Internet Draft
draft-bala-restoration-signaling-00.txt
Expires on: 8/22/2001

Bala Rajagopalan
Debanjan Saha
Tellium, Inc.
Greg Bernstein
Ciena Corp.
Vishal Sharma
Jasmine Networks

Signaling for Fast Restoration in Optical Mesh Networks

Status of this Memo

1. Abstract

Restoration of switched connections under tight time constraints is a challenging problem in optical mesh networks. This draft describes different protection modes for connections under both local and end-to-end restoration scenarios, and defines lightweight protocol mechanisms (over IP) for restoration-related signaling.

2. Introduction

Restoration of switched connections under tight time constraints is a challenging problem in optical mesh networks. Such a network

consists of optical cross-connects (OXCs) connected in a general topology [1]. Restoration typically involves the activation of an alternate path for a connection when a failure is encountered in the

primary path. A path for a connection (primary or alternate) is characterized by an ingress port, an egress port, and a set of intermediate OXCs and links through which the connection is routed. The ingress and egress port remain the same for primary and alternate paths, but the rest of the paths are typically resource-disjoint (e.g., node or link disjoint).

A bi-directional link between neighboring OXCs is usually realized as a pair of unidirectional links. The end-to-end path for a bi-directional connection therefore consists of a series of bi-directional links between the source and destination OXCs, traversing intermediate OXCs.

The manner in which restoration is accomplished depends on the type of protection afforded to the connection. In this regard, protection can be "local span" or "end-to-end". Local span protection refers to the protection of the connection segment between two neighboring switches. End-to-end protection refers to the protection of the entire connection from the ingress to the egress port. A connection may be subject to both local span protection (for each of its segments) and end-to-end protection (when local protection does not succeed). Under local span and end-to-end protection schemes, it may be required that when a failure affects any one direction of the connection, both directions of the connection are switched to a new link or path, respectively. In the following, therefore, any reference to a "link" indicates a bi-directional link (realized as a pair of uni-directional links).

Considering local-span protection, suppose a connection segment is routed over link i between two OXCs A and B. The following protection modes may be used:

1+1 (unidirectional): A dedicated alternate link j is pre-assigned to protect link i. Connection traffic is simultaneously sent on both links and received from one of the functioning link, i or j.

1+1 (bi-directional): A dedicated alternate link j is pre-assigned to protect link i. Connection traffic is simultaneously sent on both links and under normal conditions, the traffic from link i is received by OXCs A and B (in the

appropriate directions). A failure affecting link i results in both A and B switching to the traffic on link j in the respective directions.

1:M: A dedicated link j between A and B is pre-assigned to protect a set of M links (which includes i). A failure affecting any link in this set results in the corresponding traffic being restored to link j. Clearly, if more than one link in the set of M links are concurrently affected by failures, the traffic on only one of them may be restored over link j.

M:N (with pre-configured protection groups): A protection group of M+N links consists of a set of M links protected by a set of N links, with N < M. (Link i must be one of the M links in a protection group, and link j is one of the N links). A failure in any of the M links results in traffic being switched to one of the (available) N links. The number of protection groups between A and B, the value of M and N, as well as the specific links in each protection group is pre-configured. Since N < M, it is possible that not all failed links in the set of M links may be protected from the same failure event.

M:N (pooled protection links): Under this mode, a total of N links are assigned to protect a total of M other links betwen A and B, where M+N is the total number of links between A and B. (Link i must be one of the set of M links, and link j is one of the set of N links). The value of M and N is pre-configured, and the specific links in the set of N protection links may also be pre-configured. As before, since N < M, it is possible that not all failed links in the set of M links may be protected from the same failure event.

With SONET links, 1+1, 1:M and M:N with pre-configured protection groups are typically realized using SONET protocols. In this draft, we therefore focus on M:N protection with pooled protection links. This sort of protection requires a higher layer signaling protocol, as defined in this draft later.

Considering end-to-end protection, suppose a connection's primary (bi-directional) path is from an ingress port in OXC A to an egress port in OXC B over a set of intermediate OXCs. The following protection modes may be used:

1+1 (unidirectional): A dedicated, resource-disjoint alternate path

is pre-established to protect the connection. Connection traffic is simultaneously sent on both paths and received from one of the functional paths by the end OXCs, A and B.

1+1 (bi-directional): A dedicated, resource-disjoint alternate path is pre-established to protect the connection. Connection traffic is simultaneously sent on both paths and under normal conditions, the traffic from the primary path is received by OXCs A and B (in the appropriate directions). A failure affecting the primary path results in both A and B switching to the traffic on the back-up path in the respective directions

Shared: An alternate path is pre-assigned to protect the connection, but the links along the alternate path are shared among multiple connections being protected. In this case, the links are allocated in real-time for one of the protected connections whose primary path is affected by a failure. If more than one such connection is concurrently affected by a failure, only one of them will be allowed to use a shared link.

New protocols are required to realize both 1+1 (bi-directional) and shared end-to-end protection in optical mesh networks. Specifically, 1+1 (bi-directional) protection requires coordination between the end OXCs to switch to the protection path, and shared protection additionally involves the intermediate OXCs in the protection path.

The aim of this draft is to define the signaling protocols for both M:N pooled local span protection and the two end-to-end protection modes, 1+1 (bi-directional) and shared. The main requirements on these protocols are simplicity and speed. The latency requirement on switching to protection paths is typically specified in tens to hundreds of milliseconds, the performance depending on the number of hops involved. It is clear that a protocol level specification itself cannot guarantee these performance numbers, since a lot depends on the system architecture of the OXCs that implement the protocols. Indeed, this is the reason why these protocols are presently being implemented in a proprietary manner. It is, however, a reasonable goal to aim for a lightweight protocol mechanism that has a good chance of achieving the target performance. This is precisely the objective of this draft. Given the fact that IP-centric (GMPLS) protocols have been widely accepted for provisioning of primary and back-up paths, it is natural to consider standard, IP-based restoration protocols to move away from proprietary restoration solutions.

In the next section, the basic signaling mechanism proposed for fast restoration is outlined. Sections 4 and 5 describe local span and end-to-end protection protocols in detail. Section 6 presents some discussion items and Section 7 presents the conclusions.

3. Signaling Mechanism

3.1 Description

As described in the next two sections, the signaling protocols for local and end-to-end protection rely on a small set of short messages. A question that arises is whether an existing protocol framework could be used for supporting such messages. For instance, it is conceivable that RSVP-TE or CR-LDP protocols may be extended to support restoration signaling in the same way they are used for provisioning. In our view, restoration signaling must be given the highest priority and presently, there is no means to give relative priorities to different RSVP or CR-LDP messages (if used for both provisioning and restoration). Furthermore, our desire is to specify a lightweight protocol not encumbered by any of the RSVP or LDP semantics. Indeed, even if these protocols are extended for restoration signaling, they will in essence implement a logically distinct protocol framework for restoration as compared to signaling for provisioning. In this sense, it seems best to start with a new protocol dedicated for restoration signaling. This is indeed the proposal made in this draft.

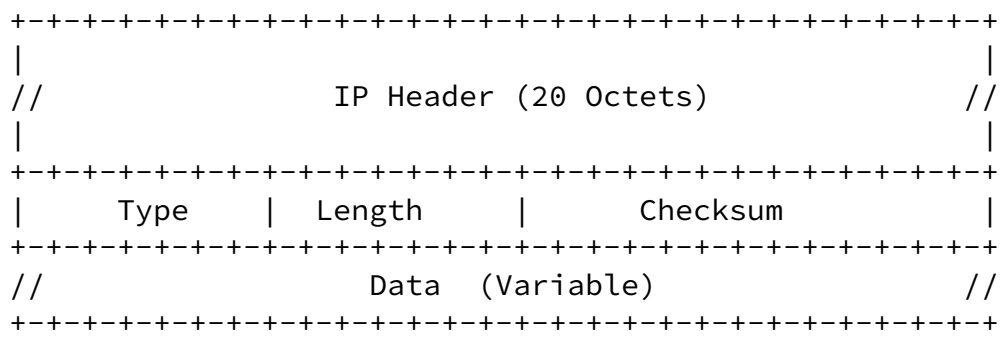
To main aspects of the proposed signaling mechanism are as follows:

- The signaling mechanism consists of new protocol running directly on top of IP
- The signaling mechanism supports a small set of messages, each of which is succinct, transported in a single short IP packet
- The signaling mechanism supports reliable delivery via a simple retransmission mechanism
- Signaling messages are sent over the IP control channel between neighboring OXCs. This control channel can be in-band or out-of-band.

As per this proposal, an implicit peering relationship exists

between protocol instances in two OXCs when the OXCs are topological neighbors in the data plane. Such a pair of OXCs must have one or more control channels between them, in-band or out-of-band. The number of protocol instances running in an OXC, and how the signaling pertaining to different connections are distributed within an OXC would depend on the system architecture. These details are beyond the scope of this draft.

The general format of signaling messages under this proposal is given below:



The source and destination address in the IP header are set to the sending OXC and the (neighboring) receiving OXC. The protocol ID in the IP header is set to the (yet unassigned) new value corresponding to the restoration protocol.

The restoration message types and data are as defined in the following sections. The 16-bit checksum field covers the entire restoration message, starting from the type field. The Length field indicates the number of 4-octet words in the message, starting with the Type field.

3.2 Applicability

The proposed signaling protocols for local span and end-to-end protection are intended to be applicable in both networks of O-E-O-type OXCs or all-optical OXCs. Because of the emphasis on bi-directional connections, the 1+1 protection procedures as described in this draft may not be relevant in general MPLS networks. Because of the need to configure intermediate OXCs in real-time for end-to-end shared protection, this procedure is also not applicable in general MPLS networks.

4. Local Span Protection

Local span protection is described with respect to two neighboring OXCs A and B. The scenario considered for local span protection is as follows:

- At any point in time, there are two sets of links between A and B, i.e., a "working" set of M (bi-directional) links carrying traffic subject to protection and a "free" set of N (bi-directional) links. A free link may have no traffic on it, or it may be carrying preemptable traffic. There is no apriori relationship between the two sets of links, but the value of M and N are pre-configured. The specific links in the free set MAY be pre-configured to be physically diverse to avoid the possibility that failure events affect a large proportion of free links (along with working links).

- When a link in the "working" set is affected by a failure, the traffic on it is diverted to a link in the "free" set, if such a link is available. Note that such traffic might consist of more than one connection, for example, an OC-192 link carrying four OC-48 connections.

- More than one link in the "working" set may be affected by the same failure event. In this case, there may not be an adequate number of "free" links to accommodate all of the affected traffic carried by failed "working" links. The set of affected "working" links that are actually restored over available "free" links is then subject to policies (e.g., based on relative priority of working traffic). These policies are not specified in this draft.

- Each OXC is assumed to have the mapping of its local link (or port) ID to the corresponding ID at the neighbor. This mapping could be configured, or obtained automatically using a neighbor discovery procedure (e.g., LMP [2]).

- When traffic must be diverted from a failed link in the "working" set to a "free" link, the decision as to which "free" link is chosen is always made by one of the OXCs, A or B. As per this draft, the OXC with the numerically larger IP address

is considered the "master" and it is required to both apply any policies and select specific "free" links to divert working traffic. The other OXC is considered the "slave". The

determination of the master and the slave may be based on configured information, or the result of running a neighbor discovery procedure.

- Failure events themselves are assumed to be detected by lower layer mechanisms (e.g., SONET). Since the bi-directional links are formed by a pair of unidirectional links, a failure in the link from A to B is typically detected by B and a failure in the opposite direction is detected by A. It is possible that a failure simultaneously affects both directions of the bi-directional link. In this case, A and B will concurrently detect failures, in the B-to-A direction and in the A-to-B direction, respectively.

The basic steps in local span protection are as follows:

1. If the master detects a failure of a "working" link, it autonomously invokes a process to allocate a "free" link to the affected traffic.

2. If the slave detects a failure of a "working" link, it must inform the master of the failure. The master then invokes the same procedure as above to allocate a "free" link. (It is possible that the master has itself detected the same failure, for example, a failure simultaneously affecting both directions of a link).

3. Once the master has determined the identity of the "free" link, it indicates this to the slave and requests the switchover of the traffic. Prior to this, if the "free" link is carrying preemptable traffic, the master stops using the link for this traffic (i.e., the traffic is dropped by the master and not forwarded into or out of the "free" link).

4. The slave sends an acknowledgement to the master. Prior to this, if the selected "free" link is carrying preemtable traffic, the slave stops using the link for this traffic (i.e., the traffic is dropped by the slave and not forwarded into or out of the "free" link). It then starts sending the (failed) "working" link traffic on the selected "free" link.

5. When the master receives the acknowledgement, it starts sending and receiving the (failed) working link traffic over the new link.

From the description above, it is clear that local span restoration requires three messages for each working link being switched: a failure indication message, a switchover request message and a switchover response message. The following identifier is also needed:

4.1 Identifiers

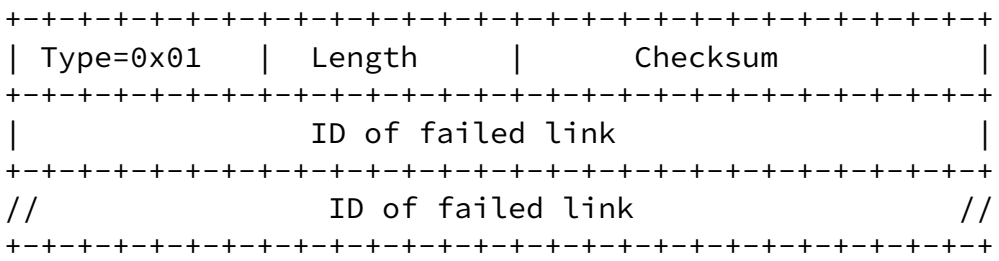
Link ID: A 32-bit identifier that uniquely identifies a bi-directional link at the sending and the receiving OXC. The link ID, for instance, could be the port ID at the receiving OXC.

The messages are as follows.

4.2 Failure Indication Message

This message is sent from the slave to the master to indicate the failure of one or more working links. (This message may not be necessary when the underlying link technology itself provides for such a notification)

The format of this message is as follows:



The number of links included in the message would depend on the number of failures detected within a window of time by the sending OXC. An OXC may choose to send separate failure indication messages in the interest of completing the restoration for a given link within an implementation-dependent time constraint.

The ID of the failed link is the identification used at the slave OXC. The master must convert this to the corresponding ID at its side.

This message is transmitted periodically by the slave, as controlled by the configurable timer, Retransmission Timer. The default value of this timer is 30ms. The slave stops transmitting this message under the following conditions:

- A corresponding Switchover Request message is received;
- The Failure Indication Timer expires (the action in the latter case is undefined in this draft. For instance, the sending OXC may notify a management system. The default value for this

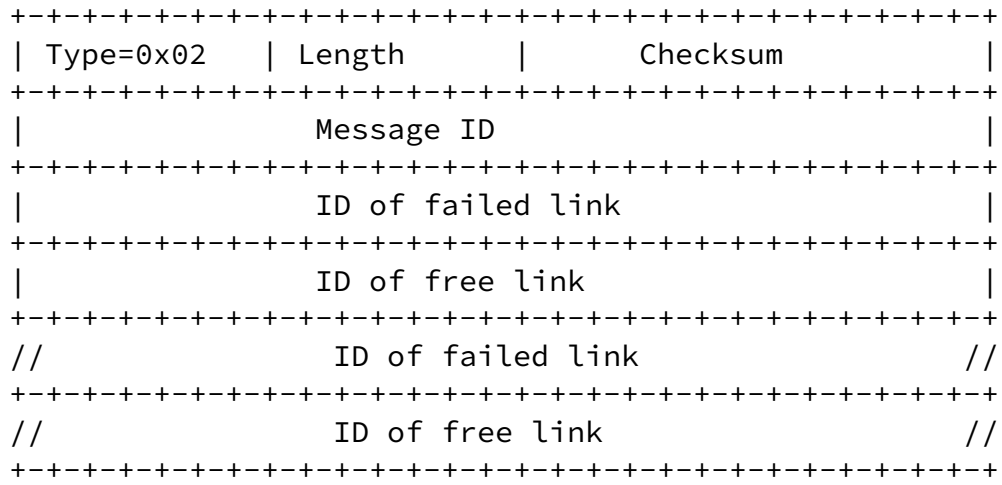
timer is 500 ms); or

- The connection is de-provisioned.

4.3 Switchover Request Message

This message is sent from the master to the slave to indicate whether the traffic on the failed working link can be switched to a free link, and if so, the ID of the free link.

The format of this message is as follows:



The link IDs are based on the identification used at the master. The slave must convert them to the corresponding local IDs. The message ID uniquely identifies the message at the master.

If the message is generated in response to a Failure Indication message from the slave then the set of failed links in the message MUST be the same as the set received in the Failure Indication message. If the ID of the free link is the same as the ID of the failed link then it is implicit that the traffic on the failed link cannot be switched to a free link.

This message is transmitted periodically by the master, as controlled by the configurable parameter, Retransmission Timer. The default value of this parameter is 30ms. Each retransmitted message has the same content, including the Message ID. The master stops transmitting this message under the following circumstances:

- The Switchover Timer expires (the action in the latter case is undefined in this draft. For instance, the master may notify a management system. The default value for this timer is 300 ms); or

- The connection is de-provisioned

- The corresponding Switchover Response is received.

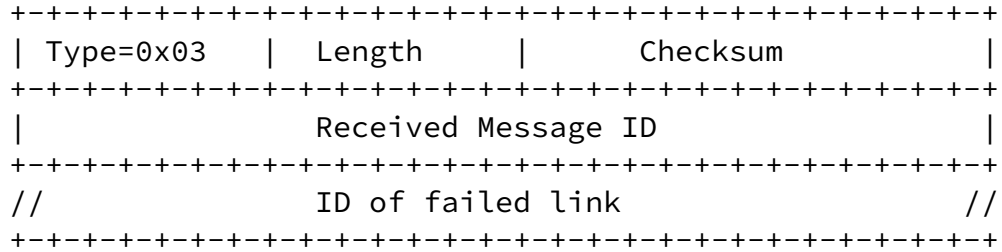
A failure event may result in the master sending more than one Switchover Request message to the same slave OXC. In this case, each of these messages will have different Message IDs and indicate

different failed links. The retransmission of each is controlled by a different instance of the Retransmission Timer.

4.4 Switchover Response Message

This message is sent from the slave to the master to indicate the receipt of the Switchover Request message.

The format of this message is as follows:



The Received Message ID field is filled with the Message ID received in the corresponding Switchover Request message. The failed link ID field, if present, indicates that the slave cannot switch over to the corresponding free link for some reason. This identification is based on the IDs used at the slave. The action to be taken by the master when there are one or more failed link IDs in the Switchover Response message is undefined (for example, the master may abort the switchover of the traffic on the failed working link, and perhaps trigger end-to-end protection).

The slave responds with the appropriate Switchover Response each time it receives a Switchover Request Message (including retransmitted Switchover Request messages). Of course, the switchover action itself must be performed only once by the slave

for a given failed link.

5. End-to-End Protection

One of the significant differences between end-to-end protection and local span protection (as considered in this draft) is that the latter is on a per-link basis while the former is on a per-connection basis. In other words, span protection switches over the entire traffic on a link which may consist of multiple connections. End-to-end protection, on the other hand, switches over individual connections. In this case, there is a "working" connection path and a "protection" path.

Another difference between end-to-end and local protection is that signaling messages may have to be transmitted multiple hops to effect restoration. The signaling messages are transmitted along the connection path, working or protection, where it is assumed that there is a control channel between each pair of intermediate OXCs.

There are two cases to be considered: signaling for bi-directional 1+1 protection and for shared protection. The description below is in the context of an end-to-end connection between a source OXC A and a destination OXC B. The source for such a connection is also considered to be the "owner" of the connection. The owner is the node which initiated the connection establishment or designated when the connection is manually provisioned.

5.1 Bi-directional 1+1 Protection

Under bi-directional 1+1 protection, the connection traffic is being sent on both working and protection links by A and B, but received only from the working link. After a failure event, signaling between A and B is required to ensure that both A and B start receiving from the protection link.

A failure event is detected by a node in the working path. Such a node sends a failure indication signal towards the owner (source) of the connection along the working path (using signaling described below, or mechanisms provided by the lower layer).

The action when the source OXC is notified of a failure is as follows:

- o Start receiving from the protection path. At the same time, send a Switchover Request message along the protection path towards the destination OXC. This message is not explicitly addressed to the destination OXC, but it carries a connection ID which aids intermediate OXCs in forwarding the message along the protection path towards the destination.

The action when the destination OXC receives a Switchover Request message is follows:

- o Start receiving from the protection path. At the same time, send a Switchover Response message along the protection path towards the source OXC. The forwarding of this message by intermediate OXCs is based on the connection identifier information.

5.1.1 Identifiers

Source ID: The IPv4 address of the source OXC (connection owner). This is assumed to be unique for OXCs within the scope of the restoration domain.

Connection index: A 32-bit identifier that uniquely identifies a connection at the source.

Together, the Source ID and the Connection index identify a connection uniquely within the scope of the restoration domain. This combination is referred to as the "Connection ID" for convenience.

5.1.2 Node Data Structure

Each OXC that is the working or protection path of a connection must maintain a table whose entries consist of the following information:

<Connection ID (source ID, connection index), Previous OXC, Next OXC, Protection Type>

The Previous and Next OXC fields are 32-bit IPv4 addresses of neighboring OXCs. These are the same addresses used in the IP header of restoration messages sent to these neighbors.

At the source and destination OXCs, the Previous and Next OXC values are both set to indicate the same neighbor towards the other endpoint of the connection.

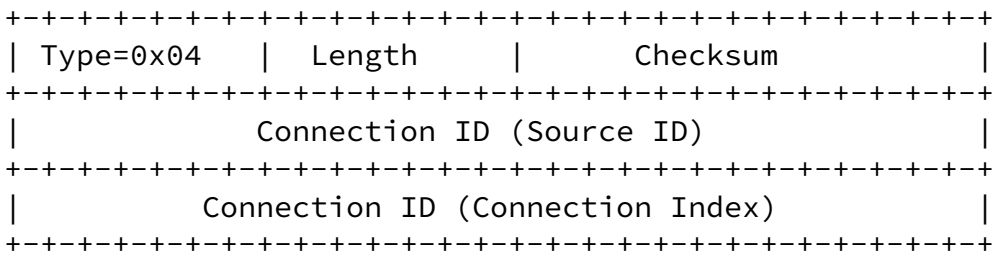
The Protection Type field indicates whether 1+1 or Shared protection is being provided for the connection.

This table may be built when the working and protection paths of the connections are provisioned using signaling, or may be configured in the case of NMS-based provisioning. The table entries must remain until the connection is explicitly de-provisioned.

5.1.3 End-to-End Failure Indication Message

This message is sent by an intermediate node towards the source of a connection along the working path of the connection. For instance, such a node might have attempted local span protection and failed (this message may not be necessary if the lower layer provides mechanisms for detection of connection failure by the endpoints).

The format of this message is as follows:



Consider an OXC detecting a link failure. Suppose the failed span is to a neighbor whose identity is N. The OXC retrieves all table entries where N is the Previous or Next OXC. Each such table entry

indicates the affected connection, and the above message is generated for each connection. The message is sent in an IP packet with the source address set to the address of the sending OXC and the destination address set to the Previous OXC.

This message is transmitted periodically by the OXC, as controlled by the configurable parameter, End-to-End Retransmission Timer. The default value of this parameter is 90ms. The OXC stops transmitting this message under the following conditions:

- A corresponding Failure Acknowledge message is received;

o The End-to-End Failure Indication Timer expires (the action in the latter case is undefined in this draft. For instance, the sending OXC may notify a management system. The default value for this timer is 1 minute); or

o The connection is de-provisioned.

Each OXC in the working path receiving such a message does the following.

- If it is the source of the indicated connection, it stops propagating the message further and generates an acknowledgement message (Sec. 5.1.4).

- If it also has originated a failure indication message for the same connection, it stops re-transmitting such a message. Furthermore, it also does not propagate the received failure indication message further. (This action is needed when a bi-directional data link failure occurs between two OXCs, say, A and B. Suppose A is upstream from B (in relation to the connection source). Also suppose that A has generated a failure indication towards the source. If it later receives a failure indication generated by B, it must suppress the retransmission of failure indication messages it originated. Also, it need not forward the first failure indication message originated by B. Finally, it must forward failure indication acknowledge messages received from the source to B (Section 5.1.4)).

- Otherwise, it looks up the table entry corresponding to the Connection ID.

    o If there is no table entry, the received message is silently discarded.

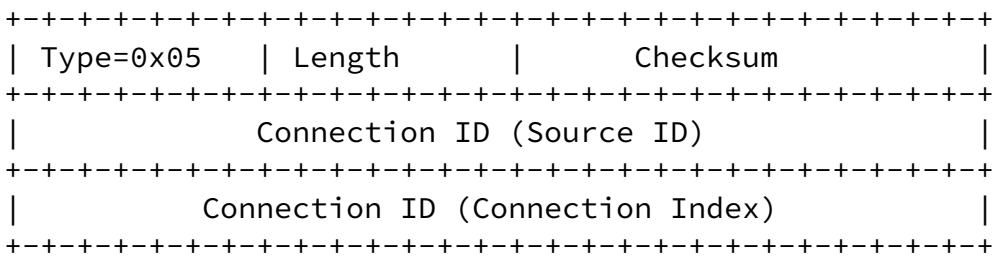
    o Otherwise, the Source Address in the IP header of the message is compared with the Next OXC field: If there is a match then the received message is sent in an IP packet to the Previous OXC, with Source Address set to the address of the sending OXC and Destination Address

set to Previous OXC. If there is no match, the received message is silently discarded.

5.1.4 End-to-End Failure Acknowledge Message

This message is sent by the source OXC in response to an End-to-End failure indication message. This message is sent towards the originator of the failure indication message along the working path of the connection.

The format of this message is as follows:



This message is transmitted in response to each End-to-End Failure Indication message. The source OXC always sends this message to the Next OXC as found in the table entry corresponding to the Connection ID.

Each intermediate OXC receiving such a message does the following.

- If it is the originator of the corresponding Failure Indication message, it stops propagating the message further.

- Otherwise, it looks up the table entry corresponding to the Connection ID.

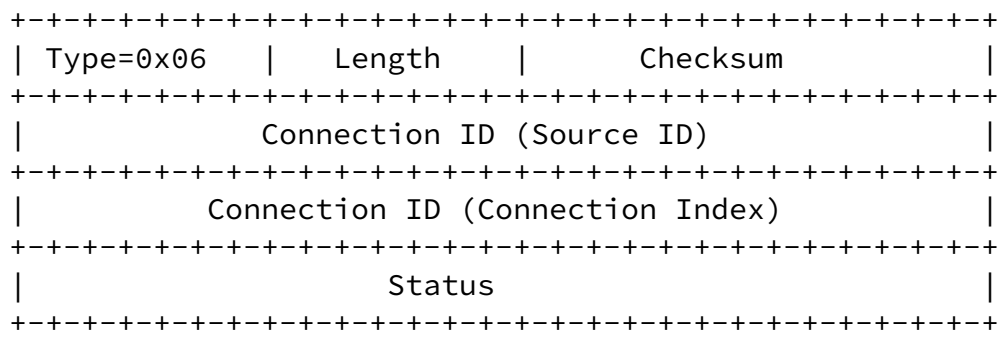
  - o If there is no table entry, the received message is silently discarded.

  - o Otherwise, the Source Address in the IP header of the message is compared with the Previous OXC field: If there is a match then the received message is sent in an IP packet to the Next OXC, with Source Address set to the address of the sending OXC and Destination Address set to Next OXC. If there is no match, the received message is silently discarded.

5.1.5 End-to-End Switchover Request Message

This message is generated by the source OXC receiving an indication of failure in a connection. It is sent to the Next OXC along the protection path.

The format of this message is as follows:



The Status field is set to 0x01 if the end OXC is able to switchover to the protection path. Otherwise, it is set to 0x00.

This message is transmitted periodically by the end OXC, as controlled by the configurable parameter, End-to-End Retransmission Timer. The default value of this parameter is 90ms. The OXC stops transmitting this message under the following conditions:

o A corresponding End-to-End Switchover Response message is received;

o The End-to-End Switchover Timer expires (the default value for this timer is 1 minute); or

o The connection is explicitly de-provisioned.

Each intermediate OXC receiving such a message does the following. It looks up the table entry corresponding to the Connection ID:

o If there is no table entry, the received message is silently discarded.

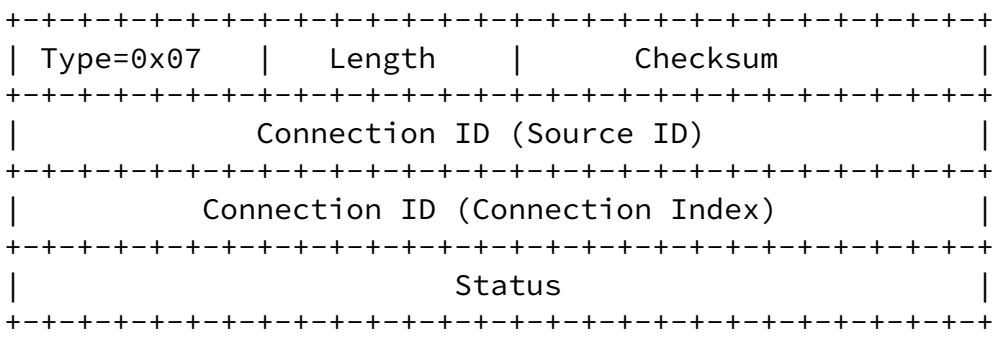
o Otherwise, if the protection type for the connection is 1+1 then

  - the Source Address in the IP header of the message is compared with the Previous OXC field: If there is a match then the received message is sent in an IP packet to the Next OXC, with Source Address set to the address of the sending OXC and Destination Address set to Next OXC. If there is no match, the received message is silently discarded.

5.1.6 End-to-End Switchover Response Message

This message is sent by the destination OXC receiving an End-to-End Switchover Request message. It is sent to the previous OXC in the protection path.

The format of this message is as follows:



This message is transmitted in response to each End-to-End Switchover Request message received. If the Status field was set to 0x00 in the Request message, it is set to 0x00 in this message also. Otherwise, the Status field is set to 0x01 or 0x00 depending on whether the sending OXC is able to switchover to the protection path or not, respectively.

Each intermediate OXC receiving such a message does the following. It looks up the table entry corresponding to the Connection ID:

- If there is no table entry, the received message is silently discarded.
- Otherwise, if the protection type for the connection is 1+1 then
  - the Source Address in the IP header of the message is compared with the Next OXC field: If there is a match then the received message is sent in an IP packet to the Previous OXC, with Source Address set to the address of the sending OXC and Destination Address set to Previous OXC. If there is no match, the received message is silently discarded.

5.2 Shared Protection

Shared protection requires prior soft-reservation of capacity along the protection path. Furthermore, after a failure event, the protection path must be explicitly activated. This requires actions at each intermediate OXC along the protection path. It is possible that a protection path may not be successfully activated when multiple, concurrent failure events occur. In this case, shared protection capacity may be claimed for more than one failed connection and the protection path can be activated only for one of them (at most).

For implementing shared protection, the identifier and data structures related to signaling along the control path are as defined for 1+1 protection in Sections 5.1.1 and 5.1.2. In addition, each OXC must also keep information needed to establish the data plane of the protection path. This information indicates the cross-connect that must be established to activate the protection path for each connection, as follows:

```
{    Connection ID, <Incoming Port, Channel etc>, <Outgoing Port,
     Channel, etc>
}
```

The precise nature of the Port, Channel, etc. information would depend on the type of OXC and connection (The Generalized MPLS signaling draft describes different type of switches [3]).

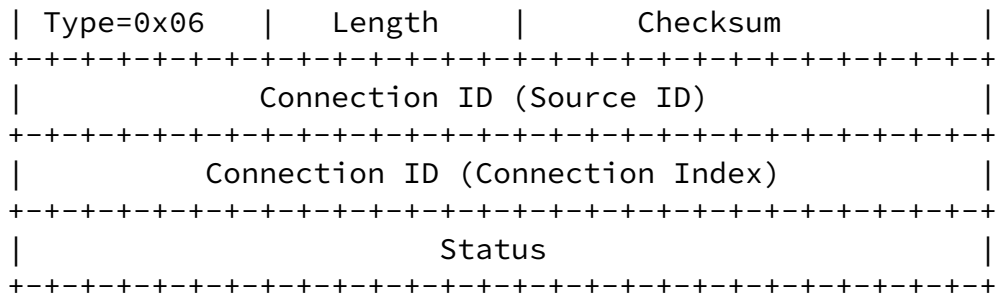
5.2.1 End-to-End Failure Indication and Acknowledgement

The End-to-End failure indication and acknowledgement procedures and messages are as defined in Sections 5.1.3 and 5.1.4.

5.2.2 End-to-End Switchover Request

This message is generated by the source OXC receiving an indication of failure in a connection. It is sent to the Next OXC in the protection path.

The format of this message is as follows:



The Status field is set to 0x01 if the end OXC is able to switchover to the protection path. Otherwise, it is set to 0x00.

This message is transmitted periodically by the end OXC, as controlled by the configurable parameter, End-to-End Retransmission Timer. The default value of this parameter is 90ms. The OXC stops transmitting this message under the following conditions:

- A corresponding End-to-End Switchover Response message is received;
- The End-to-End Switchover Timer expires; or
- The connection is explicitly de-provisioned.

Each intermediate OXC receiving such a message does the following. The OXC looks up the table entry corresponding to the Connection ID:

- If there is no table entry, the received message is silently discarded.
- Otherwise, if the Protection Type for the connection is Shared then the Source Address in the IP header of the message is compared with the Previous OXC field:
  - If there is no match then the received message is silently discarded.
  - Otherwise,
    - The value of the received Status field is checked. If it is 0x00 then the received

message is sent in an IP packet to the Next OXC, with Source Address set to the address of the sending OXC and Destination Address set to Next OXC.

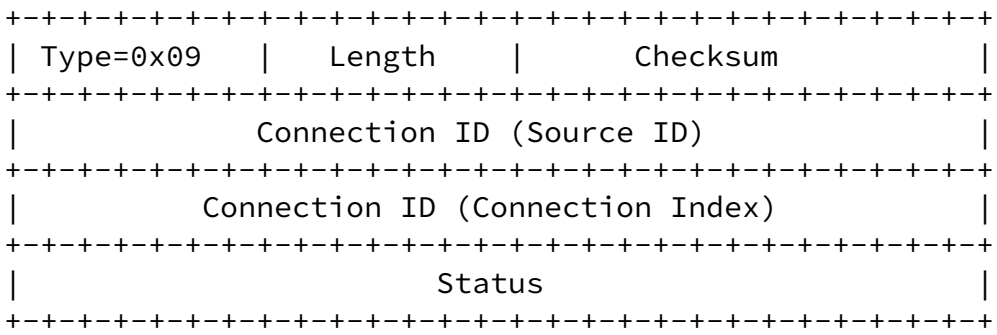
- If the received Status field is 0x01 then it is checked if the protection path for the connection can be activated. If so, the message is passed on to the Next OXC as described above.

- If the received Status field is 0x01, but the protection path for the connection cannot be activated then the message is passed on to the Next OXC as described above, with the Status field set to 0x00.

5.2.3 End-to-End Switchover Response

This message is sent by the destination OXC receiving an End-to-End Switchover Request message. It is sent to the previous OXC in the protection path.

The format of this message is as follows:



This message is transmitted in response to each End-to-End Switchover Request message received. If the Status field was set to 0x00 in the Request message, it is set to 0x00 in this message also. Otherwise, the Status field is set to 0x01 or 0x00 depending on whether the destination OXC is able to switchover to the protection path or not, respectively.

Each intermediate OXC receiving such a message does the following. The OXC looks up the table entry corresponding to the Connection ID:

- If there is no table entry, the received message is silently discarded.
- Otherwise, if the Protection Type for the connection is "shared" then the Source Address in the IP header of the message is compared with the Next OXC field:
  - If there is no match then the received message is silently discarded.
  - Otherwise,
    - The value of the received Status field is checked. If it is 0x00 then the received message is sent in an IP packet to the Previous OXC, with Source Address set to the address of the sending OXC and Destination Address set to Previous OXC. The protection path is not activated in this case.
    - If the received Status field is 0x01 then the protection path for the connection is activated and the message is passed on to the Previous OXC as described above.

6. Discussion

6.1 Relationship between Local and End-to-End Protection Procedures

In general, local protection may be attempted before invoking end-to-end protection. The exception to this is when end-to-end 1+1 protection is used for a connection. In this case, it is better to directly invoke end-to-end protection since alternate path resources are already active for the connection.

Thus, the general guideline that may be considered is to note the protection type of connections in intermediate OXCs during provisioning, and invoke local span protection only for working links carrying connections that are not 1+1 protected end-to-end. This implies that when a working link carries more than one connection, all the connections must have the same end-to-end protection type. The provisioning process must ensure this. If this is not possible then local span protection may be invoked for working links that have at least one connection that is not end-to-end 1+1 protected.

6.2 Connection Priorities During Protection

The local protection procedure described in this draft switches all the connections on a failed working link onto a protection link. The advantage of this approach is that the signaling between OXCs is at the level of links and not at the level of connections. This is beneficial if a link could potentially carry a number of connections. On the other hand, it limits flexibility, since a working link must carry connections of similar priority. Otherwise, it is not possible to ensure that higher priority connections are favored over lower priority connections when a failure event affects more than one working link and there are fewer protection links than the number of failed working links.

Also, under the above failure scenario, a decision must be made as to which working links (and therefore connections) are chosen to be protected and in what priority order. In general, an OXC might detect failures sequentially, i.e., all failed working links may not be detected simultaneously, but only sequentially. In this case, as per the proposed signaling procedures, connections on a working link may be switched over to a given protection link, but another failure (of a working link carrying higher priority connections) may be detected soon afterwards. In this case, the new connections may bump the ones previously switched over the protection link.

In the case of end-to-end shared protection, priorities may be implemented for allocating shared link resources under multiple failure scenarios. Note that shared protection works under the assumption that the primary path of connections whose backups share resources are SRLG-disjoint [1]. Under single-failure scenarios, this would ensure that exactly one connection will "claim" the allocated (shared) resources. But under multiple failure scenarios,

more than one connection can claim shared resources. If such resources are allocated to a lower priority connection, they may

have to be reclaimed and allocated to a higher priority connection. Furthermore, the lower priority connection must be de-provisioned along the protection path (this can be done using the signaling mechanisms developed for provisioning, rather than restoration signaling). The proposed signaling mechanisms can support connection-priority based allocation of shared resources during restoration signaling (specifically, during the Switchover Response step).

A way to simplify end-to-end shared protection is to allocate shared resources to connections of the same priority. This way, a connection will not be first allocated shared resources and then bumped from the protection path.

6.3 Multi-Domain Restoration

When an end-to-end connection follows a long path through multiple routing or administrative domains, it may be required to consider an intermediate form of restoration, called "intra-domain end-to-end restoration". With this approach, a failure within a domain would result in end-to-end restoration between the connection ingress and egress points within the domain (perhaps after local span restoration is attempted). When this fails, or if a failure occurs in an inter-domain link, full end-to-end restoration will be attempted.

This type of a structured approach for restoration is particularly useful in the near term when an optical internetwork may be constructed by interconnecting multi-vendor optical subnetworks [1]. In this case, intra-domain restoration may be proprietary, with standard restoration signaling implemented between border OXCs. But this type of restoration also requires some hardware support at the border nodes.

6.4 Optical mesh restoration and MPLS-based recovery

Over the past year or so, there has been considerable work on MPLS-based recovery under the auspices of the MPLS WG (see, for example, [3], [5], [6], [7], [8], [9], and [10]), with a framework document [4] being adopted as a WG document.

The terminology outlined at the start of this document is also explained in the MPLS-recovery framework document [4], in the context of MPLS LSP-based recovery.

The failure indication message of Section 4, is quite similar to the failure indication signal (FIS) defined in [5], and elaborated on in [9] and [10]. A difference between the schemes and message formats discussed in this document and those presented in [5], [9], and [10], is that these documents focus primarily on MPLS LSP

restoration. As such, the messages defined therein contain explicit label information for packet LSPs, which is not required in optical networks. Further, [5] does not specifically cover the case of the coordinated signaling required for local span protection and for M:N protection with pooled protection links, which are central to this proposal.

6.5 Implementation Considerations

As described in this draft, restoration signaling does not require any central actions (such as admission control or centralized resource allocation) within an OXC for end-to-end protection. Local span protection may require the consideration of all available protection link resources at the master. But end-to-end protection, which is more difficult from a latency perspective, can be controlled by distributing multiple, independent protocol instances in an OXC such that each instance covers a subset of connections passing through an OXC. Such optimizations would depend on the architecture of the systems implementing the proposed protocol.

6.6 Summary of Proposed Messages and Timers

The following messages were introduced in this draft:

Message Type	Message Name
0x01	Failure Indication
0x02	Switchover Request
0x03	Switchover Response
0x04	End-to-End Failure Indication
0x05	End-to-End Failure Acknowledge
0x06	End-to-End Switchover Request
0x07	End-to-End Switchover Response

The following timers were introduced in this draft:

Timer Name	Default Value	Messages Governed
Retransmission Timer	30 ms	1, 2
Failure Indication Timer	500 ms	1
Switchover Timer	500 ms	2
End-to-End Retransmission Timer	90 ms	4, 6
End-to-End Failure Indication Timer	1000 ms	4

End-to-End Switchover Timer	1000 ms	6

7. Conclusion

In this draft, a signaling mechanism for fast restoration in optical mesh networks was described. The proposed mechanism consists of a protocol running directly over IP. The proposed messages are short and the interaction amongst nodes as proposed is rather simple. This proposal deals with local span protection and end-to-end protection, 1+1 and shared.

It is likely that the structure of the proposed mechanism would allow restoration to be completed quickly. Clearly, extensive quantitative evaluation is needed before the responsiveness of the proposed mechanism can be established.

The specification of finite state machines in intermediate and end nodes for different protection modes is left for future work.

References

1. B. Rajagopalan, et al., "IP over Optical Networks: A Framework", draft-many-ip-optical-framework-02.txt.

2. J. P. Lang, et al, ""Link Management Protocol", draft-ietf-mpls-lmp-01.txt.

3. P. Ashwood-Smith, et al., "Generalized MPLS: Signaling Functional Specification," draft-ietf-mpls-generalized-signaling-01.txt.

4. Makam, et al, "Framework for MPLS-based Recovery," draft-ietf-mpls-recovery-frmwrk-02.txt, February 2001.

5. K. Owens et al, "A Path Protection/Restoration Mechanism for MPLS Networks," draft-chang-mpls-path-protection-02.txt, November 2000.

6. Kini, S., et al, "Shared Backup Label Switched Path Restoration,"

draft-kini-restoration-shared-backup-00.txt, November 2000.

7. Hellstrand, F., and Andersson, L., "Extensions to CR-LDP and RSVP-TE for setup of pre-established recovery tunnels," draft-hellstrand-recovery-merge-01.txt, November 2000.

8. D. Haskin, Krishnan, R., "A Method for Setting up an Alternative Label Switched Path to Handle Fast Reroute," draft-haskin-mpls-fast-reroute-05.txt, November, 2000.

9. K. Owens et al, "Extensions to RSVP-TE for MPLS Path Protection," draft-chang-mpls-rsvpte-path-protection-ext-01.txt, November 2000.

10. K. Owens et al, "Extensions to CR-LDP for MPLS Path Protection," draft-owens-mpls-crldp-path-protection-ext-01.txt, November 2000.

Author Information

Bala Rajagopalan, Debanjan Saha
Tellium, Inc.
2 Crescent Place
Ocean Port, NJ 07757
Ph.: +1-732-923-4237, 4264
Email: {braja, dsaha}@tellium.com

Greg Bernstein
Ciena Corp.
10480 Ridgeview Court
Cupertino, CA 94014
Ph: +1-408 366 4713
Email: gregb@ciena.com

Vishal Sharma
Jasmine Networks
3061 Zanker Rd, Suite B
San Jose, CA 95112
Ph: +1-408-895-5030
Email: vishal@jasminenetworks.com