Network Working Group
Internet Draft
Category: Standards Track

Fatai Zhang
Dan Li
Huawei
F. Javier Jimenez Chico
O. Gonzalez de Dios
Telefonica Investigacion y Desarrollo
C. Margaria. C
Nokia Siemens Networks

Expires: September 11, 2011 March 11, 2011

GMPLS-based Hierarchy LSP creation in Multi-Region and Multi-Layer Networks

draft-zhang-ccamp-gmpls-h-lsp-mln-03.txt

Status of this Memo

Abstract

This specification describes the hierarchy LSP creation models in the Multi-Region and Multi-Layer Networks (MRN/MLN), and provides the extensions to the existing protocol mechanisms described in [RFC4206], [RFC6107] and [RFC6001] to create the hierarchy LSP through multiple layer networks.

Conventions used in this document

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

Table of Contents

1. Introduction

Networks may comprise multiple layers which have different switching technologies or different switching granularity levels. The GMPLS technology is required to support control of such network.

[RFC5212] defines the concept of MRN/MLN and describes the framework and requirements of GMPLS controlled MRN/MLN. The GMPLS extension for MRN/MLN, including routing aspect and signaling aspect, is described in [RFC6001].

[RFC4206] and [RFC6107] describe how to set up a hierarchy LSP passing through multi-layer network and how to advertise the forwarding adjacency LSP (FA-LSP) created in the server layer network as a TE link via GMPLS signaling and routing protocols.

Based on these existing standards, this document further describes the provisioning of FA-LSP when the region nodes support multiple

interface switching capabilities and/or multiple switching granularities and/or adaptation functions, and then provides the extensions to the RSVP-TE protocol in order to set up hierarchy LSP according to the modes of hierarchy LSP provisioning.

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. Provisioning of FA-LSP in Server Layer Network

3.1. Selection of Switching Layers

As described in [RFC5212], the edge node of a region always has multiple Interface Switching Capabilities (ISCs), i.e., it contains multiple matrices which may be connected to each other by internal links. Nodes with multiple Interface Switching Capabilities are further classified as "simplex" or "hybrid" nodes by [RFC5212] and [RFC5339], where the simplex node advertises several TE links each with a single ISC value carried in its ISCD sub-TLV, while the hybrid node advertises a single TE link containing more than one ISCD each with a different ISC value. An example hybrid node with a link having multiple ISCs is shown in Figure 1, copied from [RFC5339].

```
                                  Network element
                           .............................
                           :            --------      :
                           :           |  PSC   |      :
                           :           |        |      :
                           :         --|#a      |      :
                           :        |  |   #b   |      :
                           :        |   --------       :
                           :        |       |          :
                           :        |  ----------      :
                           :    /|  | |    #c    |     :
                           :   | |--  |          |     :
                 Link1 ========| |    |    TDM   |     :
                           :   | |----|#d        |     :
                           :    \|     ----------      :
                           :............................
```

Figure 1 - Hybrid node (Copied from [RFC5339])

It's possible that the edge node of a region is a hybrid node which has multiple ISCs in the server layer. In this case, selection of which server layer to create the FA-LSP is necessary.

Figure 2 shows an example multi-layer network, where node B and C are region edge nodes having three switching matrices which support, for instance, PSC, TDM and WDM switching, respectively. The three switching matrices are connected to each other by the internal links. Both the link between B and E and the link between E and C support TDM and WDM switching capabilities.

```
+-------+  +------------+                   +------------+  +-------+
| +---+ |  |   +---+    |        FA         |    +---+   |  | +---+ |
| |PSC+-+--+---+PSC|....|...................|....|PSC+---+--+-+PSC| |
| +---+ |  | +-+-+-+    |                   |    +-+-+-+ |  | +---+ |
+-------+  | |   |      |                   |      |   | |  +-------+
 Node A    | |   |      |  +-------------+  |      |   | |   Node D
           | | +-+-+    |  |    +---+    |  |    +-+-+ | |
           | | |TDM|+   |  |   +|TDM|+   |  |   +|TDM| | |
           | | +-+-+|   |  |   |+-+-+|   |  |   |+-+-+ | |
           | |   |  ||\ |  | /||  |  ||\ |  | /||  |   | |
           | |   |  +| ||  || |+  |  +| ||  || |+  |   | |
           | +-+-+-+ | |====| | +-+-+ | |====| | +-+-+-+ |
           |   |WDM|-| ||  || |-|WDM|-| ||  || |-|WDM|   |
           |   +---+ |/ |  | \| +---+ |/ |  | \| +---+   |
           +------------+  +-------------+  +------------+
             Node B           Node E           Node C
```

Figure 2 - MLN with multiple ISCs at edge node

As can be seen in Figure 2, there are two choices when providing FA in the PSC layer network between node B and C: one is creating FA-LSP with TDM switching matrix through node B, E and C, the other is creating FA-LSP with WDM switching matrix through node B, E and C.

[RFC6001] introduces a new SC (Switching Capability) sub-object into the XRO (ref. to [RFC4874]), which is used to indicate which switching capability is not expected to be used. When one of the switching capabilities is selected, the SC sub-object can be included in the message to exclude all other SCs.

3.2. Selection of Switching Granularity Levels

Even in the case that the edge node only has one switching capability in the server layer, there may be still multiple choices for the server layer network to set up FA-LSP to provide new FA in the client layer network. This is because the server layer network may have the

capability of providing different switching granularity levels for the FA-LSP.

```
+-------+   +---------+                          +---------+   +-------+
| +---+ |   |  +---+  |           FA             |  +---+  |   | +---+ |
| |PSC|-+---+--+PSC|..|.......................|..|PSC+--+---+-|PSC| |
| +---+ |   |  +-+-+  |                          |  +-+-+  |   | +---+ |
+-------+   |    |    | ODU1/           ODU1/  |    |    |   +-------+
 Node A     |    |    | ODU2/ +-------+ ODU2/  |    |    |    Node D
            |  +-+-+  | ODU3  | +---+ | ODU3   |  +-+-+  |
            |  |TDM+--+-------+-+TDM+-+-------+--+TDM|  |
            |  +---+  |       | +---+ |        |  +---+  |
            +---------+       +-------+        +---------+
              Node B            Node E           Node C
```

Figure 3a - Multiple switching granularities in server layer

Figure 3a shows an example multi-region network, where the edge node B and C have PSC and TDM switching matrices, and where the TDM switching matrix supports ODU1, ODU2 and ODU3 switching levels. Therefore, when an FA between node B and C in the PSC layer network is needed, either of ODU1, ODU2 or ODU3 connection (FA-LSP) can be created in the TDM layer network.

```
 |<----------------------- ODU0 Connection ----------------------->|
 |                                                                  |
++------+   +---------+                          +---------+   +------++
| +---+ |   |  +---+  |      FA (ODU1/2/3)       |  +---+  |   | +---+ |
| |TDM|-+---+--+   |..|.......................|..|   +--+---+-|TDM| |
| +---+ |   |  |   |  |                          |  |   |  |   | +---+ |
+-------+   |  |TDM|  |       +-------+          |  |TDM|  |   +-------+
 Node A     |  |   |  | OTU3  | +---+ | OTU3     |  |   |  |    Node D
            |  |   +--+-------+-+TDM+-+-------+--+   |  |
            |  +---+  |       | +---+ |        |  +---+  |
            ++--------+       +-------+        +--------++
             |Node B            Node E           Node C |
             |                                          |
             |<--------- FA LSP (ODU1/2/3)------------>|
```

Figure 3b - TDM nested LSP provisioning

Figure 3b is another example multi-layer network within the same region. When there is a need to set up an FA between node B and C for the client layer ODU0 connection, the server layer has multiple choices, e.g., ODU1 or ODU2 or ODU3, for the FA-LSP if the multi-stage multiplexing is supported at node B and C.

```
    |<---------------- Client layer LSP (Bandwidth 1) --------------->|
    |                                                                 |
   ++------+   +---------+                       +---------+   +------++
   | +---+ |   |  +---+  |           FA           |  +---+  |   | +---+ |
   | |PSC|-+---+--+   |..|.......................|..|   +--+---+-|PSC| |
   | +---+ |   |  |   |  |                       |  |   |  |   | +---+ |
   +-------+   |  |PSC|  |       +--------+      |  |PSC|  |   +-------+
    Node A     |  |   |  |       | +---+ |      |  |   |  |    Node D
               |  |   +--+-------+-+PSC+-+-------+--+   |  |
               |  +---+  |       | +---+ |      |  +---+  |
               ++--------+       +--------+      +--------++
                |Node B           Node E           Node C |
                |                                         |
                |<--- Service layer LSP (Bandwidth 2) --->|
```

Figure 3c - PSC nested LSP provisioning

Figure 3c is a third example showing an LSP nesting scenario in a PSC signal-layer network (e.g., an MPLS-TP network). A PSC tunnel passing through node B, E and C is requested to carry the client layer LSP. There are multiple choices of the bandwidth of the tunnel, on the premise that the bandwidth of the FA-LSP is equal to or larger than the client layer LSP.

The selection of server layer switching matrix and switching granularity is based on both policy and bandwidth resources. The selection can be performed by planning tool and/or NMS/PCE/VNTM (Virtual Network Topology Manager, see [RFC5623]) and/or the network node.

3.3. Selection of Adaptation Capabilities

Adaptation function is also needed to be selected when creating the server layer connection. This is because the edge nodes may support multiple adaptation functions.

```
   +-------+   +-----------+                     +---------+   +-------+
   | +---+ |   |   +---+   |          FA         |  +---+   |   | +---+ |
   | |PSC|-+---+---+PSC|...|.....................|..|PSC+--+---+-|PSC| |
   | +---+ |   |   +---+   |                     |  +-+-+   |   | +---+ |
   +-------+   |___|_ _|___|                     |  __|__   |   +-------+
    Node A     |\_A_/ \_B_/|                     |  \_A_/   |    Node D
               |   |   |   |      +-------+      |    |    |
               |   +---+   |      | +---+ |      |  +-+-+   |
               |   |TDM+---+------+-+TDM+-+------+--+TDM|   |
               |   +---+   |      | +---+ |      |  +---+   |
               +-----------+      +-------+      +---------+
                  Node B           Node E          Node C

      _____                              _____
      \_A_/: Adaptation_Function_A;      \_B_/: Adaptation_Function_B;
```

Figure 4 - Selection of adaptation function

For example, in figure 4, the edge node B supports two adaptation functions, i.e., adaptation_function_A and adaptation_function_B, while the edge node C only supports adaptation_function_A. In this case, only adaptation_function_A can be used for the server layer connection.

The Call procedure ([RFC4974]) between edge node B and C may be used to negotiate and determine the adaptation function for the server layer if the call function is supported.

4. Signaling Extension Requirements for Server Layer Selection

[RFC5623], the framework of PCE-based MLN, provides the models of cross-layer LSP path computation and creation, which are listed below:

- Inter-Layer Path Computation Models:

  o Single PCE

  o Multiple PCE with inter-PCE

  o Multiple PCE without inter-PCE

- Inter-Layer Path Control Models:

  o PCE-VNTM cooperation

o Higher-layer signaling trigger

o NMS-VNTM cooperation (integrated flavor)

o NMS-VNTM cooperation (separate flavor)

This session keeps align with [RFC5623] except that the restriction of using PCE for path computation is not necessary (i.e., other element, such as network node, may also have path computation capability).

In this document, those models in [RFC4206] are reclassified into 3 models on the viewpoint of signaling:

- Model 1: Pre-provisioning of FA-LSP

- Model 2: Signaling trigger server layer path computation

- Model 3: Full path computation at source node

4.1. Model 1: Pre-provisioning of FA-LSP

In this model, the FA-LSP in the server layer is created before initiating the signaling of the client layer LSP. Two typical scenarios using this model are:

- Network planning and building at the stage of client network initialization.

- NMS/VNTM triggering the creation of FA-LSP when computing the path of client layer LSP. The path control models of PCE-VNTM cooperation and NMS-VNTM cooperation (both integrated and separate flavor) in [RFC5623] belong to this scenario.

In such case, the server layer selection and path computation is performed by planning tool or NMS/PCE/VNTM or the edge node. The signaling of client layer LSP and server layer FA-LSP are separated. The normal LSP creation procedures ([RFC3471] and [RFC3473]) are performed for these two LSPs and no new extension is required.

4.2. Model 2: Signaling trigger server layer path computation

In this model, the source node of client layer LSP only computes the route in its layer network. When the signaling of the client layer LSP reaches at the region edge node, the edge node performs server layer FA-LSP path computation and then creates the FA-LSP. When PCE is introduced to perform path computation in the multi-layer network, this model is the same as the model of "Higher-layer signaling trigger with Multiple PCE without inter-PCE" in [RFC5623].

In such case, the edge node will receive the client layer PATH message with a loose ERO indicating an FA is requested, and may perform the server layer selection (e.g., through the server layer PCE or the VNTM) and then compute and set up the path of the FA-LSP. The signaling procedure of client layer LSP and server layer FA-LSP is described detailedly in [RFC4206] and [RFC6107].

It's possible that the source node of the client layer LSP selects the server layer SC and/or granularity and/or adaptation function when performing path computation in the client layer, and requests or suggests the edge node to use an appointed server layer to create the FA-LSP.

The XRO including SC sub-object ([RFC6001]) is adopted for the server layer SC exclusion, which can be used indirectly to select server layer SC. Such solution is not straightforward enough and further more cannot be used for the selection of server layer granularity and adaptation function.

Therefore, in this case, new extensions for the selection of server layer SC, switching granularity and adaptation function are required.

4.3. Model 3: Full path computation at source node

In this model, the source node of the client layer LSP performs a full path computation including the client layer and the server layer routes. The server layer FA-LSP creation is triggered at the edge node by the client layer LSP signaling. When PCE is introduced to perform path computation in the multi-layer network, this model is the same as the model of "Higher-layer signaling trigger with Single PCE" or "Higher-layer signaling trigger with Multiple PCE with inter-PCE" in [RFC5623].

In such case, the server layer selection and server layer path computation is performed at the source node of the client layer LSP (e.g., through VNTM or PCE), but not at the edge node.

In [RFC4206], the ERO which contains the list of nodes and links (including the client layer and server layer) along the path is used in the client layer PATH message. The edge node can find out the tail end of the FA-LSP based on the switching capability of the node using the IGP database (see session 6.2 of [RFC 4206]).

Similar to the problem of model 2, the edge node is not aware of which switching granularity and which adaptation function to be selected for the FA-LSP because the ERO and/or XRO do not contain such information. Therefore, the edge node may not be able to create the FA-LSP, or may select another switching granularity by itself which is different from the one selected previously at the source node, which makes the creation of hierarchy LSP out of control.

Therefore, new extensions for the selection of server layer SC, switching granularity and adaptation function are also required in this model.

5. ERO Sub-Object

In order to solve the problems described in the previous sessions, a new sub-object named SERVER_LAYER_INFO sub-object is introduced in this document, which is carried in the ERO and is used to indicate which server layer to create the FA-LSP.

The SERVER_LAYER_INFO sub-object is put immediately behind the node or link (interface) address sub-object, indicating the related node is a region edge node on the LSP in the ERO.

The format of the SERVER_LAYER_INFO sub-object is shown below:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|L|    Type     |     Length    |M|         Reserved            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| LSP Enc. Type |Switching Type |            G-PID              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                      Traffic Parameters                       |
~                                                               ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

- L bit: MUST be zero and MUST be ignored when received.

- Type: The SERVER_LAYER_INFO sub-object has a type of xx (TBD).

- Length: The total length of the sub-object in bytes, including the Type and Length fields. The value of this field is always a multiple of 4.

- M (Mandatory) bit: When set, it means the edge node MUST set up the FA-LSP in the appointed server layer; otherwise, the appointed server layer is suggested and the edge node may select other server layer by local policy.

- LSP Encoding Type, Switching Type and G-PID: These 3 fields are used to point out which switching layer is requested to set up the FA-LSP. The values of these 3 fields are inherited from the Generalized Label Request Object in GMPLS signaling, referring to [RFC3471], [RFC3473] and other related standards and drafts. Note that G-PID can be used to indicate the payload type of the server layer (i.e., the client signal) as well as the adaptation function for adapting the client signal into the server layer FA-LSP.

- Traffic Parameters: The traffic parameters field is used to indicate the switching granularity of the FA-LSP. The format of this field depends on the switching technology of the server layer (which can be deduced from the LSP Encoding Type and Switching Type fields in this sub-object) and is consistent with the existing standards and drafts. For example, the Traffic Parameters of Ethernet, SONET/SDH and OTN are defined by the [RFC6003], [RFC4606] and [OTN-ctrl] respectively.

5.1. Application of SERVER_LAYER_INFO sub-object

When a node receives a PATH message containing ERO and finds that there is a SERVER_LAYER_INFO sub-object immediately behind the node or link address sub-object related to itself, the node determines that it's a region edge node. Then, the edge node finds out the server layer selection information from the sub-object:

- Determine the switching layer by the LSP Encoding Type and Switching Type fields;

- Determine the switching granularity of the FA-LSP by the Traffic Parameters field;

- Determine the adaptation function for adapting the client signal into the server layer FA-LSP by the G-PID field.

The edge node MUST then determine the other edge of the region, i.e., the tail end of the FA-LSP, with respect to the subsequence of hops of the ERO. The node that satisfies the following conditions will be treated as the tail end of the FA-LSP:

- There is a SERVER_LAYER_INFO sub-object that immediately behind the node or link address sub-object which is related to that node;

- The LSP Encoding Type, Switching Type, G-PID and the Traffic Parameters fields of this SERVER_LAYER_INFO sub-object is the same as the SERVER_LAYER_INFO sub-object corresponding to the head end;

- The node is the first one that satisfies the two conditions above in the subsequence of hops of the ERO.

If a match of tail end is found, the head end now has the clear server layer information of the FA-LSP and then initiates an RSVP-TE session to create the FA-LSP in the appointed server layer between the head end and the tail end.

6. Security Considerations

TBD.

7. IANA Considerations

TBD.

8. Acknowledgments

TBD.

9. References

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

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

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

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

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

[RFC5212] K. Shiomoto et al, "Requirements for GMPLS-Based Multi-Region and Multi-Layer Networks (MRN/MLN)", RFC5212, July 2008.

[RFC5339] JL. Le Roux et al, "Evaluation of Existing GMPLS Protocols against Multi-Layer and Multi-Region Networks (MLN/MRN)", RFC5339, September 2008.

[RFC4206] K. Kompella et al, "Label Switched Paths (LSP) Hierarchy with Generalized Multi-Protocol Label Switching (GMPLS) Traffic Engineering (TE)", RFC4206, October 2005.

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

[RFC4974] D. Papadimitriou and A. Farrel, "Generalized MPLS (GMPLS) RSVP-TE Signaling Extensions in Support of Calls", RFC4974, August 2007.

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

[RFC5623] E. Oki et al, "Framework for PCE-Based Inter-Layer MPLS and GMPLS Traffic Engineering", RFC 5623, September 2009.

[RFC4606] E. Mannie, D. Papadimitriou, "Generalized Multi-Protocol Label Switching (GMPLS) Extensions for Synchronous Optical Network (SONET) and Synchronous Digital Hierarchy (SDH) Control", RFC 4606, August 2006.

[OTN-ctrl] Fatai Zhang et al, "Generalized Multi-Protocol Label witching (GMPLS) Signaling Extensions for the evolving G.709 Optical Transport Networks Control", draft-zhang-ccamp-gmpls-evolving-g709-04.txt, February 27, 2010.

[RFC6003] D. Papadimitriou, "Ethernet Traffic Parameters", RFC6003, October, 2010.

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

10. Authors' Addresses

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

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

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

Phone: +86-755-28970230
Email: huawei.danli@huawei.com

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

Phone: +86-755-28972914
Email: yi.lin@huawei.com

Francisco Javier Jimenez Chico
Telefonica Investigacion y Desarrollo
Emilio Vargas 6
Madrid, 28043 Spain

Phone: +34 913379037
Email: fjjc@tid.es

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