Internet Draft   Debanjan Saha
Expiration Date: Sept., 1, 2000   Bala Rajagopalan
                                                   Bo Tang
                                                    Tellium Inc.

RSVP Extensions for Signaling Optical Paths

draft-saha-rsvp-optical-signaling-00.txt

1. Status of this Memo

This document is an Internet-Draft and is in full conformance with
all provisions of Section 10 of RFC2026.

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

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

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

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

2. Abstact

This document has two objectives. First, we identify signaling
requirements for setting up and maintaining Optical Switched Paths
(OSPs) in mesh-connected optical networks. We then propose
extensions to RSVP to satisfy some of these requirements within the
context of MPLS-based traffic engineering framework.

3. Introduction

The IETF is currently in the process of defining and standardizing a
control architecture for creating and managing switched OSPs. A
number of contributions [2,3,5] submitted to IETF, and other
standards bodies have already addressed some of the important
aspects of this architecture. In this document we have identified
several key requirements for signaling OSPs within the framework
proposed in [2,3,5] that take into account some of the intricacies
of an optical network. In light of these observations, we have
proposed extensions to RSVP-TE to satisfy the signaling requirements
of OSPs.

4. Optical Network Architecture

The optical network model considered in this draft consists of
multiple Optical Cross-Connects (OXCs) interconnected by optical
links in a general topology referred to as an "optical mesh
network". In general, it can be expected that topologically adjacent
OXCs in an optical mesh network will be connected via multiple,
parallel bi-directional links. Each link can carry one or more
optical channels or wavelengths. Each wavelength can be further
demultiplexed into multiple Sub-channels. We also assume that one or
more control channels exist between neighboring OXCs for signaling
purposes.

Each OXC is assumed to be capable of switching a wavelength from a
given input port to a given output port. This switching function is
controlled by appropriately configuring a cross-connect table. In
the most general case, the cross-connect table consists of entries
of the form <input port i, wavelength j, sub-channel k, output port
m, output wavelength n, output sub-channel p>, indicating that the
Sub-channel k, in wavelength j entering input port i will be
switched to Sub-channel p in output wavelength n at output port j.
An "optical path" from an ingress port in an OXC to an egress port
in a remote OXC is established by setting up suitable cross-connects
in the ingress, the egress and a set of intermediate OXCs such that
a continuous physical path exists from the ingress to the egress
port.

Automated establishment of optical paths involves setting up the
cross-connect table entries in the appropriate OXCs in a coordinated
manner such that the desired physical path is realized. The request
to establish an optical path may arise from either a router (or some
other device) connected to the OXCs or from a management system.
Such a request must identify the ingress and the egress OXC as
endpoints of the optical path. In addition, it may also optionally
specify the input and output ports, wavelengths, and Sub-channels.
The request may also include bandwidth parameters and channel type
(e.g., OC-48/OC-192, 10 GE/N), reliability parameters, restoration
options, setup and holding priorities for the path etc. On receipt
of the request, the ingress OXC computes a suitable route for the
requested path, following applicable policies and constraints. Once
the route has been computed, the ingress invokes RSVP to set up the
path. The purpose of this draft is to identify special requirements
for signaling OSPs and propose extensions to RSVP protocol to
satisfy those requirements.

5. Signaling Requirements

5.1 Support for Bi-directional OSPs

A OSP can be either unidirectional or bi-directional. However, bi-directional OSPs are far more typical than unidirectional OSPs. Although a bi-directional OSP can be modeled as a pair of unidirectional OSPs that can be setup independently, there are many limitations to this approach. Most importantly, two separately signaled unidirectional OSPs may follow two different paths

altogether. Even if they follow the same path, it is quite likely that they will use two different ports/wavelengths/Sub-channels (on the same OXC) for the forward and the backward directions. Either scenario makes it difficult for the network management system to manage and maintain the OSPs, especially in the event of failures that trigger automatic fault isolation and restoration events such as SONET APS (Automatic Protection Switching). The ability to setup bi-directional OSPs also reduces the signaling complexity and setup time.

Since RSVP is designed to setup unidirectional paths, extensions are required for signaling bi-directional paths. The tricky part here is the handling of the potential race condition. Since different nodes may autonomously initiate the signaling for optical paths, it is possible that two path set-up attempts are in progress at the same time. Specifically, while setting up an optical path, an OXC A may select output port i, which is connected to input port j of the "next" OXC B. Concurrently, OXC B may select output port j for setting up a different optical path, where the next downstream OXC is A. If they also happen to choose the same wavelength and Sub-channel (when applicable) this results in a "collision". There are two ways to deal with such collisions. First, collisions may be detected and the involved paths may be torn down and re-established. Or, collisions may be avoided altogether. The latter solution is preferred, if the complexity involved is acceptable. In Section 5 we propose extensions to RSVP-tunnel to handle this race condition which avoids collision.

5.2 Support for Protection Paths

Another important requirement for lighpaths is protection. We consider two primary protection modes:

1. Span protection: In this mode the optical channel between

every two adjacent OXCs along the primary path is protected
individually. There are different modes of span protection,
such as 1+1, M:N etc.

2. Path protection: In this mode end-to-end backup paths are
provisioned for the primary paths. In path protection backup
paths are typically link or link and node disjoint from the
primary paths. There could be various levels of sharing among
backup paths, such as completely dedicated backup paths (1+1
style), backup paths that are shared by multiple primary paths
(M:N style), and backup paths that share resources
(wavelengths) among themselves. In the first two cases the
backup paths span between the same source and the destination
pair and can be pre-setup. The backup paths that share
resources among themselves can be pre-provisioned but can not
be pre-setup. In this scenario the backup paths can span
between different source and destination pairs.

RSVP-TE has limited support for setting up protection paths. In
section 7 we introduce new C-types to signal protection requirements
for both span and path protection. We also introduce a new shared
reservation style to enable sharing of resources among backup paths.

Internet Draft   draft-saha-rsvp-optical-signaling-00.txt

5.3 Support for Permanent OSPs

RSVP relies on soft-state mechanisms to maintain OSPs. If the Path
and Resv states corresponding to a OSPs are not refreshed within the
configured timeout value, they are timed out and the OSP is deleted.
The requirements for OSPs are quite different. Most OSPs are long-
lived and should not be torn down unless explicitly requested. Also,
the impact of transient partial failures must be minimized in an
optical network. Specifically, optical paths that are not directly
affected by a failure must not be torn down due to the failure. For
example, the control processor in an OXC may fail, affecting
signaling and other internodal control communication. Similarly, the
control channel between OXCs may be affected temporarily by a
failure. These failures may not affect already established OSPs
passing through the OXC fabric. These requirements can be summarized
as follows:

1. During setup OSPs can be configured as permanent OSPs. Permanent
OSPs should not be torn down unless explicitly requested.

2. The Path and Resv states for permanent OSPs need not be refreshed
and should not be timed out.

3. We need explicit mechanisms for tearing down permanent OSPs. This
requires enhancements to PathTear and ResvTear messages since these

messages are not refreshed and are not transported reliably.

6. Extensions for Bi-directional OSPs

To avoid the potential race condition and collision in label
assignment we propose a solution where the OXC with the higher node
ID (IP address) is always the owner of the label space between two
adjacent OXCs. For the purpose of this discussion we assume that a
label consists of three components: (1) the port ID, (2) the
wavelength ID, and (3) the sub-channel ID. We also assume that if
port i on OXC1 is connected to port j on OXC2, both OXCs have that
information available to them via link layer discovery protocol. The
wavelength ID and the sub-channel ID are assumed to be the same on
both OXCs.

We define a new C_type for the label request object defined in [2]
as follows:

Label Request with OSP

Class = 19, C_type = 10   (OSP_Tunnel)

```
    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
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |    Reserved     |    Flags      |            L3PID            |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                           Port ID                            |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |          Wavelength ID        |        Sub-channel ID        |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```
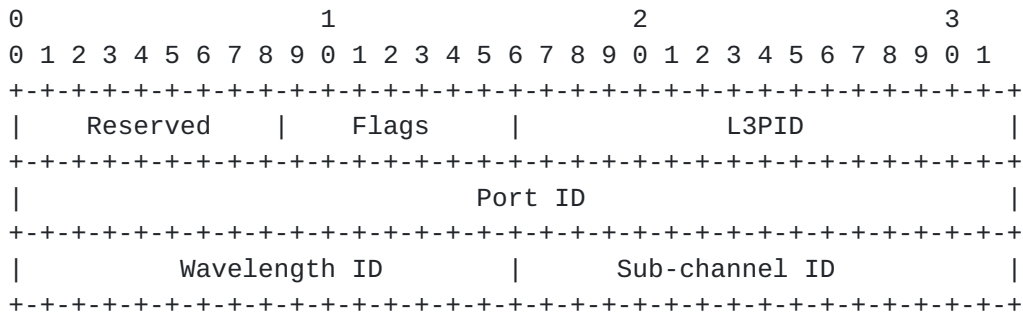
Reserved: This field is reserved. It MUST be set to zero on
transmission and MUST be ignored on receipt.

Flags (8 bits):
xxxxxxx0 û If the upstream OXC is the slave and no label has been
picked.
xxxxxxx1 û If the upstream OXC is the master has chosen the label
for the OSP.

L3PID (16 bits): An identifier of the layer 3 protocol using this
path. Standard Ethertype values are used.

Port ID (32 bits): Valid only when the LSB of the Flags is set. It

is the Port ID of the downstream OXC.

Wavelength ID (16 bits): Valid only when the LSB of the Flags is
set. It is the wavelength ID within the port ID.

Sub-channel ID (16 bits): Valid only when the LSB of the Flags is
set. It is the sub-channel ID within the wavelength ID.

The basic scheme is as follows. The upstream OXC sends the
Label_Request object with of C_type 4 to the downstream OXC. If the
upstream OXC is the master (has higher node ID) it picks the label
and sets the LSB of Flags to 1. If the upstream OXC is the slave, it
sets the LSB of the Flags to 0 and the label is picked by the
downstream OXC. Since the master irrespective of the direction of
the message flow always picks the label, no collision occurs.

We also add a new C_Type (10: OSP_Tunnel) to the Label object. The
LABEL object has the following format:

LABEL class = 16, C_Type = 10

```
    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
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                                                               |
   //                     (Object contents)                       //
   |                                                               |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                          Port ID                              |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |        Wavelength ID       |          Sub-channel ID          |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```
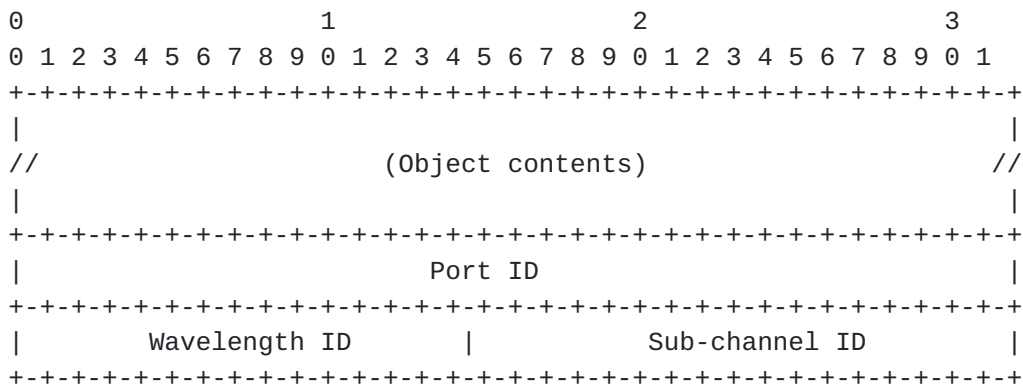
The contents of a LABEL object are a stack of labels, where each
label is encoded in 8 octets as shown above. The top of the stack
is in the right 8 octets of the object contents. A LABEL object
that contains no labels is illegal.

Internet Draft   draft-saha-rsvp-optical-signaling-00.txt

7. Extensions for Protection Paths

To handle this issue we propose a new C_type for the SESSION_ATTRIBUTE
object.

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   OSP type    |  Local Prot.  | Local Prot. Extension         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Setup Prio   | Holding Prio  |          Reserved             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
//                     Optional Subobjects                     //
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```
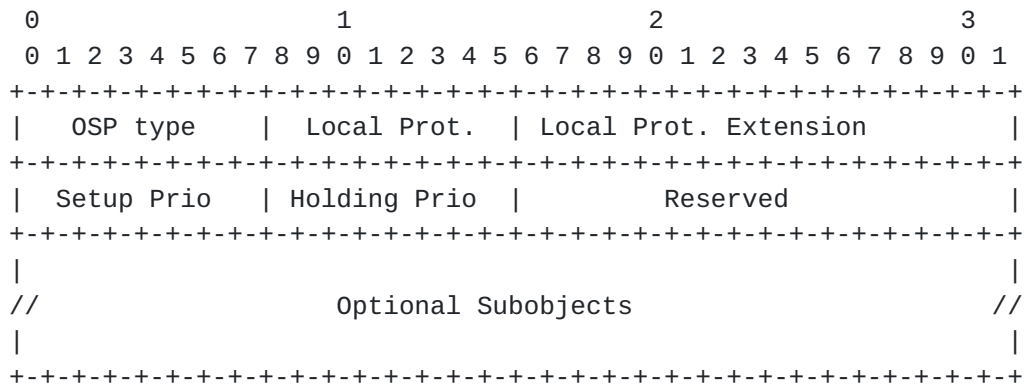
 Class-Num: The Class-Num indicates that the object is 207. (TBD)

  C-Type: The C-Type OSP_Tunnel (10). (TBD)

  OSP Type (8 bits):

01 Permanent Uni-directional Primary Path
02 Primary Bi-directional Primary Path
03 Permanent Uni-directional Backup Path
04 Permanent Bi-directional Backup Path
05 Permanent Uni-directional Shared Backup Path
06 Permanent Bi-directional Shared Backup Path
07 Soft-state Uni-directional Primary Path
08 Soft-state Bi-directional Primary Path
09 Soft-state Uni-directional Backup Path
10 Soft-state Bi-directional Backup Path

Local Protection (8 bits):
01 1+1
02 1:N

Local Protection Extension (16 bits): Protection scheme specific
information.

Setup Priority: The priority of the session with respect to taking
resources, in the range of 0 to 7. The value 0 is the highest
priority. The Setup Priority is used in deciding whether this
session can preempt another session.

Holding Priority: The priority of the session with respect to
holding resources, in the range of 0 to 7. The value 0 is the
highest priority. Holding Priority is used in deciding whether this
session can be preempted by another session.

Besides the mandatory fields described above, the session attribute
object may contain one or more optional subobjects. The subobjects
are formatted as TLVs and are defined below.

```
  0                   1
  0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+------------//----------------+
 |   Type      |    Length   | (Subobject contents)           |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+------------//----------------+
```

Type: The Type indicates the type of contents of the subobject.
Currently defined values are:
            0   Reserved
            1   PRIMARY_OSP subobject
            2   SHARING_POLICY subobject

The PRIMARY_OSP subobject may be included in the SESSION_ATTRIBUTE
object while setting up a backup path. This subobject includes
information about the primary OSP and may be used for functions such
as local admission control, sharing of backup etc. The format and
content of this subobject is TBD.

SHARING_POLICY subobject may be included in the SESSION_ATTRIBUTE
object while signaling to setup shared backup paths. This
information can be used to facilitate local sharing decision. The
format and content of this subobject is TBD.

Note that for the shared backup paths the cross-connects can not be
setup during the signaling for the backup path since multiple backup
paths may share the same resource and can over-subscribe it. The
idea behind shared backups is to make soft reservations and to claim
the resource only when it is required. A shared backup path can be
converted to a primary path by sending Path refresh messages with
OSP Type set to 01 or 02 depending on the type of the backup path.

[8](8). **Extensions for Permanent OSPs**

In order to satisfy the requirements for a permanent OSP we need to
the followings:

1. Add a new session attribute that identifies a OSP as a permanent
OSP. If this new attribute is not present, by default an OSP is
considered a ôsoft-stateö OSP. For permanent OSPs, the Path State
Block (PSB) and the Resv State Block (RSB) are never timed out. The
only way to delete a permanent OSP is through explicit signaling,
that is via a PathTear or a ResvTear message.

2. Since the permanent OSPs are never timed out and the PathTear and
ResvTear messages are not refreshed, we need additional mechanisms
that can be used to tear down the permanent OSPs in a reliable way.
One way to achieve this is to add reliability to PathTear and

ResvTear messages as suggested in [4]. The other option is to change
the OSP_TYPE of a permanent OSP to a soft-state OSP and then rely on
RSVP soft-state mechanisms along with PathTear to tear down the OSP.

9. Summary

This draft described some issues that arise in developing signaling
procedures for path provisioning and restoration in optical
networks. The following signaling requirements were identified:
support for support for bi-directional optical path set-up with
collision avoidance features, support for protected paths, and
support for permanent OSPs. Other requirements may arise as work
proceeds in this area. In light of these observations, we propose
extensions to RSVP-TE to satisfy these requirements.

10. **References**

1. Awduche, D., Rekhter, Y., Drake, J., Coltun, R., "Multi-Protocol
Lambda Switching: Combining MPLS Traffic Engineering Control With
Optical Crossconnects", draft-awduche-mpls-te-optical-01.txt.

2. Awduche, et al, "Extensions to RSVP for LSR Tunnels," draft-ietf-
mpls-rsvp-lsp-tunnel-04.txt, Work in progress, September 1999

3. Basak, D., Awduche, D., Drake, J., Rekhter, Y., "Multi-protocol
Lambda Switching: Issues in Combining MPLS Traffic Engineering
Control With Optical Crossconnects", draft-basak-mpls-oxc-issues-
00.txt

4. Berger L., Gan D., Swallow G., Pan P., Tommasi F., Molendini S.,
ôRSVP Refresh Overhead Reduction Extensionsö, draft-ietf-rsvp-
refresh-reduct-02.txt.

5. Rajagopalan B., Saha D., Tang B., Krishna B., ôSignaling Framework
for Automated Establishment and Restoration of Paths in Optical Mesh
Networksö, dratf-rstb-optical-signaling-framework-00.txt, March
2000.

11. **Author Information**

Debanjan Saha, Bala Rajagopalan, Bo Tang
Tellium Inc.
2 Crescent Place
P.O Box 901

Ocean Port, NJ 07757
Email: {dsaha, braja, btang}@tellium.com

         ********** This draft expires on Sept. 1, 2000 ***********