	3.4

All sub-TLVs defined here may occur at most once in any given Optical Node TLV. "At most once" means that if there is sub-TLV related information, it should be always included. These restrictions need not apply to future sub-TLVs. Unrecognized sub-TLVs are ignored.

2.1. Sub-TLV Details

Among the sub-TLVs defined above, the Resource Pool State sub-TLV and Block Shared Access Wavelength Availability are dynamic in nature while the rest are static. As such, they can be separated out from the rest and be advertised with multiple TE LSAs per OSPF router, as described in [RFC3630] and [RFC5250].

2.1.1. Resource Block Information

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

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

Value   Length   Sub-TLV Type

TBA	variable	Input Modulation Format List
TBA	variable	Input FEC Type List
TBA	variable	Input Bit Range List
TBA	variable	Input Client Signal List
TBA	variable	Processing Capability List
TBA	variable	Output Modulation Format List
TBA	variable	Output FEC Type List

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

Name	Section [WSON-Encode]
Input Modulation Format List	4.2
Input FEC Type List	4.3
Input Bit Range List	4.4
Input Client Signal List	4.5
Processing Capability List	4.6
Output Modulation Format List	4.7
Output FEC Type List	4.8

2.1.2. Resource Pool Accessibility

This sub-TLV describes the structure of the resource pool in relation to the switching device. In particular it indicates the ability of an ingress port to reach a resource block and of a resource block to reach a particular egress port.

2.1.3. Resource Block Wavelength Constraints

Resources, such as wavelength converters, etc., may have a limited input or output wavelength ranges. Additionally, due to the structure of the optical system not all wavelengths can necessarily reach or leave all the resources. Resource Block Wavelength Constraints sub-TLV describe these properties.

2.1.4. Resource Pool State

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

2.1.5. Block Shared Access Wavelength Availability

Resources blocks may be accessed via a shared fiber. If this is the case then wavelength availability on these shared fibers is needed to understand resource availability.

3. WSON Specific Scalability and Timeliness

This document has defined five sub-TLVs specific to WSON. The examples given in [WSON-Encode] show that very large systems, in terms of channel count, ports, or resources, can be very efficiently encoded. However there has been concern expressed that some possible systems may produce LSAs that exceed the IP Maximum Transmission Unit (MTU) and that methods be given to allow for the splitting of WSON specific LSA into smaller LSA that are under the MTU limit. This section presents a set of techniques that can be used for this purpose.

3.1. Different Sub-TLVs into Multiple LSAs

Five sub-TLVs are defined in this document:

1. Resource Block Information
2. Resource Pool Accessibility
3. Resource Block Wavelength Constraints
4. Resource Pool State
5. Block Shared Access Wavelength Availability

All these are carried in an Optical Node Property TLV (see Section 2 for detail) of which there can be at most one in an LSA. Of these sub-TLVs the first three are relatively static, i.e., only would change with hardware changes or significant system reconfiguration. While the fourth and fifth are dynamic, meaning that they may change with LSP setup or teardown through the system. The most important technique for scalability and OSPF bandwidth reduction is to separate the dynamic information sub-TLVs from the static information sub-TLVs
and advertise them in OSPF TE LSAs, each with the Optical Node Property TLV at the top level ([RFC3630 and RFC5250]).

For LSA overhead reduction it is recommended to group as many of the three static sub-TLVs into the same LSA (within the Optical Node Property TLV). If the size of this LSA is greater than the MTU, then these sub-TLV can be packed into separate LSAs. From the point of view of path computation, the presence of the Resource Block Information sub-TLV indicates that resources exist in the system and may have signal compatibility or other constraints. The other four sub-TLVs indicate constraints on access to, and availability of those resources.

Hence the "synchronization" procedure from a path computation point of view is quite simple. Until a Resource Block Information sub-TLV is received for a system path cannot make use of the other four sub-TLVs since it does not know the nature of the resources, e.g., are the resources wavelength converters, regenerators, or something else. Once this sub-TLV is received path computation can proceed with whatever of the additional types of sub-TLVs it may have received (there use is dependent upon the system type). If path computation proceeds with out of date or missing information from these sub-TLVs then there is the possibility of either (a) path computation computing a path that does not exist in the network, (b) path computation failing to find a path through the network that actually exists. Both situations are currently encountered with GMPLS, i.e., out of date information on constraints or resource availability.

Note that the connection establishment mechanism (signaling or management) is ultimately responsible for the establishment of the connection, and this implies that such mechanisms must insure signal compatibility.

3.2. Separating a Sub-TLV into Multiple OSPF TE LSAs

In the highly unlikely event that a WSON sub-TLV by itself would result in an LSA exceeding the MTU, all five WSON specific sub-TLVs in this document provide mechanisms that allow them to be subdivided into smaller sub-TLVs that can be sent in separate OSPF TE LSAs.

What is suggested as below is the only option allowed when dividing up the current set of sub-TLVs into separate OSPF TE LSAs. This means each sub-TLV will be packaged as the sole element in an OSPF TE LSA with a unique Link State ID. When such division is implemented, then the source node must flush the existing LSA (i.e., the original OSPF TE LSA with all sub-TLV’s packaged together as described in Section 2). This will avoid duplicating the same information being advertised across multiple LSAs.

Sub-Division by Sets

All five sub-TLVs currently make use of one or more RB Set Fields [WSON-Encode] or Link Set Fields [Gen-Encode]. Long set fields can be decomposed into multiple smaller set fields resulting in multiple sub-TLVs that can be sent in multiple OSPF TE LSAs. The interpretation of the separate pieces is quite natural and reviewed in the following:

Resource Block Information
Information about different resources of similar types would get sent separately (LSAs). Path computation would not know a resource exists until it receives the instance of a sub-TLV that mentions that instance.

Resource Pool Accessibility
Information about accessibility to resources to/from ports would be in as separate pieces base on port or resource set separation. All pieces are combined to give complete resource/port accessibility view. Late/missing pieces would imply resources are not accessible to/from given ports.

Resource Block Wavelength Constraints
Information about resource wavelength constraints can be sent in separate pieces based on resource sub-sets. Late/missing pieces (LSAs) would imply resources accessible when they might not be.

Resource Pool State
Information about resource state can be sent in separate pieces based on resource sub-sets. Late/missing pieces (LSAs) could imply incorrect resources availability.

Block Shared Access Wavelength Availability
Information about resource shared access wavelength can be sent in separate pieces based on resource sub-sets. Late/missing pieces (LSAs) could imply incorrect shared wavelength availability.

Due to the reliability mechanisms in OSPF the phenomena of late or missing pieces for relatively static information (first three types of sub-TLVs) would be relatively rare.

4. Security Considerations

This document does not introduce any further security issues other than those discussed in [RFC3630], [RFC4203].

5. IANA Considerations

This document introduces a new Top Level Node TLV (Optical Node Property TLV) under the OSPF TE LSA defined in [RFC3630].

Value	TLV Type
TBA	Optical Node Property

IANA is to allocate a new TLV Type and its Value for this Top Level Node TLV.

This document also introduces the following sub-TLVs associated with the Optical Node Property TLV as defined in Section 2.1 as follows:

Value	Length	Sub-TLV Type
TBA	variable	Resource Block Information
TBA	variable	Resource Pool Accessibility
TBA	variable	Resource Block Wavelength Constraints
TBA	variable	Resource Pool State
TBA	variable	Block Shared Access Wavelength Availability

IANA is to allocate new sub-TLV Types and their Values for these sub-TLVs defined under the Optical Node Property TLV.

There are seven nested sub-TLVs defined in the Resource Block Information sub-TLV as follows:

Value	Length	Sub-TLV Type
TBA	variable	Input Modulation Format List
TBA	variable	Input FEC Type List
TBA	variable	Input Bit Range List
TBA	variable	Input Client Signal List
TBA	variable	Processing Capability List
TBA	variable	Output Modulation Format List

TBA   variable   Output FEC Type List

IANA is to allocate new Sub-TLV Types and their Values for these Sub-TLVs defined under the Resource Block Information Sub-TLV.

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.

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

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

[RFC4203] Kompella, K., Ed., and Y. Rekhter, Ed., "OSPF Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4203, October 2005.

[WSON-Encode] G. Bernstein, Y. Lee, D. Li, W. Imajuku, "Routing and Wavelength Assignment Information Encoding for Wavelength Switched Optical Networks", draft-ietf-ccamp-rwa-wson-encode, work in progress.

[Gen-Encode] G. Bernstein, Y. Lee, D. Li, W. Imajuku, "General Network Element Constraint Encoding for GMPLS Controlled Networks", draft-ietf-ccamp-general-constraint-encode, work in progress.

6.2. Informative References

[WSON-Info] Y. Lee, G. Bernstein, D. Li, W. Imajuku, "Routing and Wavelength Assignment Information Model for Wavelength Switched Optical Networks", draft-ietf-ccamp-rwa-info, work in progress.

[RFC6250] T. Otani, Ed., D. Li, Ed., "Generalized Labels for G.694 Lambda-Switching Capable Label Switching Routers", RFC 6250, March 2011.

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

[RFC5250] Berger, L., et al., "The OSPF Opauqe LSA option", RFC 5250, July 2008.

7. Authors and Contributors

Authors' Addresses

Young Lee (ed.)
Huawei Technologies
1700 Alma Drive, Suite 100
Plano, TX 75075
USA

Phone: (972) 509-5599 (x2240)
Email: ylee@huawei.com

Greg M. Bernstein (ed.)
Grotto Networking
Fremont California, USA

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

Intellectual Property Statement

Disclaimer of Validity

Acknowledgment

Funding for the RFC Editor function is currently provided by the Internet Society.

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: January 3, 2012

W. Sun
SJTU
T. Nadeau
Lucidvision
M. Morrow
Cisco Systems
G. Zhang
CATR
W. Hu
SJTU
July 2, 2011

Label Switched Path (LSP) Provisioning Performance Management Information Base for Generalized MPLS (GMPLS) / MPLS-TE networks

draft-sun-ccamp-gmpls-perf-mib-00.txt

Abstract

This memo defines Management Information Bases (MIBs) for performances of provisioning Label Switched Paths (LSPs) in Generalized MPLS or MPLS-TE networks.

When Generalized MPLS/MPLS-TE is used to provision LSPs, it is useful to record the performance of the provisioning process, such as the delay in creating and deleting the LSPs. The managed information may be retrieved by the Management System and visualized on the GUI, so that the performance of dynamic provisioning may be monitored in a timely manner.

This work is a continuation of the work in [RFC5814] and [I-D.ietf-ccamp-dpm], where the provisioning performance values are obtained through active measurements.

Status of this Memo

This Internet-Draft will expire on January 3, 2012.

Copyright Notice

Table of Contents

1. Introduction

When Label Switched Paths (LSPs) are provisioned dynamically within an operational network, it is helpful to monitor and record the related performance information, such as the experienced provisioning delay and error events. Such information may help operators to ensure correct operation of dynamic LSP provisioning in their network, or possibly identify performance degradation in the control plane.

This memo defines a set of objects that can reveal the performance of an operational network in terms of dynamic LSP provisioning. It is intended to complement the performance objects, such as the number of packets received and sent, per LSP tunnel, in [RFC3812] and [RFC4802].

Unlike the work in [RFC5814] and [I-D.ietf-ccamp-dpm], where the performance values are obtained through active measurements, this document focuses on the performance values in operational environments. The actual value of the performance in this document is recorded only when an LSP is provisioned, and is thus collected passively. Hence such information reflects only the performance at specific and discrete times. However, when properly used, they can be helpful in identifying performance degradation, or even malfunctioning, in the network control plane.

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. 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 [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].

4. Brief Description of LSP performance MIB Objects

4.1. gmplsPerfMaxEntries

Defines the maximum number of rows stored in the gmplsPerfTable. An implementation MUST start assigning gmplsPerfEntryIndex values at 1 and wrap after exceeding the maximum possible value, as defined by the limit of this object.

4.2. gmplsPerfTunnelConfigured

Defines the The total number of tunnels configured.

4.3. gmplsPerfTable

The performance of past LSP provisioning process is stored in this table. To handle possible provisioning failures, start and complete timestamp of a provisioning operation is recorded. For example, for LSP creation process, the timestamps of creation initiation and completion are recorded seperatly. It is up to the users to determine the actual performance value, or identify a possible creation/deletion failure. The maximum number of entries stored in this table is determined by the value of gmplsPerfMaxEntries.

5. GMPLS Performance MIB Module

```
GMPLS-PROV-PERF-STD-MIB DEFINITIONS ::= BEGIN

IMPORTS
    gmplsTeStdMIB
        FROM GMPLS-TE-STD-MIB
    mplsStdMIB,
    MplsTunnelIndex,
    MplsExtendedTunnelId
        FROM MPLS-TC-STD-MIB            -- RFC 3811
    TimeStamp
        FROM SNMPv2-TC
    MODULE-IDENTITY, OBJECT-TYPE,
    Gauge32, Unsigned32
        FROM SNMPv2-SMI
    OBJECT-GROUP
        FROM SNMPv2-CONF;

gmplsPerfMIB MODULE-IDENTITY
    LAST-UPDATED "201104180654Z"    -- Apr 18, 2011 6:54:00 AM
    ORGANIZATION "IETF Common Control and Measurement Plane Working
Group"
    CONTACT-INFO
        "Weiqiang Sun
        Shanghai Jiao Tong University (SJTU)
        Email: sunwq@mit.edu

        Thomas D. Nadeau
        Email: thomas.nadeau@huawei.com"
    DESCRIPTION
        "Copyright (C) The Internet Society (2011).  This version of
        this MIB module is part of RFC XXX; see the RFC itself for
        full legal notices.

        This MIB module defines managed object definitions
        for dynamic LSP provisioning."
    REVISION "201104180654Z"        -- Apr 18, 2011 6:54:00 AM
    DESCRIPTION
        "Initial version."
    -- 1.3.6.1.2.1.10.166.13.1
    ::= { gmplsTeStdMIB 1 }


gmplsPerfTunnelConfigured OBJECT-TYPE
    SYNTAX  Unsigned32
    MAX-ACCESS read-only
    STATUS  current
```

```
        DESCRIPTION
            "The total number of tunnels configured."
        -- 1.3.6.1.2.1.10.166.13.1.3
        ::= { gmplsPerfMIB 3 }


    gmplsPerfMaxEntries OBJECT-TYPE
        SYNTAX  Gauge32
        MAX-ACCESS read-only
        STATUS  current
        DESCRIPTION
            "An implementation MUST start assigning gmplsPerfEntryIndex
            values at 1 and wrap after exceeding the maximum possible
            value, as defined by the limit of this object.

            A value of 0 for this object disables creation of
            gmplsPerfEntry."
        -- 1.3.6.1.2.1.10.166.13.1.2
        ::= { gmplsPerfMIB 2 }


    --Performance Table

    gmplsPerfTable OBJECT-TYPE
        SYNTAX  SEQUENCE OF GmplsPerfEntry
        MAX-ACCESS not-accessible
        STATUS  current
        DESCRIPTION
            "Defines a table for storing the results of LSP
            provisioning operations. It allows the provisioning
            performance be retrieved later for monitoring or
            diagnostic purposes. The recorded performance information
            is intended to complement the existing performance
            statistics in the MPLS-TE-STD-MIB and GMPLS-TE-STD-MIB.

            Note that the creation and tear-down operation performances
            are stored in one table, ie., gmplsPerfTable. When an LSP
            tunnel creation operation is  initiated, an entry MUST be
            added in this table and Tunnel ID as well as the time of
            initiation MUST be recorded. Upon completion of the creation
            process, ie., a positive signaling feedback is received by
            the ingress LSR,  this complete time object in this entry
            MUST be updated.

            When an LSP tunnel deletion process is initiated, the
            corresponding entry with the same tunnel ID MUST be located
            and updated with time of the deletion initiation time. When
            the deletion operation is complete, the entry MUST again
```

```
            be updated with the completion time.

            Under circumstances that the creation or deletion operation
            may fail, an entry may be partially updated. Eg., when a
            creation operation timeouts without a positive signaling
            feedback, the creation completion time may never be updated.
            When a tear-down operation is caused by nodes other than the
            Ingress LSR, the tear-down start time may not be known to
            the ingress LSR. In such cases, the user of the MIB MUST be
            aware of such events and treat the performance information
            accordingly.

            The number of entries in this table is limited by the value
            of the corresponding gmplsPerfMaxEntries object. An
            implementation MUST start assigning gmplsPerfEntryIndex at
            1 and wrap after exceeding the maximum possible value, as
            defined by the limit of gmplsPerfMaxEntries. An
            implementation of this MIB will remove the oldest entry
            in the gmplsPerfTable to allow the addition of a new entry
            once the number of rows in the gmplsPerfTable reaches the
            value specified by gmplsPerfMaxEntries."
        -- 1.3.6.1.2.1.10.166.13.1.1
        ::= { gmplsPerfMIB 1 }


    gmplsPerfEntry OBJECT-TYPE
        SYNTAX  GmplsPerfEntry
        MAX-ACCESS not-accessible
        STATUS  current
        DESCRIPTION
            "Defines an entry in the gmplsPerfTable. An entry can be
            created when an LSP tunnel is signaled. An implementation
            of this MIB MAY choose to disable the creation of
            performance entry, when an LSP is provisioned through
            SNMP."
        INDEX {
            gmplsPerfEntryIndex,
            gmplsPerfTunnelID }
        -- 1.3.6.1.2.1.10.166.13.1.1.1
        ::= { gmplsPerfTable 1 }


    GmplsPerfEntry ::= SEQUENCE {

        gmplsPerfEntryIndex          Gauge32,
        gmplsPerfTunnelID            MplsTunnelIndex,
        gmplsPerfCurrentStatus       INTEGER,
        gmplsPerfSrcID               MplsExtendedTunnelId,
```

```
        gmplsPerfDstID              MplsExtendedTunnelId,
        gmplsPerfCreateStartTime    TimeStamp,
        gmplsPerfCreateCompleteTime TimeStamp,
        gmplsPerfDeleteStartTime    TimeStamp,
        gmplsPerfDeleteCompleteTime TimeStamp }


    gmplsPerfEntryIndex OBJECT-TYPE
        SYNTAX  Gauge32
        MAX-ACCESS not-accessible
        STATUS  current
        DESCRIPTION
            "The index of the performance entry. The number of entries
            in this table is limited by the value of the corresponding
            gmplsPerfMaxEntries object. An implementation MUST start
            assigning gmplsPerfEntryIndex at 1 and wrap after exceeding
            the maximum possible value, as defined by the limit of
            gmplsPerfMaxEntries. An implementation of this MIB will
            remove the oldest entry in the gmplsPerfTable to allow the
            addition of an new entry once the number of rows in the
            gmplsPerfTable reaches the value specified by
            gmplsPerfMaxEntries."
        -- 1.3.6.1.2.1.10.166.13.1.1.1.1
        ::= { gmplsPerfEntry 1 }


    gmplsPerfTunnelID OBJECT-TYPE
        SYNTAX  MplsTunnelIndex
        MAX-ACCESS read-create
        STATUS  current
        DESCRIPTION
            "The ID of the tunnel being provisioned."
        REFERENCE
            "RFC 3812"
        -- 1.3.6.1.2.1.10.166.13.1.1.1.2
        ::= { gmplsPerfEntry 2 }


    gmplsPerfCurrentStatus OBJECT-TYPE
        SYNTAX  INTEGER {
            CreationInProgress(0),
            Up(1),
            DeletionInProgress(2),
            Deleted(3) }
        MAX-ACCESS read-create
        STATUS  current
        DESCRIPTION
            "This object defines the current status of the LSP tunnel.
```

```
            CreationInProgress
              The corresponding LSP tunnel is being created, but
              the creation operation has not finished yet.

            Up
              The corresponding LSP tunnel has been created
              successfully.

            DeletionInProgress
              The corresponding LSP tunnel is being deleted, but
              the deletion process has not finished yet.

            Deleted
              The corresponding LSP tunnel has been deleted."
        -- 1.3.6.1.2.1.10.166.13.1.1.1.3
        ::= { gmplsPerfEntry 3 }


    gmplsPerfSrcID OBJECT-TYPE
        SYNTAX  MplsExtendedTunnelId
        MAX-ACCESS read-create
        STATUS  current
        DESCRIPTION
            "The address of the ingress LSR ID."
        -- 1.3.6.1.2.1.10.166.13.1.1.1.5
        ::= { gmplsPerfEntry 5 }


    gmplsPerfDstID OBJECT-TYPE
        SYNTAX  MplsExtendedTunnelId
        MAX-ACCESS read-create
        STATUS  current
        DESCRIPTION
            "The address of the egress LSR ID."
        -- 1.3.6.1.2.1.10.166.13.1.1.1.6
        ::= { gmplsPerfEntry 6 }


    gmplsPerfCreateStartTime OBJECT-TYPE
        SYNTAX  TimeStamp
        MAX-ACCESS read-create
        STATUS  current
        DESCRIPTION
            "The time when the tunnel setup operation is
            initiated."
        -- 1.3.6.1.2.1.10.166.13.1.1.1.7
        ::= { gmplsPerfEntry 7 }
```

```
    gmplsPerfCreateCompleteTime OBJECT-TYPE
        SYNTAX  TimeStamp
        MAX-ACCESS read-create
        STATUS  current
        DESCRIPTION
            "The time when the LSP tunnel create operation
            is complete."
        -- 1.3.6.1.2.1.10.166.13.1.1.1.8
        ::= { gmplsPerfEntry 8 }


    gmplsPerfDeleteStartTime OBJECT-TYPE
        SYNTAX  TimeStamp
        MAX-ACCESS read-create
        STATUS  current
        DESCRIPTION
            "The time when the LSP Tunnel tear-down operation
            is initiated."
        -- 1.3.6.1.2.1.10.166.13.1.1.1.9
        ::= { gmplsPerfEntry 9 }


    gmplsPerfDeleteCompleteTime OBJECT-TYPE
        SYNTAX  TimeStamp
        MAX-ACCESS read-create
        STATUS  current
        DESCRIPTION
            "The time when an LSP tear-down operation
            is complete."
        -- 1.3.6.1.2.1.10.166.13.1.1.1.10
        ::= { gmplsPerfEntry 10 }


    gmplsPerfGroups OBJECT IDENTIFIER
        -- 1.3.6.1.2.1.10.166.13.1.4
        ::= { gmplsPerfMIB 4 }

    gmplsDeletionGroup OBJECT-GROUP
        OBJECTS {
            gmplsPerfTunnelID,
            gmplsPerfCurrentStatus,
            gmplsPerfSrcID,
            gmplsPerfDstID,
            gmplsPerfDeleteStartTime,
            gmplsPerfDeleteCompleteTime }
        STATUS  current
        DESCRIPTION
            "The group of object that constitute the LSP tunnel
```

```
            deletion performance."
        -- 1.3.6.1.2.1.10.166.13.1.4.1
        ::= { gmplsPerfGroups 1 }

    gmplsCreationGroup OBJECT-GROUP
        OBJECTS {
            gmplsPerfTunnelID,
            gmplsPerfCurrentStatus,
            gmplsPerfSrcID,
            gmplsPerfDstID,
            gmplsPerfCreateStartTime,
            gmplsPerfCreateCompleteTime }
        STATUS  current
        DESCRIPTION
            "The group of object that constitute the LSP tunnel
            creation performance."
        -- 1.3.6.1.2.1.10.166.13.1.4.2
        ::= { gmplsPerfGroups 2 }

    gmplsPerfBasicGroup OBJECT-GROUP
        OBJECTS {
            gmplsPerfEntryIndex,
            gmplsPerfTunnelID,
            gmplsPerfMaxEntries,
            gmplsPerfCurrentStatus,
            gmplsPerfCreateStartTime,
            gmplsPerfCreateCompleteTime,
            gmplsPerfDeleteStartTime,
            gmplsPerfDeleteCompleteTime,
            gmplsPerfDstID,
            gmplsPerfSrcID,
            gmplsPerfTunnelConfigured,
            gmplsPerfErrThreshold }
        STATUS  current
        DESCRIPTION
            "Basic objects."
        -- 1.3.6.1.2.1.10.166.13.1.4.3
        ::= { gmplsPerfGroups 3 }

    END
```

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.

[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.

[RFC3410] Case, J., Mundy, R., Partain, D., and B. Stewart, "Introduction and Applicability Statements for Internet-Standard Management Framework", RFC 3410, December 2002.

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

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

6.2. Informative References

[I-D.ietf-ccamp-dpm]
Sun, W. and G. Zhang, "Label Switched Path (LSP) Data Path Delay Metrics in Generalized MPLS/ MPLS-TE Networks", draft-ietf-ccamp-dpm-03 (work in progress), May 2011.

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

Authors' Addresses

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

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

Thomas D. Nadeau
Lucidvision

Email: tnadeau@lucidvision.com

Monique Morrow
Cisco Systems
Richistrasse 7
CH-8304 Zurich-Wallisellen
Switzerland

Phone: +41 44 878 9412
Email: mmorrow@cisco.com

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

Phone: +86-1062300106
Email: zhangguoying@mail.ritt.com.cn

Weisheng Hu
Shanghai Jiao Tong University
800 Dongchuan Road
Shanghai 200240
China

Phone: +86 21 3420 5419
Email: wshu@sjtu.edu.cn

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: April 20, 2012

Q. Wang
X. Fu
G. Liu
ZTE Corporation

Oct 18, 2011

Requirements for GMPLS Routing for ASON
draft-wang-ccamp-rfc4258bis-00

Abstract

The Generalized Multi-Protocol Label Switching (GMPLS) suite of protocols has been defined to control different switching technologies as well as different applications. These include support for requesting Time Division Multiplexing (TDM) connections including Synchronous Optical Network (SONET)/Synchronous Digital Hierarchy (SDH) and Optical Transport Networks (OTNs).

Along with the development of technology, some new routing requirements which are addressed in G.7715.1 emerge. This document concentrates on the new routing requirements addressed in G.7715.1 and placed on the GMPLS suite of protocols in order to support the capabilities and functionalities of an Automatically Switched Optical Network (ASON) as defined by the ITU-T.

Status of this Memo

Copyright Notice

Table of Contents

1. Introduction

The Generalized Multi-Protocol Label Switching (GMPLS) suite of protocols provides, among other capabilities, support for controlling different switching technologies. These include support for requesting TDM connections utilizing SONET/SDH (see [T1.105] and [G.707], respectively) as well as Optical Transport Networks (OTNs, see [G.709]). However, there are certain capabilities that are needed to support the ITU-T G.8080 control plane architecture for an Automatically Switched Optical Network (ASON). Therefore, it is desirable to understand the corresponding requirements for the GMPLS protocol suite. The ASON control plane architecture is defined in [G.8080]; ASON routing requirements are identified in [G.7715] and in [G.7715.1] for ASON link state protocols. These Recommendations apply to all [G.805] layer networks (e.g., SDH and OTN), and provide protocol-neutral functional requirements and architecture.

RFC4258 focuses on the routing requirements for the GMPLS suite of protocols to support the capabilities and functionality of ASON control planes. Along with the development of technology, some new routing technology emerged in G.7715.1, whereas not in RFC4258. So this document mainly focuses on these new routing requirements which are not addressed in RFC4258. This document does not address GMPLS applicability or GMPLS capabilities. Any protocol (in particular, routing) applicability, design, or suggested extensions are strictly outside the scope of this document.

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].

Although [RFC2119] describes interpretations of these key words in terms of protocol specifications and implementations, they are used in this document to describe design requirements for protocol extensions.

2. New Routing Requirements

Routing for transport networks is performed on a per-layer basis, where the routing paradigms MAY differ among layers and within a layer. Not all equipment supports the same set of transport layers or the same degree of connection flexibility at any given layer. A difference of the definition of layer exists between G.800 and G.8080 and each <Switching Type, Encoding Type, Signal Type> SHOULD be treated as a separate layer. While RFC4258 addresses the routing requirement that signal types can also be advertised, different attributes (e.g., link weight, resource class, SRG) SHOULD also be able to be advertised per signal type.

While RFC4258 describes in section 5.1 how to identify links that go between "technology regions" (cf. RFC4206), it doesn’t handle the identification of layers that exist within regions. For example, when refer to TDM technology region, multiple layers (e.g., VC3 vs VC4 within the TDM/SDH) MAY exist within one technology region. In order to explicitly identify these layers, Switching/termination capabilities of a link end SHOULD be introduced to address this.

Multiple adaption methods MAY exist between layers (e.g., X.86 vs GFP, 802.3 vs Ethernet V2), so there SHOULD be a mechanism introduced to identify the specific client adaption methods supported on a link end. This is a new routing requirement that is not addressed by RFC4258.

3. New Routing Attributes

The new routing requirements and attributes described in section 3 mainly pertain to link attributes, so in this draft, this section mainly focuses on the updates to the routing requirements of link attributes on the basis of RFC4258. That is to say, except the routing requirements of link attributes described in this section, the other routing requirements are in accordance with RFC4258. So before you read this draft, make sure you are familiar with RFC4258.

The section 4.1.4 of RFC4258 describes the link attributes of ASON routing architecture and requirements. Link attributes include local SNPP link ID, remote SNPP link ID and layer specific characteristic. This document has nothing to do with local SNPP link ID and remote SNPP link ID, while new routing requirements are only added to the layer specific characteristics.

Table 3 of RFC4258 lists the link characteristics which is also included followed. New routing requirements described here that are related to the corresponding link characteristics are Signal Type, Local Adaptation Support and Termination/Switching Identification.

Link Characteristics	Capability	Usage
Signal Type	REQUIRED	OPTIONAL
Link Weight	REQUIRED	OPTIONAL
Resource Class	REQUIRED	OPTIONAL
Local Connection Types	REQUIRED	OPTIONAL
Link Capacity	REQUIRED	OPTIONAL
Link Availability	OPTIONAL	OPTIONAL
Diversity Support	OPTIONAL	OPTIONAL
Local Adaptation Support	OPTIONAL	OPTIONAL
Termination/Switching Identification	OPTIONAL	OPTIONAL

Figure 1: Format of the sub-TLV

- Signal Type: This identifies the characteristic information of the layer network. Attributes of one signal SHALL be able to be advertised following the relational signal type. The signal attributes indicated here is a new routing requirement to RFC4258.

- Link Weight: This is the metric indicating the relative desirability of a particular link over another, e.g., during path computation.

- Resource Class: This corresponds to the set of administrative groups assigned by the operator to this link. A link MAY belong to zero, one, or more administrative groups.

- Local Connection Types: This attribute identifies whether the local SNP represents a Termination Connection Point (CP), a Connection Point (CP), or can be flexibly configured as a TCP.

- Link Capacity: This provides the sum of the available and potential bandwidth capacity for a particular network transport layer. Other capacity measures MAY be further considered.

- Link Availability: This represents the survivability capability such as the protection type associated with the link.

- Diversity Support: This represents diversity information such as the SRLG information associated with the link.

- Local Adaptation Support: This indicates the set of client layer adaptations supported by the TCP associated with the local SNPP. This is applicable only when the local SNP represents a TCP or can be flexibly configured as a TCP. A new requirement is that a mechanism SHOULD be introduced to identify the specific client adaption methods supported on a link end when multiple adaption methods exist between layers.

- Termination/Switching Identification: Capability introduced to handle the identification of layers that exist within regions (e.g., VC3 vs VC4 within the TDM/SDH). This is a new routing requirement to RFC4258.

4. Security Considerations

TBD

5. References

5.1. Normative References

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

[RFC5787] Papadimitriou, D., "OSPFv2 Routing Protocols Extensions for Automatically Switched Optical Network (ASON) Routing", RFC 5787, March 2010.

5.2. Informative References

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

[G.7715.1] "ASON Routing Architecture and Requirements for Link State Protocols", ITU-T Draft Rec. G.7715.1/Y.1706.1 , December 2007.

[RFC4258] Brungard, D., "Requirements for Generalized Multi-Protocol Label Switching (GMPLS) Routing for the Automatically Switched Optical Network (ASON)", RFC 4258, November 2005.

Authors' Addresses

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

Email: wang.qilei@zte.com.cn
URI: http://wwwen.zte.com.cn/

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

Phone: +8613798412242
Email: fu.xihua@zte.com.cn
URI: http://wwwen.zte.com.cn/

Guoman Liu
ZTE Corporation
No.68 ZiJingHua Road,Yuhuatai District
Nanjing 210012
P.R.China

Email: liu.guoman@zte.com.cn
URI: http://wwwen.zte.com.cn/

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: April 24, 2012

Fatai Zhang
Xiaobing Zi
Huawei
Ramon Casellas
CTTC
O. Gonzalez de Dios
Telefonica
D. Ceccarelli
Ericsson

October 24, 2011

GMPLS OSPF-TE Extensions in support of Flexible-Grid in DWDM Networks

draft-zhang-ccamp-flexible-grid-ospf-ext-00.txt

Status of this Memo

Abstract

This memo describes the OSPF-TE extensions in support of GMPLS control for flexible-grid in DWDM networks.

Conventions used in this document

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

Table of Contents

1. Introduction

[G.694.1v1] defines the Dense Wavelength Division Multiplexing (DWDM) frequency grids for 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, i.e. the frequency spacing between two allowed nominal central frequencies could be 12.5 GHz, 25 GHz, 50 GHz, 100 GHz and integer multiples of 100 GHz as defined in [G.694.1v1]. All of the wavelengths on a fiber should use different central frequencies and occupy a fixed bandwidth of frequency.

[G.FLEXIGRID], an updated version of [G.694.1v1] will be consented in December 2011 in support of flexible-grids. The terms "frequency slot (The frequency range allocated to a channel and unavailable to other channels within a flexible-grid)" and "slot width" (the full width of a frequency slot in a flexible-grid) are introduced to address flexible-grids. A channel is represented as a LSC (Lambda

Switching Capable) LSP in the control plane, i.e. a LSC LSP should occupy a frequency slot on each fiber it traverses. In the case of flexible-grid, different LSC LSPs may have different slot width on a fiber, i.e. the slot width is flexible on a fiber.

[WSON-OSPF] defines the OSPF-TE extensions for WSON networks, which focuses on the fixed grids of DWDM. [GEN-OSPF] defines OSPF-TE extensions in support of the general network element constraints under the control of GMPLS. This document describes the additional requirements and extensions of routing protocol brought by flexible-grid.

This document uses the fiber link model which is shown in [FLEXIBLE-REQ] to describe the requirement and extensions for routing. The flexible-grid related terminologies can also refer to [FLEXIBLE-REQ].

2. Terminology

Flexible Grid: See [FLEXIBLE-REQ].

Frequency Slot Width: See [FLEXIBLE-REQ].

Frequency Range: See [FLEXIBLE-REQ].

SSON: Spectrum-Switched Optical Networks; See [FLEXIBLE-REQ].

LSC SS-LSP or flexi-LSP (Lambda Switch Capable Spectrum-Switched Label Switched Path): a control plane construct that represents a data plane connection in which the switching involves a frequency slot of a variable (flexible) slot width. The mapped client signal is transported over the slot width, using spectrum efficient modulations such as Coherent Optical Orthogonal Frequency Division Multiplexing (CO-OFDM).

3. Requirements for Flexible-grid Routing

As described in [FLEXIBLE-REQ], the main changes for routing brought by flexible-grid are related to the DWDM links.

3.1. Available Frequency Ranges on the Flexible-Grid DWDM Links

In the case of flexible-grids, the central frequency steps from 193.1 THz with 6.25 GHz granularity. The central frequency is calculated as follows:

Central Frequency = 193.1 THz + n * 0.00625 THz

Where n is a positive or negative integer including 0.

Different LSC LSPs could occupy frequency slots with different slot width. The frequency slot width of a LSC LSP is defined as follows:

Slot width = 0.0125 THz * m

Where m is a positive integer.

The frequency slot of a LSP can be determined by the slot width and central frequency as follows.

```
Lowest frequency = (central frequency) - (slot width)/2

                 = (193.1 + n * 0.00625) - (0.0125 * m)/2

                 = (193.1 + (n - m) * 0.00625) THz;

Highest frequency = (central frequency) + (slot width)/2

                 = (193.1 + n * 0.00625) + (0.0125 * m)/2

                 = (193.1 + (n + m) * 0.00625) THz;
```

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.

```
                      Frequency Slot 1   Frequency Slot 2
                        +-----------+-----------------------+
                        |           |                       |
   -9 -8 -7 -6 -5 -4 -3 -2 -1 0  1  2  3  4  5  6  7  8  9 10  11
...+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--...
                        ------------ -----------------------
                              ^                 ^
                 Central F = 193.1THz    Central F = 193.1375 THz
                  Slot width = 25 GHz    Slot width = 50 GHz
```

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 LSC LSPs.

Hence, the available frequency ranges should be advertised for the flexible-grid DWDM links. A set of non-overlapping available frequency ranges SHOULD be disseminated in order to allow efficient resource management of Flexible-grid DWDM links and RSA procedures which are described in section 4 of [FLEXIBLE-REQ].

3.2. Comparison with Fixed-grid DWDM Links

In case of fixed-grid DWDM links, each wavelength has a pre-defined central frequency and all the wavelengths occupy the same frequency range (channel spacing). Hence all the wavelengths in the DWDM links can be identified uniquely and the status (available or not) of the wavelengths can be advertised through routing protocol.

```
     W(-2)  |    W(-1)  |    W(0)   |    W(1)   |     W(2)  |
...---------+-----------+-----------+-----------+-----------+----...
50 GHz  |   50 GHz  |  50 GHz   |  50 GHz   |   50 GHz  |

    n=-2        n=-1        n=0         n=1         n=2
...---+-----------+-----------+-----------+-----------+---------...
                              ^
                 Central F = 193.1THz
```

Figure 2 - A Link supports Fixed Wavelengths with 50 GHz Channel Spacing

Figure 2 shows a link that supports fixed-grid with 50 GHz channel spacing. The central frequencies of the wavelengths are pre-defined by ’n’ and each wavelength occupies a fixed 50 GHz frequency range as described in [G.694.1v1].

Different from the fixed-grid DWDM links, the slot width of the wavelengths are flexible on a flexible-grid DWDM link as described in section 2.1, i.e., the value of m in the formula is uncertain before a frequency slot is allocated. So, the available frequency ranges instead of the specific "wavelengths" should be advertised for a flexible-grid DWDM link.

4. Extensions

As described in [FLEXIBLE-REQ], the network connectivity topology constructed by the links/nodes and node capabilities are the same as

WSON which can be advertised by GMPLS routing protocol (refer to section 6.2 of [RFC6163]. In case of flexible-grid, the available frequency ranges instead of the specific ''wavelengths'' should be advertised for the link, which is different from the fixed grid DWDM. This section defines the GMPLS OSPF-TE extensions in support of advertising the available frequency ranges for the flexible-grid DWDM links.

4.1. Available Labels Set sub-TLV

As described in section 2.1, the available frequency ranges other than the available frequency slots should be advertised for the flexible-grid DWDM links. The Available Labels Set sub-TLV defined in [GEN-OSPF] can be re-used to advertise the available frequency ranges for the flexible-grid DWDM links.

To make the encoding efficiently, the inclusive/exclusive label ranges format of Available Labels Set sub-TLV defined in [GEN-OSPF] can be used for specifying the frequency ranges of the flexible-grid DWDM links.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|2 or 3 | Num Labels(not used)  |            Length             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                    Start Label                                |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     End Label                                 |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Note that it needs multiple Available Labels Set sub-TLVs if there are multiple discontinuous frequency ranges on a link.

The fields of Start Label and End Label specify the lowest frequency and highest frequency of a frequency range. The label format defined in [FLEXIBLE-SIG] shown below can be used to encode the Start Label and End Label:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|Grid | C.S.  |    Identifier   |              n                |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

In case of Grid=1 (ITU-T DWDM), a new value of C.S. is defined for flexible 6.25 GHz grid.

If the C.S. is 6.25 GHz in an Available Labels Set sub-TLV, it means that the corresponding link supports flexible-grid and the Start Label/End Label specifies the frequency range of the link.

[Editors' Note: the other formats of Label set (e.g., Inclusive/Exclusive Label Lists and Bitmap Label Set) can also be used to specify the frequency ranges for the flexible-grid DWDM links.]

4.2. Examples

Figure 3 shows an example of a flexible-grid DWDM link which is traversed by two LSC LSPs.

```
                      Frequency Slot 1     Frequency Slot 2
                        -------------     -------------------
                        |           |     |                 |
   -9 -8 -7 -6 -5 -4 -3 -2 -1 0  1  2  3  4  5  6  7  8  9 10  11
...+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--...
                        -------------     -------------------
                              ^                    ^
                 Central F = 193.1THz    Central F = 193.14375 THz
                  Slot width = 25 GHz    Slot width = 37.5 GHz
```

Figure 3 - Two Frequency Slots on a Link

The available frequency resource of the link could be advertised as follows:

<Available Labels> sub-TLV:

o Exclusive Range 1: [Start label = 193.1 + (-2)*0.00625,

End Label = 193.1 + 2*0.00625]

o Exclusive Range 2: [Start label = 193.1 + 4*0.00625,

End Label = 193.1 + 10*0.00625]

It is noted that the central frequency denoted by n=3 is available for a LSC LSP with 12.5 GHz slot width request but unavailable for a LSC LSP with a wider slot width request.

- o If a LSC LSP which requires a 12.5 GHz width frequency slot is requested on this link, the central frequency denoted by n=3 is available because the corresponding frequency slot [n=2, n=4] dose not overlap the existing LSPs (the unavailable frequency ranges is [n=-2, n=2] and [n=4, n=10]).

- o If a LSC LSP which requires a 25 GHz width frequency slot is requested on this link, the central frequency denoted by n=3 is unavailable because the corresponding frequency slot [n=1, n=5] overlaps the unavailable central frequencies (the unavailable frequency ranges is [n=-2, n=2] and [n=4, n=10]).

5. IANA Considerations

TBD.

6. Security Considerations

This document does not introduce any further security issues other than those discussed in [RFC3630], [RFC4203].

7. References

[RFC2119] S. Bradner, "Key words for use in RFCs to indicate requirements levels", RFC 2119, March 1997.

[G.694.1v1] ITU-T Recommendation G.694.1, Spectral grids for WDM applications: DWDM frequency grid, June 2002.

[G.FLEXIGRID] Draft revised G.694.1 version 1.3, Unpublished ITU-T Study Group 15, Question 6.

[WSON-PCE] Y. Lee, G. Bernstein, Jonas Martensson, T. Takeda and T. Tsuritani, "PCEP Requirements for WSON Routing and Wavelength Assignment", draft-ietf-pce-wson-routing-wavelength-05, July 2011.

[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-02, September 2011.

[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-06, September 2011.

[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-02, September 2011.

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

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

[FLEXIBLE-REQ] F.Zhang et al, "Requirements for GMPLS Control of Flexible-grids",draft-zhang-ccamp-flexible-grid-requirements, in progress.

[FLEXIBLE-SIG] F.Zhang et al, " RSVP-TE Signaling Extensions in support of Flexible-grid",draft-zhang-ccamp-flexible-grid-rsvp-te-ext-00, in progress.

8. Authors' Addresses

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

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

Xiaobing Zi
Huawei Technologies
F3-5-B R&D Center, Huawei Base
Bantian, Longgang District
Shenzhen 518129 P.R.China
Phone: +86-755-28973229
Email: zixiaobing@huawei.com

Intellectual Property

Disclaimer of Validity

Full Copyright Statement

Network Working Group
Internet-Draft
Intended status: Standards Track

Fatai Zhang
Xiaobing Zi
Huawei
O. Gonzalez de Dios
Telefonica
Ramon Casellas
CTTC

Expires: April 27, 2012
October 27, 2011

Requirements for GMPLS Control of Flexible Grids

draft-zhang-ccamp-flexible-grid-requirements-01.txt

Status of this Memo

Abstract

A new flexible grid of DWDM is being developed within the ITU-T Study Group 15 to allow more efficient spectrum allocation. This

memo describes the requirements of GMPLS control of flexible grid DWDM networks.

Conventions used in this document

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

Table of Contents

1. Introduction

[G.694.1v1] defines the DWDM frequency grids for 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, i.e. the frequency spacing between two allowed nominal central frequencies could be 12.5 GHz, 25 GHz, 50 GHz, 100 GHz and integer multiples of 100 GHz as defined in [G.694.1v1]. The frequency spacing of the channels on a fiber is fixed.

The speed of the optical signal becomes higher and higher with the advancement of the optical technology. In the near future, high-speed signals (beyond 100 Gbit/s or even 400 Gbit/s) will be deployed in optical networks. These signals may not be accommodated in the channel spacing specified in [G.694.1v1]. On the other hand, ''mixed rate'' scenarios will be commonplace, and bandwidth requirements of the optical signals with different speed will probably be quite different. As a consequence, when the optical signals with different speed are mixed to be transmitted on a fiber, the frequency allocation needs to be more flexible to promote the efficiency.

An updated version of [G.694.1v1] will be consented in December 2011 in support of flexible grids. The terms ''frequency slot (the frequency range allocated to a channel and unavailable to other channels within a flexible grid)'' and ''slot width'' (the full width of a frequency slot in a flexible grid) are introduced to address flexible grid. A channel is represented as a LSC (Lambda Switching Capable) LSP in the control plane and it means a LSC LSP should occupy a frequency slot on each fiber it traverses. In the case of flexible grid, different LSC LSPs may have different slot widths on a fiber, i.e. the slot width is flexible on a fiber.

WSON related documents are being developed currently with the focus of the GMPLS control of fixed grid optical networks. This document describes the new characteristics of flexible grids and analyses the requirements of GMPLS control for the new ''flexible grid'' based optical transmission.

2. Terminology

Flexible Grid: a new WDM frequency grid defined with the aim of allowing flexible optical spectrum management, in which the Slot Width of the frequency ranges allocated to different channels are flexible (variable sized).

Frequency Range: a frequency range is defined by a lowest frequency and a highest frequency.

Frequency Slot: The frequency range allocated to a channel and unavailable to other channels within a flexible grid. A frequency slot is defined by its nominal central frequency and its slot width.

Slot Width: the full width (in Hz) of a frequency slot in a flexible grid. A slot width can be expressed as a multiple (m) of a basic slot width (e.g. 12.5 GHz)

SSON: Spectrum-Switched Optical Network. An optical network in which a data plane connection is switched based on an optical spectrum frequency slot of a variable (flexible) slot width, rather than based on a fixed grid. Note than a wavelength switched optical network (WSON) can be seen as a particular case of SSON in which all slot widths are equal and depend on the used channel spacing.

LSC SS-LSP or flexi-LSP (Lambda Switch Capable Spectrum-Switched Label Switched Path): a control plane construct that represents a data plane connection in which the switching involves a frequency slot of a variable (flexible) slot width. The mapped client signal is transported over the frequency slot, using spectrum efficient modulations such as Coherent Optical Orthogonal Frequency Division Multiplexing (CO-OFDM) and Forward Error Correction (FEC) techniques. Although still in the scope of LSC, the term flexi-LSP is used, when needed, to differentiate from regular WSON LSP in which switching is based on a nominal wavelength.

3. Characteristics of Flexible Grid

Per [G.FLEXIGRID], a flexible grid is defined for the DWDM system. Compared with the fixed grids (i.e. traditional DWDM), flexible grid has a smaller granularity for the central frequency and the slot width of the LSC LSPs is more flexible on a fiber.

3.1. Central Frequency

According to the definition of flexible DWDM grid in [G.FLEXIGRID], the step granularity for the central frequency of flexible grid is 6.25 GHz. The allowed nominal central frequencies are calculated as in the case of flexible grid:

Central Frequency = 193.1 THz + n * 0.00625 THz

Where 193.1 THz is ITU-T ''anchor frequency'' for transmission over the C band and n is a positive or negative integer including 0.

3.2. Slot Width

A slot width is defined by:

12.5 GHz * m, where m is a positive integer.

Note that, when flexi-grid is supported on a fiber or DWDM link, the slot width of different flexi-LSPs may be different.

4. Impact on WSON

Wavelength Switched Optical Networks (WSONs) are constructed from subsystems that include Wavelength Division Multiplexing (WDM) links, tunable transmitters and receivers, Reconfigurable Optical Add/Drop Multiplexers (ROADMs), wavelength converters, and electro-optical network elements. WSON framework is described in [RFC6163]. The introduced flexible grid brings some changes on WSON.

The concept of WSON is extended to SSON, to highlight that such subsystems are extended with flexible and/or elastic capabilities (i.e. flexi-grid). Note that, when modeling SSONs, a WSON can be seen as a particular case of SSON where all LSC LSP use a fixed (and equal) slot width which depends on the used channel spacing.

Transceivers may be able to fully leverage flexible optical channels with advanced modulation formats, and ROADMs may need to be extended to allow flexible spectrum switching, based in, for example, Spectrum Selective Switches (SSS).

4.1. Fiber Links

The nominal (central) frequencies for the flexible grid are defined with a granularity of 6.25 GHz and the allocated frequency slot widths are defined as a multiple of 12.5 GHz. The fiber link for flexible grid can be modeled as shown in figure 1.

```
  -9 -8 -7 -6 -5 -4 -3 -2 -1  0  1  2  3  4  5  6  7  8  9  10 11
...+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--...

                              ^
                           193.1 THz
               <--+-->
              6.25 6.25 m=1, n=-4
```

Figure 1 Fiber link model for flexible grid

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. The nominal central frequency is 193.1 THz when n equals zero.

Because the resource allocated to each flexi-LSP is a frequency range on a fiber link, the following information is needed as parameters to perform resource allocation for the LSPs:

o Available frequency ranges: The set or union of frequency ranges that are not allocated (i.e., available or unused) to flexi-LSPs crossing the DWDM link. The relative grouping and distribution of available frequency ranges in a fiber is usually referred to as ''fragmentation'' and it is common design criterion for optical resource control and management.

4.2. Optical Transmitters and Receivers

In WSON, the optical transmitter is the wavelength source and the optical receiver is the wavelength sink of the WDM system. In each direction, the wavelength used by the transmitter and receiver along a path shall be consistent if there is no wavelength converter in the path.

In the case of flexible grids, the central frequency utilized by a transmitter or receiver may be fixed or tunable. The slot width needed by different transmitters or receivers may be different. Hence, the changes introduced by flexible grid on fundamental modeling parameters for optical transmitters and receivers from the control plane perspective are:

o Available central frequencies: The set of central frequencies which can be used by an optical transmitter or receiver.

o Slot width: The slot width needed by a transmitter or receiver.

Similarly, information on transmitters and receivers capabilities, in regard to signal processing is needed to perform efficient RSA, much like in WSON [WSON-ENCODE]. Additional modeling parameters are:

o Supported Input/Output Modulation formats and spectral efficiency and reach, as well as Input/Output client signals.

o Supported FEC techniques.

5. Routing and Spectrum Assignment

A LSC flexi-LSP occupies a frequency slot, i.e. a range of frequency, on each link the LSP traverses. The route computation and frequency slot assignment could be called RSA (Routing and Spectrum Assignment).

Similar to fixed grids network, if there is no (available) wavelength converter in an optical network, a flexible grid LSC LSP (flexi-LSP) resource allocation will be subject to the ''wavelength continuity constraint'', which is described as section 4 of [RFC6163].

Because of the high cost of the wavelength converters, an optical network is generally deployed with limited or without wavelength converters (sparse translucent optical network). Hence, the wavelength/spectrum continuity constraint should always be considered, and the possibility of wavelength conversion will not be taking into account during the RSA process. When available, information regarding spectrum conversion capabilities at the optical nodes MAY be used by RSA mechanisms

The RSA should determine a route and frequency slot for a flexi-LSP. Note that the mapping between client signals data rates (10, 40, 100... Gbps) and optical slot widths (which are dependent on modulation formats and other physical layer parameters) is out of the scope of the document. The frequency slot can be deduced from the central frequency and slot width parameters as follows:

Lowest frequency = (central frequency) - (slot width)/2;

Highest frequency = (central frequency) + (slot width)/2.

Hence, when a route is computed (by the routing assignment process or subprocess, RA) the spectrum assignment process (SA) should determine the central frequency for a flexi-LSP based on the slot width and available central frequencies information of the transmitter and receiver, and the available frequency ranges information of the links that the route traverses.

Figure 2 shows two LSC LSPs that traverse a link.

```
                     Frequency Slot 1     Frequency Slot 2
                       -------------     -------------------
                       |           |     |                 |
  -9 -8 -7 -6 -5 -4 -3 -2 -1  0  1  2  3  4  5  6  7  8  9  10 11
...+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--...
                       -------------     -------------------
                             ^                     ^
                Central F = 193.1THz    Central F = 193.14375 THz
                 Slot width = 25 GHz    Slot width = 37.5 GHz
```

Figure 2 Two LSC LSPs traverse a Link

The two wavelengths shown in figure 2 have the following meaning:

Flexi-LSP 1: central frequency = 193.1 THz, slot width = 25 GHz. It means the frequency slot [193.0875 THz, 193.1125 THz] is assigned to this LSC LSP.

Flexi-LSP 2: central frequency = 193.14375 THz, slot width = 37.5 GHz. It means the frequency slot [193.125 THz, 193.1625 THz] is assigned to this LSC LSP.

Note that the frequency slots of two LSC flexi-LSPs on a fiber MUST NOT overlap with each other.

5.1. Architecture 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.

5.1.1. 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 capability, etc.

The computation entity could reside on the following elements, which depends on the implementation:

o PCE: PCE get the detailed network information and implement the RSA algorithm for RSA requests from the PCCs.

o Ingress node: Ingress node gets the detailed network information through routing protocol and implements the RSA algorithm when a LSC LSP request is received.

5.1.2. 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.

5.1.3. 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 wavelength continuity constraint should be collected hop by hop along the route. This procedure can be implemented by the GMPLS signaling protocol.

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 central frequencies that meet the slot width requirement of the LSC LSP, i.e. the frequency slot which is determined by the central frequency and slot width MUST NOT overlap with the existing LSC LSPs.

6. Requirements of GMPLS Control

According to the different architecture approaches to RSA some additional requirements have to be considered for the GMPLS control.

6.1. Routing

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.

6.1.1. Available Frequency Ranges of DWDM Links

In the case of flexible grids, channel central frequencies span from 193.1 THz towards both ends of the spectrum with 6.25 GHz granularity. Different LSC LSPs could make use of different slot widths on the same link. Hence, the available frequency ranges should be advertised.

6.1.2. Tunable Optical Transmitters and Receivers

The slot width of a LSC LSP is determined by the transmitter and receiver. The transmitters and receivers could be mapped to ADD/DROP interfaces in WSON. Hence, the slot width of an ADD/DROP interface should be advertised.

The central frequency of a transmitter or receiver could be fixed or tunable. Hence, the available central frequencies should be advertised.

6.2. Signaling

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].

In the case of routing and distributed SA, GMPLS signaling can be used to allocate the frequency slot to a LSC LSP. This brings the following changes to the GMPLS signaling.

6.2.1. Slot Width Requirement

In order to allocate a proper frequency slot for a LSC LSP, the signaling should specify the slot width requirement of a LSC LSP. Then the intermediate nodes can collect the acceptable central frequencies that meet the slot width requirement hop by hop.

The tail node also needs to know the slot width of a LSC LSP to assign the proper frequency resource. Hence, the slot width requirement should be specified in the signaling message when a LSC LSP is being set up.

6.2.2. Frequency Slot Representation

The frequency slot can be determined by the two parameters, which are central frequency and slot width as described in section 5. Hence, the signaling messages should be able to specify the central frequency and slot width of a LSC LSP.

6.3. PCE

[WSON-PCE] describes the architecture and requirements of PCE for WSON. In the case of flexible grid, RSA instead of RWA is used for routing and frequency slot assignment. Hence PCE should implement RSA for flexible grids. The architecture and requirements of PCE for flexible grids are similar to what is described in [WSON-PCE]. This section describes the changes brought by flexible grids.

6.3.1. RSA Computation Type

A PCEP request within a PCReq message MUST be able to specify the computation type of the request:

- o Combined RSA: Both of the route and frequency slot should be provided by PCE.

- o Routing Only: Only the route is requested to be provided by PCE.

The PCEP response within a PCRep Message MUST be able to specify the route and the frequency slot assigned to the route.

RSA in SSON MAY include the check of signal processing capabilities, which MAY be provided by the IGP. A PCC should be able to indicate additional restrictions for such signal compatibility, either on the endpoint or any given link (such as regeneration points).

A PCC MUST be able to specify whether the PCE MUST also assign a Modulation list and / or a FEC list, as defined in [WSON-ENCODE] and [WSON-PCE].

A PCC MUST be able to specify whether the PCE MUST or SHOULD include or exclude specific modulation formats and FEC mechanisms.

In the case where a valid path is not found, the response MUST be able to specify the reason (e.g., no route, spectrum not found, etc.)

6.3.2. RSA path re-optimization request/reply

For a re-optimization request, the PCEP request MUST provide the path to be re-optimized and include the following options:

o Re-optimize the path keeping the same frequency slot.

o Re-optimize spectrum keeping the same path.

o Re-optimize allowing both frequency slot and the path to change.

The corresponding PCEP response for the re-optimized request MUST provide the Re-optimized path and frequency slot.

In case the path is not found, the response MUST include the reason (e.g., no route, frequency slot not found, both of route and frequency slot not found, etc.)

6.3.3. Frequency Constraints

PCE for flexible grids should consider the following constraints brought by the transmitters and receivers:

o Available central frequencies: The set of central frequencies that can be used by an optical transmitter or receiver.

o Slot width: The slot width needed by a transmitter or receiver.

This constraints may be provided by the requester (PCC) in PCReq or reside within the PCE’s TEDB which stores the transponder’s capabilities.

PCC may also specify the frequency constraints for policy reasons. In this case, the constraints should be specified in the PCReq message sent to the PCE. In any case, PCE will compute the route and assign the frequency slot to meet the constraints specified in the PCReq message. Then return the result to the PCC.

7. Security Considerations

This document does not introduce any further security issues other than those described in [RFC6163] and [RFC5920].

8. References

8.1. Normative References

[RFC2119] S. Bradner, "Key words for use in RFCs to indicate requirements levels", RFC 2119, March 1997.

[WSON-PCE] Y. Lee, G. Bernstein, Jonas Martensson, T. Takeda and T. Tsuritani, "PCEP Requirements for WSON Routing and Wavelength Assignment", draft-ietf-pce-wson-routing-wavelength-05, July 2011.

[WSON-ENCODE] G. Bernstein, Y. Lee, Dan Li and W. Imajuku, "Routing and Wavelength Assignment Information Encoding for Wavelength Switched Optical Networks", draft-ietf-ccamp-rwa-wson-encode, August 2011.

[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.

[G.FLEXIGRID] Draft revised G.694.1 version 1.3, Unpublished ITU-T Study Group 15, Question 6.

8.2. Informative References

[G.694.1v1] ITU-T Recommendation G.694.1, Spectral grids for WDM applications: DWDM frequency grid, June 2002.

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

9. Authors' Addresses

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

Oscar Gonzalez de Dios
Telefonica Investigacion y Desarrollo
Emilio Vargas 6
Madrid,   28045
Spain
Phone: +34 913374013
Email: ogondio@tid.es

Ramon Casellas
CTTC
Av. Carl Friedrich Gauss, 7
Castelldefels, 08860, Spain
Phone: +34 936452900
Email: ramon.casellas@cttc.es

Xiaobing Zi
Huawei Technologies
F3-5-B R&D Center, Huawei Base
Bantian, Longgang District
Shenzhen 518129 P.R.China
Phone: +86-755-28973229
Email: zixiaobing@huawei.com

Felipe Jimenez Arribas
Telefonica Investigacion y Desarrollo
Emilio Vargas 6
Madrid,   28045
Spain
Email: felipej@tid.es

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: April 16, 2012

Fatai Zhang
Huawei
Oscar Gonzalez de Dios
Telefonica
D. Ceccarelli
Ericsson

October 16, 2011

RSVP-TE Signaling Extensions in support of Flexible Grid

draft-zhang-ccamp-flexible-grid-rsvp-te-ext-00.txt

Status of this Memo

Abstract

This memo describes the signaling extensions of GMPLS control of flexible grid 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

[G.694.1v1] defines the DWDM frequency grids for 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, i.e. the frequency spacing between two allowed nominal central frequencies can be 12.5 GHz, 25 GHz, 50 GHz, 100 GHz and integer multiples of 100 GHz as defined in [G.694.1v1]. All of the wavelengths on a fiber SHALL use different central frequencies and occupy a fixed bandwidth of frequency.

[G.FLEXIGRID], an updated version of [G.694.1v1] will be consented in December 2011 in support of flexible grids. 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 to define flexible grid. A channel is represented as an LSC (Lambda Switching Capable) LSP in the control plane and SHOULD occupy a frequency slot on each fiber it traverses. In the case of flexible grid, the different LSC LSPs may have different slot width on a fiber, i.e. the slot width is flexible on a fiber.

[WSON-SIG] describes the requirements and extensions for WSON signaling. It focuses on the fixed grids control. This document describes the additional requirements and extensions for signaling control brought by flexible grids.

2. Requirements for Flexible Grid Signaling

An LSC LSP SHOULD occupy a frequency slot, i.e. a range of frequency. The route computation and frequency slot assignment could be called RSA (Routing and Spectrum Assignment).

[FLEXIGRID-REQ] describes three types of architecture approaches to RSA, which are: combined RSA, separated RSA and distributed SA. In the case of combined RSA and separated RSA, both the routing and the spectrum (frequency slot) are provided by the RSA algorithm before the signaling procedure. It could be called "centralized SA". In the case of distributed SA, only the route is provided before the signaling procedure and the spectrum assignment is done during the signaling procedure.

In the case of centralized SA, the frequency slot SHOULD be specified in the Path message. In the case of distributed SA, the slot width of the LSC LSP SHOULD be specified in the Path message for the purpose of frequency slot assignment.

Similar to fixed grid network, if there is no wavelength converter in an optical network, there is "wavelength continuity constraint" of a LSC LSP which is described as section 4 of [RFC 6163].

2.1. Slot Width

The slot width is an end-to-end parameter representing how much spectrum resource is requested for a LSC LSP. Since different LSPs may request different amounts of spectrum portion in flexible grid networks, the slot width SHOULD be carried in the signaling message, so that all the nodes along the LSP can know how much spectrum portion will be allocated for the LSP.

2.2. Frequency Slot

The frequency slot information represents which part of the spectrum portion is allocated on each link for an LSC LSP. This information SHOULD be carried hop-by-hop in signaling message so that each node can indicate its neighbor the resource reservation on the link between them.

The frequency slot can be represented by the two parameters: central frequency and slot width, as follows:

```
      Frequency slot = [(central frequency) - (slot width)/2] ~
                       [(central frequency) + (slot width)/2]
```

Since the slot width information is carried in the signaling message (as described in Section 2.1), also the central frequency parameter SHOULD be carried in the signaling message for frequency slot determination.

Figure 1 shows an example of two LSC LSPs traversing a link and illustrates how to determine the frequency slot based on the central frequency and slot width information.

```
                         Frequency Slot 1     Frequency Slot 2
                           -------------     -------------------
                           |           |     |                 |
      -9 -8 -7 -6 -5 -4 -3 -2 -1 0  1  2  3  4  5  6  7  8  9 10  11
   ...+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--...
                           -------------     -------------------
                                 ^                    ^
                    Central F = 193.1THz    Central F = 193.14375 THz
                     Slot width = 25 GHz    Slot width = 37.5 GHz
```

Figure 1 - Two LSC LSPs traverse a Link

The two wavelengths shown in figure 1 have the following meaning:

LSC LSP 1: central frequency = 193.1 THz, slot width = 25 GHz. It means the frequency slot [193.0875 THz, 193.1125 THz] is assigned to this LSC LSP.

LSC LSP 2: central frequency = 193.14375 THz, slot width = 37.5 GHz. It means the frequency slot [193.125 THz, 193.1625 THz] is assigned to this LSC LSP.

Note that the frequency slots of two LSC LSPs on a fiber MUST NOT overlap with each other.

3. Extensions

This section defines the extensions of signaling for flexible grid.

3.1. WSON Traffic Parameters

As described in Section 2, the slot width represents how much spectrum resource is requested for an LSC LSP, i.e., it describes the end-to-end traffic profile of the LSP. Therefore, the slot width SHOULD be regarded as a traffic parameter for an LSC LSP.

The WSON traffic parameters are organized as follows:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     m         |                     Reserved                  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

m (8 bits): the slot width is specified by m*12.5 GHz.

Note that the slot width of fixed grid defined in [G.694.1v1] can be also specified by m because the defined channel spacing (12.5 GHz, 25 GHz, 50 GHz, 100 GHz and integer multiples of 100 GHz) are also the multiple of 12.5 GHz. Therefore, the traffic parameters are general for WSON including both fixed grid and flexible grid.

The WSON traffic parameters SHOULD be carried in SENDER_TSPEC or FLOWSPEC objects:

WSON SENDER_TSPEC: Class = 12, C-Type = to be assigned by IANA, preferred 8.

WSON FLOWSPEC: Class = 9, C-Type = to be assigned by IANA, preferred 8.

3.2. Generalized Label

In the case of flexible grid, the allowed central frequency is calculated as follows:

Central Frequency = (193.1 + n * 0.00625) THz

Where n is a two’s-complement integer (positive, negative, or 0).

The Label object is used to indicate the resource reserved on a link. In Flexible Grid networks, it is used to indicate which frequency slot is allocated on a link for the given LSC LSP.

Since the frequency slot assigned to an LSC LSP can be determined by the combination of [central frequency, slot width], while the slot width of an LSC LSP is specified in the traffic parameters, the Label object just needs to carry the assigned central frequency. Therefore, the wavelength label format defined in [RFC6205] can be reused to specify the central frequency of an LSC LSP, without any change on the label format.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|Grid | C.S.  |    Identifier   |              n                |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

The meaning of Grid, Identifier and n fields are not changed. The usage of the label format is also not changed.

According to [G.FLEXIGRID], flexible grid still belongs to DWDM, so there is no need to introduce a new type of Grid, i.e., Grid=1 (ITU-T DWDM) SHOULD be used for flexible grid.

In case of Grid=1 (ITU-T DWDM), a new value of C.S. is defined for flexible 6.25 GHz grid. The C.S.(Channel Spacing) field is defined as follows:

C.S. (GHz)	Value
Reserved	0
100	1
50	2
25	3
12.5	4
6.25	5 (TBA)
Future use	6 ~ 15

The value for flexible 6.25 GHz is to be assigned by IANA, preferred 5.

3.3. Signaling Procedures

This section describes the signaling procedures for distributed SA and centralized SA (See [FLEXIGRID-REQ]).

3.3.1. Distributed SA

In this case, only the route is provided by a PCE or ingress node before the signaling procedure. The available central frequencies will be collected hop by hop and the egress node SHOULD select a proper central frequency for the LSP.

After the route is computed, the ingress node SHOULD find out the available central frequencies for the LSP on the next link of the route. If the frequency slot which is determined by a central frequency and slot width of the LSC LSP (See section 2.2) does not overlap with the existing LSC LSPs, the central frequency is considered to be available for the requesting LSC LSP.

Then a Path message is sent to the next node on the route. The Path message MUST contain a Flexible Grid SENDER_TSPEC object to specify the slot width of the LSC LSP. A LABEL_SET object SHALL be added to the Path message, which contains the available central frequencies for the LSP on the next link.

When an intermediate node receives a Path message, it can get the slot width from the Flexible Grid SENDER_TSPEC object. Then it SHOULD find the available central frequencies for the LSP on the next link of the route similar to the ingress node. The common part of the two available central frequency sets, i.e. the set received from the Path message and the set of the next link, SHALL be selected as the new available central frequency set for the LSP. If the new set is null, the Path message SHALL be rejected by a PathErr message. Otherwise, the LABEL SET object in the Path message SHALL be updated according to the new set and the Path message is forwarded to the next node on the route.

When an egress node receives a Path message, it SHOULD select an available central frequency from the LABEL SET object based on local policy and determine the frequency slot based on the slot width and the selected central frequency (See section 2.2). Then a Resv message is responded so that the nodes along the LSP can establish the optical cross-connect based on the frequency slot determined by the slot width in the traffic parameters and the central frequency in the label.

3.3.2. Centralized SA

In this case, both of the routing and frequency slot are provided by PCE or ingress node. When signaling the LSP, the slot width is carried in the traffic parameters, and the assigned central frequency is carried in the Label ERO. When the nodes along the LSP receive the Path message carrying these information, they can determine the frequency slot by the slot width and the central frequency and then establish the optical cross-connect based on the central frequency. The procedures of ERO and Label ERO are the same as described in [RFC3209] and [RFC3473].

4. IANA Considerations

TBD.

5. Security Considerations

TBD.

6. References

[RFC2119] S. Bradner, "Key words for use in RFCs to indicate requirements levels", RFC 2119, March 1997.

[G.694.1v1] ITU-T Recommendation G.694.1, Spectral grids for WDM applications: DWDM frequency grid, June 2002.

[WSON-PCE] Y. Lee, G. Bernstein, Jonas Martensson, T. Takeda and T. Tsuritani, "PCEP Requirements for WSON Routing and Wavelength Assignment", draft-ietf-pce-wson-routing-wavelength-05, July 2011.

[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-02, September 2011.

[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.

[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.

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

[FLEXIGRID-REQ] F.Zhang et al, "Requirements for GMPLS Control of Flexible Grids",draft-zhang-ccamp-flexible-grid-requirements, in progress.

[G.FLEXIGRID] Draft revised G.694.1 version 1.3, Unpublished ITU-T Study Group 15, Question 6.

7. Authors' Addresses

Fatai Zhang
Huawei Technologies
F3-5-B R&D Center, Huawei Base
Bantian, Longgang District
Shenzhen 518129 P.R.China

Phone: +86-755-28972912
Email: zhangfatai@huawei.com

Oscar Gonzalez de Dios
Telefonica Investigacion y Desarrollo
Emilio Vargas 6
Madrid, 28045
Spain

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

Felipe Jimenez Arribas
Telefonica Investigacion y Desarrollo
Emilio Vargas 6
Madrid,   28045
Spain
Email: felipej@tid.es

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

Xiaobing Zi
Huawei Technologies
F3-5-B R&D Center, Huawei Base
Bantian, Longgang District
Shenzhen 518129 P.R.China

Phone: +86-755-28973229
Email: zixiaobing@huawei.com

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

Intellectual Property

Disclaimer of Validity

Full Copyright Statement

MPLS Working Group
Internet-Draft
Intended status: Standards Track
Expires: March 19, 2012

F. Zhang
ZTE Corporation

September 16, 2011

RSVP-TE Extentions to Exchange MPLS-TP Tunnel Numbers
draft-zhang-ccamp-mpls-tp-rsvpte-ext-tunnel-num-00

Abstract

The MPLS Transport Profile (MPLS-TP) identifiers document [I-D.ietf-mpls-tp-identifiers] introduce two tunnel numbers, A1-Tunnel_Num and Z9-Tunnel_Num, which allow a compact format for Maintenance Entity Point Identifier (MEP_ID). For some Operation, Administration and Maintenance (OAM) functions, such as Connectivity Verification (CV) [I-D.ietf-mpls-tp-cc-cv-rdi], source MEP_ID MUST be inserted in the OAM packets, so that the peer endpoint can compare the received and expected MEP_IDs to judge whether there is a mismatch, which means that the two MEP nodes need to pre-store each other’s MEP_IDs.

The specification of setting up co-routed bidirectional LSP is described in the document [RFC3473], which does not introduce the locally configured tunnel number on the tunnel endpoint. This document defines the Connection object to exchange the tunnel numbers.

Status of this Memo

Copyright Notice

Table of Contents

1. Introduction

The MPLS Transport Profile (MPLS-TP) identifiers document [I-D.ietf-mpls-tp-identifiers] introduce two tunnel numbers, A1-Tunnel_Num and Z9-Tunnel_Num, which are locally assigned and allow a compact format for Maintenance Entity Point Identifier (MEP_ID). For a co-routed bidirectional LSP, the format of A1-MEP_ID is A1-Node_ID::A1-Tunnel_Num::LSP_Num, and the format of Z9-MEP_ID is Z9-Node_ID::Z9-Tunnel_Num::LSP_Num. In order to realize some Operation, Administration and Maintenance (OAM) functions, such as Connectivity Verification (CV) [I-D.ietf-mpls-tp-cc-cv-rdi], source MEP-ID MUST be inserted in the OAM packets, in this way the peer endpoint can compare the received and expected MEP-IDs to judge whether there is a mismatch. Hence, the two MEP nodes must pre-store each other's MEP-IDs before sending the CV packets.

Although the exchange of MEP_IDs can be accomplished by Network Management System (NMS) if it is deployed, it is still complex when the LSPs cross different adiminstration domains, which needs the cooperation of NMSs. So when the LSPs are set up by control plane, Resource ReserVation Protocol Traffic Engnieering (RSVP-TE) signaling will be more suitable to realize the exchange of MEP_IDs.

The specification of setting up co-routed bidirectional LSP is described in the document [RFC3473], which does not introduce the locally configured tunnel number on the tunnel endpoint. This document defines the Connection object to exchange the tunnel numbers.

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. Operation

MPLS-TP co-routed bidirectional LSPs can be deloyed across one or more administration domains, and NMS may exist in some administration domains, which knows the tunnel spaces of every node in it's responsible domain. Consider that LSP1 is initialized at A1 node with Connection object inserted in LSP1's Path message , the following modes may happend.

Modes 1: L bit is set, and the Z9-Tunnel_Num is designated in the "Destination Tunnel Num" field. If the Z9 node finds that this tunnel number is occupied, or it can not be used because of some local policies, a PathErr message must be sent with "Unavailable tunnel number" error. Otherwise, the designated tunnel number must be adopted, and the Connection object may be inserted in the Resv message without any change.

Modes 2: L bit is not set, and a recommended Z9-Tunnel_Num may be filled in the "Destination Tunnel Num" field. If the Z9 node finds that the recommended value can be used, the Connection object must be inserted in the Resv message without any change; if the recommended value can not be used or the "Destination Tunnel Num" field is empty, a new tunnel number will be allocated and filled into the Connection object that must be inserted in the Resv message.

Each mode has its own pros and cons and how to determine the right mode for a specific network mainly depends on the operators' preference. For example, for the operators who are used to operate traditional transport network and familiar with the Transport-Centric operational model may prefer mode 1. The second mode is more suitable for the operators who are familiar with the operation and maintenance of IP/MPLS network, or the MPLS-TP LSPs cross multiple administration domains.

4. Connection Object

The format of Connection Object (Class-Num of the form 11bbbbbb with value = TBA, C-Type = TBA) is as follow:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|            Length             | Class-Num(TBA)|  C-Type (TBA) |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|L|           Reserverd         |       Destination Tunnel Num  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Connection Object

L

The L bit is set if the initiating node enforces the peer endpoint to configure the value carried in the field of "Destination Tunnle Num".

If the bit is not set, the peer endpoint firstly tries to use the recommended tunnel number; it can use any other unoccupied tunnnel numbers when the recommended tunnel number is unavailable.

Reserverd

Must be set to 0 on transmit and ignored on receive.

Destination Tunnel Num

If the L bit is set, it indicates that the peer endpoint must configure the value carried in this field.

If the L bit is not set, this field can be empty or filled by the recommended value.

The Connection object may appear in Path or Resv message, and a midpoint that does not support this object is required to pass it on unaltered, as indicated by the C-Num and the rules defined in [RFC2205].

5. IANA Considerations

TBD.

6. Security Considerations

TBD.

7. Acknowledgement

This document was prepared based on the discussion with George Swallow, valuable comments and input was also received from Venkatesan Mahalingam and Muliu Tao.

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.

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

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

8.2. Informative References

[I-D.ietf-mpls-tp-cc-cv-rdi] Allan, D., Swallow, G., and J. Drake, "Proactive Connectivity Verification, Continuity Check and Remote Defect indication for MPLS Transport Profile", draft-ietf-mpls-tp-cc-cv-rdi-06 (work in progress), August 2011.

[I-D.ietf-mpls-tp-identifiers] Bocci, M., Swallow, G., and E. Gray, "MPLS-TP Identifiers", draft-ietf-mpls-tp-identifiers-07 (work in progress), July 2011.

Author's Address

Fei Zhang
ZTE Corporation

Email: zhang.fei3@zte.com.cn

Xiao Bao
ZTE Corporation

Email: bao.xiao1@zte.com.cn

Network Working Group   Fatai Zhang
Internet-Draft   Dan Li
Intended status: Standards Track   Huawei
                                                    O. Gonzalez de Dios
                                    Telefonica Investigacion y Desarrollo
                                                          C. Margaria. C
                                                   Nokia Siemens Networks
Expires: April 30, 2012   October 31, 2011

RSVP-TE Extensions for Configuration SRLG of an FA

draft-zhang-ccamp-srlg-fa-configuration-04.txt

Status of this Memo

Abstract

This memo provides extensions for the Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) for the support of the automatic discovery of SRLG of an LSP.

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

As described in [RFC4206], H-LSP (Hierarchical LSP) can be used for carrying one or more other LSPs. [RFC6107] further mentions the implementation of H-LSP. In packet networks, e.g. MPLS networks, H-LSP mechanism can be implemented by MPLS label stack. In non-packet networks where the label is implicit, label stacks are not possible, and H-LSPs rely on the ability to nest switching technologies. Thus, for example, a lambda switch capable (LSC) LSP can carry a time division multiplexing (TDM) LSP, but cannot carry another LSC LSP.

S-LSP (LSP Stitching), which is defined in [RFC5150], is an LSP that represents a segment of another LSP, i.e., the S-LSP is viewed as one hop by another LSP. As described in [RFC6107], in the data plane the LSPs are stitched so that there is no label stacking or nesting.

Thus, an S-LSP must be of the same switching technology as the end-to-end LSP that it facilitates.

Therefore, H-LSP mechanism can be used in both multi-domain and multi-layer scenarios and S-LSP mechanism can only be used in multi-domain scenario.

Both of the H-LSP and S-LSP can be advertised as a TE link in a GMPLS routing instance for path computation purpose. As described in [RFC6107], if the LSP (H-LSP or S-LSP) is advertised in the same instance of the control plane that advertises the TE links from which the LSP is constructed, the LSP is called an FA.

In multi-domain or multi-layer context, the path information of an LSP may not be provided to the ingress node for confidential reasons and the ingress node may not run the same routing instance with the intermediate nodes traversed by the path. In such scenarios, the ingress node can not get the SRLG information of the path information which the LSP traverse.

Even if the ingress node has the same routing instance with the intermediate nodes traversed by the path, the path information of the H-LSP or S-LSP may not be provided to the ingress node. Hence the ingress node may also not know the SRLG of the path the LSP traverses.

In the case that the ingress node does not get the SRLG of the path the LSP traverses(i.e. H-LSP or S-LSP), there are disadvantages as follows:

o SRLG-disjoint path, for instance in case of end-to-end path protection, cannot be calculated

o Intermediate nodes of a pre-planned shared restoration LSP cannot correctly decide on the SRLG-disjointness between two PPRO (PRIMARY_PATH_ROUTE Object)

o In case that an LSP is advertised as a TE-Link, the ingress node cannot provide the correct SRLG for the TE-Link automatically

In case that an LSP is advertised as a TE-Link, the SRLG information of the TE link needs to be configured manually or automatically. However, for manual configuration, there are some disadvantages (e.g., require configuration coordination and additional management; manual errors may be introduced) mentioned in Section 1.3.4 of [RFC6107].

In addition, Section 1.2 of [RFC6107] describes it is desirable to have a kind of automatic mechanism to advertise the FA (i.e., to signal an LSP and automatically coordinate its use and advertisement in any of the ways with minimum involvement from an operator).

Thus, in order to provide the SRLG information to the TE link automatically when an LSP (H-LSP or S-LSP) is advertised as a TE link, allow disjoint path calculation at ingress and allow correct pre-planned shared LSP to correctly share resource, this document provides an automatic mechanism to collect the SRLG used by a LSP automatically.

2. RSVP-TE Requirements

2.1. SRLG Collection Indication

The head nodes of the LSP must be capable of indicating whether the SRLG information of the LSP should be collected during the signaling procedure of setting up an LSP.

2.2. SRLG Collection

The SRLG information can be collected during the setup of an LSP. Then the endpoints of the LSP can get the SRLG information and use it for routing, sharing and TE link configuration purposes.

2.3. SRLG Update

When the SRLG information changes, the endpoints of the LSP need to be capable of updating the SRLG information of the path. It means that the signaling should be capable of updating the newly SRLG information to the endpoints.

3. RSVP-TE Extensions

3.1. SRLG Collection Flag

In order to indicate nodes that SRLG collection is desired, this document defines a new flag in the Attribute Flags TLV, which is carried in an LSP_REQUIRED_ATTRIBUTES Object:

o Bit Number (to be assigned by IANA, recommended bit zero): 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 [RFC3209].

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|      Type     |     Length    |            Reserved           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                 SRLG ID 1 (4 bytes)                           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                           ......                              ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                 SRLG ID n (4 bytes)                           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Type

The type of the sub-object, to be assigned by IANA, which is recommended 34.

Length

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

SRLG Id

The 32-bit identifier of the SRLG.

Reserved

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

The rules of the processing of the LSP_REQUIRED_ATTRIBUTES Object and ROUTE_RECORD Object are not changed.

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 in an LSP_REQUIRED_ATTRIBUTES Object.

When a node receives a Path message which carries an LSP_REQUIRED_ATTRIBUTES Object and the SRLG Collection Flag is set, if local policy determines that the SRLG information should not be provided to the endpoints, it must return a PathErr message to reject the Path message. Otherwise, it must add an SRLG sub-object to the RRO to carry the local SRLG information. Then it forwards the Path message to the next node in the downstream direction.

Following the steps described above, the intermediate nodes of the LSP can collect the SRLG information in the RRO during the forwarding of the Path message hop by hop. When the Path message arrives at the tail node, the tail node can get the SRLG information from the RRO.

Before the Resv message is sent to the upstream node, the tail node adds an SRLG sub-object to the RRO. The collected SRLG information can be carried in the SRLG sub-object. Therefore, during the forwarding of the Resv message in the upstream direction, the SRLG information is not needed to be collected hop by hop.

Based on the above procedure, the endpoints can get the SRLG information automatically. Then the endpoints can for instance advertise it as a TE link to the routing instance based on the procedure described in [RFC6107] and configure the SRLG information of the FA automatically.

It is noted that a node (e.g. the edge node of a domain) may edit the RRO to remove the route information (e.g. node, interface identifier information) before forwarding it due to some reasons (e.g. confidentiality or reduce the size of RRO), but the SRLG information should be retained if it is desirable for the endpoints of the LSP.

4.2. 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] MUST be used to refresh the SRLG information.

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:

o Whether the SRLG IDs of the domain or specific layer network can be exposed to the nodes outside the domain or layer network.

o If the SRLG IDs must not be exposed to the nodes outside of the domain or specific layer network by policy, the border node must reject the Path message desiring SRLG recording and send a PathErr message with the defined error code ''Policy Control Failure''/''Inter-domain policy failure''.

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.

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. IANA Considerations

6.1. RSVP Attribute Bit Flags

The IANA has created a registry and manages the space of attributes bit flags of Attribute Flags TLV as described in section 11.3 of [RFC5420]. It is requested that the IANA makes assignments from the Attribute Bit Flags.

This document introduces a new Attribute Bit Flag:

- Bit number: TBD (0)

- Defining RFC: this I-D
- Name of bit: SRLG Collection Flag
- The meaning of the Attribute Flags TLV on a Path is defined in this I-D

6.2. ROUTE_RECORD Object

IANA has made the following assignments in the "Class Names, Class Numbers, and Class Types" section of the "RSVP PARAMETERS" registry located at http://www.iana.org/assignments/rsvp-parameters. We request that IANA make assignments from the ROUTE_RECORD [RFC3209] portions of this registry.

This document introduces a new RRO sub-object:

Type	Name	Reference
TBD (34)	SRLG sub-object	This I-D

7. Security Considerations

TBD.

8. Acknowledgements

The authors would like to thank Igor Bryskin and Ramon Casellas for their useful comments to the document.

9. References

[RFC2119] S. Bradner, "Key words for use in RFCs to indicate requirements levels", RFC 2119, March 1997.

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

[RFC3477] K. Kompella, Y. Rekhter, "Signalling Unnumbered Links in Resource ReSerVation Protocol - Traffic Engineering (RSVP-TE)", rfc3477, January 2003.

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

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

[RFC4874] CY. Lee, A. Farrel, S. De Cnodder, " Exclude Routes - Extension to Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) ", RFC 4874, April 2007.

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

[RFC5420] A. Farrel, D. Papadimitriou, J.P, and A. Ayyangar, "Encoding of Attributes for MPLS LSP Establishment Using Resource Reservation Protocol Traffic Engineering (RSVP-TE)", RFC 5420, February 2009.

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

10. Authors' Addresses

Fatai Zhang
Huawei Technologies
F3-5-B R&D Center, Huawei Base
Bantian, Longgang District
Shenzhen 518129 P.R.China

Phone: +86-755-28972912
Email: zhangfatai@huawei.com

Dan Li
Huawei Technologies
F3-5-B R&D Center, Huawei Base
Bantian, Longgang District
Shenzhen 518129 P.R.China

Phone: +86-755-28970230
Email: danli@huawei.com

Oscar Gonzalez de Dios
Telefonica Investigacion y Desarrollo
Emilio Vargas 6
Madrid, 28045
Spain

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

Cyril Margaria
Nokia Siemens Networks
St Martin Strasse 76
Munich, 81541
Germany

Phone: +49 89 5159 16934
Email: cyril.margaria@nsn.com

Xiaobing Zi
Huawei Technologies
F3-5-B R&D Center, Huawei Base
Bantian, Longgang District
Shenzhen 518129 P.R.China

Phone: +86-755-28973229
Email: zixiaobing@huawei.com

Intellectual Property

Network Working Group
Internet-Draft
Intended status: Informational
Expires: April 26, 2012

J. Zhang
YL. Zhao
ZY. Yu
BUPT

October 24, 2011

OSPF-TE Protocol Extension for Constraint-aware RSA in Flexi-Grid Networks

draft-zhangj-ccamp-flexi-grid-ospf-te-ext-00

Abstract

ITU-T Study Group 15 has introduced a new flexible grids technology of DWDM network which is an effective solution to improve the efficiency of spectrum resource utilization. This memo extends the OSPF-TE protocol to support constraint-aware routing and spectrum assignment (RSA) in flexi-grid networks.

Status of this Memo

Copyright Notice

Table of Contents

1. Introduction

To enable the dynamic and effective allocation of spectrum resource based on the demand of the client LSP's requests, the latest revision of ITU-T Recommendation [G.694.1] has introduced a flexible grid technique in DWDM optical networks. The flexible grid has a finer granularity (i.e. according to the definition of flexible grid in [G.694.1], the data channel can be selected on a channel spacing of 6.25 GHz with a variable slot width measured in units of 12.5 GHz) for the spectrum slot.

In the dynamic flexi-grid networks, except for selecting an appropriate route for the client LSP, the appropriate width of spectrum slot is also needed to choose and assigned to the client LSP. The spectrum bandwidth assigned to the client LSP is made up of an appropriate number of consecutive spectrum slots from end-to-end, which is determined by the used modulation format, according to the client LSPs data rate requests and physical constraints of the selected path.

The routing and spectrum assignment (RSA) of flexi-grid networks need to consider some constraints. In this memo two of those constraints (other constraints are left for future considered) that are necessary for RSA are discussed in detail, and then describes the OSPF-TE protocol extension for these constraints related to RSA in flexi-grid networks.

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. Terminologies

CSSW: Consecutive spectrum slots weight

GB: Guard band

RSA: Routing and spectrum assignment

WSON: Wavelength switched optical networks

4. Motivation for Routing Protocol Extension

In this section we introduce the RSA constraints and the motivation of routing protocol extension for of flexi-grid networks

4.1. Constraints Considerations for RSA in Flexi-Grid Networks

When processing RSA in flexi-grid networks, the constraints information (such as the information of spectrum bandwidth in a network link and so on.) are necessary for computing and selecting an appropriate backup route and a certain number of consecutive spectrum slots for the client LSPs effectively.

Some of the necessary constraints are listed as follows:

- Spectral consecutiveness constraint
- Variable guard band constraint
- Spectral continuity constraint
- Impairments constraint
- Other constraints

All the constraints can generate important impacts for the performance of the client LSPs, even for the entire network. The first two constraints are mainly talked about in this memeo.

Just like the wavelength continuity constraint in WSON, the spectral continuity constraint means allocation of the same spectrum slots on each link along a path because not all of the nodes in optical networks have the ability of wavelength conversion.

The degradation of the optical signals due to impairments that accumulate along the path (without 3R regeneration), can result in unacceptable bit error rates or even a complete failure to demodulate and/or detect the received signal[draft-ietf-ccamp-wson-impairments-07]. So it is necessary to consider about the impairments constraint within flexi-grid networks. The impairments constraint in flexi-grid networks will be studied in future in this memo.

Also, there may be some other constraints for RSA, other than the four kinds above, such as the modulation levels constraint, which are left for future researching.

4.2. Consecutive Spectrum Slots Information

The spectral consecutiveness constraint is that the allocated spectrum slots must be chosen from consecutive spectrum slots in the spectrum space on each link of flexi-grid networks.

Compared with the technology of WSON, the number of spectrum slots in flexi-grid networks will be much larger than the number of wavelength in WSON. After a long running time, the situation of available spectrum slots will be much complex, especially the situation of the available consecutive spectrum slots.

After selecting a route, the appropriate consecutive spectrum slots need to be assigned for the client LSP. When we choose one of the backup routes for the client LSP without considering the situation about the available consecutive spectrum slots information, the route may have no enough consecutive spectrum slots which means that the selected route have no available resource for the LSP’s request, and then the client LSP will be rejected or trigger another path computation process which will increase the blocking rate of the network or increase network resources consumed by communication and computing of new route.

When computing a route with the knowledge of the consecutive spectrum slots information of the network link (for example, the number of ten available consecutive spectrum slots in a network link, or the number of twenty available consecutive spectrum slots in a network link.), it will be very useful to select a better route which has higher probability of enough available consecutive spectrum slots for the client LSP. And this will improve the success rate of setting up new client LSPs.

4.3. Variable Guard Band Information

Some spectrum slots need to be reserved as Guard Band(GB) between two adjacent client LSPs to avoid bad impact of non-linear impairments and other network elements. Since the granularity of the flexi-grid networks will be very small, the spectrum interval, i.e., GB need to be considered more carefully to avoid poor quality impact of the adjacent client LSPs. Which means with the changing of network environment and the operating of the network, the bandwidth of the GB also need to change.

In flexi-grid networks, with the increasing of the total transportation power and the smaller of the channel space, the channel crosstalk that results from non-linear effects will become the important factor that affects the performance of the network. The impact between two adjacency client LSPs may be changing based on the change of crosstalk and other changes of network. With the changing of those parameters, the interferences between two adjacency client LSPs may be increasing, if the Guard Band is fixed, the quality of the adjacent client LSPs and also the network's will be decreased. If the GB can be varied based on the network environment changing, then the bad impact can be avoided.

5. OSPF-TE Protocol Extension

In this section, we define the enhancements to the Traffic Engineering (TE) properties of flexi-grid networks' TE links that can be announced in OSPF-TE LSAs.

The TE LSA, which is an opaque 10 LSA with area flooding scope [RFC3630], has only one top-level and has one or more nested sub-TLVs for extensibility. [RFC3630] also defines two top Type/Length/Value (TLV) triplet to support traffic engineering of OSPF, i.e. (1) Router Address TLV and (2) Link TLV. In this memo, we enhance the sub-TLVs for the Link TLV in support of flexi-grid networks. Specifically, we add the following sub-TLVs to the Link TLV:

- o Consecutive spectrum slots weight sub-TLV

- o Variable Guard Band sub-TLV

5.1. Consecutive Spectrum Slots Weight Sub-TLV

In distribution networks, we propose the CSSW as a sub-TLV of OSPF-TE Link TLV which represents the situation of the available consecutive spectrum slots in a link of the flexi-grid networks for example the percentage of the total bandwidth of the number of five consecutive spectrum slots, the percentage of the total bandwidth of the number of ten consecutive spectrum slots ...). With knowing the weight of available consecutive spectrum slots in a link, the spectrum resource assignment in the flexi-grid networks can be working more efficiently.

The format of the CSSW sub-TLV is as follows:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|      Type = TBD               |         Length = variable     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
|         Value = Consecutive Spectrum Slots Weight             |
//                                                             //
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Type: TBD.

The Type of CSSW sub-TLV is left for future to define.

Length: Variable.

The length of CSSW sub-TLV is based on its define of the value which is variable based on different implementation ways.

Value: TBD

The content of the CSSW sub-TLV is left for future researching.

5.2. Variable Guard Band Sub-TLV

The Guard Band sub-TLV (which is also short for GB sub-TLV) describes the spectrum interval between two client LSPs to avoid crosstalk and other network elements(such as impairment elements) that can affect the transmission performance of each client LSP.

The format of the GB sub-TLV is as follows:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|      Type = TBD               |          Length = TBD         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                Value = Variable Guard Band                    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Type: TBD.

The Type of GB sub-TLV is left for future to define.

Length: TBD.

The length of CSSW sub-TLV is based on the define of the value of it.

Value: TBD.

The content of the CSSW sub-TLV and it is left for future researching.

6. Security Considerations

TBD.

7. Acknowledgments

TBD.

8. References

8.1. Normative References

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

[RFC2328] Moy, J., "OSPF Version 2", RFC 2328, April 1998.

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

8.2. Informative References

[draft-ietf-ccamp-wson-impairments-07]
Lee, Y., Bernstein, G., Li, D., and G. Martinelli, "A Framework for the Control of Wavelength Switched Optical Networks (WSON) with Impairments", July 2011.

Authors' Addresses

Jie Zhang
BUPT
No.10,Xitucheng Road,Haidian District
Beijing 100876
P.R.China

Phone: +8613911060930
Email: lgr24@bupt.edu.cn
URI: http://www.bupt.edu.cn/

Yongli Zhao
BUPT
No.10,Xitucheng Road,Haidian District
Beijing 100876
P.R.China

Phone: +8613811761857
Email: yonglizhao@bupt.edu.cn
URI: http://www.bupt.edu.cn/

Ziyan Yu
BUPT
No.10,Xitucheng Road,Haidian District
Beijing 100876
P.R.China

Phone: +8615116984347
Email: yzhziyan@gmail.com
URI: http://www.bupt.edu.cn/