Network Working Group Jonathan P. Lang (Chromisys)
Internet Draft Krishna Mitra (Chromisys)
Expiration Date: September 2000 John Drake (Chromisys)

Extensions to RSVP for optical networking

draft-lang-mpls-rsvp-oxc-00.txt

1. Status of this Memo

2. Abstract

Dynamically provisionable optical crossconnects (OXCs) will play an active role in future optical networks and the MPL(ambda)S control plane will be used to establish, teardown, and reroute optical trails through the network. This document specifies extensions to RSVP to address some of the unique requirements of such optical trails. Specifically, we propose extensions to RSVP that allow an upstream node to make a Label suggestion to a downstream node when establishing an optical trail and allow both directions of a bi-directional optical trail to be established simultaneously. A new message type is also defined so that an RSVP node can notify (possibly non-adjacent) RSVP nodes when network failures occur, without affecting the RSVP states of intermediate RSVP nodes. Finally, we propose a modification to allow bundle messages to be sent to non-adjacent RSVP nodes.

3. Conventions used in this document

In this document, we follow the naming convention of [2] and use OXC to refer to all categories of optical crossconnects, irrespective of the internal switching fabric. We use the term source node to refer to an RSVP node that initiates the optical trail establishment and the term destination node to refer to the RSVP node that terminates the trail at the other end of the network. Furthermore, we call the message path from the source node to the destination node the downstream direction and the reverse path from the destination node back to the source node the upstream direction. Note that for bi-directional connections our terminology is such that there is only one source node and one destination node.

4. Introduction

Future optical networks will consist of label switched routers (LSRs) and optical crossconnects (OXCs) that internetwork using the MPL(ambda)S control plane. Support for provisioning and restoration of end-to-end optical trails within this type of network imposes new requirements on the signaling protocols. Specifically, optical trails will require small setup latency, support for bi-directional trails, and rapid restoration of trails in case of network failures. This document builds on work already done for traffic engineering in MPLS and proposes solutions for these requirements.

The modifications proposed in this document enhance the extensions of RSVP-TE [3] to support the following functions:

1. Reduction of trail establishment latency by allowing resources to be configured in the downstream direction.
2. Establishment of bi-directional trails as a single process instead of establishing two uni-directional trails, one in each direction, each being a separate process. Normally, both directions of a bi-directional trail have the same traffic engineering requirements and need to be routed over the same physical route. As a result, they cannot be treated as two separate trail requests.
3. Fast failure notification to a node responsible for trail restoration can be achieved so that restoration techniques can be quickly initiated. For example, for end-to-end path restoration, the source is responsible for rerouting failed trails, and must be notified when the trail's resources are involved in a failure.

The organization of the remainder of this document is as follows. In Section 5, we propose a Label suggestion to reduce the trail establishment latency. In Section 6, we present modifications to RSVP so that both directions of a bi-directional trails can be provisioned simultaneously. In Section 7, we introduce a new Notify message that is to notify nodes when failures occur in the network. Finally, in Section 8, we discuss a modification to the bundle message [3] to allow transmission between non-adjacent nodes.

5. Label suggestion

Currently in RSVP, the Label object for an optical trail is returned in the Resv message. A unique feature of OXCs is that selecting a (fiber, lambda) Label (see [4]) for a trail requires configuring the OXC so that an input is switched to an output, and all data that arrives over the input must go to the same output. This is different from traditional LSRs where multiple flows from the same input maybe be assigned different Labels and subsequently switched to different outputs. A consequence of this is that when an OXC is initially configured, Labels can be assigned to each input and output and protocols (such as LMP [5]) can be used to exchange Label mappings between adjacent nodes.

A consequence of the inherent receiver-oriented nature of RSVP is that the internal configuration of an OXC in the downstream direction cannot be initiated until it receives the Resv message from the downstream node. The ability to begin configuring an OXC before receiving a Label object in the Resv message can provide a significant reduction in the setup latency, especially in OXCs with non-negligible configuration time.

To accomplish this, we propose that an upstream OXC suggest a (fiber, lambda) Label for the downstream node to use by including the suggested Label object in the Label Request object [3] of the Path message. The Label object will contain the downstream nodeÆs Label for the bearer channel, which can be obtained through the Link Management Protocol (LMP) [5]. This will allow the upstream OXC to begin its internal configuration before receiving the Resv message from the downstream node. If, however, the downstream node ignores the suggested Label and passes a different Label upstream, the upstream OXC must reconfigure itself so that it uses the label specified by the downstream node.

5.1. Label Request

The LABEL_REQUEST object format is shown below, where we have defined a new C_Type for a suggested Label.

Class = 19, C_Type = 5 (suggested 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|       Link Media Type         |             L3PID             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                          Label Object                         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Link Media Type:
The Link Media Type is the two-octet media type values in IS-IS/OSPF Link Media Type TLV defined in [4].

L3PID:
The L3PID is an identifier of the layer 3 protocol using this path. Standard Ethertype values are used.

Label Object:
The Label object is the suggested Label for the downstream node.

6. Bidirectional Optical Paths

In future optical networks, it may be desirable to establish bi-directional optical paths across the network. Using RSVP-TE [3], this requires establishing two unidirectional paths: an initial path from the source to the destination and a subsequent path from the destination back to the source. This approach has two disadvantages: the latency to establish the bi-directional path requires three source/destination transit times, and the time window between reserving the resources in the downstream direction and reserving them in the upstream direction may be large (as much as two times the source/destination transit time), decreasing the probability of successfully establishing the overall bi-directional path.

To address the disadvantages of establishing bi-directional paths using current techniques in RSVP, we propose that a Label object is added to the Path message in the downstream direction. In this way, the upstream direction of the bi-directional path is established on the first pass from the source to the destination, reducing the latency of the reservation process. Furthermore, if suggested Labels are used for the downstream direction of the bi-directional path (see Section 5), then the time between reserving resources in the upstream and downstream directions can be eliminated, increasing the overall probability of success for the bi-directional path.

The format of the Path message is as follows (where we assume the extensions of [3] are also implemented):

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

<sender descriptor> ::= <SENDER_TEMPLATE> [<SENDER_TSPEC>]
                   [<ADSPEC>] [<RECORD_ROUTE>]
```

Note that for bi-directional trails, if a Label suggestion is also used, there will be two Labels in the Path message: the upstream Label in the Label object and the suggested Label in the Label_Request object.

6.1. Contention Resolution

It is possible that a pair of uni-directional bearer channels (one in each direction) is routed together through the fiber optic infrastructure and that it is desirable to allocate a bi-directional trail to the pair. (This would be known through local configuration at a given node.) In this situation, it is possible to get a collision between two bi-directional path requests traveling in opposite directions between two adjacent nodes. For example, this could occur if each node selects for its upstream direction a unidirectional bearer channel from the same pair.

To address this situation, we propose that the node with the higher IP address wins the contention. The node with the higher IP address will send a PATH_ERROR message to the adjacent node. To ensure proper interpretation of the error, a new Error Code is defined (Error Code 25) to indicate that the resources could not be obtained due to a collision. This message is intended for the adjacent node only, and should not be propagated further. When the adjacent node receives the PATH_ERROR message with Error Code 25, it will try to allocate a different (fiber, lambda) Label to the bi-directional path. If at that point, no other resources are available, the adjacent node will send a new PATH_ERROR message upstream towards the source node, indicating that resources were not available for the bi-directional path.

7. Failure Notification

A requirement of reliable optical networks is that reaction to network failures must be quick, and rerouting decisions must be made intelligently. A critical functionality of networking protocols that satisfy this requirement is the ability to rapidly detect and isolate failures as well as to notify the nodes responsible for restoring the failed optical trails. Failure detection should be handled at the layer closest to the failure, the optical layer, and a number of techniques are being developed to facilitate rapid failure detection. Failure isolation requires coordination between adjacent nodes, and protocols such as the LMP [5] can be used to rapidly isolate network failures using a lightweight messaging scheme. Failure notification involves exchanging information between nodes that may or may not be adjacent, and signaling protocols should be used to exchange the information. It should be clear that the methods used to detect and isolate network failures can be independent of the signaling protocol that is used to notify nodes of failed optical trails, and we will not consider failure detection/isolation techniques further in this document.

As part of failure notification, a node passing transit trails (i.e., trails that neither originate nor terminate at that node) should be able to notify the nodes responsible for restoring the trails when failures occur (e.g., the source node needs to be notified if end-to-end path restoration is used), without intermediate nodes processing the messages or modifying the state of the affected trails. This is important because restoration procedures may reuse segments of the original trail in a "make-before-break" fashion. As part of the notification process, the affected trail and the failed resource must be identified in the message.

We propose a new RSVP message, called the Notify message, which can be used to notify (possibly non-adjacent) RSVP nodes when failures occur. The Notify message will be transmitted with the router alert option turned off so that intermediate nodes will not process or modify the message, but only perform standard IP forwarding of the message.

7.1. Notify message

The Notify message is sent to the unicast address of an RSVP host. Notify messages do not modify the state of any node through which they pass and are only reported to the addressed RSVP host.

```
<Notify message> ::= <Common Header> <SESSION> <ERROR_SPEC>
                     [<sender descriptor>]

<sender descriptor> ::=  <SENDER_TEMPLATE> [<SENDER_TSPEC>]
                         [<ADSPEC>] [<RECORD ROUTE]>
```

The ERROR_SPEC object specifies the error and includes the IP address of either the node that detected the error or the link that has failed.

8. Bundle message modification

A recent Internet draft [3] proposed an extension to RSVP to allow the use of bundle messages to reduce the overall message-handling load. An RSVP bundle message consists of a bundle header followed by a body consisting of a variable number of standard RSVP messages. Support for the bundle message is optional, and currently, bundle messages can only be sent to adjacent RSVP nodes.

Future optical networks will require high reliability, which can be ensured by using fast protection and restoration techniques. These techniques may be applied at the local level (e.g., span protection/restoration) and/or at the path level (e.g., path protection/restoration) and may require rerouting multiple optical paths simultaneously. To enable fast protection/restoration techniques, an RSVP node may need to send multiple simultaneous messages to a nonadjacent RSVP node. For example, an intermediate node may need to send multiple Notify messages (see Section 7.1) to a particular source RSVP node if it detects a failure affecting multiple trails with that node as the source.

In order to effectively restore the network to a stable state, nodes that are running restoration algorithms should consider as many failed trails as possible before making restoration decisions. To improve performance and ensure that the nodes are provided with as many of the affected paths as possible, it is useful to include the entire set of Notify messages in a single bundle message and send it to the responsible RSVP node directly, without message processing by the intermediate RSVP nodes. This can be accomplished by addressing the bundle message to the source RSVP node and turning off the router alert option in the IP header.

Intermediate RSVP nodes should perform standard IP forwarding of this message (i.e., they should not process the message) and must not alter its contents. The source RSVP node must disable the RSVP neighbor check, which would normally cause it to discard the message.

The source RSVP node will send an RSVP MESSAGE ACK (see [6]), containing the message IDs of all of the messages in the bundle message, addressed to the node that sent it the bundle message. Intermediate RSVP nodes should perform standard IP forwarding of this message (i.e., not process it) and must not alter its contents.

9. Security Considerations

No new security issues are raised in this document. See [7] for a general discussion of security issues for RSVP.

10. Referenc

[1] Bradner, S., "The Internet Standards Process -- Revision 3," BCP 9, RFC 2026, October 1996.

[2] Awduche, D. O., Rekhter, Y., Drake, J., Coltun, R., "Multi-Protocol Lambda Switching: Combining MPLS Traffic Engineering Control with Optical Crossconnects," Internet Draft, draft-awduche-mpls-te-optical-00.txt, October 1999.

[3] Awduche, D. O., Berger, L., Gan, D.-H., Li, T., Swallow, G., Srinivasan, V., "Extensions to RSVP for LSP Tunnels," Internet Draft, draft-ietf-mpls-rsvp-lsp-tunnel-04.txt, September 1999.

[4] Kompella, K., Rekhter, Y., Awduche, D. O., et al, "Extensions to IS-IS/OSPF and RSVP in support of MPL(ambda)S," Internet Draft, draft-kompella-mpls-optical.txt, February 2000.

[5] Lang, J. P., Mitra, K., Drake, J., et al, "Link Management Protocol," Internet Draft, draft-lang-mpls-lmp-00.txt, March 2000.

[6] Berger, L., Gan, D.-H., Swallow, G., Pan, P., Tommasi, F., "RSVP Refresh Overhead Reduction Extensions," Internet Draft, draft-ietf-rsvp-refresh-reduct-02.txt, January 2000.

[7] Braden, R., Zhang, L., Berson, S., et al, "Resource ReserVation Protocol -- Version 1 Functional Specification," RFC 2205, September 1997.

11. Author's Addresses

Jonathan Lang
Chromisys, Inc.
421 Pine Avenue
Santa Barbara, CA 93117

Krishna Mitra
Chromisys, Inc.
1012 Stewart Drive
Sunnyvale, CA 94086

Email: jplang@Chromisys.com   email: krishna@Chromisys.com

John Drake
Chromisys, Inc.
1012 Stewart Drive
Sunnyvale, CA 94086
email: jdrake@Chromisys.com